



# Enhance Action Recognition Modeling Long-Term Interactions to

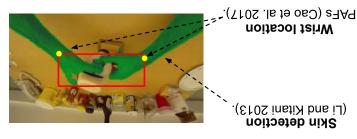
Alejandro Cartas, Petia Radeva, and Mariella Dimiccoli



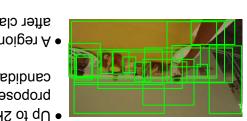


## Approach

Primary action region: Hands



after classification. A region is chosen



#### Method GTEA 61\* GTEA 71 GAZE 25\* GAZE 40\* GAZE+ **Comparative Results**

Turn on burner Cut tomato with knife Put lettuce on plate Put plas

Open Fridge Put plast

True Prediction Examples

Open peanut Open peanut Make sandwich Take Cheese

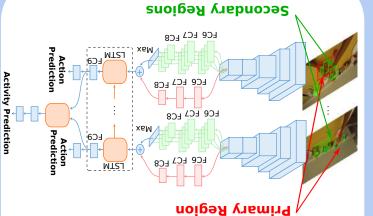
False Prediction Examples

-	-	-	41.87	6£.67	LSTA Sudhakaran et al. [2019b]
£1.09	-	-	LL	69.77	Sudhakaran and Lanz [2018]
04.88	43.42	04.29	73.24	8.27	[8102] .ls fe sM
-	-	-	91.99	41.47	LSTA-RGB Sudhakaran et al. [2019b]
55.25	-	-	62.73	97.78	Femporal Segments Network Wang et al. [2016]
96.65	52.75	85.28	72.95	69.07	Ours (Hierarchical LSTM)
14.82	9t'6t	68.89	40.17	£8.e9	(MTS_level t) shuO
52.75	47.25	¢6 <sup>.</sup> 99	⊅Z.4a	76.88	Ours (frame level)
£8.64	9Z.04	43.44	96.8₽	∠9 <sup>-</sup> †9	CNN Baseline †
	01. 331/0	07.7710		10 4710	DOUISON.

# Up to 2K regions are Secondary action regions



# Primary Region Secondary action regions



# Motivation and goal

Action & activities have spatio-temporal

	regi		
Action Close Mayonnaise	Prim acti	<b>noi</b> esisnno	] <b>Act</b> γεΜ τυο9
ndwich	coufext:		

action region

Secondary

### Key contribution

their tempors evolutio	+	Egocentric contextual cues (Hands+Objects)



Gaze

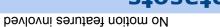


(Li et al., 2015)

Gaze+

и әшел





### Datasets

Frame 1



L	-	۷	#Activities
07	S2 & 40	17.8.18	#Actions
9	۲l	Þ	#People
75	۷۱	21	soəpi∧#
CONTRACTOR OF THE	AND DESCRIPTION OF THE PERSON NAMED IN COLUMN TWO IS NOT THE PERSON NAMED IN COLUMN TWO IS NAMED IN COLUM	•	•