

Continuous Online Action Detection from Egocentric Videos

Riccardo Santambrogio¹, Chiara Plizzari², Francesca

Palermo, Simone Mentasti¹, Matteo Matteucci¹

¹ Politecnico di Milano, Milan, Italy

² Università Bocconi, Milan, Italy

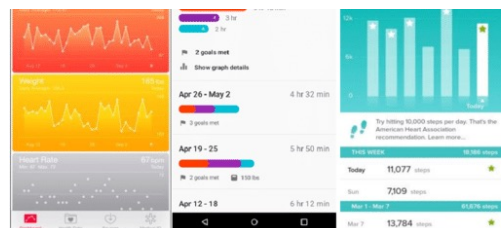
ICPR 2026

Egocentric Action Understanding

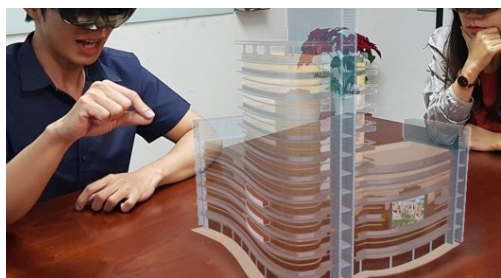
Assistive Technologies



Life-Logging



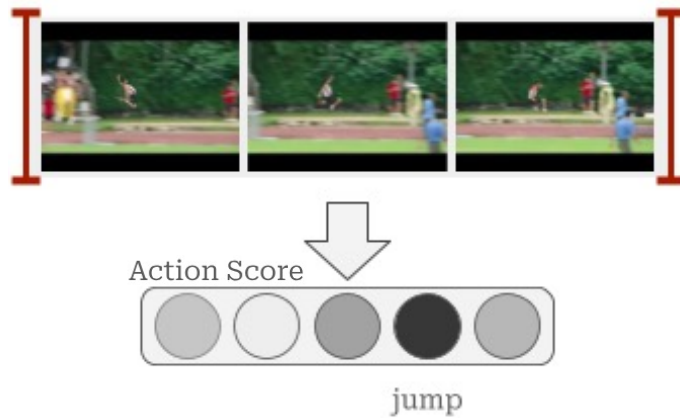
AR/XR



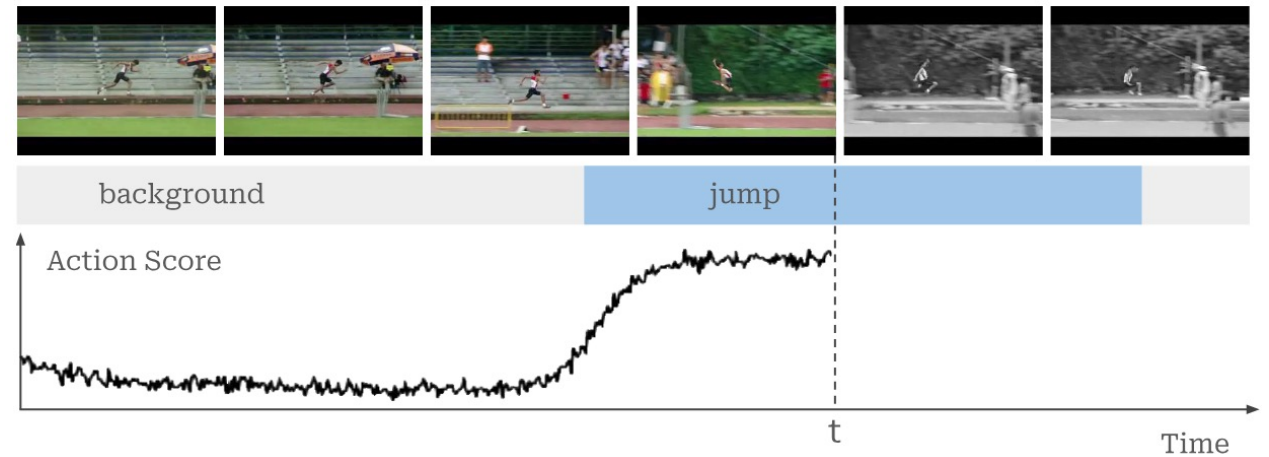
Source: "Ego4D: Teaching AI to perceive the world through your eyes"

