SEATTLE, WA CVPR JUNE 17-21, 2024

MemoNav: Working Memory Model for Visual Navigation

CVPR'24 Highlight

Hongxin Li, Zeyu Wang, Xu Yang, Yuran Yang, Shuqi Mei, **Zhaoxiang Zhang**

University of Chinese Academy of Sciences (UCAS)

New Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences

Center for Artificial Intelligence and Robotics, HKISI, CAS

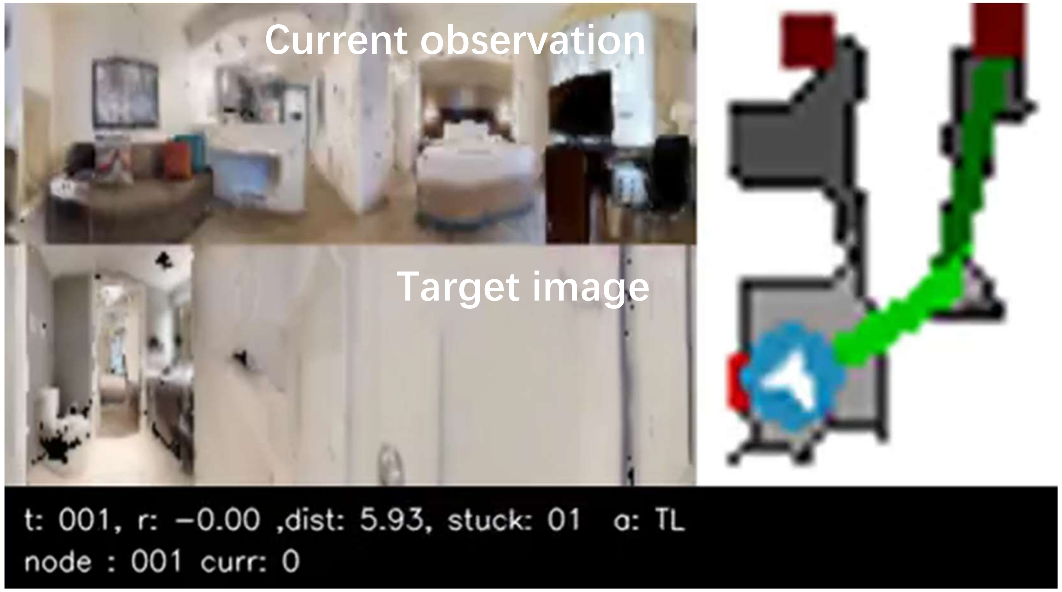
Shanghai AI Lab Tencent

Repo: https://github.com/ZJULiHongxin/MemoNav

1 Background

Task: Image-goal visual navigation

Task requirements: The agent navigates to the goal area specified by an image with the fewest number of steps



Navigation Example

Target-driven visual navigation in indoor scenes using deep reinforcement learning, ICRA 2017

2 Related Works

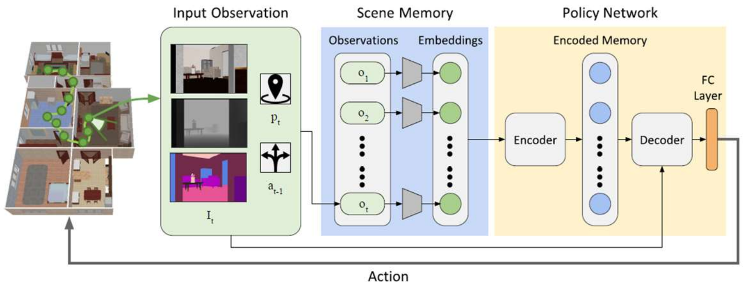
Basic approach:

- Build scene memory for navigation decision-making
- Use IL or RL to train agents

➢ Existing method 1： SMT

Memory implementation: stacking navigation history information

Drawback: The storage and computational complexity are high



➢ Existing method 2： VGM

Memory implementation： Employ topology maps to selectively store landmark features

Drawbacks: (1) Too many redundant nodes→ too much noise

(2) Lack of scene-level features → Inferior decision-making

Scene Memory Transformer for Embodied Agents in Long-Horizon Tasks, CVPR2019

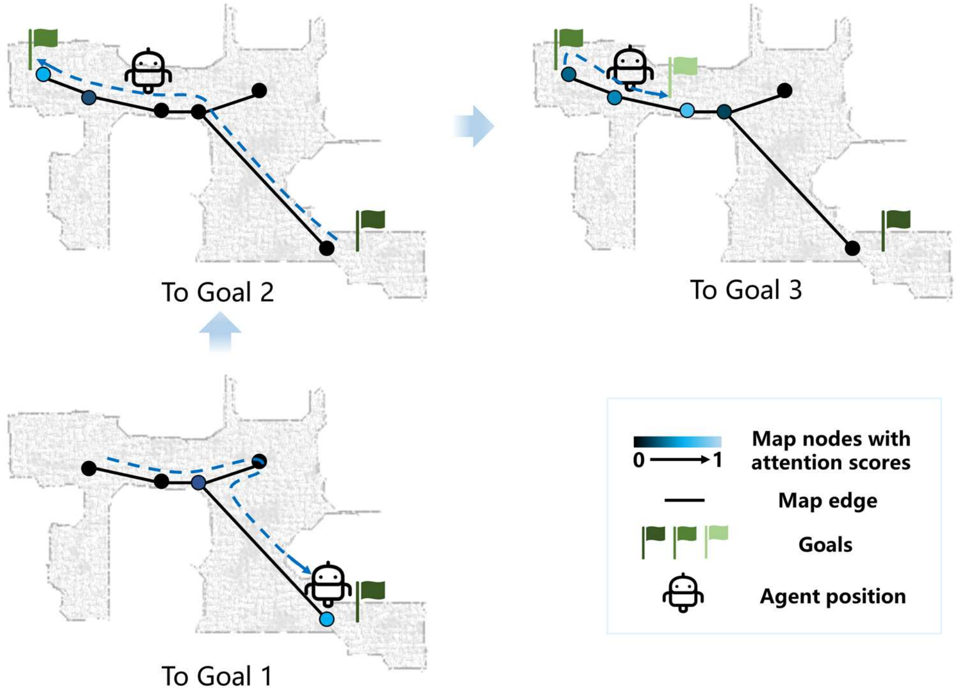
Visual Graph Memory With Unsupervised Representation for Visual Navigation, ICCV2021

2 Related Works

Common shortcomings of existing methods

Typically test only on single-goal datasets → The role of memory mechanisms is hard to be adequately evaluated

Our opinion: Multi-goal navigation tasks are more suitable, as scene memory should help the agent quickly return to the explored area



Examples of multi-goal navigation tasks

Scene Memory Transformer for Embodied Agents in Long-Horizon Tasks, CVPR2019

Visual Graph Memory With Unsupervised Representation for Visual Navigation, ICCV2021

3 MemoNav: Agent Design

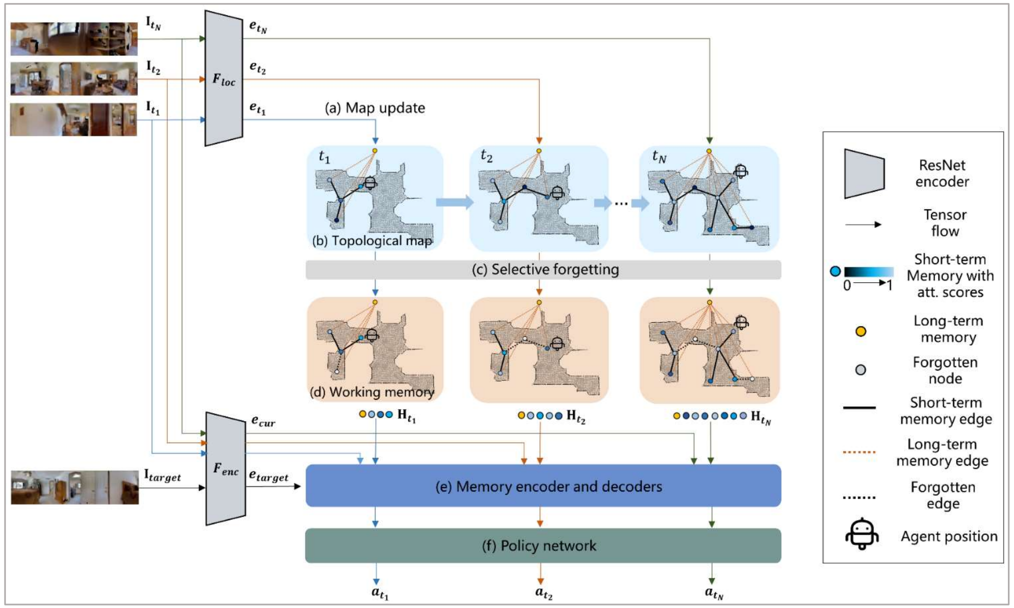
MemoNav: A navigation agent that mimics the working memory of the human brain

