Human vs. Machine Minds: Ego-Centric Action Recognition Compared

Sadegh Rahmani[1], Filip Rybansky[2], Quoc Vuong[2], Frank Guerin[1], Andrew Gilbert[1]
[1] University of Surrey, [2] University of Newcastle

Scan to read

Funded By LEVERHULME TRUST

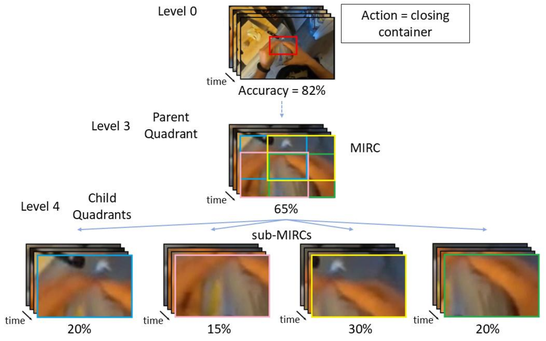
Motivation

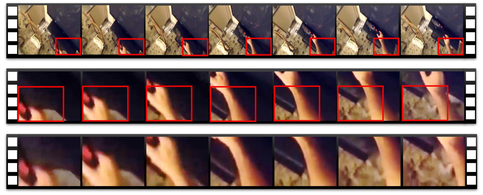
- **Is the Recognition of action different between humans and AI models?**
  - ✓ We test their performance in challenging real-world ego-centric scenarios involving occlusion, clutter, and low-resolution.
  - ✓ This study seeks to understand *how* and *why* human and machine recognition diverge by introducing a new benchmark, **Epic ReduAct**

- **What is the takeaway?**
  - ✓ Humans excel in complex recognition but fail with minimal input, while AI stays more robust.
  - ✓ Future work will explore temporal cues and multimodal data to better align AI with human perception.

Epic ReduAct dataset

- **What is Epic ReduAct?**
  - ✓ A *systematically reduced* version of videos from the **Epic-Kitchens-100** dataset, designed to study **minimal visual requirements** for recognizing ego-centric human actions.
  - ✓ Identify the smallest recognisable quadrants or **Minimal Recognisable Configurations (MIRCs)**
  - ✓ Actions span common kitchen tasks (e.g., *cut*, *pour*, *wash*)
  - ✓ Created two sets; Easy/Hard, based on AI model (MOFO) performance