Online Action Detection (OAD)

Offline Action Recognition



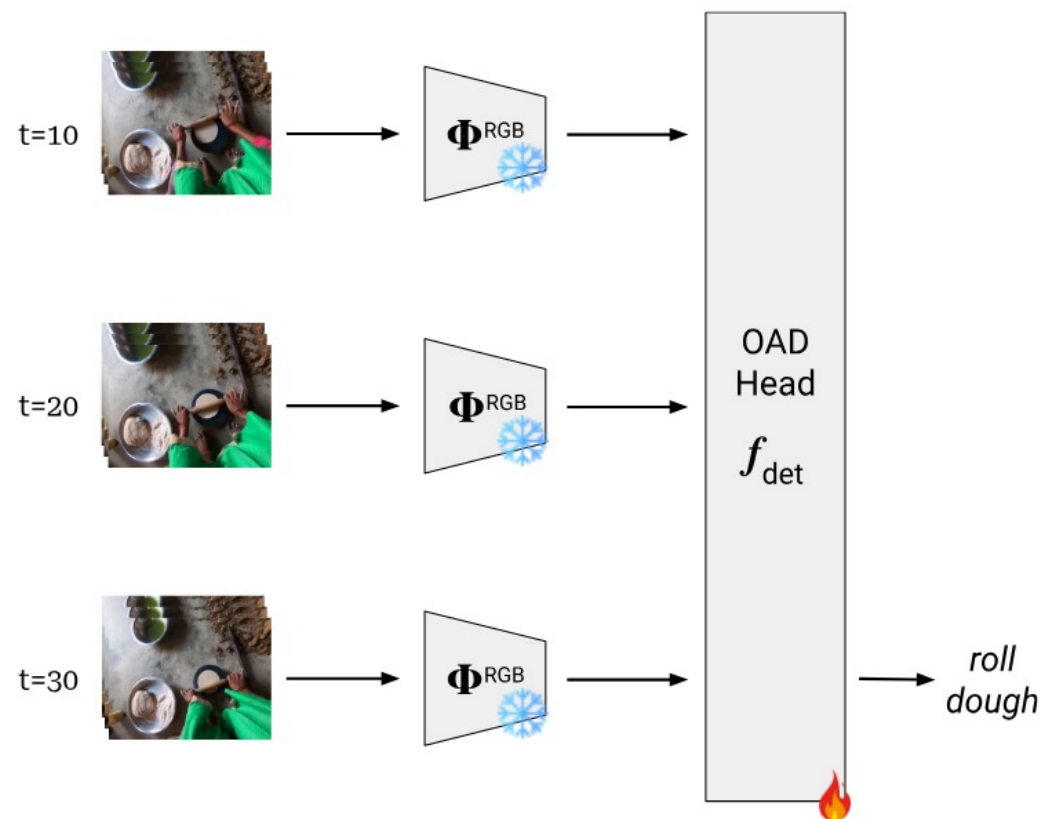
Online Action Detection



- Online Action Detection: task of recognizing ongoing actions from streaming video
- Egocentric OAD has received limited attention outside of kitchen activities

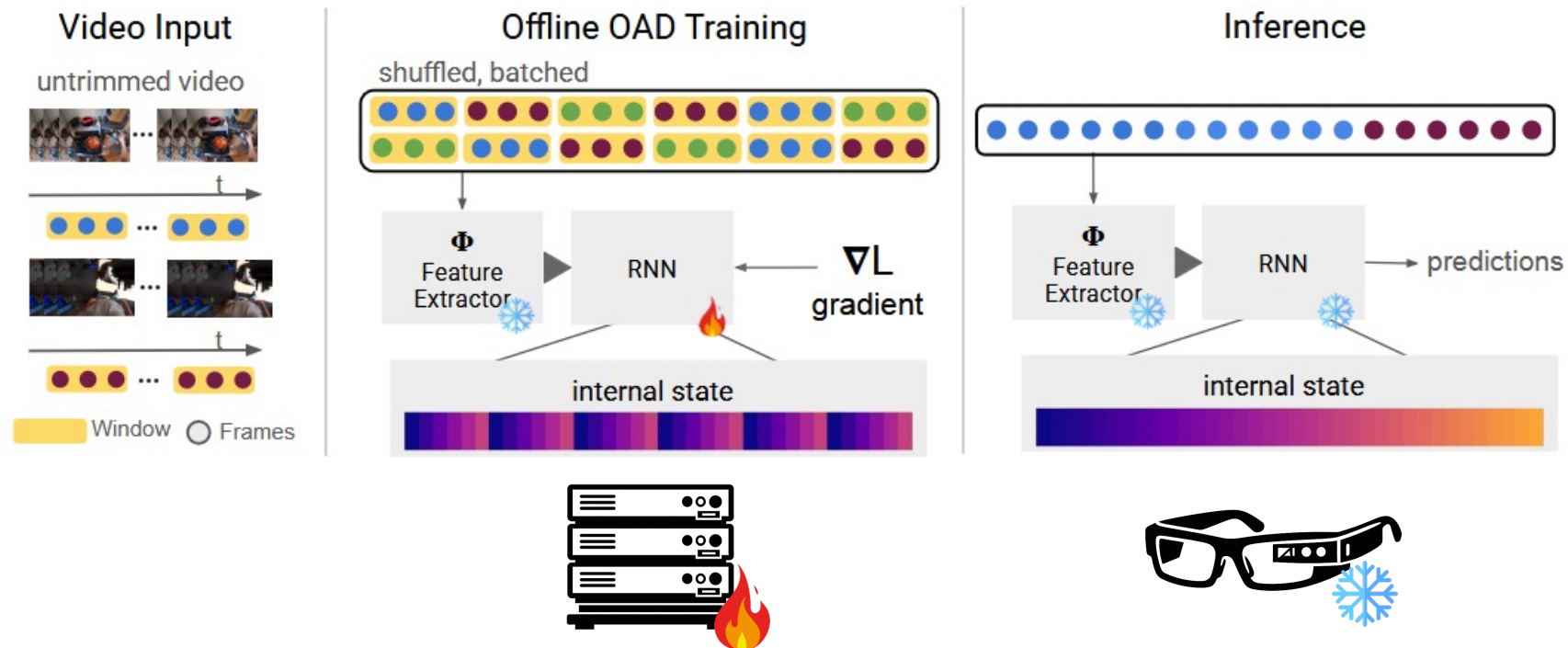
OAD Architecture

- OAD methods follow a 2-stage architecture:
 1. Pretrained local feature extractor
 2. OAD head for long-term aggregation
- We adopt a lightweight RNN head