We introduce 3 types of scene memories:

1. **Short-term memory (STM):** Local nodes in a topology map
2. **Long-term memory (LTM):** A global map node
3. **Working memory (WM):** STM retained by our proposed forgetting module and LTM

"Working memory is essential for the organization of ***goal-directed behavior****, as it maintains task-relevant information."*

——Farshad A. Mansouri et al.



MemoNav architecture

Working Memory in the Service of Executive Control Functions, Front Syst Neurosci, 2015

3 MemoNav: Agent Design

Adaptive Forgetting Module

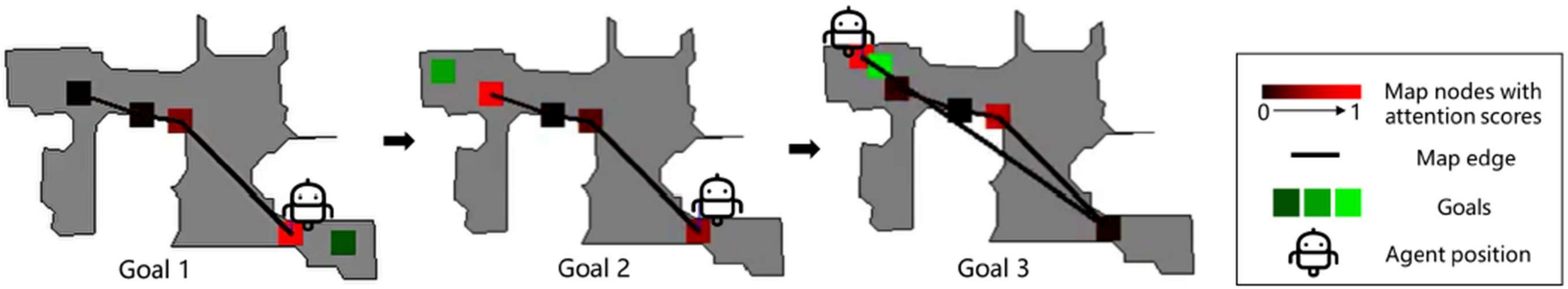
① Decode the topological map while assigning each short-term memory attention score$\{\alpha_i\}_{t_{i=1}}^{N}$

③ Temporarily remove (forget) the bottom 20% of the short-term memory

④ Before the next navigation goal, the forgotten memories are restored to the topological map

Effect

- Retain goal-relevant information and exclude noise from irrelevant areas
- Reduce computation



Visualization of attention scores for STM (map nodes)

