SG-Nav: Online 3D Scene Graph Prompting for LLM-based Zero-shot Object Navigation

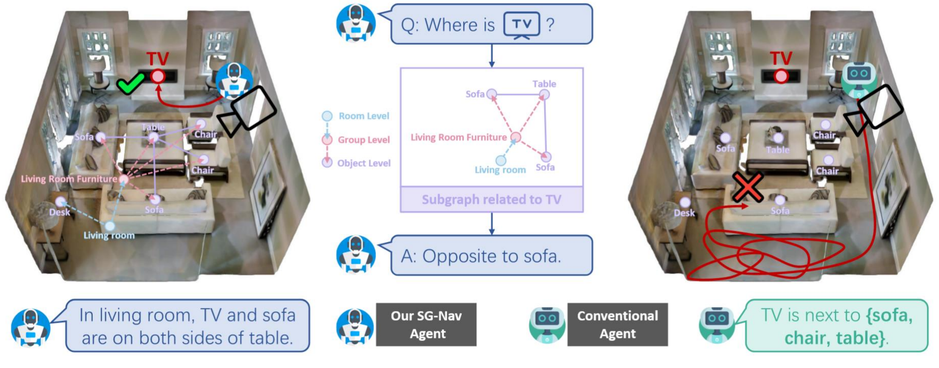
Hang Yin[1], Xiuwei Xu[1], Zhenyu Wu[2], Jie Zhou[1], Jiwen Lu[1]

[1] Tsinghua University, China [2] Beijing University of Posts and Telecommunications, China

NEURAL INFORMATION PROCESSING SYSTEMS

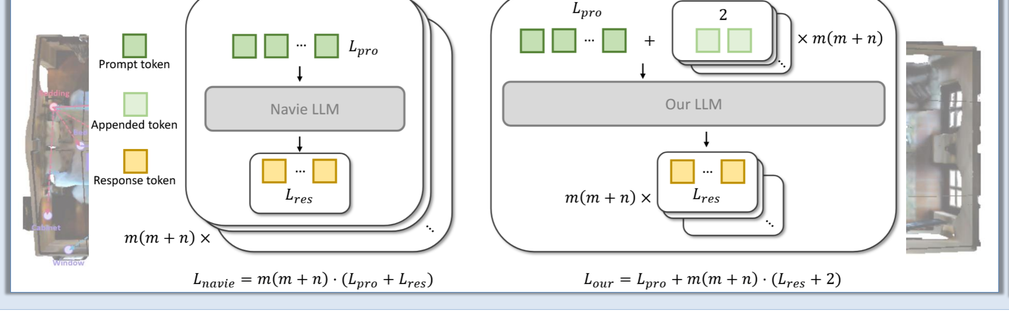
Highlight

- Conventional object-goal navigation methods rely on time-consuming training. Our SG-Nav online construct hierarchical scene graph and prompt LLM to reason the position of object goal in a zero-shot manner, with higher generalization ability. We even surpass some learning-based methods.



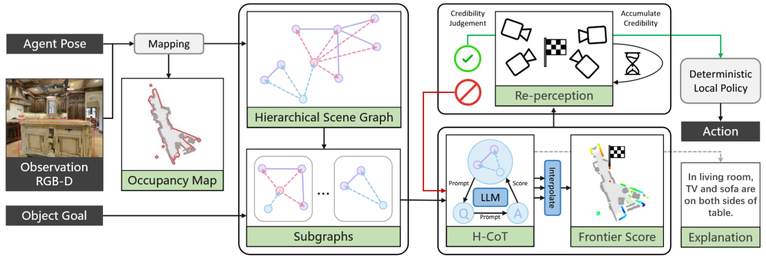
Analysis

- To speed up the dense connecting process for real-time processing, we propose a new form of prompt. This enables LLM to generate relationships between all pairs of nodes in one-shot with much less computational cost, comparing to navie generation approach.

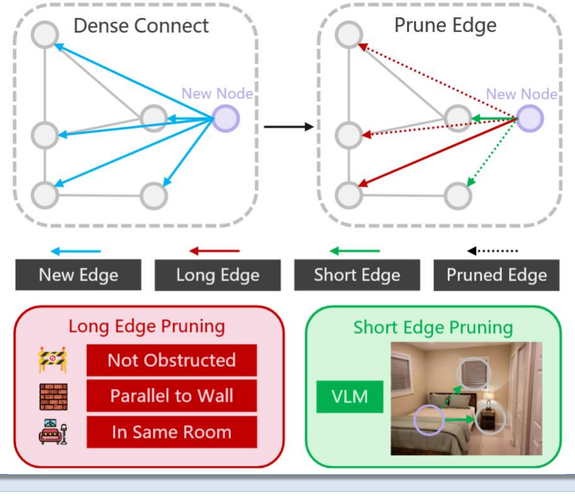


Approach

- We propose to online construct a hierarchical scene graph and decompose the scene graph and use the subgraphs to prompt LLM in a chain-of-thought manner. Then we score the frontier points and select a long-term goal. Besides, we propose re-perception to solve the problem of perception errors.



- We densely connect the newly detected nodes to previous nodes in an incremental manner. We further prune the dense connected edges to make the 3D scene graph more precise.



Experiments

- Our SG-Nav achieves leading performance comparing to most zero-shot methods.

Method	Unsupervised	Zero-shot	MP3D		HM3D		RoboTHOR	
			SR	SPL	SR	SPL	SR	SPL
CoW [10]	Yes	Yes	7.4	3.7	–	–	26.7	16.9
ESC [51]	Yes	Yes	28.7	14.2	39.2	22.3	38.1	22.2
L3MVN [45]	Yes	Yes	34.9	14.5	48.7	23.0	41.2	22.5
OpenFMNav [18]	Yes	Yes	37.2	15.7	52.5	24.1	44.1	23.3
VLFM [44]	Yes	Yes	36.2	15.9	52.4	**30.3**	42.3	23.0
SG-Nav-LLaMA	Yes	Yes	40.1	16.0	53.9	24.8	47.3	23.7
SG-Nav-GPT	Yes	Yes	**40.2**	**16.0**	**54.0**	24.9	**47.5**	**24.0**

- Visualization of Navigation Process: