Query: Find the door on the right of the chair near a table.

Agents

scene scans
object locations

online
Chairs are at …, tables are at …
Their **distances** are …
So the chairs are may be chair_1 and chair_2. chair_1 is nearer to table_1, but no object is on the right. Now check chair_2…
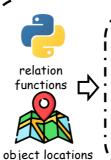Result: The target object is door_1. ✅

- 😊 accurate
- 😢 long context
- 😢 online LLM generation

Visprog.

relation functions
object locations

executor
```
> Loc("table")
[table_1]
> Near(table_1)
[chair_1]
> right(chair_1)
[]
Result: Empty. ❌
```
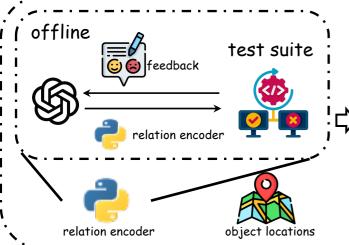
- 😊 short context
- 😢 not accurate
- 😢 function annotation

Ours

offline
feedback
relation encoder

test suite

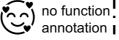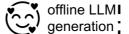relation encoder
object locations

executor
```
> Near(table_1, chair_1)
0.9
> Right(door_1, chair_1)
0.1
> Near(table_1, chair_1)
0.7
> Near(table_1, chair_1)
1.0
> 0.9 * 0.0 < 0.7 * 1.0
True
Result: The target object is door_1. ✅
```

- 😊 short context
- 😊 accurate
- 😊 no function annotation
- 😊 offline LLM generation

Grounding Performance

Time Cost / Sec. | Token Cost | Accuracy

ZSVG3D | VLMGrounder | EaSe