3 MemoNav: Agent Design

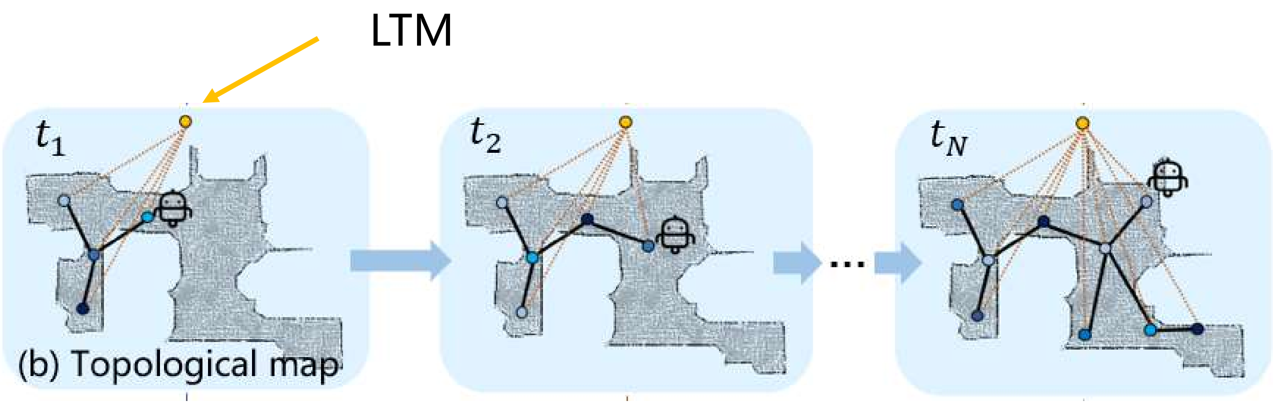
Long-term memory (LTM) generation

① On top of the topology map, we add a global node as the LTM

② Graph convolution is used to aggregate STM features into LTM

Effect

- Store scene-level features
- Facilitate feature fusion among long-distance graph nodes
- Assist in the forgetting module



The LTM connects and aggregates all node features

3 MemoNav: Agent Design

Generate WM for decision-making

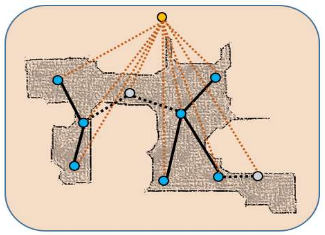
① The retained STM and LTM are further encoded into working memory (WM) by the graph attention mechanism

② WM is input to the policy module → Navigation actions

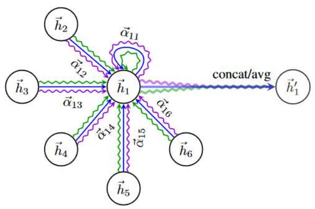
Effect

① Both local and global information is used for decision-making

② Use the information that is most beneficial to the goals

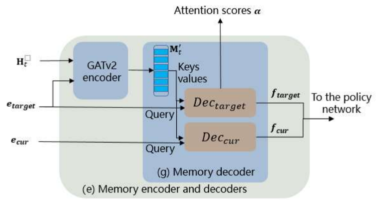


Retained STM + LTM



GATv2 Encoding

Working memory



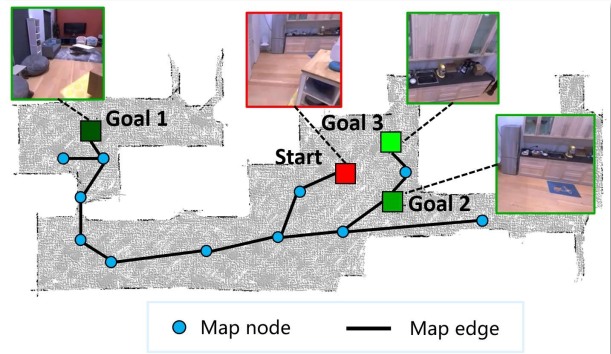
Decoding

Policy Network → Navigation actions

Decision-making

3 MemoNav: Experiments

Quantitative comparison on multi-goal navigation tasks



Scene	Methods	1-goal SR	1-goal SPL	2-goal PR	2-goal PPL	3-goal PR	3-goal PPL	4-goal PR	4-goal PPL
G	ANS [10]	30.0	11.0	-	-	-	-	-	-
	NTS [11]	43.0	26.0	-	-	-	-	-	-
	CNNLSTM [44]	53.1	39.2	31.5	10.6	18.0	2.8	12.4	1.6
	TSGM [22]	70.3	50.0	27.8	16.1	17.4	**10.4**	13.4	4.6
	VGM [23]	70.0	55.4	42.9	17.1	29.5	7.0	21.5	4.1
	MemoNav (ours)	**74.7**	**57.9**	**50.8**	**20.1**	**38.0**	9.0	**28.9**	**5.1**
M	CNNLSTM [44]	16.2	9.8	10.8	2.6	7.7	1.4	-	-
	TSGM [22]	24.0	14.6	13.5	**6.2**	7.8	**3.8**	-	-
	VGM [23]	25.1	**16.6**	16.7	5.0	11.8	2.5	-	-
	MemoNav (ours)	**26.1**	16.3	**19.5**	5.6	**13.6**	2.9	-	-

(SR/PR: Success Rate, SPL/PPL: Path length-weighted success rate)

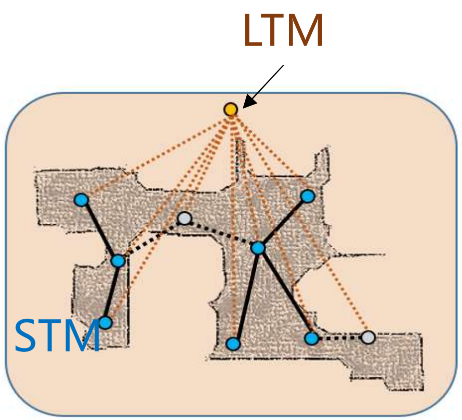
Analysis:

① The SR of MemoNav on multi-goal tasks outperforms the others significantly.

