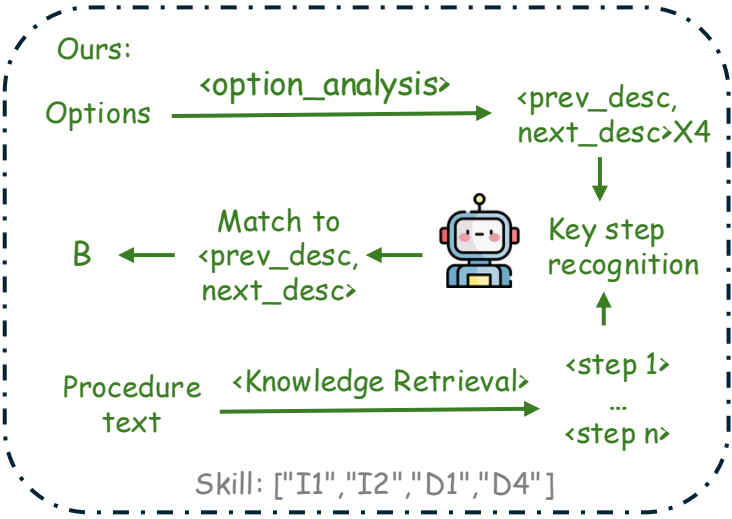


PSR: In the context of making Breakfast Burritos, based on the intermediate key-step shown in this video clip, what is the key-step sequence based on procedure text? Choose the correct combination of the immediate previous step and the immediate next step.



- A: "Previous action: Warm the tortilla in the microwave. Next action: Fold the burrito tightly."
- B: "Previous action: Mix the contents of the bowl well. Next action: Pour egg mixture on top of the tortilla."
- C: "Previous action: Mix the contents of the bowl well. Next action: Fold the burrito tightly."
- D: "Previous action: Pour egg mixture on top of the tortilla. Next action: Sprinkle 1 tbsp shredded cheddar cheese on top of the egg."

Qwen2.5-VL: A Vinci: D LLaVA-OneVision: D
Video-LLaMA2: D

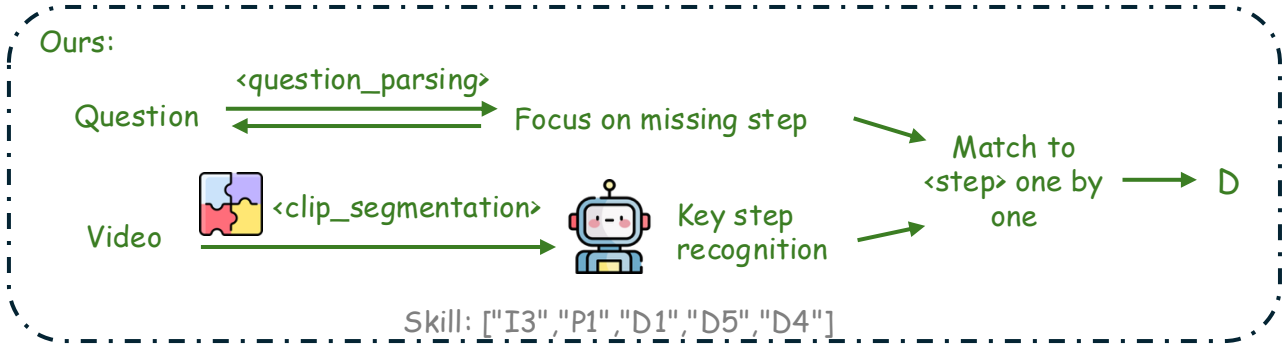


KSM: I have concatenated two video clips from a procedure. The full procedure involves 3 consecutive steps: 1. Add chilli to the pan, 2. Cook for 1 minute, mixing everything, 3. Add 1/8 tsp of turmeric to the pan. Based on the visual content, which step is missing from the provided video?



- A: "Add chilli to the pan"
- B: "Add 1/8 tsp of turmeric to the pan"
- C: "pour-Slowly pour the whisked eggs into the pan"
- D: "Cook for 1 minute, mixing everything"

Qwen2.5-VL: A Video-LLaMA2: B
Vinci: A LLaVA-OneVision: B

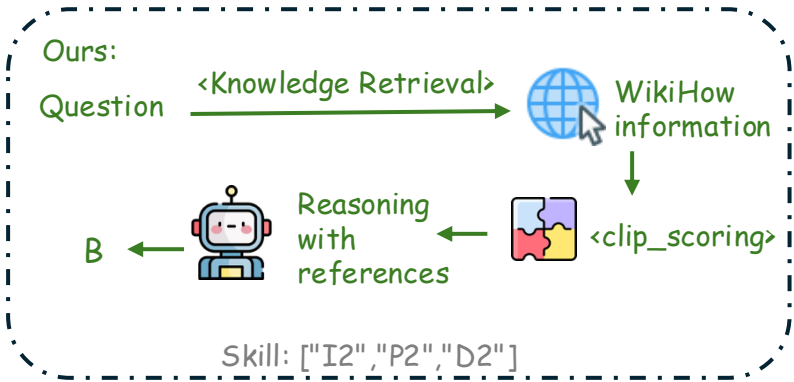


TC: The task is: Tent Assembly. The video clip shows one step from this task. Is the task completed?11

Qwen2.5-VL: A Video-LLaMA2: A
Vinci: A LLaVA-OneVision: A



- A: "Not completed"
- B: "Completed"