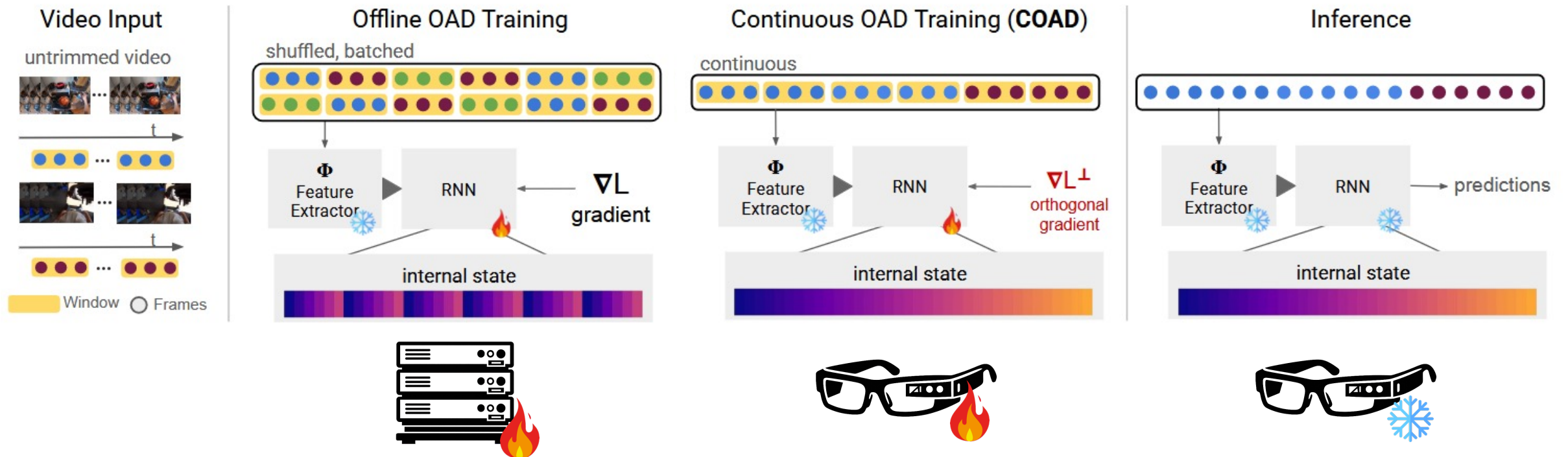
Continuous Learning for OAD

- OAD models are trained offline and deployed as fixed predictors
- We address continuous post-deployment learning for OAD



Continuous OAD (COAD)

- We formulate COAD and propose a training scheme that enables learning from streaming video in a deployment setting



