SFU

BOREALIS AI

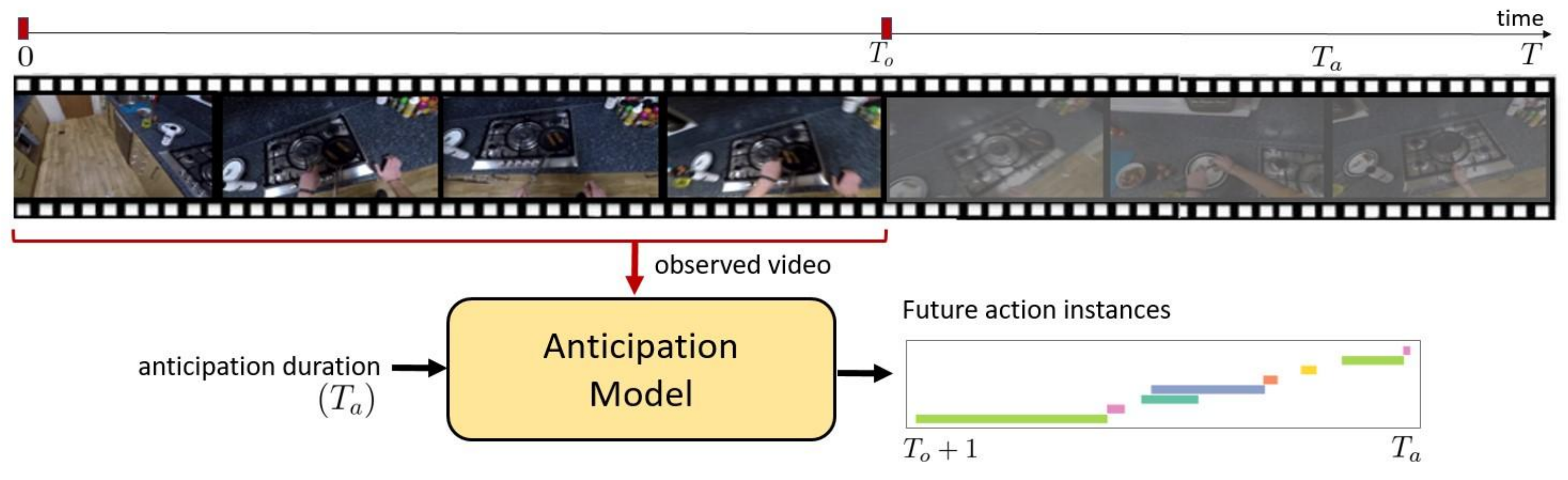
Rethinking Learning Approaches for Long Term Action Anticipation

Megha Nawhal, Akash Abdu Jyothi, Greg Mori

meghanawhal.github.io/projects/anticipatr.html

Long Term Action Anticipation

Given a partial video and an anticipation duration, we predict a set of future action instances over the given duration.



Why predict set of instances?

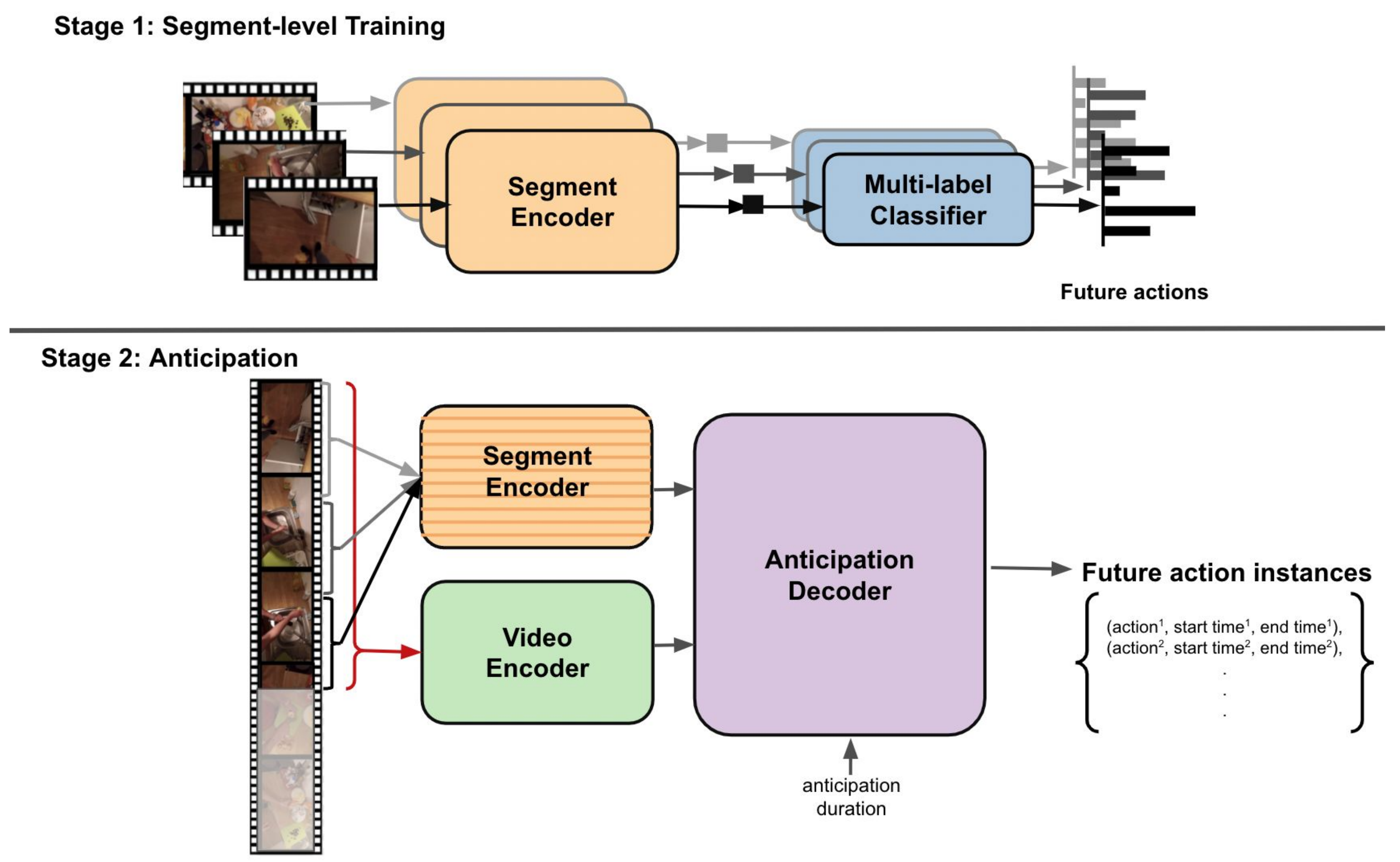
- *Generic* for all types of anticipation outputs (action sequence, labels only)
- *Single-shot* prediction for all timestamps over the given anticipation duration

