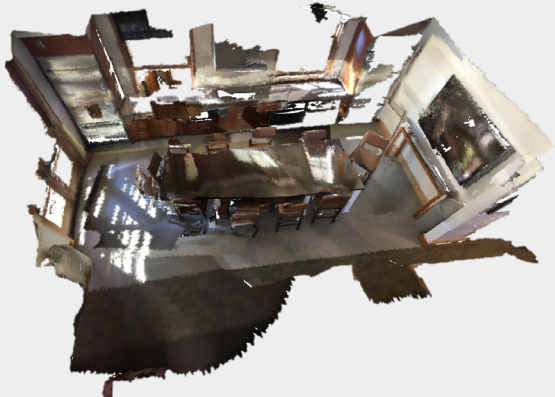


Input

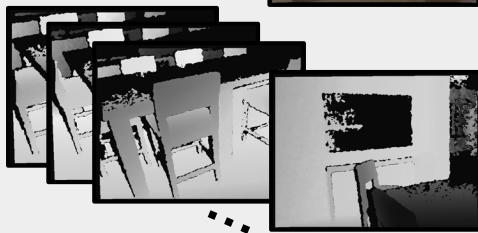
3D
Point
Cloud



RGB
Frames



Depth
Frames



Class-agnostic Proposal Module

1

Pre-trained 3D
Instance
Segmenter



2

SAM-based 3D
Instance
Segmenter

3

Best Viewpoint

Bounding
Box

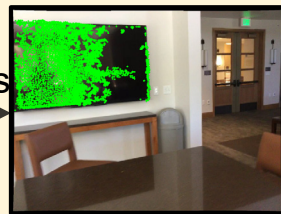


CoT
Prompt

4

Multi-
modal
LLM

Landmarks



CoT
Prompt

Multi-
modal
LLM

SAM Mask



CoT
Prompt

Multi-
modal
LLM

Semantic Reasoning Module