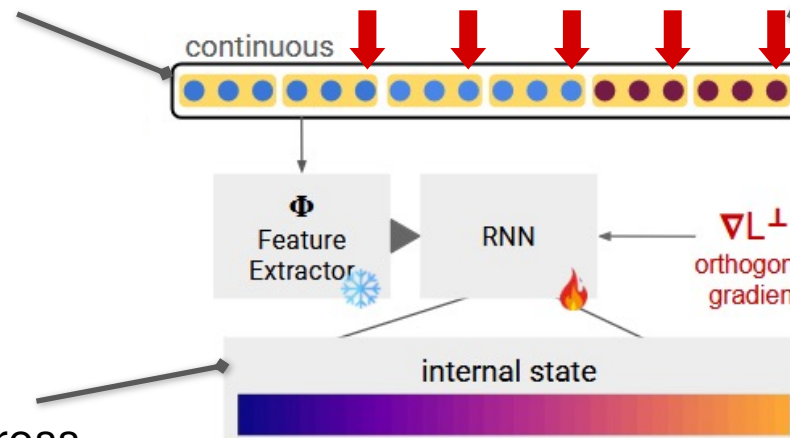
COAD Components

Single-Pass training

Single forward pass on the continuous stream in causal order and no replay

Non-uniform Loss

Use sparse supervision by computing loss only at the final step



State Continuity

Preserve temporal state across time for consistent long-term modeling

Orthogonal Gradient

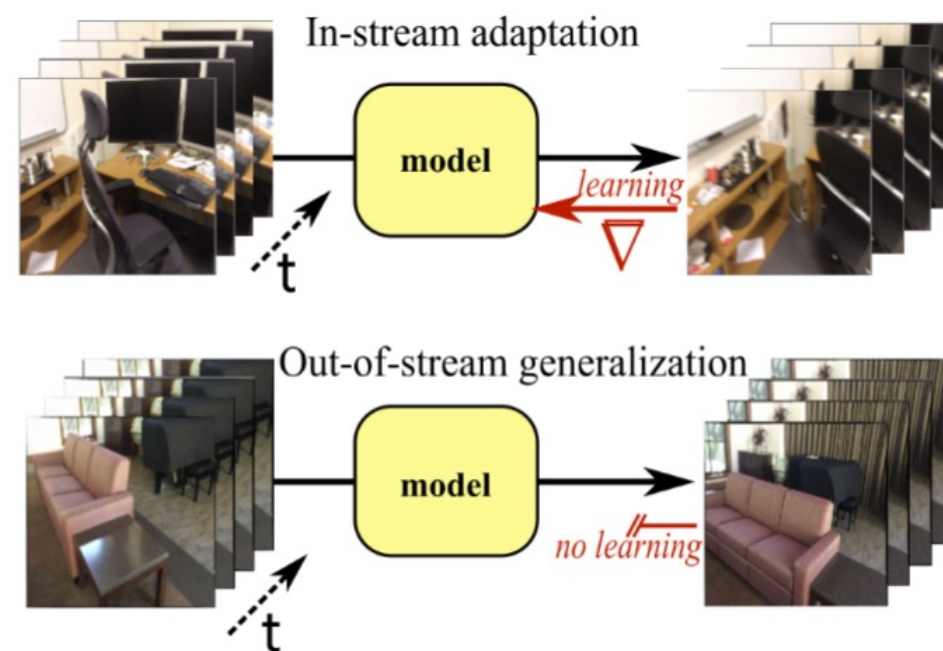
Decorrelate gradients between consecutive steps to stabilize training

$$g_t^\perp = g_t - \frac{\langle g_t \cdot g_{t-1} \rangle}{|g_{t-1}|^2} g_{t-1}$$

T. Han, et al., Learning from Streaming Video with Orthogonal Gradients, CVPR 2025.

Evaluation Protocol

- We follow previous works on streaming video learning (Carrera et al., Han et al.) to evaluate COAD
- Protocol comprises 3 phases:
 1. Offline pretraining
 2. Continuous training on *in-stream* data
 3. Testing on *out-of-stream* data

