

# Supplementary Materials: Error Detection in Egocentric Procedural Task Videos

Shih-Po Lee<sup>1</sup>    Zijia Lu<sup>1</sup>    Zekun Zhang<sup>2</sup>    Minh Hoai<sup>2</sup>    Ehsan Elhamifar<sup>1</sup>

<sup>1</sup>Northeastern University    <sup>2</sup>Stony Brook University

<sup>1</sup>{lee.shih, lu.zij, e.elhamifar}@northeastern.edu    <sup>2</sup>{zekzhang, minhhoai}@cs.stonybrook.edu

## S.1. Data Collection

In this section, we provide more details of our data collection process. Figure 1, 2, 3, 4, and 5 are the task graphs of *Pinwheels*, *Quesadilla*, *Tea*, *Oatmeal*, and *Coffee*. We first created transcripts for each task based on the task graph. The participants then executed the steps in the transcripts one-by-one. During data collection, the participants receive instructions about the next action (including errors) to perform from an earphone (to prevent the recorded environmental audio being mixed with these verbal instructions). Before every recording session, we randomly changed the type, number, and location of the objects on the desk to increase the variability of our dataset. The objects could be irrelevant from the tasks. For example, there might be a microwave when making pinwheels. In addition, we changed the environment by rotating the table and changing the lighting condition to simulate every possible condition.

## S.2. Ablation Studies

Table 1 shows the ablation of using CSPL module, effect of different video and relational feature extractor. As the results demonstrate, our method i) with GCN achieves 2.8 % more AUC than with MLP, ii) with CSPL improves the AUC by 1.2 %, iii) with I3D feature achieves higher score than Hiera feature [1] under our error detection setting.

## S.3. Qualitative Analysis

Figure 6, 7 and 8 visualize the ground-truths and action segmentation predictions of our method with different components. Note that the input videos are error-free since we aim at visualizing how effective our proposed method can be on procedure understanding. Our proposed method with contrastive step prototype learning module and active object detection (AF + CSPL + AOD) can capture more accurate step segments than the backbone action segmentation model (AF), especially for recognizing background segments (in purple), see the last two rows of Figure 7 and 8.

## S.4. Annotations

**Error Steps.** Table 2, 3, 4, 5, and 6 show the details of

normal and error steps in EgoPER.

– Step Omission: When step omission occurs, some following steps may not happen. For example, in *Pinwheels*, if "Forget to use knife to scoop nut butter" occurs, "Spread butter onto tortilla" will not happen. In *Oatmeal*, if "Forget to put bowl in the microwave" occurs, the following "Microwave for X seconds" and "Remove the bowl from the microwave" will not happen.

– Step Addition: Step addition will happen at any place before some certain steps. For example, in *Quesadilla*, "Pour sugar on the tortilla" always happen before "Roll the tortilla". In *Tea*, "Sprinkle cinnamon" always occur before "Stir using spoon".

– Step Modification: To execute some step modification, the participants had to omit some steps and execute the modified steps. For example, in *Tea*, the participants executed "Forget to measure 12 ounces of cold water" and later executed "Directly pour water to kettle". In addition, the step "Add sugar" in *Tea* and *Oatmeal* belongs to modification but in *Pinwheels* it belongs to addition. This is because the recipes of previous two tasks have the step "Add honey", meaning that they need a step to add sweet. Therefore, instead of "Add honey", we modify the step by using an alternative step "Add sugar".

– Step Slip: Step slip is not necessary followed by correction steps, especially in the coffee task. The participants sometimes continued the actions without correction if the errors were not serious, e.g., spill out coffee beans or weigh incorrect amount of water. Since the errors are not serious, it becomes more challenging to detect them.

– Step Correction: Step correction sometimes contains both correction step and normal step. For example, in *Pinwheels*, both "Discard tortilla and place a new one on cutting board" and "Unfold tortilla and roll it" include correction step and normal step.

**Normal Steps.** The participants executed the actions according to the transcripts. Some repetitions are required in some tasks, some are not. For example, "Insert toothpick"



