3D Scene Scan



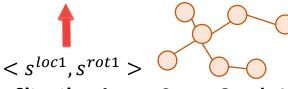
Images of Objects



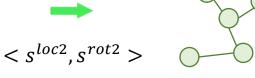












Situation 2 Scene Graph 2

More Situations More Scene Graphs



Situation Description and Situated QA Generation

Messages

- 1. System prompt
- 2. Human demonstrations
- 3. Query: situated scene graph

Responses

Situation description: I am facing some storage boxes. There is a <jacket-3-IMG> in front of me.

Situated QA:

Q: How many monitors are on my 10 o' clock? A: 1.



Modality Replacement

There is a red <jacket-3-IMG> in front of me. !

There is a in front of me.



Refinement Procedure

Counting: Q: How many monitors are on my left? A: $\frac{1}{2}$. \rightarrow 1.

Existence: Is there a toilet in the room? A: $\frac{\text{Yes.}}{\text{Yes.}} \rightarrow \text{No.}$

Negative responses: What is the material of the TV? Unknown.