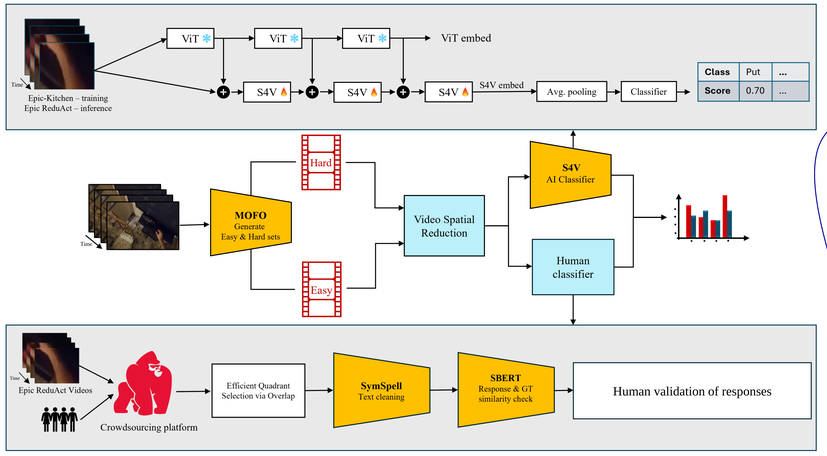


a) Original video – No reduction – GT *put*

b) Level 2 – The MIRC level

c) Level 3 – The sub-MIRC level of Fig. b

Methodology & Pipeline

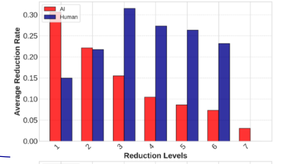


Experiments & Results

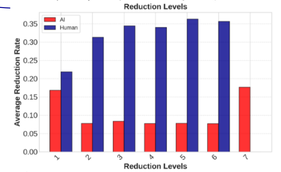
Average Reduction Rate

Measures impact of spatial reduction on recognition accuracy

Metric: Average reduction rate as a function of reduction level for all parent-child quadrant pairs.
AI Model Performance:
✓ Accuracy declines gradually as spatial reduction increases.
Human Performance:
✓ Recognition stable at initial levels of spatial reduction.
✓ Drops sharply at higher levels of reduction.

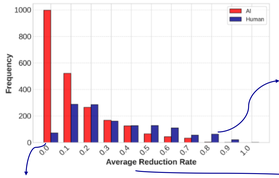


(a) easy set



(b) hard set

combined



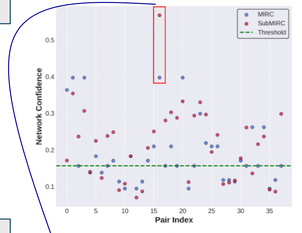
~ 100 videos with a high drop

Around 1000 videos with the lowest drop

Metric: Distribution of pairs at all reduction levels.
Pattern: Humans' recognition drops significantly while AI model remains more robust in recognition

Classwise Recognition Gap

The difference in average classification accuracy between full MIRCs and their sub-MIRC quadrants.

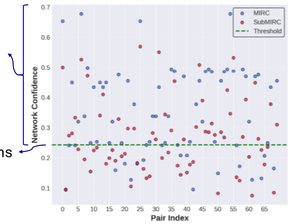


(b) Class put

subMIRC got a better accuracy than its parent quadrant (MIRC) in AI model
Video depicted in dataset section (bottom figure)

59% of MIRCs above threshold
As human acc. for MIRC is 59%

Threshold at 24% (which means 59% of AI predictions are above)



(d) Class take

Therefore, the recognition gap for that class will be the difference between the average accuracy of MIRCs of that class (59%) and those sub-MIRCs above the threshold line (36%), which is 59% - 36% = 23%

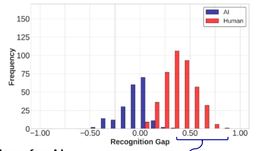
Classifiers	Sets	hang	serve	take	open	remove	turn-off	turn-on	wash	peel	close	cut	put	pour	insert
Human	Easy	+38.65	N/A	+37.46	+37.86	+27.85	+34.18	+31.11	+31.89	N/A	+44.75	+36.59	+36.84	+39.70	N/A
	Hard	+41.70	+42.87	+40.78	+38.36	+37.50	+60.55	N/A	+31.05	+38.20	+33.75	N/A	+34.31	+24.70	+40.00
AI	Easy	+0.02	N/A	-19.25	-5.91	-9.20	-7.36	-16.36	-0.96	N/A	+4.26	-2.68	+4.12	+0.14	N/A
	Hard	-0.04	0.00	-9.00	-3.18	-0.49	-0.21	N/A	-2.08	-0.01	-0.07	N/A	+2.63	0.67	-0.02

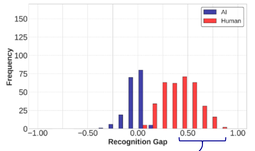
While human recognition performance declines across all classes, AI model exhibits improvements in both sets

Recognition Gap

Metric: Frequency distribution of recognition gap.

(a) easy set



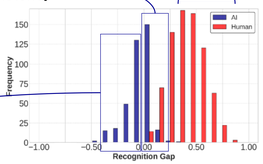


(b) hard set

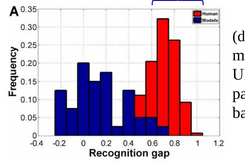
No sign of major drop for AI model, sign of consistency

Humans' detection falls of a cliff

(c) combined



Sign of improvement by reduction in AI predictions



(d) The same metric from Ullman's 2018 paper on image-based recognition