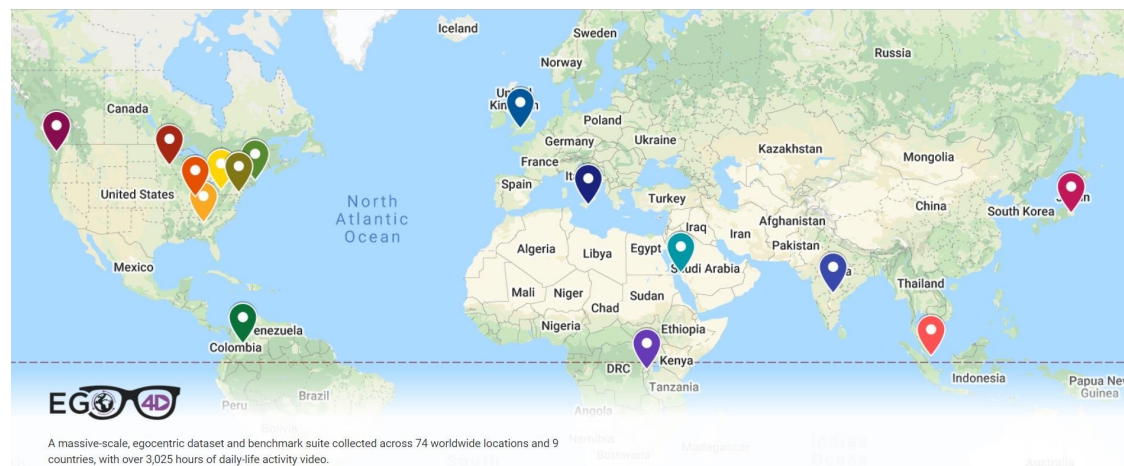


J. Carrera, et al., Learning from One Continuous Video Stream, CVPR 2024.

T. Han, et al., Learning from Streaming Video with Orthogonal Gradients, CVPR 2025.

The Ego4D Dataset

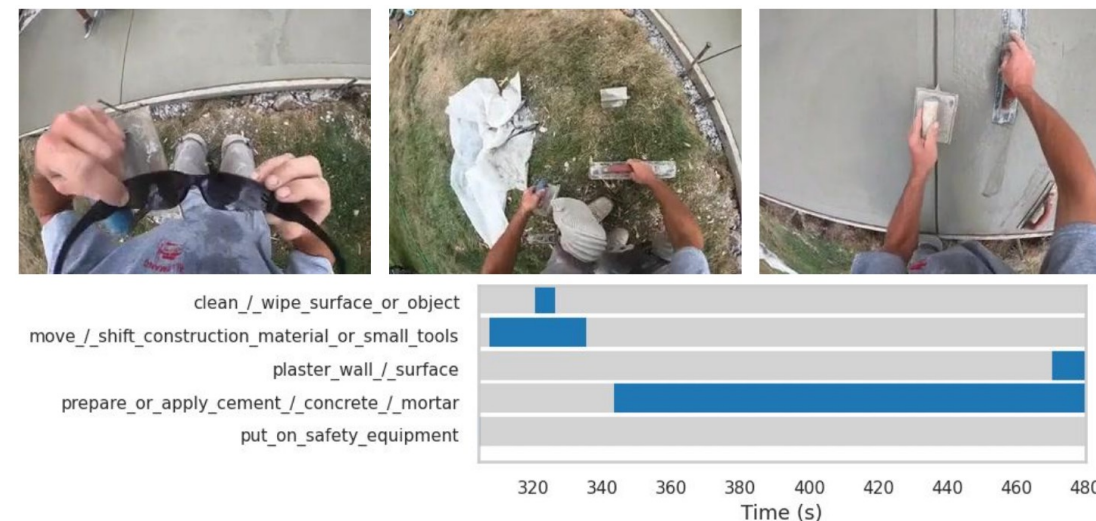
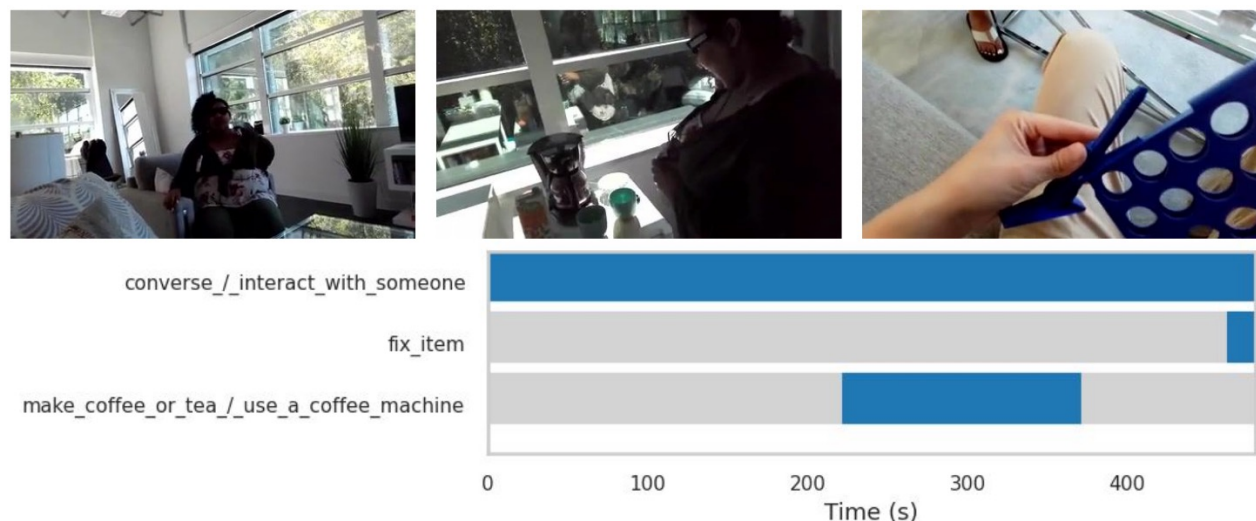
- Massive-scale egocentric dataset
- 3670 hours of recording
- 74 worldwide locations
- Diverse settings and actions



