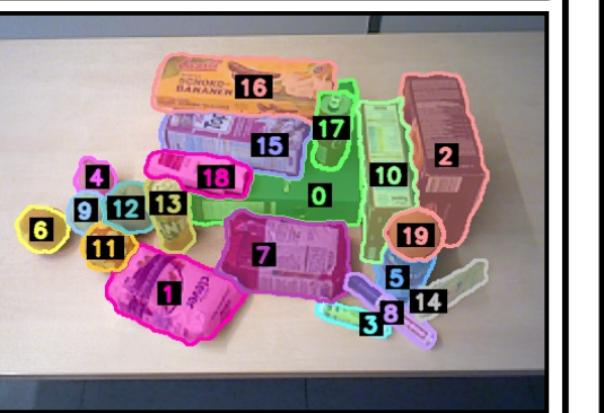


## Open-Ended Referring Segmentation

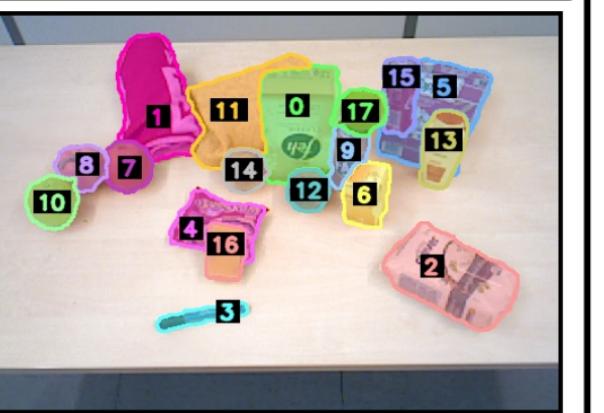


**User:** I broke my mug, can you get me something to fix it?

**GPT-4v:** To find the object that best matches the description: "I broke my mug, can you get me something to fix it?", I will look for an item that could be used to repair a broken mug, which would most likely be some form of adhesive. In the first image, there are several items on the table, but the one that stands out as the potential adhesive is the tube near the bottom of the image. This tube could be a glue or a similar substance that could be used to fix a broken mug. In the second image, the tube is marked with the color blue and has the ID number 3 inside a black rectangle.

My final answer is: [3]

## Grounded Grasp Planning



**Prompt:** [...] determine a plan to ensure a target object is graspable. [...]

Respond with a numbered list of objects to manipulate, by referring to their numeric IDs [...]. Task instruction: "Grasp object [9]".

**GPT-4v:** To grasp [9], it appears that objects [17] and [6] are directly breaking access to it. These objects must be removed first to clear the way for a successful grasp of [9].

My final answer is:

1. remove [17]
2. remove [6]
3. pick [9]

## Grasp Ranking via Contact Reasoning



**Prompt:** [...] Provide a list of the grasp IDs sorted from best to worst.

**GPT-4v:** The object of interest outlined with a red line appears to be a banana [...]. The grasp poses closest to surrounding objects are [1], [2] and [9] [...]. Good grasp poses are those whose centers are inside the body of the banana and not close to surrounding objects. These would likely be: [4], [5], [6] and [7]. The next best would be those that are inside the body but closer to the edges, which might be [3] and [8]. The least favorable grasp poses are the closest to surrounding objects or at the very ends of the banana, which are [1], [2] and [9].

My final answer is: [4, 5, 6, 7, 3, 8, 1, 2, 9]