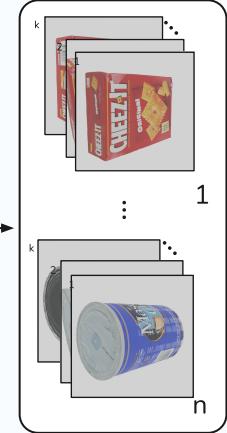


# Input



3D Object Database

## Onboarding



Rendered Views

VLM

```
"cracker_box":  
    view_1.png: "Red box of cheez-it",  
    view_2.png: "Cracker box",  
    ...,  
    view_n.png: "Cheez-it cracker";  
  
"master_chef_can":  
    view_1.png: "Coffee",  
    view_2.png: "Cleaning supplies",  
    ...,  
    view_n.png: "Food";  
...
```

Image Descriptions



RGB Image

"Yellow mustard"

Language Query

## Candidate Objects

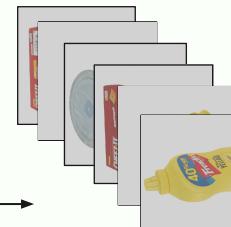
GSAM

CLIP

CLIP

Cosine Similarity

Thresholding



Candidate images

## Object Retrieval

DINO2

Cosine Similarity

DINO2



Matching CAD Model

Segmentation