Main Idea: Use video-level & segment-level representations to predict actions

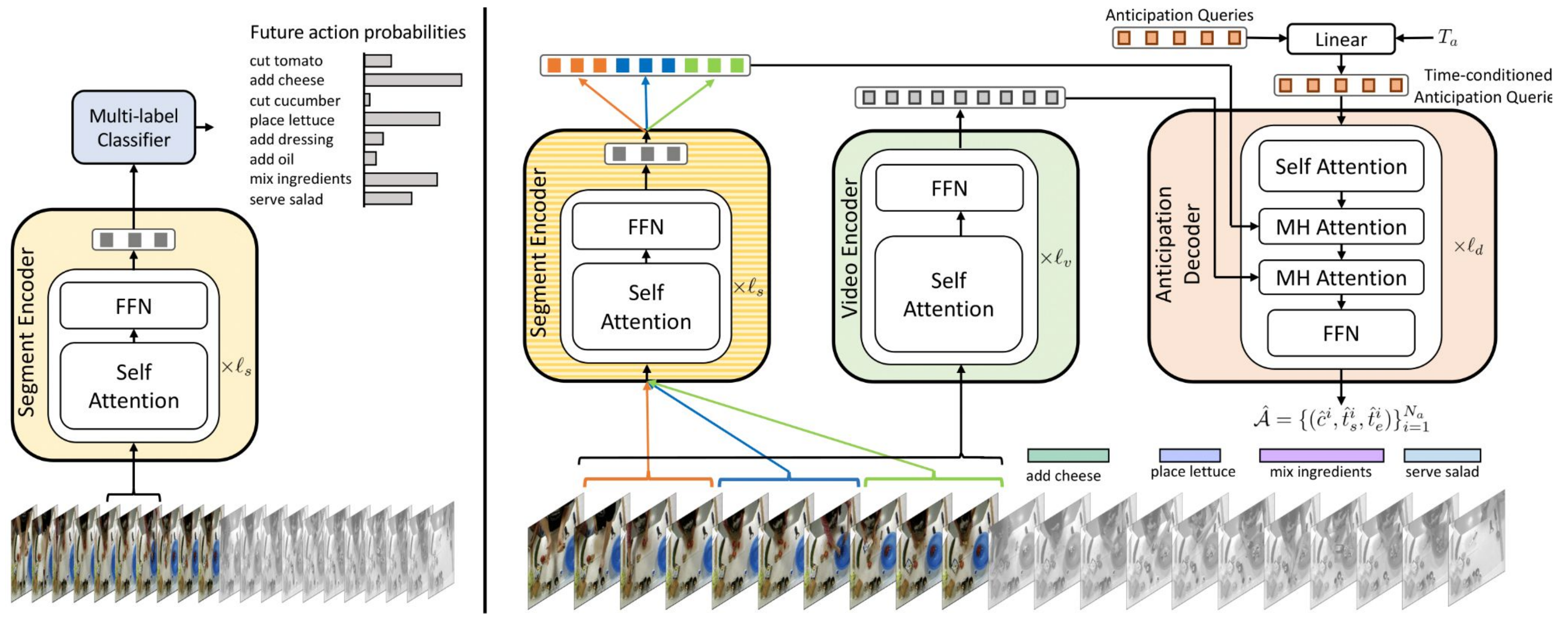
- Context of the ongoing activity in the video ⇒ video-level representations
- Cross-activity information from individual actions ⇒ segment-level representations

Anticipation Transformer (ANTICIPATR)

Two-stage learning approach to first learn segment-level representations and then use them along with video-level representations



Encoder-decoder model for long term action anticipation for a given anticipation duration

