# ACTIONS-TO-ACTION: INDUCTIVE ATTENTION FOR EGOCENTRIC VIDEO ACTION ANTICIPATION – SUPPLEMENTARY MATERIAL

#### Anonymous authors

Paper under double-blind review

# A INPUT PREPROCESSING

For the TSN baseline utilized in the three action anticipation benchmarks shown in Tables 1, 3, and 4 of the main paper, we sourced the pre-extracted features from the official EPIC-Kitchens websites<sup>1</sup>. TSN employs the BN-Inception backbone, extracting RGB frame input of size (456, 256). The implementation and pretrained weights of Swin and ConvNeXt, used to present results in Tables 1, 2, 3, and 4, are adopted from the open-source implementation<sup>2</sup>. The frame inputs for Swin and ConvNeXt are of size (224, 224), with values rescaled to the range [-1, 1].

## **B** INDUCTIVE ATTENTION ANALYSIS

Table 7 provides an analysis of the components comprising our novel inductive attention mechanism when integrated with Swin, ConvNeXt, and TSN backbones on the EPIC-Kitchens-100 datasets. This inductive attention is characterized by two key enhancements: (a) the utilization of prior predictions as the attention query Q and keys K, and (b) the expansion of the attention context window from first-order to higher-order historical states. Our results demonstrate that this advanced inductive attention consistently elevates performance across various backbones, as evidenced by improvements in overall, unseen, and tail action Mean Top-5 Recall (MT5R) metrics.

Table 7: Analysis of models with and without the implementation of inductive attention, focusing on their performance in overall, unseen, and tail Mean Top-5 Recall (MT5R) metrics within the EPIC-Kitchens-100 dataset.

Proposed Change	MT5R (%)	$\Delta$ MT5R (%)
With Inductive Attention (Swin-IAM)	18.1/16.6/17.6	-/-/-
- (a) Prior predictions serve as attention query (Q) and keys (K)	17.2/15.9/17.4	-0.9/-0.7/-0.2
- (b) Extend to higher-order (S=1 to S=30)	16.8/15.8/17.0	-0.4/-0.1/-0.4
Without Inductive Attention	16.8/15.8/17.0	-1.3/-0.8/-0.6
Inductive Attention on (ConvNeXt-IAM)	17.6/12.0/17.1	_/_/-
- (a) Prior predictions serve as attention query (Q) and keys (K)	16.9/11.9/16.4	-0.8/-0.1/-0.7
- (b) Extend to higher-order (S=1 to S=30)	16.7/11.2/16.2	-0.2/-0.7/-0.2
Without Inductive Attention	16.7/11.2/16.2	-1.0/-0.8/-0.9
Inductive Attention on (TSN-IAM)	17.5/11.9/16.8	_/_/-
- (a) Prior predictions serve as attention query (Q) and keys (K)	17.0/11.1/16.4	-0.5/-0.8/-0.4
- (b) Extend to higher-order (S=1 to S=30)	16.4/10.9/16.0	-0.6/-0.2/-0.4
Without Inductive Attention	16.4/10.9/16.0	-1.1/-1.0/-0.8

We expand our analysis to identify the most significant positive and negative inductive priors in relation to the predictions of each model. This identification is accomplished through the calculation of partial gradients in relation to the attention keys. These keys are conditioned by previous predictions, which in turn maximize the prediction being assigned. Such methodology effectively reveals the sensitivity of the priors. It is crucial to highlight that this depth of analysis is a distinctive

<sup>&</sup>lt;sup>1</sup>https://github.com/epic-kitchens/C3-Action-Anticipation

<sup>&</sup>lt;sup>2</sup>https://github.com/huggingface/pytorch-image-models, v0.5.4

