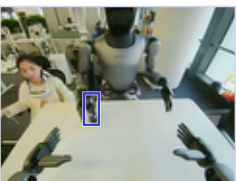


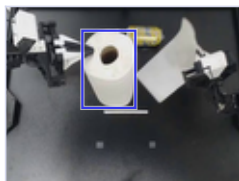
Question type: Referring Task grounding Multi-instance Absence and judgment Cross-view

Referring point point



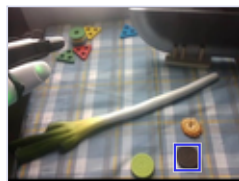
Show me the pepsi drink with a point.

Referring box box



Draw a bounding box around the toilet paper roll.

Attribute point



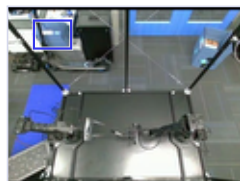
Where is the chocolate cookie? Point to it.

Ordinal point



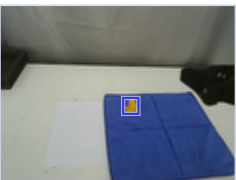
Point to the bottommost shoe.

Relation point



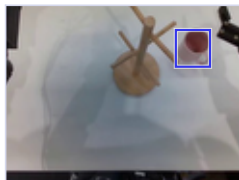
Point to the blue bin to the left of the pencil.

Small object point



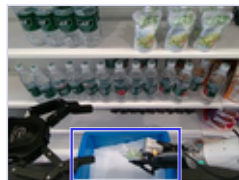
Where is the yellow plastic part? Point to it.

Task target point



Task: "Hang the mug on the cup rack." Which object should the robot act on? Point to it.

Task destination point



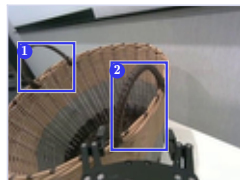
For the task "Pick snacks from the snack shelf and place them into the snack-purchase bin.", which place is the destination? Point to it.

Every instance 3 points



Point to every soda can in the image.

Counting 2



Count the basket handles in the image.

Absent object none



Is there a red box in the image? If so, point to it; if not, answer "none".

Absent attribute none



Point to the red box in the image. If there is no red box in the image, answer "none".

Spatial yes/no yes



In the image, is the plate to the left of the toaster?

Cross-view correspondence point in image 2



Image 1: third-person

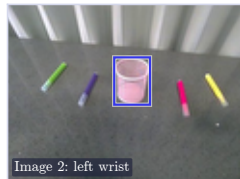


Image 2: left wrist

The red box in image 1 marks an object. Image 2 is another camera view of the same scene at the same time. Where is that object in image 2? Point to it.