② MemoNav achieves leading performances consistently on two popular scene datasets

3 MemoNav: Experiments

Ablation Study of the proposed working memory model



	Components			1-goal		2-goal		3-goal		4-goal	
	Forget	LTM	WM	SR	SPL	PR	PPL	PR	PPL	PR	PPL
1				52.1	46.7	42.9	17.1	29.5	7.0	21.5	4.1
2	✓			55.1	46.1	44.9	17.5	29.4	6.5	21.5	4.2
3		✓		58.9	49.7	43.8	17.8	29.6	6.9	25.1	4.0
4	✓	✓		60.6	49.9	48.1	19.5	37.5	**9.1**	28.8	4.9
5		✓	✓	61.1	48.9	47.6	17.8	33.7	7.9	27.4	5.0
6	✓	✓	✓	**62.4**	**50.7**	**50.8**	**20.1**	**38.0**	9.0	**28.9**	**5.1**

(SR/PR: Success Rate, SPL/PPL: Path length-weighted success rate)

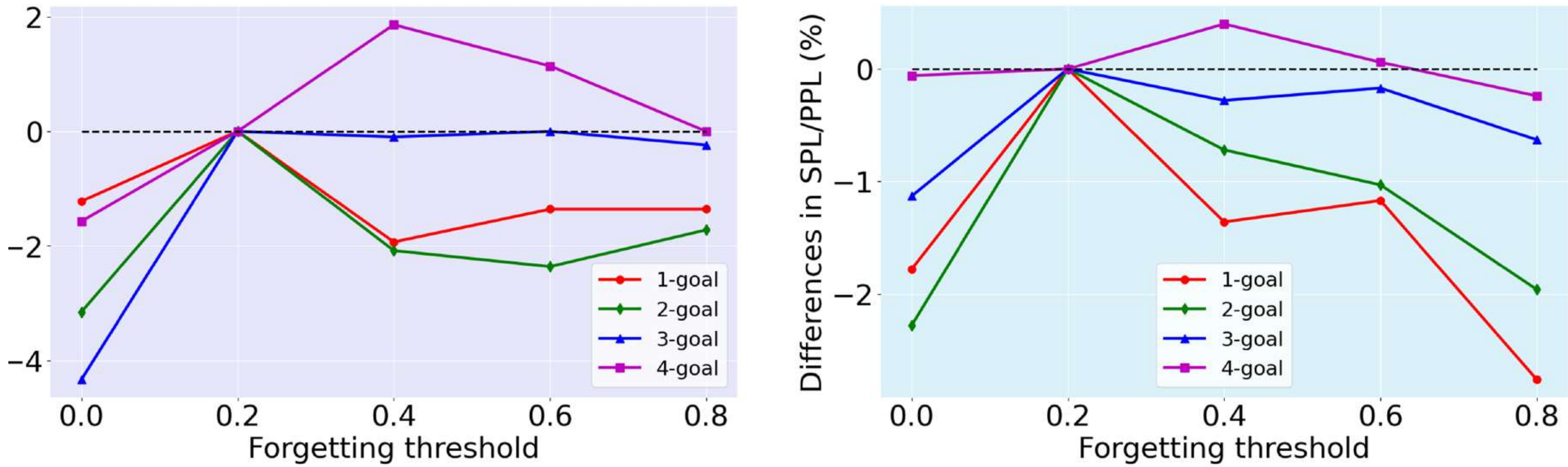
Analysis:

① Applied independently, the forgetting module and LTM both improves performance. The combination of the two brings larger gains

② The synergy among the three components leads to the best performance

3 MemoNav: Experiments

Ablation study of forgetting threshold p



Analysis:

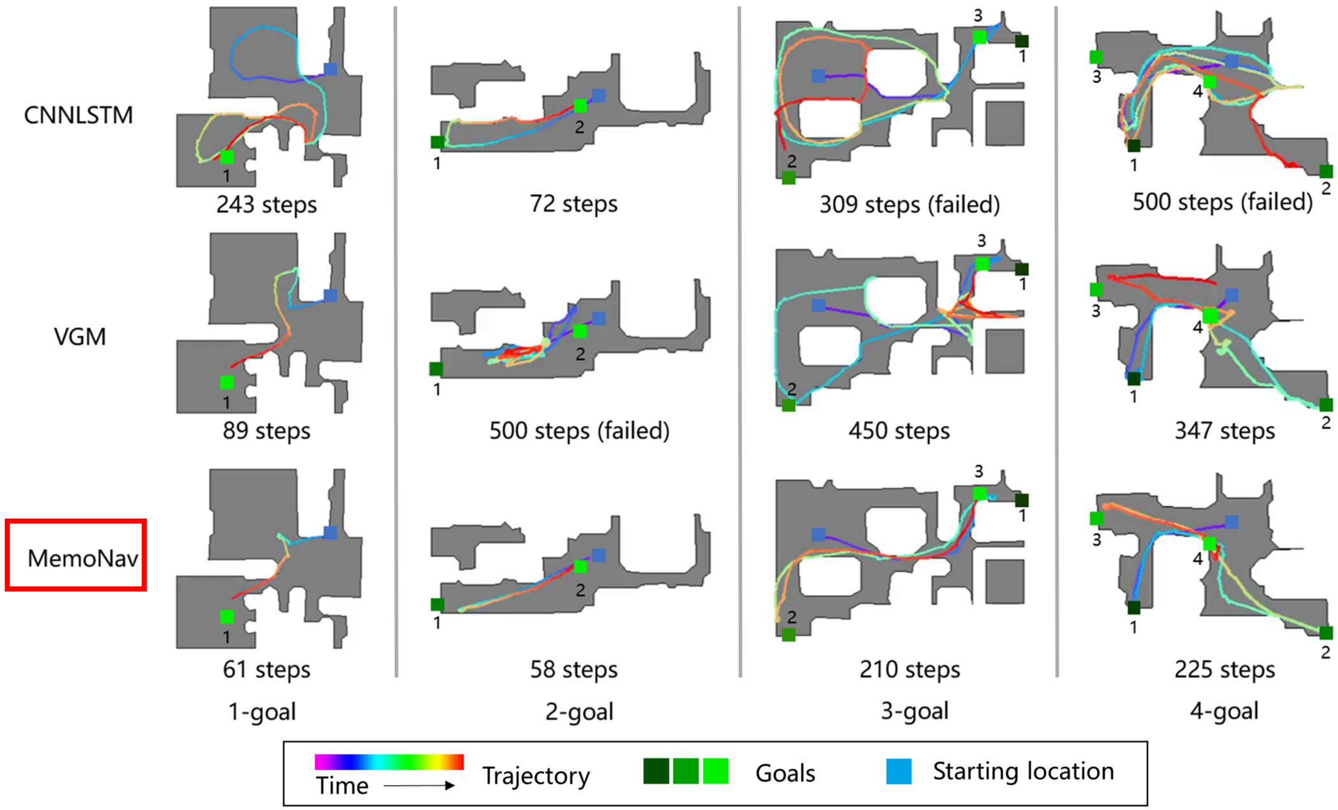
① MemoNav performs the best on easier tasks with a lower p but a higher p is more beneficial for harder tasks.

② MemoNav maintains high SR while forgetting **40%** of STM on the 4-goal tasks.

3 MemoNav: Experiments

Qualitative comparison

Analysis: Our MemoNav explores the scenes more efficiently, plans faster paths, and owns greater ability to get rid of deadlocks.