			5 (5)		3	3 6	100
Tup-5 Positive squeeze cloth wash sink unroll cloth gather robbish put sink Tup-5 Negative apply bowl gather broccoli insert beargreen close freezer	Tup-5 Positive wash cloth wash top unroll cloth dry hand wash ping Tup-5 Negative letgo sponge mik lettuce eat pork check box soak plate	Top-5 Positive wash cloth dry hand wash hand wash hand wash top Top-5 Negative gather meat gather broccoli eat broccoli wash food insert pandoust	Top-5 Positive squeeze cloth umoil cloth wash plug puil cup hang hand Top-5 Negative sort broccoli eat pork coat oil brush dough eat broccoli	Top-5 Positive wash cloth wash sink dry hand take cloth put cloth <b>Top-5 Negative</b> throw pear put sheets pat omelette unroll omelette	Top-5 Positive squeeze cloth wash plog wash cloth pull cop hang hand Top-5 Negative roll omelette wrap box e at broecoli turn ring:onion pat noolle	Top-5 Positive squeeze cloth wask cloth hang hand pul cup squeeze sponge Top-5 Negative roll lemon pat omelette eat pork turn-on liquid:washing roll omelette	wash sink squeeze cloth wash top wash cloth squeeze sponge
Top-5 Positive wash top wash cloth throw food take cloth throw rubbish <b>Top-5 Negative</b> shake rubbish wash ping shake straw dry cloth hang hand	Tup-5 Positive squeeze cloth um01 cloth wash oven shake hand gather rubbish Tup-5 Negative peel squash insert nut take nut cut apple look bag	Top-5 Positive squeeze cloth wash sink wash floor drink beer remove rubbish Top-5 Negative insert banana peel sausage unroll onelette cut banana hang hand	Top-5 Positive wash sink wash cloth squeeze sponge wash top wash sponge Top-5 Negative shake peach squeeze garlic carry bag wash straw squeeze caper	Top-5 Positive squeeze songe wash top wash oven put sponge <b>Top-5 Negative</b> dry scissors search salt turn-on water throw squash hang cup	Tup-5 Positive wash sink dry hand take cloth wash oven gather rubbish Top-5 Negative fill liquid hang hand squeeze can shake liquid shake courgette	Top-5 Positive wash sink dry hand wash hand Top-5 Negative adjust plate dry sponge move funnel entpty container filter plate	wash sink squeeze cloth wash top wash cloth squeeze sponge
Top-5 Positive wash top wash con- supeers club supeers sponge take club Top-5 Negative shake staw shake subbish throw kale hang mad move stand	Top-5 Positive wash top squeeze sponge wash cloth wash sponge take cloth Top-5 Negative shake straw shake straw shake caper wash plug pat napkin	Top-5 Positive wash top unr01 cloth wash vor dry hand squeeze sponge Top-5 Negative shake straw brash sink adjust straw shake caper take alarm	Top-5 Positive wash sink ump iclosed positive states and output in the second wash hand dry hand Top-5 Negative threw squash small candle let-go sponge soak plate shake caper	Top-5 Positive squeeze sponge squeeze cloth wash sponge tum-off tap take cloth Top-5 Negative shake straw throw squash shake rubbish wash straw drop sponge	Top-5 Positive wash top squeeze sponge wash cloud dy hand take cloth Top-5 Segutive shake straw shake straw shake straw shake straw shake straw shake straw shake straw shake straw	Top-5 Positive wash top wash cloth squeeze sponge take cloth dry hand <b>Top-5 Negative</b> hang hand move coffee shake straw squeeze caper shake rubbish	wash sink squeeze cloth <b>wash clot</b> h wash cloth squeeze sponge
Top-5 Positive wash sink wash cloth wash sponge throw rubbish move rack:drying Top-5 Negative cut corn pat napkin pour yeast drop sponge pour beer	Top-5 Positive wash top unroll cloth wash oven dry hand squeeze water Top-5 Negative press glass put dishwasher remove pasta filter plate look can	Top-5 Positive squeeze cloth wash sink unroll cloth wash floor Top-5 Negative put jujcer serve pan move paste press drawer open rosemary	Top-5 Positive wash sink wash top squeece sponge wash hand dry hand Top-5 Negative shake straw shake caper squeece caper wash straw filter plate	Top-5 Positive squeeze cloth put sink unroll cloth squeeze water pull cup Top-5 Negative turn ringennion roll lemon close cooker-slow put oven insert scale	Top-5 Positive wash top squeeze sponge wash oven dy hand take cloth Top-5 Negative shake straw shake caper shake cuper shake cuper shake traw	Top-5 Positive squeeze cloth hang hand dry hand wash top pour liquid-washing Top-5 Negative throw pasta scrub cutlery drop sponge crush leaf throw squash	wash sink squeeze cloth wash top <b>wash cloth</b> squeeze sponge
Top-5 Positive squeeze cloth wash sink dry hand throw rubbish squeeze sponge Top-5 Negative move opener:bottle scrub cutpy adjust fork scoop spoon adjust tray	Top-5 Positive wash sink squeeze sponge wash top take eloth wash sponge Top-5 Negative shake rubbish shake caper dry eloth squeeze can shake straw	Top-5 Positive wash sink wash top wash oven wash sponge squeeze sponge Top-5 Negative shake caper throw cork filter plate pross glass check kettle	Top-5 Positive squeeze cloth wash plug unroll cloth squeeze water put sink Top-5 Negative insert beer put juicer coat cill look can move paste	Top-5 Positive squeeze cloth wash sink put sink unroll cloth Top-5 Negative cost oil insert beer scrape spreads pour extractivalilla shake blender	Top-5 Positive squeeze cloth wash oven wash top gather rubbish Top-5 Negative open tray choose pear shake caper pull jug hang cup	Top-5 Positive squeeze cloth wash sink squeeze sponge wash oven wash plug Top-5 Negative remove pepper put lemon scoop pork shake board:chopping throw onion:spring	wash sink squeeze cloth wash top wash cloth squeeze sponge