K. Grauman, et al., Ego4D: Around the World in 3,000 Hours of Egocentric Video, CVPR 2022.

Ego-OAD Benchmark

- We curate Ego-OAD from the Ego4D Moments Queries dataset
- 87 action classes with 22,991 annotated instances spanning 263h of video
- Realistic and general benchmark for OAD in egocentric vision



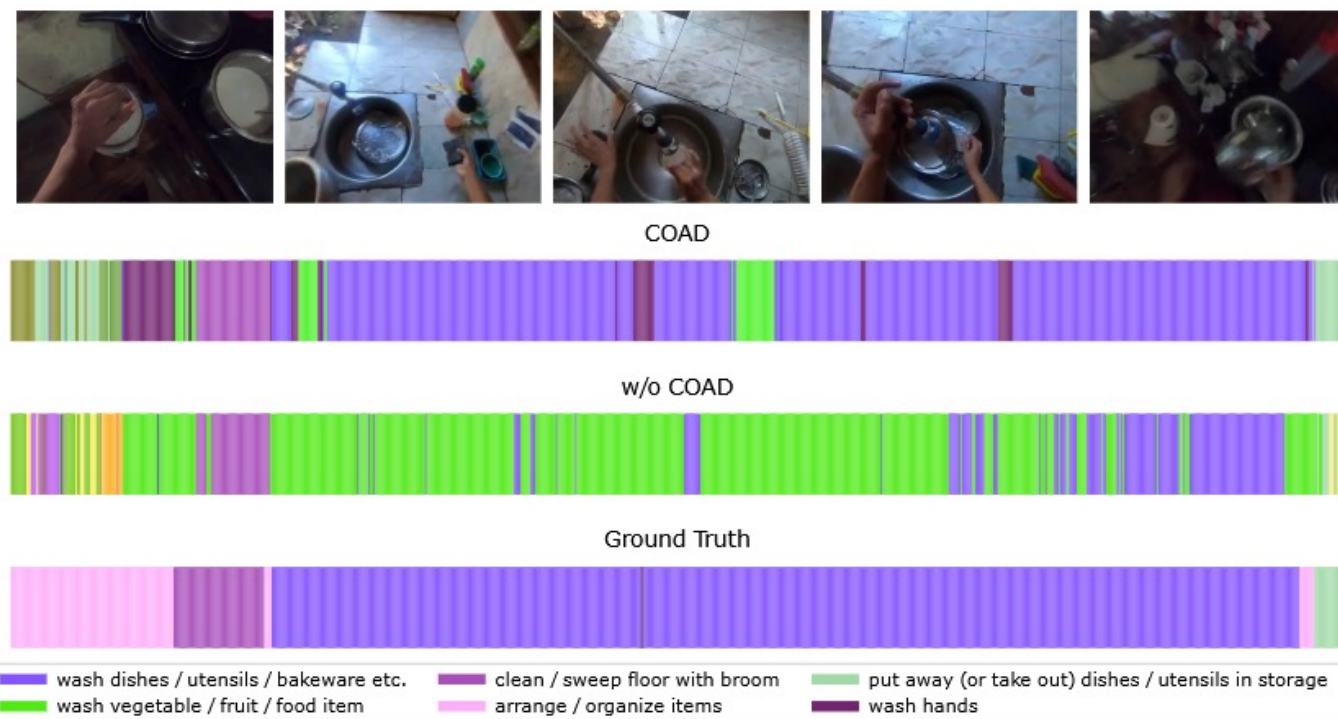
COAD Results

- We use TimeSformer backbone under both exo (*Kinetics400*) and ego (*EgoVLP*) pretraining
- COAD enables effective learning from streaming video, **improving out-of-stream generalization** while maintaining equivalent in-stream adaptation performance

Backbone	Method	Adaptation	Out-of-stream				In-stream	
			mAP		Top-5 Recall		mAP	Top-5 Recall
Ego	Pretrained Only	X	20.1	–	69.1	–	24.1	73.3
	w/o COAD	✓	25.5	+5.4	71.6	+2.5	39.0	86.7
	COAD	✓	26.0	+5.9	76.0	+6.9	36.8	89.3
	Offline Training	–	30.0		82.9		–	–
Exo	Pretrained Only	X	15.8	–	55.5	–	15.3	57.5
	w/o COAD	✓	19.0	+3.2	57.8	+2.3	31.0	76.2
	COAD	✓	20.5	+4.7	62.0	+6.5	31.0	80.0
	Offline Training	–	24.3		70.9		–	–

COAD Results

- COAD enables effective learning from streaming video, **improving out-of-stream generalization** while maintaining equivalent in-stream adaptation performance



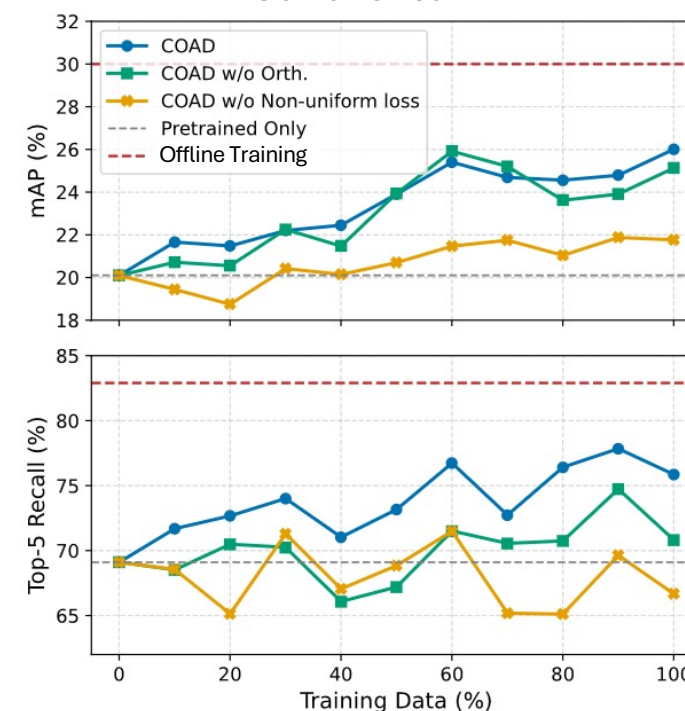
COAD Ablations

- All components of the proposed training scheme contribute to the final performance

OUT / IN -stream evaluation

State Cont.	Orth. Grad.	Non-uniform Loss	Adapt	mAP	Top-5 Recall
✓	✓	✓	✓	26.0 / 36.8	76.0 / 89.3
✓	✓	✗	✓	21.8 / 42.4	67.7 / 88.0
✓	✗	✓	✓	25.3 / 37.4	71.5 / 87.9
✗	✓	✓	✓	25.9 / 36.7	75.8 / 89.2
✗	✗	✗	✓	25.5 / 39.0	71.6 / 86.7
✗	✗	✗	✗	20.1 / 24.1	69.1 / 73.3

Out-of-stream



Conclusions

- We introduced **Continuous Online Action Detection (COAD)** for post-deployment training of OAD models
- We curated **Ego-OAD**, a large-scale and diverse benchmark for egocentric Online Action Detection
- Our COAD training strategy improves performance (**up to +4.5% Top-5 Recall** out-of-stream) over baseline continuous training

Continuous Online Action Detection from Egocentric Videos

Riccardo Santambrogio¹, Chiara Plizzari², Francesca

Palermo, Simone Mentasti¹, Matteo Matteucci¹

¹ Politecnico di Milano, Milan, Italy

² Università Bocconi, Milan, Italy

contact: riccardo.santambrogio@polimi.it

ICPR 2026