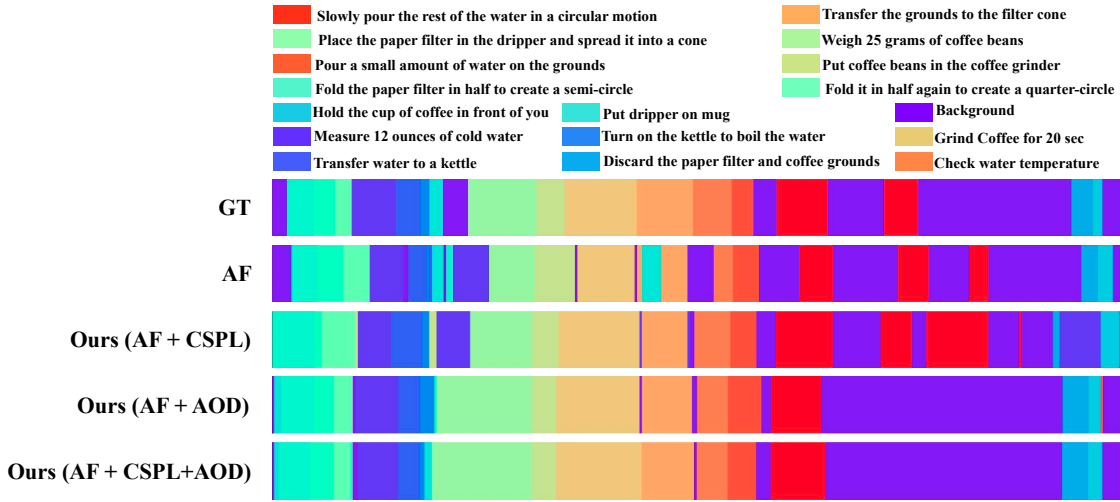


Figure 6. Comparison of action segmentation results on *Coffee* in EgoPER.

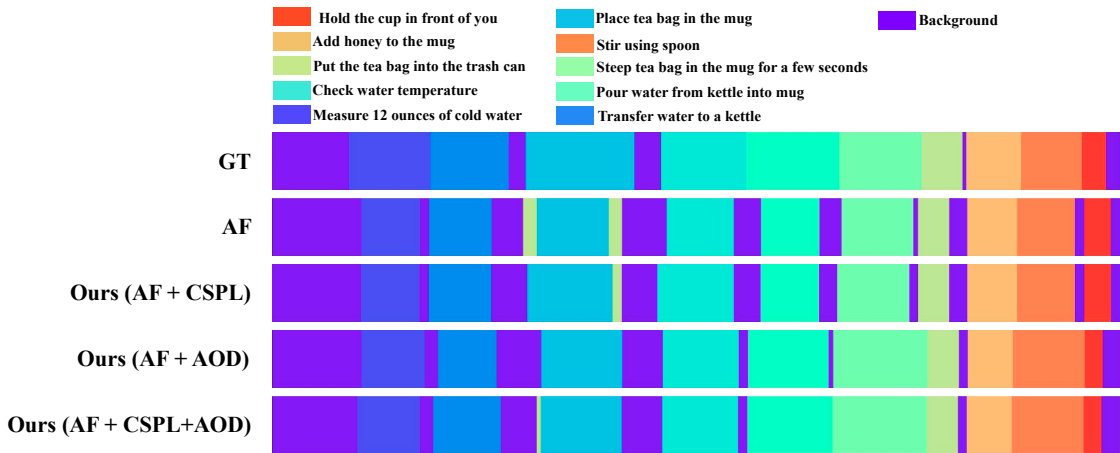


Figure 7. Comparison of action segmentation results on *Tea* in EgoPER.

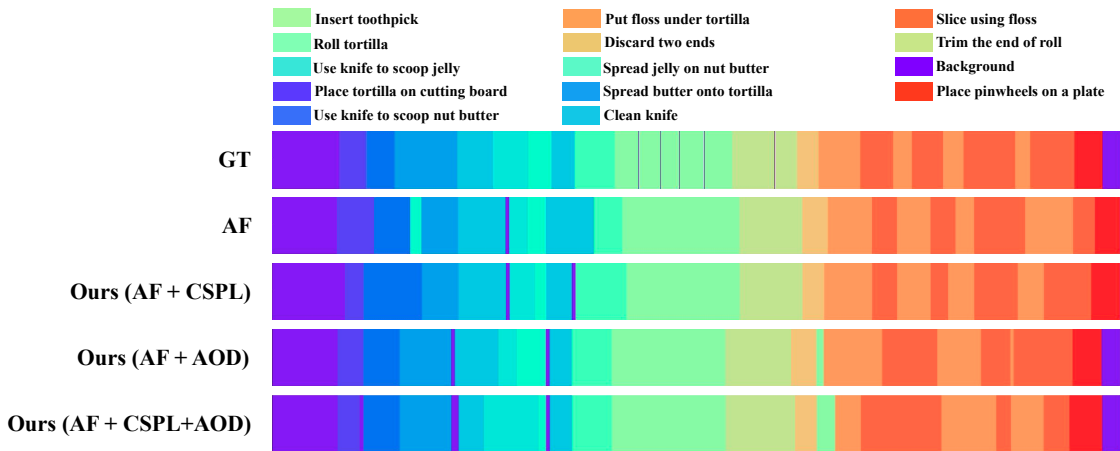


Figure 8. Comparison of action segmentation results on *Pinwheels* in EgoPER.