Figure 7: Illustration of a sample with correct model prediction. The top row displays the final eight frames of the observed video sequence. Directly beneath the last frame, top-5 prediction of inductive attention model is presented, with the ground truth highlighted in green. Subsequently, each of the following rows showcases the topmost positive and negative priors, computed from our inductive attention, corresponding to each of the top-5 predictions.

feature of the inductive attention mechanism we propose. This uniqueness stems from our method of explicitly representing both the query and the keys within the action space context. Thereby, this approach not only enhances the interpretability and precision of the model but also contributes to making the decision-making process of the model more transparent. This analysis is based on the Swin-IAM model on EPIC-Kitchens-100 samples.

Figure 7 presents an example of the top prior predictions contributing to each model prediction. The correct answer ranks second among the top-5 predicted actions. In this sample, a washing action is identified, followed by a squeezing action. This sequence is further corroborated by the priors contributing most to the predictions. Conversely, the topmost negative priors appear noticeably irrelevant to the described action.

An intriguing observation is the sensitive prior action "wash sink", which correlates with the targeted prediction "squeeze cloth". This correlation indicates a potential causal relationship, where "wash sink" precedes "squeeze cloth". While this pattern does not universally apply in our model analysis, it uncovers a promising avenue for future research, exploring the potential causal relationships in model predictions.

