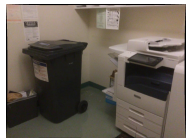
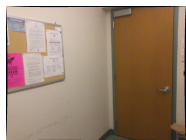




Question: If I am standing in front of the door. Is the trash bin on my right side?

Input views & Camera Poses



Vision Encoder

Image Tokens

Plücker Encoder

Camera Tokens

MLLM

Yes, the trash bin is on your right.

$\mathcal{L}_{\text{lang}}$

Ground Truth Answer



Pose Adapter

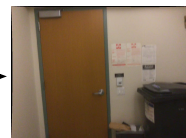
Predicted Camera Pose

Ground Truth Camera Pose

$\mathcal{L}_{\text{pose}}$

Imagined Target View

Image Decoder



Pose Fusion

Fusion MLP