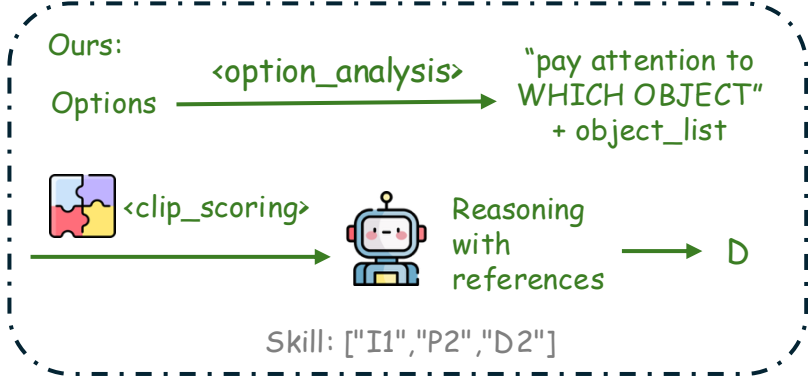


KSR: What is the key step being performed by the camera wearer in the video clip from 360.4 seconds to 409.0 seconds?

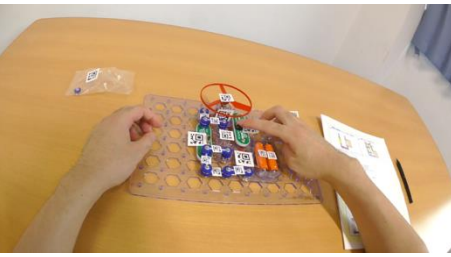
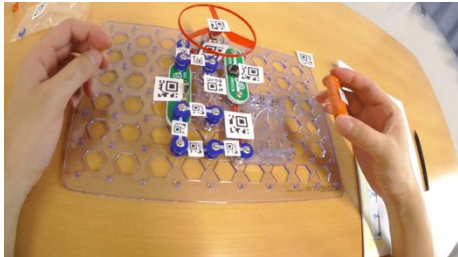
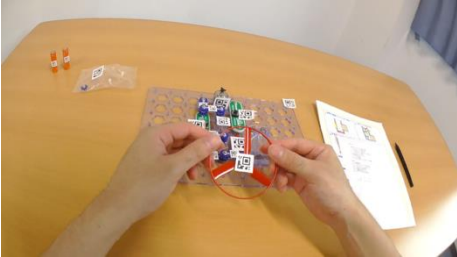


- A: "Open the lid of the spice jar"
- B: "Sprinkle spices on the tortilla"
- C: "Whisk the eggs in the bowl"
- D: "Sprinkle oregano in the bowl"

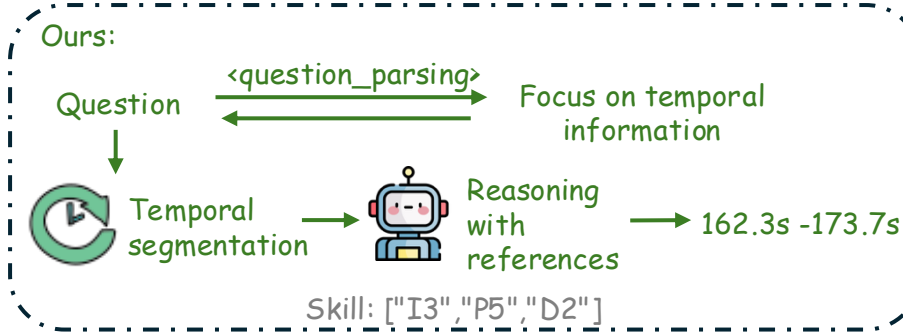
Qwen2.5-VL: B Vinci: A LLaVA-OneVision: A
Video-LLaMA2: B



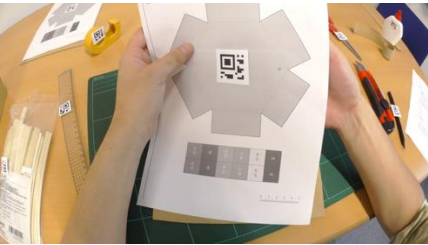
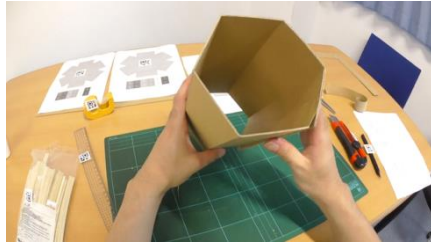
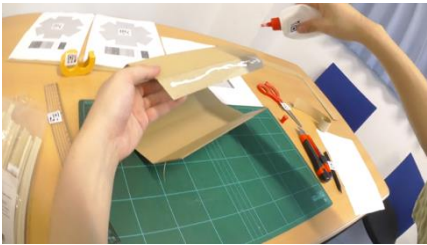
KSG: The provided video clip covers the period from 150.9s to 185.1s. It contains the following steps:\n1. Put a propeller on the motor.\n2. Put two battery in the battery box.\n3. Turn on the switch S2.\nQuestion: At what time interval does the action 'Put two battery in the battery box.' occur?



GT: 157.8s - 177.5s
Qwen2.5-VL: 176.0s - 181.6s
Vinci: 150.9s - 185.1s
LLaVA-OneVision: 165.9s - 170.3s
Video-LLaMA2: 150.9s - 185.1s



KSO: I have provided 3 video clips (Clip 1, Clip 2, Clip 3) extracted from the activity 'cardboard'. Procedure:... They are presented in a random order. Please arrange them in the correct chronological order.



- A: "Clip 1 -> Clip 2 -> Clip 3"
- B: "Clip 2 -> Clip 3 -> Clip 1"
- C: "Clip 2 -> Clip 1 -> Clip 3"
- D: "Clip 3 -> Clip 1 -> Clip 2"

Qwen2.5-VL: A Video-LLaMA2: B
Vinci: D LLaVA-OneVision: B