-	A set		CEAR A	(B)	(PA)	A CAL		
1	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Rank 28 of 3806
(	apply olive	take pizza	take pizza	put pizza	put pizza	put pizza	apply olive	Marik 20 07 5000
	put mushroom	apply olive	apply olive	flatten dough	take pizza	apply olive	cut pizza	
	put potato	apply basil	apply basil	apply olive	apply basil	open pizza	remove pizza	
	throw rubbish	take basil	check pizza taka basil	Told bag	apply olive take basil	put truit	put pin:rolling	put pizza
	unow skin	open oven	take bash	par potato	take basi	crose on	parioou	take pizza
	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	cut pizza
	look bottle	check board:chopping	look sauce	turn-on dishwasher	eat bread	pat spatula	look can	remove pizza
	sort tork	search book	remove jar	sort fork	press cornander take candle	insert ring:onion	take backpack	
	put parsley	insert ring:onion	insert holder	scrub scissors	open corn	check pie	insert cap	
	insert oven	insert cloth	lift dishwasher	look box	look box	put pie	pull box	
1	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	
(	put pizza	put pizza	insert pizza	flatten dough	insert pizza	take pizza	put pizza	
	take pizza	insert pizza chack pizza	remove pizza	cut pizza	flatten dough	insert pizza	take pizza	
	close bin	open pizza	flatten dough	remove pizza	apply olive	put sausage	drink beer	
	open pizza	wrap plate	open oven	put fruit	put fruit	remove pizza	open pizza	take pizza
								insert pizza
	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	
	pat garlic	press onion	put pear	divide pizza	look bottle	throw oil	sprinkle tomato	
	press onion	remove stalk	take toaster	unroll clothes	sort broccoli	throw butter	close oregano	
	check ginger	apply onion	eat broccoli	move fridge	pour bean:green	take pear	put pear	
$\sim$	check onion	sort onion	remove cucumber	dry food	brush dough	close paper	scrape skin	
/	Top. 5 Positive	Ton 5 Positive	Top. 5 Positive	Top. 5 Positive	Top. 5 Positive	Ton-5 Positive	Ton 5 Positive	
1	apply olive	take pizza	put pizza	flatten dough	remove pizza	put pizza	apply olive	
	cut pizza	insert pizza	insert pizza	insert pizza	insert pizza	open pizza	take pizza	
	remove pizza	check pizza	check pizza	take pizza	take pizza	close bin	cut pizza	
	apply basil	open pizza	cut pizza	fold bag	cut pizza	drink beer	put fruit	
	take aubergine	open oven	open pizza	divide dough	appry basi	remove salad	put tood	take pizza
	Top-5 Negative	Top-5 Negative	Top-5 Negative	<b>Top-5 Negative</b>	Top-5 Negative	Top-5 Negative	Top-5 Negative	<u>insert pizza</u>
	squeeze garlic	cut corn	cut corn	search plate	cut corn	cut com	check board:chopping	remove pizza
	pat burger	press onion	press onion	lift foil	attach tray	close pepper	unroll clothes	
	squeeze sausage mix sausage	sort onion	sort onion	sort pizza	rub sauce	insert shell:egg	scrane salmon	
	sort onion	squeeze garlic	remove stalk	hold tray	shake tomato	press onion	insert oven	
1	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	
(	put pizza	cut pizza	take pizza	put pizza	put pizza	take pizza	apply olive	
	take pizza	remove pizza	apply olive	insert pizza	insert pizza	insert pizza	take pizza	
	cut pizza	apply basil	flatten dough	take pizza	put tray	check pizza	put fruit	
1	close bin	take basil	apply dough	cut pizza	take dough	remove pizza	put potato	take pizza
		The States	The Chinese Street	The Streetler	The Physical Section	The Chinese Street	The Chinese Street	insert pizza
	rop-5 Negative	take backnack	sort plate	press onion	norss onion	empty container	turn ring:onion	remove pizza
	cut com	throw tea	insert foil	sort onion	turn-on liquid:washing	divide tomato	unroll clothes	
	check onion	cook kettle	rip wrap:plastic	remove stalk	divide broccoli	carry bag	sort fork	
	attach tray	look sauce	hold towel	apply onion	cut corn	pat burger	scrape salmon	
~	pat garne	unroll clothes	таке васкраск	insert grass:iemon	remove stark	empty pan:dust	put noor	
1	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	Top-5 Positive	
	insert pizza	put pizza	take pizza	put pizza	insert pizza	insert pizza	put pizza	
	put pizza	apply olive	apply olive	divide dough	take pizza	flatten dough	put food	
1	put potato	insert pizza	flatten dough	remove pizza	fold bag	apply basil	take tongs	nut nizza
	throw skin	put fruit	put fruit	open pizza	put fruit	put fruit	open oven	take pizza
	Ton-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	Top-5 Negative	insert pizza
	empty pan:dust	cook oil	adjust pan	open rubbish	open rosemary	scrape salmon	divide bread	remove pizza
1	close paper	cook pan	turn ring:onion	press onion	close oregano	remove bean:green	insert cloth	
	insert foil	pour omelette	brush dough	remove stalk	apply leaf	remove rice	wrap cake	
	remove floor check milk	search oven	sort plate	remove pepper	wash aubergine wait sausage	open kale	unroll paper	
~	CHECK IIIIK	Search Oren	acte pinte	remove popper	man annouge	орон ване	amon baber	

Figure 8: Illustration of a sample with incorrect model prediction. The top row displays the final eight frames of the observed video sequence. Directly beneath the last frame, top-5 prediction of inductive attention model is presented, with the ground truth highlighted in green. Subsequently, each of the following rows showcases the topmost positive and negative priors, computed from our inductive attention, corresponding to each of the top-5 predictions.

In another example, illustrated in Figure 8, the predictions of the model are notably erroneous, with the correct prediction ranking only 29th out of 3806. A closer examination of inductive prior of our model reveals a significant bias towards the noun "pizza", as seen in the video. From a human perspective, accurately interpreting this video hinges on recognizing the action of transferring a pizza from paper to a plate. This highlights a key challenge in action anticipation: discerning the underlying intention amidst distracting elements. It is noteworthy that the actual ground truth of this sample is unrelated to the pizza, but instead concerns the paper. Despite the incorrect prediction in this instance, the model output remains plausible and aligns with the inductive attention's learned priors for actions.

The per-frame top-5 action anticipation for these two cases is presented in Figure 9 (a) for the correct example and Figure 10 (a) for the incorrect one.