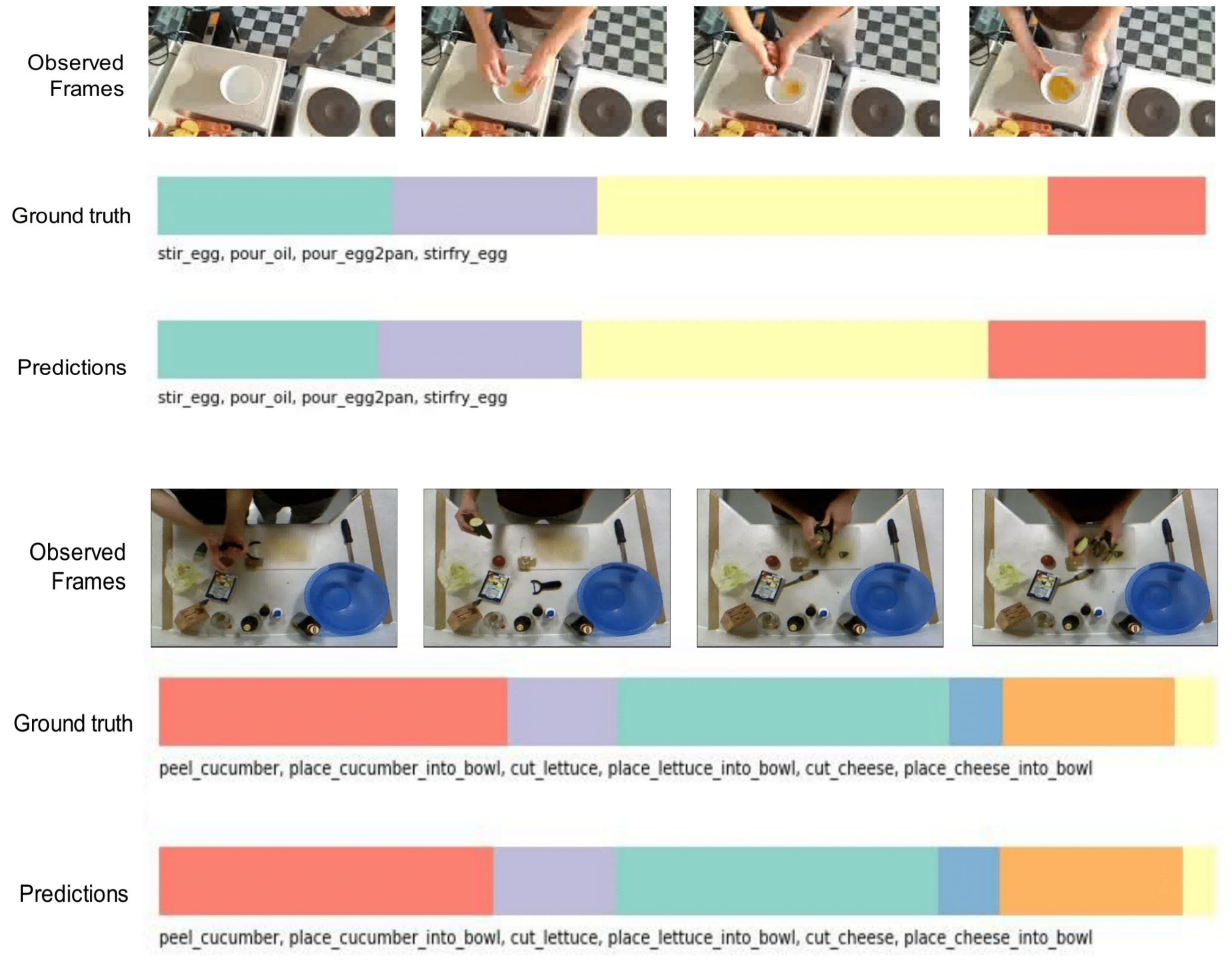


Results

State-of-the-art performance on benchmarks: Breakfast, 50 Salads, Epic-Kitchens-55, EGTEA+

	Breakfast				50 Salads			
Anticipation duration →	10%	20%	30%	50%	10%	20%	30%	50%
RNN	18.1	17.2	15.9	15.8	30.1	25.4	18.7	13.5
CNN	17.9	16.3	15.3	14.5	21.2	19.0	15.9	9.8
Ke *et al.*, CVPR'19	18.4	17.2	16.4	15.8	32.5	27.6	21.3	15.9
Sener *et al.*, ECCV'20	24.2	21.1	20.0	18.1	25.5	19.9	18.2	15.1
ANTICIPATR (OURS)	**37.4**	**32.0**	**30.3**	**28.6**	**41.1**	**35.0**	**27.6**	**27.3**

Method	Epic-Kitchens-55			EGTEA+		
	ALL	FREQ	RARE	ALL	FREQ	RARE
RNN	32.6	52.3	23.3	70.4	76.6	54.3
I3D	32.7	53.3	23.0	72.1	79.3	53.3
ActionVLAD	29.8	53.5	18.6	73.3	79.0	58.6
Timeception	35.6	55.9	26.1	74.1	79.7	59.7
EGO-TOPO	38.0	56.9	**29.2**	73.5	80.7	54.7
ANTICIPATR(OURS)	**39.1**	**58.1**	29.1	**76.8**	**83.3**	**55.1**



ECCV
EUROPEAN CONFERENCE ON COMPUTER VISION
TEL AVIV 2022
October 23-27, 2022, Tel Aviv