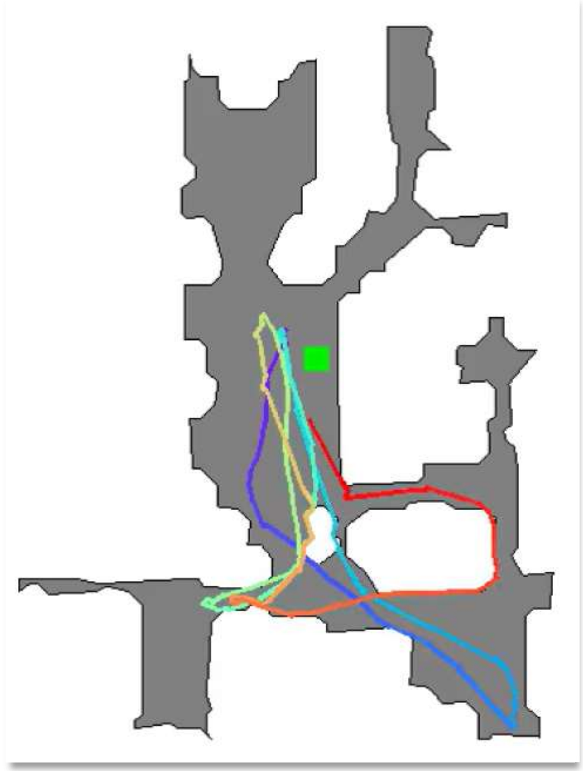


3 MemoNav: Experiments

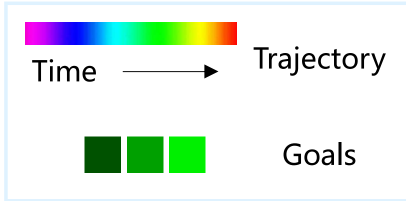
Qualitative comparison: visualization of multi-goal trajectories

Our MemoNav (Faster)

V.S.



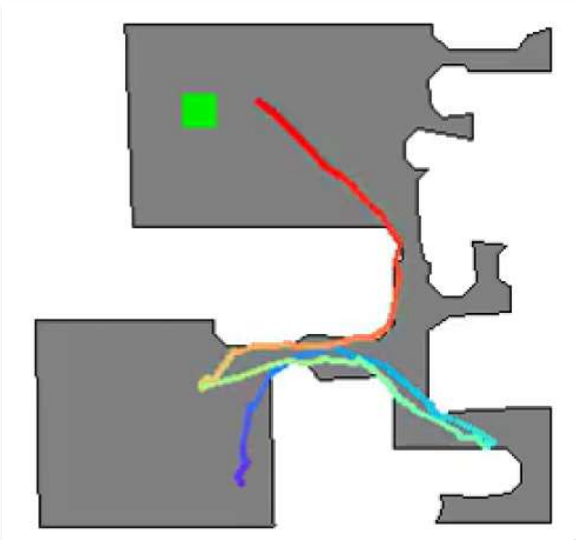
Baseline method (VGM)



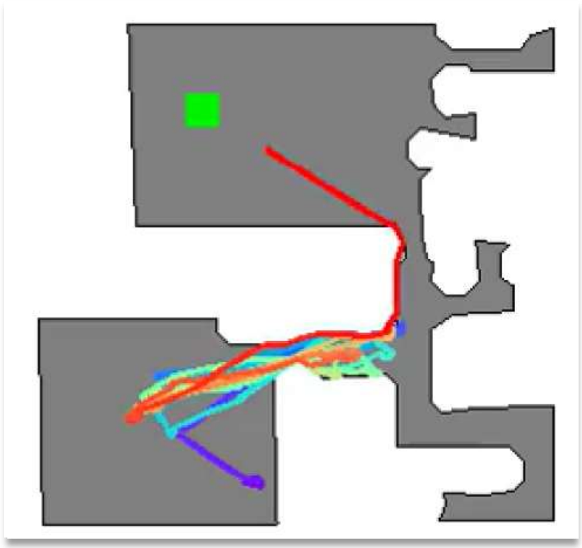
Visual Graph Memory With Unsupervised Representation for Visual Navigation, ICCV2021

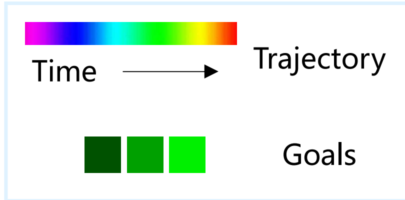
3 MemoNav: Experiments

Qualitative comparison: visualization of multi-goal trajectories



V.S.





Our MemoNav (Faster)

Baseline method (VGM)

Visual Graph Memory With Unsupervised Representation for Visual Navigation, ICCV2021

Thanks for watching!