# C QUALITATIVE ANALYSIS

We derive success and failure cases of our model from the EPIC-Kitchens-100, EPIC-Kitchens-55, and EGTEA Gaze+ datasets for qualitative analysis. Four video clips, labeled (a) to (d), illustrate

instances where the ground truth is within the top-5 predictions of model at the last frame. Additionally, another four examples, labeled (e) to (h), depict cases where the model failed to accurately anticipate the target action. For each video sample, we present the last eight frames along with the corresponding top-5 predictions per frame. The figures are most effectively viewed horizontally and with zoom-in.

## C.1 EPIC-KITCHENS-100

Figures 9 and 10 display samples from the EK100 dataset. The model demonstrates proficiency in refining predictions to specific verbs of activities based on context. For instance, in video (**a**), predictions converge to verbs "*wash*" or "*squeeze*". Moreover, video (**b**) initially recognizes "*milk*" and subsequently shifts to the plausible resultant object "*cereal*". Video (**c**) contains minimal movements, with the model maintaining relevant predictions but facing challenges in anticipating the object until the final moment. Video (**d**) represents a challenging scenario with ambiguous subject intention throughout the observation.

Figure 10 uncovers instances where our model encountered difficulties. In video (e), despite accurately identifying the verb, an incorrect noun "*pizza*" is consistently predicted. Other cases reflect mispredictions due to inadequate visual observations (e.g., (f)) or an abundance of potential objects (e.g., (g), (h)).

### C.2 EPIC-KITCHENS-55

Figure 11 showcases successful cases, noting that the target noun objects in videos (**a**) and (**b**) are not visible. Nonetheless, the model accurately infers based on past predictions propagated by inductive attention. Additionally, video (**c**) demonstrates the model's capacity to identify the object "*plate*" appearing only in the initial three frames. Video (**d**) maintains predictions pertinent to cooking pasta.

Conversely, sample (e) lacks any indication of the anticipated object "*tofu*", predicting "*container*" instead. Mispredicted nouns and verbs are also observed in videos (f), (g), and (h).

### C.3 EGTEA GAZE+

Figure 13 highlights video (b), demonstrating correct prediction through awareness of *tomato:container* visible only in the first two frames. Video (c) involves subtle movements about the *bread:container*, depicted as being taken out and replaced in the final three frames.

Figure 14 reveals incorrect predictions, primarily attributed to insufficient observable evidence (e.g., (e) and (f)) or an array of possibilities (e.g., (g), with notable confidence on verbs), or a combination as in (h).

I	(a)		(q)		<b>(</b> )		(q)	ا <sup>ر</sup> د د ــ ر
	wash sink	squeeze cloth wash top wash cloth throw rubbish		pour milk open milk insert milk close milk put milk		open salt mix food put tongs put spoon mix pasta		wash spatula put spatula wash board:chopping wash pan scoop food
	wash sink	squeeze cloth wash cloth wash top wnroll cloth		shake milk insert cap insert milk open milk pour milk		mix food mix sauce open salt put tongs put spoon		wash spatula wash pan wash sponge put spatula pour liquid:washing
	wash sink	squeeze cloth wash cloth unroll cloth wash top	-	insert cap insert milk take milk take cereal put milk		mix food mix sauce put spoon pour sauce put ladle		wash spatula wash pan pour liquid:washing move tap wash sponge
	squeeze cloth	wash sink wash cloth squeeze sponge unroll cloth	4	insert milk put milk close milk shake milk pour milk		mix sauce mix food mix pasta put spoon scoop sauce		wash spatula wash pan wash ladle put spatula squeeze sponge
	wash sink	squeeze cloth squeeze sponge wash cloth put sink		insert milk shake milk put milk take cereal take milk		mix food put sauce put spoon mix sauce scoop sauce		wash spatula move tap pour liquid:washing wash sponge wash hand
	wash sink	squeeze cloth wash cloth wash top squeeze sponge		take cereal move floor put cereal insert cereal shake milk		mix food mix sauce put spoon scoop sauce mix pasta		throw food wash top close bin open bin pour liquid:washing
	wash sink	squeeze cloth wash cloth wash top dry hand		take cereal put cereal insert cereal put squash eat squash		mix sauce mix food put spoon scoop sauce mix pasta		throw food close bin open bin dry hand wash top
Time	wash sink	squeeze cloth wash top wash cloth squeeze sponge		put squash put cereal put mushroom insert cereal take cereal		mix food put spatula put tongs pour squash		close bin throw food throw bin open bin throw rubbish

Figure 9: Four video clips from the **EPIC-Kitchens-100** validation set, illustrating **correct** predictions, are presented in the figure. Each clip displays the last eight frames along with the corresponding top-5 action anticipations at  $\tau_a = 1s$ . The ground truth is prominently highlighted in bold green.