Normal step	Error step				
	Omission	Addition	Modification	Slip	Correction
Place tortilla on cutting board Use knife to scoop nut butter Spread butter onto tortilla Place tortilla on cutting board Clean knife Use knife to scoop jelly Spread jelly on nut butter Roll tortilla Insert toothpick Trim the end of roll Discard two ends Put floss under tortilla Slice using floss Place pinwheels on a plate	Forget to clean the knife Forget to use knife to scoop nut butter Forget to trim the ends Forget to use knife to scoop jelly	Pour sugar on the tortilla Remove the toothpicks from the tortilla pieces	Place tortilla on the table Use spoon to scoop nut butter Use spoon to scoop jelly Clean spoon Slice using knife	Drop the tortilla Fold tortilla	Discard tortilla and place a new one on cutting board Unfold tortilla and roll it

Table 2. Normal and error steps in *Pinwheels*.

Normal step	Error step				
	Omission	Addition	Modification	Slip	Correction
Place tortilla on cutting board Use knife to scoop Nutella Spread Nutella onto tortilla Put banana slices on tortilla Sprinkle cinnamon onto tortilla Fold tortilla Slice using knife Place tortilla wedges on a plate	Forget to use knife to scoop Nutella Forget to put banana slices Forget to sprinkle cinnamon	Add a handful of raisins to tortilla Pour a handful of oatmeals on tortilla	Place tortilla on the table Place tortilla wedges into a bowl Rip tortilla by hands	Drop the tortilla Fold tortilla into quarter-circle	Discard tortilla and place a new one on cutting board Unfold tortilla

Table 3. Normal and error steps in *Quesadilla*.

Normal step	Error step				
	Omission	Addition	Modification	Slip	Correction
Measure 12 ounces of cold water Transfer water to a kettle Place tea bag in the mug Check water temperature Pour water from kettle into mug Steep tea bag in the mug for a few seconds Put the tea bag into the trash can Add honey to the mug Stir using spoon Hold the cup in front of you	Forget to discard tea bag Forget to add honey Forget to check water temperature Forget to measure 12 ounce of cold water	Sprinkle cinnamon Put mug into microwave for 5 seconds	Directly pour water to kettle Stir using knife Add sugar to the mug	Drop tea bag Pour water into another mug	Discard tea bag and place a new one Move the tea bag to the new mug

Table 4. Normal and error steps in *Tea*.

Normal step	Error step				
	Omission	Addition	Modification	Slip	Correction
Measure 4 Tablespoons of quick-cook oats Measure 1/2 cup water Pour water to the bowl Put bowl in the microwave Microwave for X seconds Remove the bowl from the microwave Add handful of raisins Stir using spoon Put bananas Sprinkle cinnamon on the bowl Drizzle honey in the bowl	Forget to add bananas Forget to put bowl in the microwave Forget to stir Forget to measure water	Weigh the bowl containing the oats using scale Clean the spoon after stirring using paper towel Stir using knife	Directly pour quick oats into bowl Pour sugar instead of honey	Add bananas to another empty bowl Add water to a different bowl	Pour bananas from the second bowl to the first one Pour water from the second bowl to the first one

Table 5. Normal and error steps in *Oatmeal*.

Normal step	Error step				
	Omission	Addition	Modification	Slip	Correction
Measure 12 ounces of cold water Transfer water to a kettle Turn on the kettle to boil the water Discard the paper filter and coffee grounds Hold the cup of coffee in front of you Put dripper on mug Fold the paper filter in half to create a semi-circle Fold it in half again to create a quarter-circle Place the paper filter in the dripper and spread it into a cone Weigh 25 grams of coffee beans Put coffee beans in the coffee grinder Grind Coffee for 20 sec Transfer the grounds to the filter cone Check water temperature Pour a small amount of water on the grounds Slowly pour the rest of the water in a circular motion	Forget to discard paper filter Forget to fold the paper filter in half to create a semi-circle		Pour water without circular motion Squeeze the paper filter into cone without folding	Knock over the dripper Tear the paper filter Measure incorrect amount of water Remove dripper before drained Measure incorrect number of coffee beans Spill out coffee grounds Spill out water Spill out coffee beans Pour too much water to overflow the mug Grind coffee beans without enough time	Put the dripper back to mug Discard the filter and take a new one

Table 6. Normal and error steps in *Coffee*.