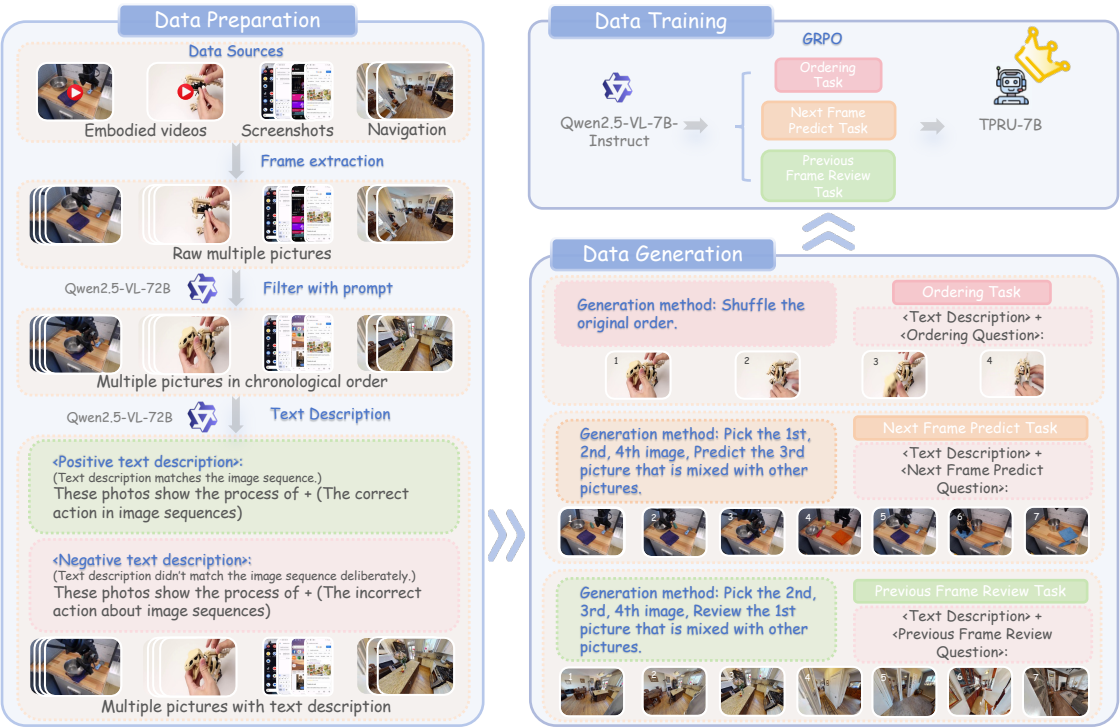
Data Preparation
Data Sources
Embodied videos
Screenshots
Navigation
Frame extraction
Raw multiple pictures
Qwen2.5-VL-72B
Filter with prompt
Multiple pictures in chronological order
Qwen2.5-VL-72B
Text Description
<Positive text description>:
(Text description matches the image sequence.)
These photos show the process of + (The correct action in image sequences)
<Negative text description>:
(Text description didn't match the image sequence deliberately.)
These photos show the process of + (The incorrect action about image sequences)
Multiple pictures with text description
Data Training
GRPO
Qwen2.5-VL-7B-Instruct
Ordering Task
Next Frame Predict Task
Previous Frame Review Task
TPRU-7B
Data Generation
Generation method: Shuffle the original order.
Ordering Task
<Text Description> + <Ordering Question>:
Generation method: Pick the 1st, 2nd, 4th image, Predict the 3rd picture that is mixed with other pictures.
Next Frame Predict Task
<Text Description> + <Next Frame Predict Question>:
Generation method: Pick the 2nd, 3rd, 4th image, Review the 1st picture that is mixed with other pictures.
Previous Frame Review Task
<Text Description> + <Previous Frame Review Question>: