

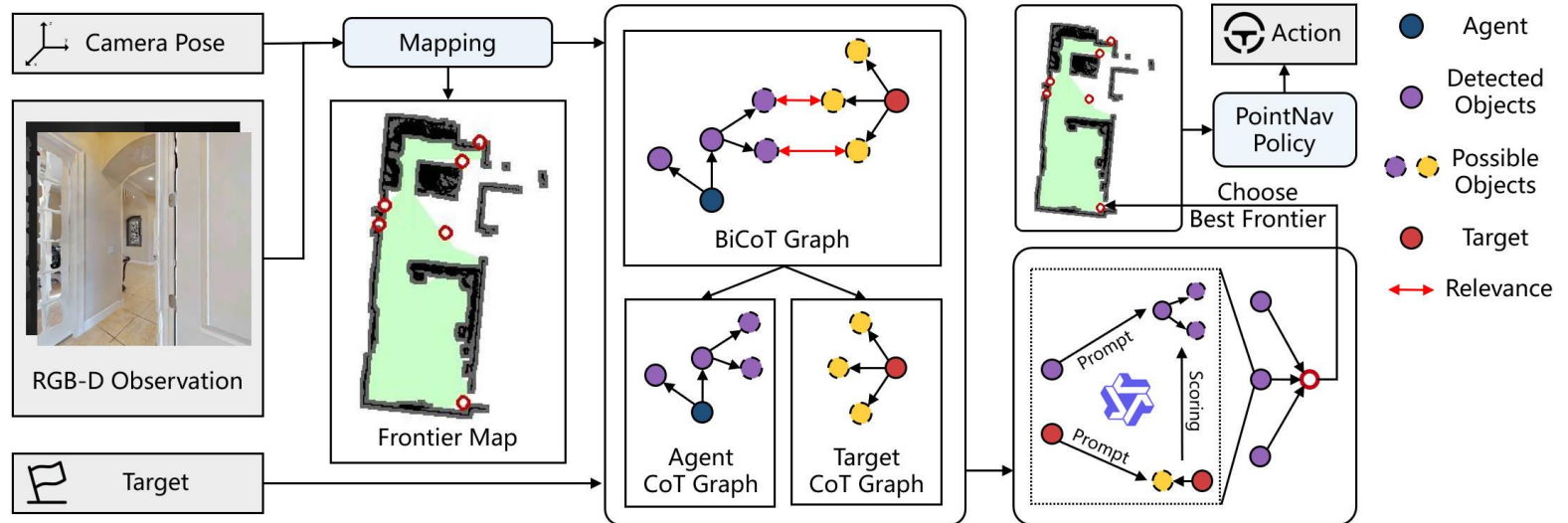
# Bidirectional Chain-of-Thought for Zero-shot Object Navigation

**Haonan LUO, Sijia LI, Yijie ZENG, Zihang WANG  
Botao JIANG, Xiruo JIANG**

Frontiers of Computer Science, DOI: [10.1007/s11704-025-41283-7](https://doi.org/10.1007/s11704-025-41283-7)

# Problems & Ideas

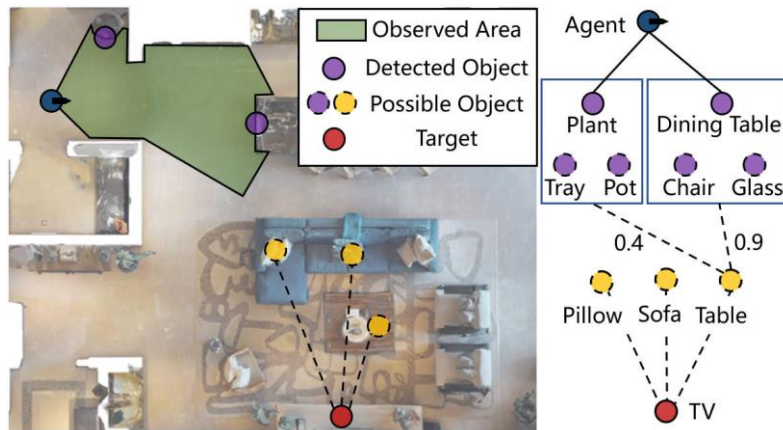
- Problems of object goal navigation:
  - Target is invisible from the starting point, the agent needs reasoning ability like human.
  - Previous learning-based methods struggle with generalization to novel environment.
- Ideas: A bidirectional chain-of-thought deduces what landmark may occur near the target or the objects in current observation.



The overall navigation process. At the beginning of an episode, the LLM is prompted to reason objects that may appear near the target to build the target CoT graph. During the navigation process, the agent detects objects from the RGB images and build the agent CoT graph. By assessing the relevance of the target CoT graph and agent CoT graph, we can choose which frontier to explore based on the agent CoT graph around it.

# Main Contributions

- Contributions:
  - A bidirectional chain-of-thought (BiCoT) approach for zero-shot object navigation that deduces what landmark may occur near the target (target CoT) or the objects (agent CoT) in the current observation;
  - A navigation schema that decides where to explore based on the possibilities among the target CoT graph and the agent CoT graphs via a large language model;
  - A comprehensive experiment that demonstrates the effectiveness of our proposed method.



We reason the potential route to the target from both sides, the target, and the objects detected in observations.

	Training	MP3D		HM3D	
		SR↑	SPL↑	SR↑	SPL↑
PONI [7]	ObjectNav	12.1	31.8	-	-
PIRLNav [8]	ObjectNav	-	-	27.1	<b>64.1</b>
ZSON [3]	ImageNav	4.8	15.3	12.6	25.5
CoW [2]	None	3.7	7.4	-	-
ESC [6]	None	14.2	28.7	22.3	39.2
BiCoT-Qwen (Ours)	None	17.1	33.2	27.8	44.3
BiCoT-GPT (Ours)	None	<b>17.4</b>	<b>33.6</b>	<b>28.1</b>	<b>44.8</b>

Zero-shot Object Navigation results on MP3D and HM3D benchmarks. Qwen and GPT denotes the type of LLM.