Figure 10: The figure displays four video clips from the **EPIC-Kitchens-100** validation set, showcasing **incorrect** predictions. Each clip features the last eight frames and the corresponding top-5 action anticipations at  $\tau_a = 1s$ . The ground truth is revealed in the final frame, with the correct verb and noun of the action highlighted in bold green.



Figure 11: The figure displays four video clips from the **EPIC-Kitchens-55** validation set, illustrating **correct** predictions. Each clip features the last eight frames along with the corresponding top-5 action anticipations at  $\tau_a = 1s$ . The ground truth is prominently highlighted in bold green.



Figure 12: The figure presents four video clips from the **EPIC-Kitchens-55** validation set, showcasing **incorrect** predictions. Each clip depicts the last eight frames and the associated top-5 action anticipations at  $\tau_a = 1s$ . The ground truth is revealed at the end, with the correct verb and noun of each action highlighted in green.

**()** 

Ξ

£

e



Figure 13: The figure displays four video clips from the **EGTEA Gaze+** validation set, highlighting **correct** predictions. Each clip features the last eight frames along with the corresponding top-5 action anticipations at  $\tau_a = 0.5s$ . The ground truth is emphasized in bold green.

Time	Pour water Pour water Turn off faucet Wash pot Move Around pot Put pot (GT: Take pan)	Take tomato:container Take tomato:container Put tomato:container Take grocery:bag ainer Take tomato (GT: Put lettuce)	the put sessoning:container Take easting:container Take cooking:utensil Take pan (GT: Take bowl)	Pour condiment trainertower condiment trainertower condiment trainertower condiment for a dit container but dit container Mix pasta (GT: but container)
	Pour water Turn off fauc Wash pot Put pot	Open fridge Close fridge put containent:contai put containent:contai	Put sesoning:cont Take esting:utent Take esting:utent Inspect/Read reci Take plate	Pur condiment rectors condiment and parts Close oil:container put oil:container
	Pour water Turn off faucet Wish pot Move Around pot Put pot	Open fridge but containt: Close fridge put cheese:container but bread:container	Open fridge Put sesoning:container Inspect/Read recipe Take sesoning Pour seasoning	Pour condiment Mix pasta Close oil:container Dut oil:container Put oil:container
	Pour water Turn off faucet Wish pot Move Around pot Put pot	Open fridge Open fridge Ptt container: Ptt cheese: container Ptt cheese: container Close fridge	r T Take pasta: container Take oil: container Take oil: container Pur seasoning: container Pour seasoning	Pour condiment close condiment close oil: container but oil: container Wix pasta
	Pour water Turn off faucet Wash pot Move Around pot Put pot	Open fridge Open fridge Take bread:container Take bread:container Take tomato:container Put tomato:container	Take sesoning: container Take sesoning: container Pake pasta: container Pake oil: container Take oil: container Inspect/Read recipe	Pour condiment Close oil:container Dis oil:container Put oil:container Mix pasta
	Pour water Turn off faucet Wash pot Move Around pot Put pot	Open fridge Close fridge Close fridge tut condainer Put bread:container	Put seasoning:container Take oil:container Pour seasoning Take oil:container	Pour condiment: Intercondiment: Mix pasta Close oil:container Mix mixture
	Pour water Turn off faucet Wash pot Wur pot Put pot	Open fridge but container Put container Put cheese:container Put cheese:container Dispect/Read recipe	Put sasoning: container pour seasoning Mix pasta put trash	Pour condiment close condiment close oil: container Mix mixture but oil: container
	Pour water Turn off faucet Wash pot Nove Around pot Put pot	Open fridge Open fridge Open fridge drawer Open fridge drawer Put grocery:bag	Put sessoning: container Put sessoning: container Pour oil Put trash	Pour condiment Close condiment Close of L: container Purt oil: container Mix pasta
	e	E)	6)	ц.

Figure 14: The figure depicts four video clips from the **EGTEA Gaze+** validation set, showcasing **incorrect** predictions. Each clip presents the last eight frames and their corresponding top-5 action anticipations at  $\tau_a = 0.5s$ . The ground truth is revealed in the final frames, with the correct verb and noun of each action highlighted in green.