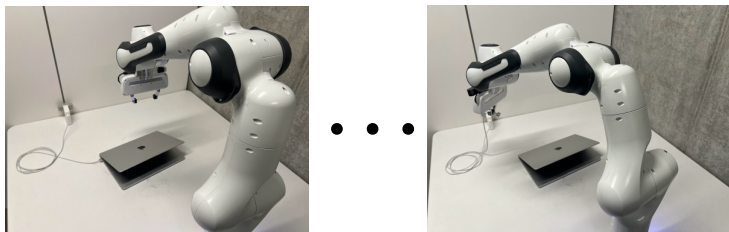


# Input: Example Episodes with Instruction



[ $\mathcal{I}$ ]: “unplug the laptop”

## Step 1: Identify Keyframes



$t_1$

$t_2$

$t_T$

## Step 2: Estimate Initial Object Poses and Extract Robot Actions



$[\mathcal{O}_1][\mathcal{O}_2]$

$[\mathcal{A}_{t_2}]$

$[\mathcal{A}_{t_T}]$

$[\mathcal{O}_1]$ : “laptop”: [55, 49, 4, 0, 36, 19]

$[\mathcal{O}_2]$ : “cable”: [35, 55, 4, 0, 28, 13]

$[\mathcal{A}_{t_2}]$ : [45, 52, 12, 0, 36, 22, 1]

⋮

$[\mathcal{A}_{t_T}]$ : [71, 38, 2, 0, 36, 37, 1]



## Step 3: Create ICL Prompt

You are a Franka Panda robot with a parallel gripper. We provide you with some demos in the format of observation>[action\_1, action\_2, ...]. Then you will receive a new observation and you need to output a sequence of actions that match the trends in the demos. Do not output anything else.

$\{[\mathcal{O}_1], [\mathcal{O}_2], [\mathcal{I}]\} > \{[\mathcal{A}_{t_2}], \dots, [\mathcal{A}_{t_T}]\}$

$\{[\mathcal{O}_1^{\text{Test}}], [\mathcal{O}_2^{\text{Test}}], [\mathcal{I}]\} >$

### Test Sample:



$[\mathcal{O}_1^{\text{Test}}]$   
 $[\mathcal{O}_2^{\text{Test}}]$   
 $[\mathcal{I}]$

Robot Action 

$\{[\hat{\mathcal{A}}_{t_2}^{\text{Test}}], \dots, [\hat{\mathcal{A}}_{t_T}^{\text{Test}}]\}$

Off-the-shelf  
LLM