CAMON: Cooperative Agents for Multi-Object Navigation with LLM-based Conversations

Pengying Wu[1], Yao Mu[2], Kangjie Zhou[1], Ji Ma[1], Junting Chen[3], Chang Liu[1]

[1]Peking University, [2]The University of Hong Kong, [3]National University of Singapore

Introduction

Task Definition:

Multiple collaborative robots *jointly* **explore** and **approach** target objects in indoor environments.

Motivation:

1. Previous studies have been limited to **single**-target tasks and **single**-robot settings.
2. What **content** should be communicated to achieve better collaboration among multi-agents.
3. How to efficiently **divide** complex tasks and **make** collaborative decisions.

Method

Perception:

➢Semantic Mapping:

$$M_t \leftarrow \text{Mapping}(\text{ObjDet}(RGB), Depth, pose, M_{t-1})$$

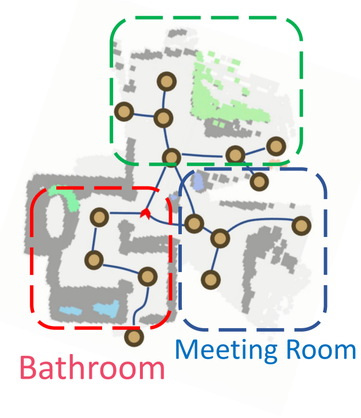
➢Room Description:

$$\text{RoomDesc} \leftarrow \text{LLM}(Images, \text{RoomSeg}(M_t))$$

➢Topological Mapping

ESDF

$$\mathcal{V} = \{x \in \mathcal{X} \backslash \Omega | \exists \omega_i \neq \omega_j \in \Omega, d(x, \omega_i) = d(x, \omega_j) = f(x)\}$$



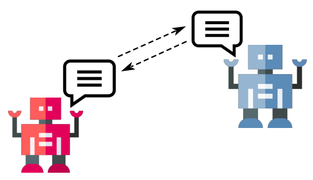
Communication:

➢Initial Decision Making:

$$Ps_i = \text{LLM}(Pr(i, P_i, S_i, G, H_i))$$

➢Teamwork Coordination:

$$Re^* = \text{LLM}(Pr^*(Ps_i, P_g, S_g, G))$$

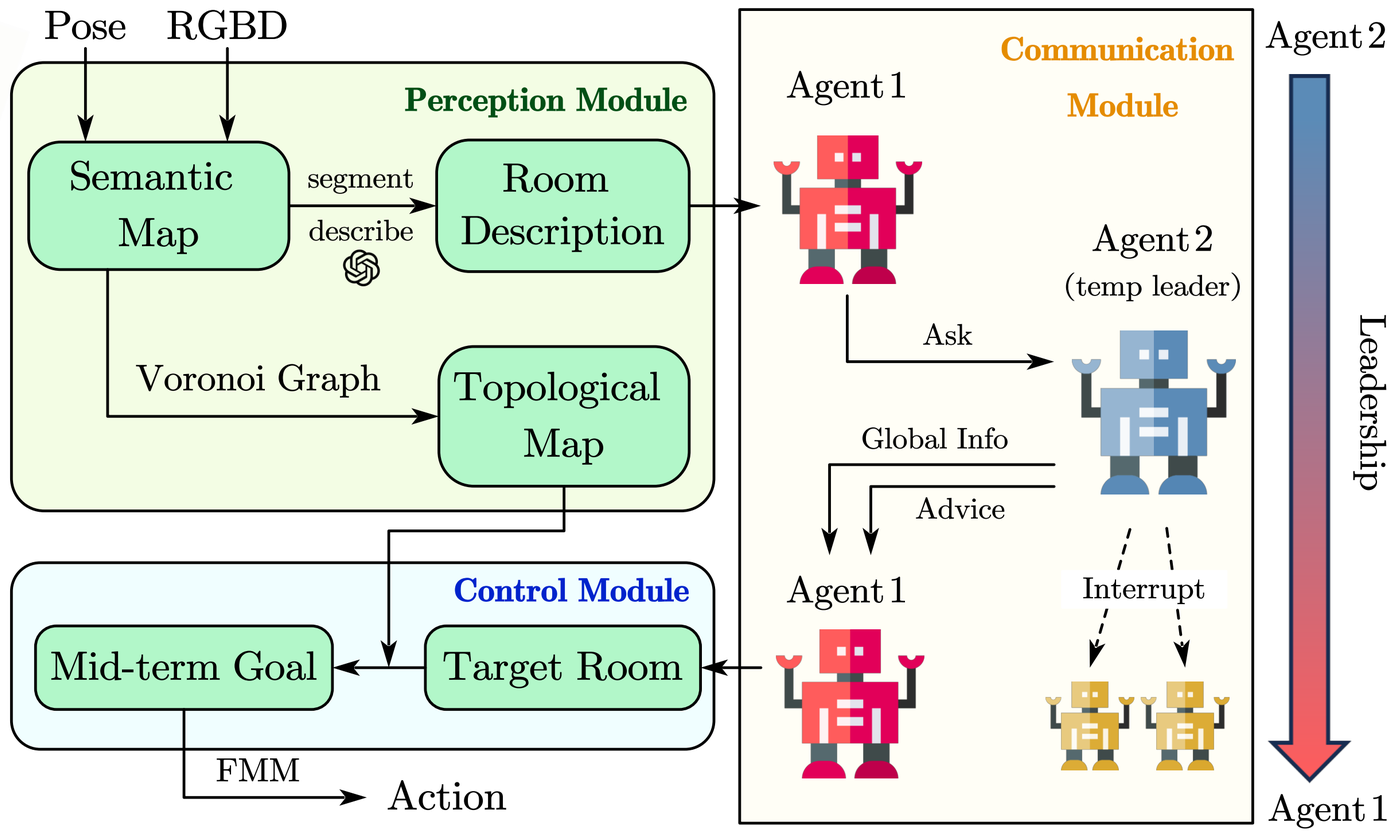


Motion Planning:

➢Mid-term Goal Point

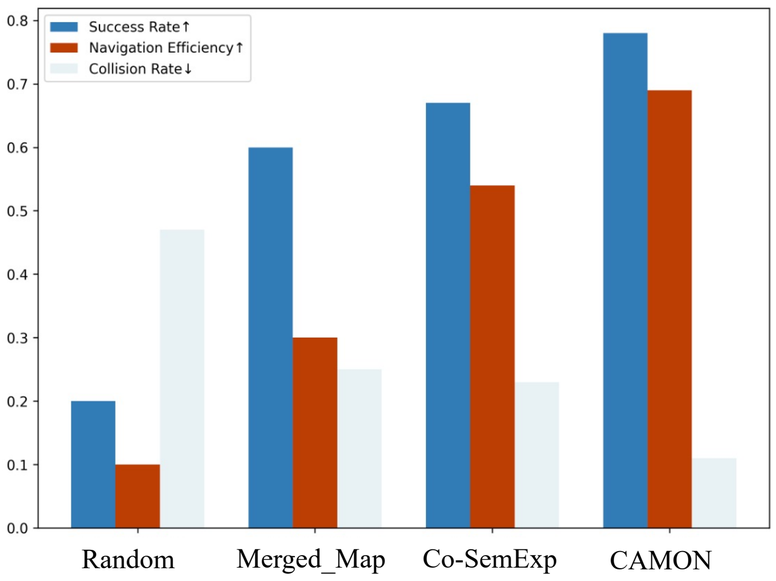
$$MidGoal \leftarrow \text{Select}(M_t, \mathcal{V}, Re^*)$$

➢Point-to-point Planning (FMM method)



Result

SOTA performance among all comm paradigms



Visualization of simulation experiments