This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Explore with Long-term Memory: A Benchmark and Multimodal LLM-based Reinforcement Learning Framework for Embodied Exploration

Sen Wang[1*] Bangwei Liu[1] Zhenkun Gao[1] Lizhuang Ma[1] Xuhong Wang[2] Yuan Xie[1] Xin Tan[1,2†]

[1]East China Normal University [2]Shanghai AI Laboratory

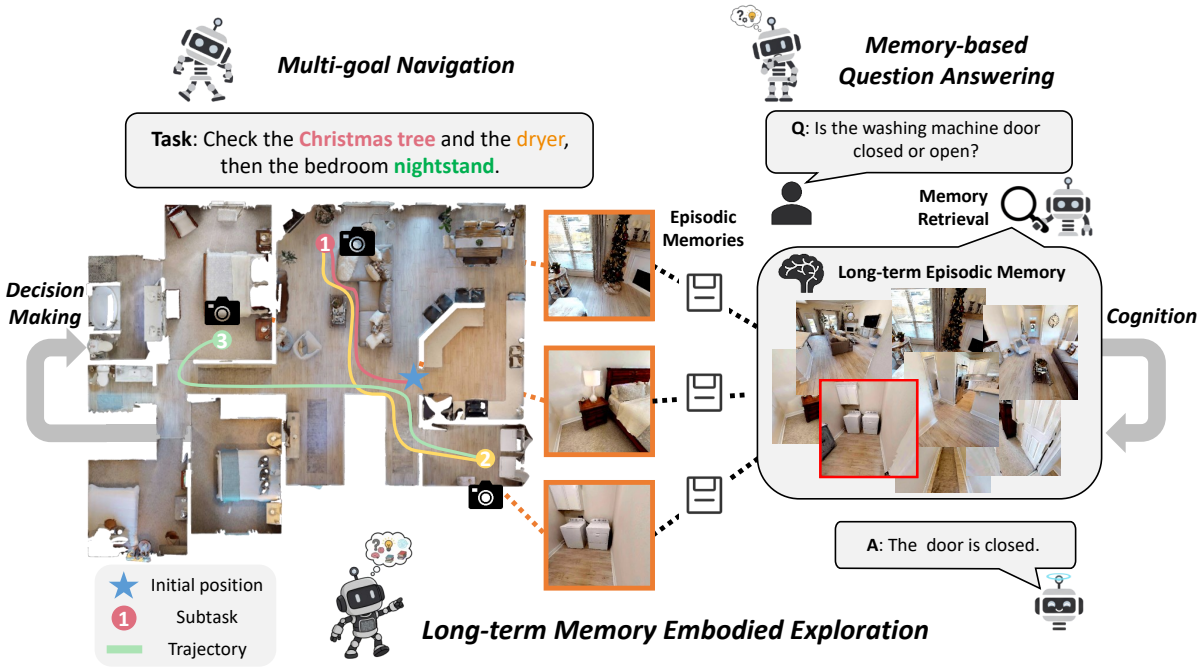

Figure 1. We propose Long-term Memory Embodied Exploration, which aims to collect episodic memories during Multi-goal Navigation and introduces Memory-based Question Answering to unify and evaluate the model's cognitive and decision-making abilities.

Abstract

An ideal embodied agent should possess lifelong learning capabilities to handle long-horizon and complex tasks, enabling continuous operation in general environments. This not only requires the agent to accurately accomplish given tasks but also to leverage long-term episodic memory to optimize decision-making. However, existing mainstream one-shot embodied tasks primarily focus on task completion results, neglecting the crucial process of exploration and memory utilization. To address this, we propose Long-term Memory Embodied Exploration (LMEE), which aims to unify the agent's exploratory cognition and decision-making behaviors to promote lifelong learning. We further construct a corresponding dataset and benchmark, LMEE-Bench, incorporating multi-goal navigation and memory-based question answering to comprehensively evaluate both the process and outcome of embodied exploration. To enhance the agent's memory recall and proactive exploration capabilities, we propose MemoryExplorer, a novel method that fine-tunes a multimodal large language model through reinforcement learning to encourage active memory querying. By incorporating a multi-task reward function that includes action prediction, frontier selection, and question answering, our model achieves proactive exploration. Extensive experiments against state-of-the-art embodied exploration models demonstrate that our approach achieves significant advantages in long-horizon embodied tasks. Our dataset and code will be released at https://wangsen99.github.io/papers/lmee/.

†Corresponding author.

*This work was done by Sen Wang during an internship at Shanghai AI Laboratory.

1. Introduction

A key objective in embodied intelligence is to empower agents with lifelong learning capabilities, enabling them to perform complex tasks and operate continuously in dynamic and unfamiliar environments. For example, as illustrated in Fig. 1, consider the instruction "*Check the Christmas tree and the dryer, then the bedroom nightstand.*" After completing this sequence of tasks, the agent should have developed a comprehensive understanding of the explored environment. Later, when asked a follow-up question such as "*Is the washing machine door closed or open?*", the agent should be able to quickly retrieve its stored memory and respond, "*The door is closed.*" This capability represents more than simple task completion, as it demonstrates the agent's ability to construct dynamic and context-aware memory representations that support efficient recall and reasoning for future interactions. Such an ability is essential for developing embodied agents that can continuously learn, adapt, and evolve in complex real-world environments [18].

Embodied exploration aims to enable agents to proactively explore unknown environments. As shown in Tab. 1, current research paradigms mainly focus on tasks such as goal navigation [3, 19, 49, 54] and embodied question answering [16, 22, 31, 57]. However, these one-shot tasks tend to emphasize outcomes while neglecting the exploration process itself. Although multi-goal navigation [18] focuses on long-horizon embodied tasks, it still overlooks how the exploration process contributes to the agent's scene understanding and decision-making. Achieving a unified integration of cognition and decision-making is therefore crucial for developing general embodied intelligence.

Multimodal Large Language Models (MLLMs) [17, 20, 23, 37] have shown remarkable potential in embodied exploration, particularly for complex scenes [39, 40, 47, 48, 58]. However, existing approaches still struggle to make effective use of memory. Many methods treat memory passively, limiting the agent's autonomy and reasoning capabilities. For example, imitation learning-based approaches [58, 62] train agents to replicate expert trajectories. This passive learning paradigm restricts generalization to unseen scenarios and, crucially, prevents the agent from developing its own proactive exploration strategies. Other vision-language exploration methods [52, 57] that depend on memory snapshots use filtering strategies to mitigate the constraints of limited context windows but fail to harness the active querying capability inherent in MLLMs. Likewise, models with long-term spatiotemporal memory [15] mainly perform post-exploration reasoning, missing the opportunity to use memory to proactively guide exploration.

To address these challenges, firstly, we introduce Long-term Memory Embodied Exploration (LMEE), which emphasizes not only the outcome (goal) of the embodied task but also the exploration process (memory). As illustrated in Fig. 1, LMEE consists of two key components: Multi-goal Navigation and Memory-based Question Answering. During navigation, the agent dynamically constructs on-the-fly memories. LMEE further provides a large collection of exploration-related questions that the agent must answer based on its memories. To evaluate LMEE, we construct a comprehensive dataset and establish a benchmark, LMEE-Bench, which assesses the agent from two perspectives: (1) success rate and efficiency in multi-goal navigation, and (2) accuracy in memory-based question answering. This assesses the agent's ability to use episodic memory. The dataset encompasses 246 object categories, over 9,000 goals and questions, and 1,982 exploration trajectories.

Secondly, we propose MemoryExplorer, an MLLM-based framework for active exploration and memory retrieval trained via Reinforcement Fine-Tuning (RFT). The model proactively queries and invokes memory retrieval tools to access multimodal memory information based on current task instructions, multi-view observations, and goal-related questions. By analyzing its answers and assessing task progress, the agent gains an understanding of the current situation and plans subsequent actions accordingly. Furthermore, we design a Multi-Task Reward function that integrates action and frontier prediction with question answering, effectively unifying scene understanding, memory utilization, and planning-based decision-making. This enables the model to tackle challenging tasks in complex environments. Extensive experiments demonstrate that our approach achieves superior performance in long-term memory embodied exploration, significantly enhancing the model's capacity for autonomous exploration and active memory retrieval, which are key abilities for realizing lifelong learning in embodied agents.

In summary, our contributions are as follows:

- We introduce LMEE, a new paradigm for developing autonomous agents by unifying exploration with memory-based reasoning. We also present its corresponding benchmark, LMEE-Bench, to holistically evaluate agents using multi-goal navigation and question answering to assess the crucial abilities of memory utilization, cognitive understanding, and decision-making.
- We propose MemoryExplorer, which uses reinforcement learning to enable active exploration and memory retrieval in unknown environments. By combining frontier prediction, action planning, and question answering, we design a multi-task reward function that optimizes the model's policy for effective reasoning in complex scenes.

2. Related Work

2.1. Embodied Navigation and Question Answering

Goal-driven Navigation and Vision-Language Question Answering are mainstream tasks in embodied intelligence.

Table 1. Comparison to popular embodied exploration benchmarks.

Dataset	Train		Test		Multi-goal	Memory	Question Num	Question Type	
	Categories	Goals	Categories	Goals				Choices	Open Vocab
ObjectNav-MP3D [3]	21	7509	-	-	✗	✗	-	-	-
ObjectNav-HM3D [49]	6	5216	-	-	✗	✗	-	-	-
InstanceImageNav-HM3D [19]	6	2516	-	-	✗	✗	-	-	-
HM3D-OVON [54]	280	10987	49	1278	✗	✗	-	-	-
GOAT-Bench [18]	193	13025	36	1282	✓	✗	-	-	-
MP3D-EQA [42]	-	-	-	-	✗	✗	1136	✗	✗
HM-EQA [31]	-	-	-	-	✗	✗	500	✓	✗
OpenEQA [22]	-	-	-	-	✗	✗	1600+	✗	✓
EXPRESS-Bench [16]	-	-	-	-	✗	✗	2044	✗	✓
MemoryEQA [57]	-	-	-	-	✗	✓	1587	✓	✗
LMEE (Ours)	208	8880	38	828	✓	✓	9286 (406)	✓	✓

Early navigation tasks mostly targeted single targets [5, 14, 29, 43], and due to the limitations of current model generalization capabilities, navigation methods are based on modular designs and only employ modality-specific encoders. Similarly, the questions in embodied question answering tasks are relatively simple [8, 42, 56]. In recent years, MLLMs have shown impressive results in the field of embodied intelligence [11, 12, 21, 46, 60]. Whether in open question answering [22] or multimodal navigation [2, 18, 53], embodied tasks are moving towards a long-term, generalized paradigm [7, 27, 58]. However, current tasks focus on the outcome, such as whether the target object has been found [16], or whether the answer to the embodied question is accurate [31]. This lack of attention to the process is not conducive to building lifelong learning agents [36]. We introduce Long-term Memory Embodied Exploration, aiming to focus on the preservation of contextual memory during long-range task exploration, enabling the agent to think about problems based on experience like a human, ultimately achieving lifelong learning and self-evolution.

2.2. Memory-based Agents

Several works have explored memory mechanisms for language-based agents. For example, MemGPT [25], MemAgent [55], and Mem-α [41] provide large language models with extended context via diverse memory systems, while ReasoningBank [24] builds experience pools to support long-term tasks such as web browsing and software engineering. However, these text-based or discrete-state-based memory models are not directly applicable to embodied AI, where memory must capture spatio-temporal information from the physical world.

Memory is crucial for long-horizon embodied tasks, which are growing increasingly complex [51]. Methods like MTU3D [62], 3D-Mem [52], and 3DLLM-Mem [15] construct memory banks or snapshots to support navigation, object localization, and spatio-temporal reasoning. Yet, these approaches mostly rely on passive memory usage, *e.g.*, imitating trajectory data [15, 62]. In contrast, we aim to enable active memory querying to enhance proactive exploration and task handling in embodied settings.

2.3. Reinforcement Learning for Embodied AI

The advantage of reinforcement learning lies in the agent's ability to learn actively [34], and the strong generalization capability of LLMs further amplifies this benefit. Recent work leverages LLMs to automatically learn reward models from interaction data without manual annotation [6, 32]. Reinforcement learning for training multimodal LLMs to solve embodied tasks has also gained increasing attention. [35] trains perception and action jointly through real-time environmental interaction, enabling generalization across embodied tasks. [30] allows agents to ask clarifying questions for ambiguous instructions, while [38] introduces reinforcement fine-tuning to support self-evolution of embodied agents. However, early approaches rely on overly simple task instructions [35], and active questioning remains limited in scope [30], restricting applicability to complex scenarios. Other methods emphasize only task completion [10, 26, 38] and overlook the task process, which hinders lifelong learning. In contrast, we emphasize the memory retrieval capability of embodied agents for complex long-horizon tasks, enabling better scene understanding and more efficient decision-making.

3. Data Construction of LMEE

To enable agents to proactively explore and memorize in unknown environments, we construct a Long-term Memory Embodied Exploration dataset. Based on multi-goal navigation tasks, it dynamically builds a memory bank by collecting observations during the exploration process, thereby enabling memory-based embodied exploration training and evaluation. As shown in Fig. 2, the data construction process consists of three parts: task instruction generation, ex-

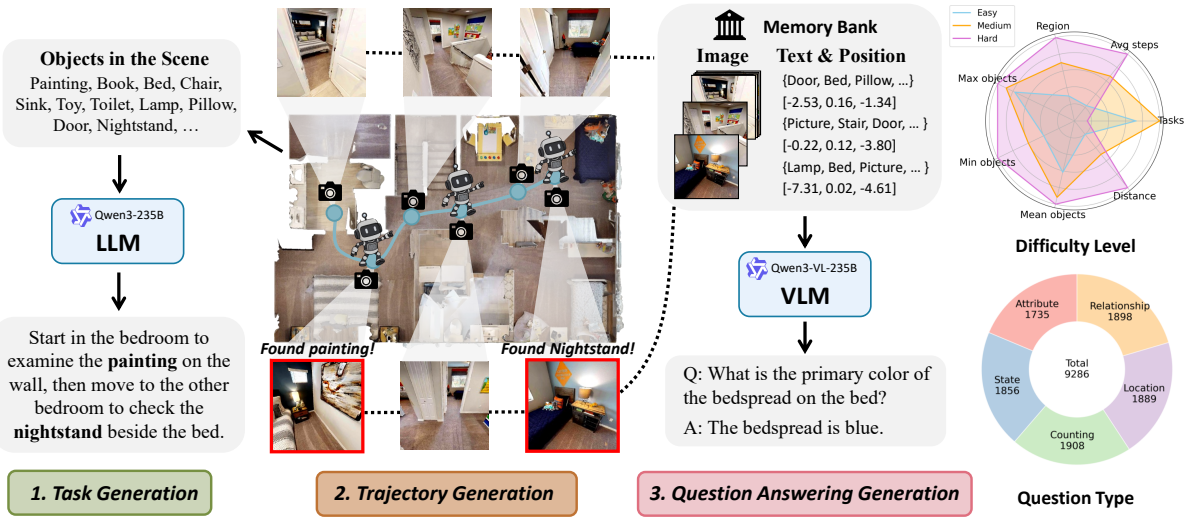

Figure 2. The construction process of Long-term Memory Embodied Exploration and data statistics.

ploration trajectory generation, and memory-based question answering generation.

Task Instruction Generation. We use the real-world HM3DSem [50] dataset, including 145 training scenes and 36 test scenes with semantic labels. From these scenarios, we collected over 200 different categories of targets based on the Goat-Bench [18] dataset. Then, different regions and their corresponding objects are fed into the Large Language Model (LLM), which combines them to generate task instructions and goals.

Exploration Trajectory Generation. Based on the agent's initial position and the positions of different targets, we used the Habitat-Sim [33] to plan the exploration path, thereby generating a multi-goal, step-by-step exploration sequence trajectory, including the corresponding action, observation, position, and rotation for each step. This is beneficial for the model to learn action-based step-by-step exploration in long-horizon planning. Simultaneously, we utilize an image tagging model [59] to label object information in each image as a text description, thereby constructing a multi-modal memory bank including text, position, and image.

Memory Bank. To facilitate efficient data sampling, we construct a memory bank:

$$\mathcal{M} = \{(p_i, f_i, o_i) \mid i = 1, \ldots, n\}, \tag{1}$$

where each entry stores the position p_i, the text features f_i, and the image features o_i for each step i. We employ CLIP [28] to extract both o and f, and compute their pairwise similarity using dot products. The overall similarity between the current state (p_c, f_c, o_c) and a memory entry (p_i, f_i, o_i) is defined as:

$$s_i = \omega_f (f_c^\top f_i) + \omega_o (o_c^\top o_i) + \omega_p \, \mathrm{dist}(p_c, p_i), \tag{2}$$

where ω_f, ω_o, and ω_p are the weighting coefficients for text, visual, and distance similarities, respectively, and $\mathrm{dist}(p_c, p_i)$ denotes an exponential function of the Euclidean distance between locations. To maintain temporal consistency, we further aggregate similarity scores from the k most recent samples by computing their mean and standard deviation, and dynamically filter contextual memories at each step based on an adaptive similarity threshold.

Difficulty Level. To evaluate the model's exploration capabilities and long-term memory, we categorize tasks into three levels: easy, medium, and difficult, based on the number of regions and goals to be explored and the distance from the initial position to the target object. Detailed statistics are presented in Fig. 2.

Question Answering Generation. We leverage a Vision Language Model (VLM) to generate question-answer pairs based on observation images associated with the navigation targets. This design serves two main purposes. First, questions are focused on navigation targets to avoid asking about objects an agent may not have observed due to non-unique trajectories, thus providing a more reliable assessment of its memory. Second, goal-oriented questions help the model understand the progress of multi-goal long-term tasks during training, determine the next goal to be found, and thus plan actions and paths. As shown in Fig. 2, the types of questions include five categories: attribute, counting, location, relationship, and state. The types of answers include two forms: open-ended and optional.

Continuous Actions. We found that in many cases, the simulator cannot execute multi-step continuous actions, which greatly increases the difficulty of the model predicting actions. Therefore, we utilize a continuous action window to sample x consecutively occurring identical actions, using one of such samples as training data.

Dataset details. We generate step-by-step trajectory data using the training and testing scenes from HM3DSem,

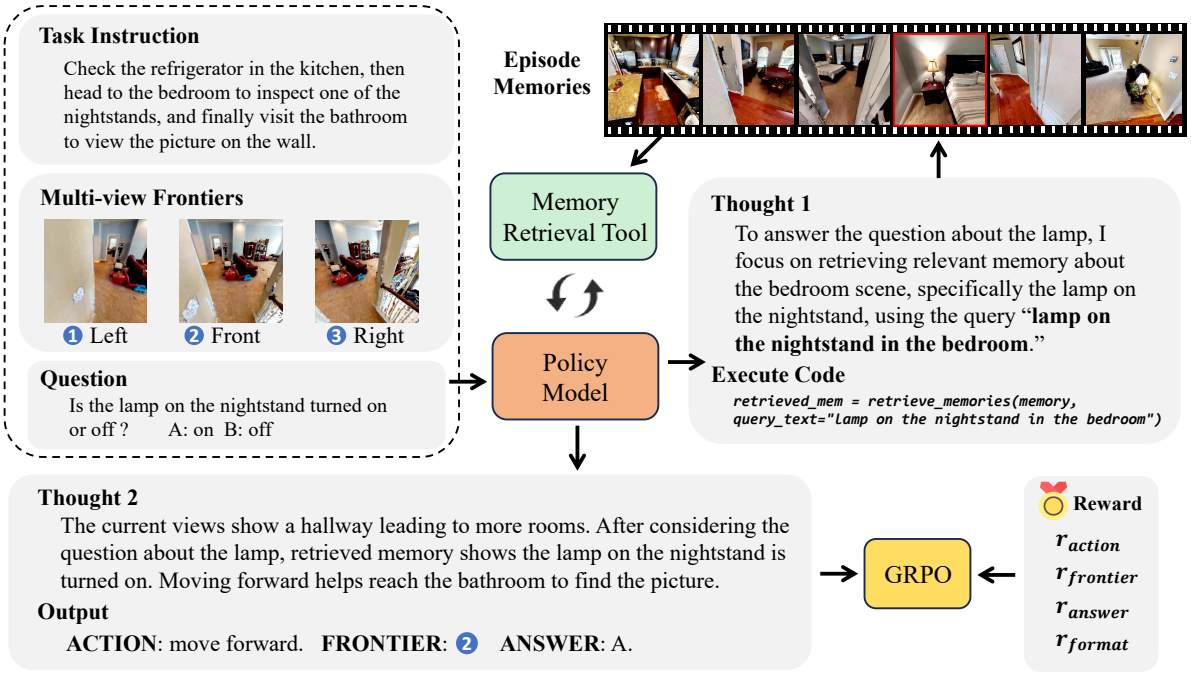

Figure 3. Illustration of training in MemoryExplorer. Given a task instruction, the multi-view observations, and a goal-oriented question. Model retrieves relevant multimodal memories from the episodic memory bank using tools, analyzes the current information alongside the retrieved memories to understand the progress of the long-term task, and performs `ACTION` prediction, `FRONTIER` selection, and question `ANSWER`. The policy model output response calculates the reward using a Multi-Task Reward function and is fine-tuned using GRPO.

which serve as the training and testing sets, respectively, as shown in Tab. 1. We use Qwen3-235B-A22B-Instruct to generate task instructions and Qwen3-VL-235B-A22B-Instruct to generate question answering. The model's action space includes moving forward (0.25 m) and turning left or right (30°), so each step contains views from these three directions. The full dataset comprises 1,982 tasks with a total of 377,311 entries, of which 1,816 tasks are allocated for training and 166 tasks for testing. After sampling, we obtain 11,684 instances for training. For more details, please refer to the *Supplementary Material*.

4. Method

Unlike previous modular learning-based vision-language navigation [4, 18, 62], we utilize an MLLM to train an end-to-end embodied agent with proactive exploration awareness. In long-term embodied exploration tasks, MLLM needs to integrate task instructions, observed images, and long-term memories for scene planning and decision-making. However, due to the limitations of the context window, the model cannot access all long-term memories at once, and supervision of long-term multi-step actions may weaken the agent's autonomous exploration ability. Therefore, we propose MemoryExplorer, an embodied exploration model based on reinforcement learning with memory retrieval as shown in Fig. 3.

Task Definition. For LMEE, the model policy can be represented as $\pi_\theta(I, O, Q; M)$ of parameterized model weights θ. Model inputs include task instructions I, current multi-view observations O, a memory-based goal-oriented question Q, and externally stored long-range memory M including image-text pairs. The model's final response can be represented as $y = (S, F, A)$, with outputs including a single-step action S, a frontier F, and an answer A.

Memory Retrieval. The proposed Long-term Memory Embodied Exploration dataset provides the multimodal contextual memories during the exploration process. The model is guided to generate code that invokes the external memory retrieval tool $\mathcal{R}$, subsequently obtaining the corresponding memories for reasoning and answering the given question. The memories obtained from invoking $\mathcal{R}$ are fed back to the model as additional input, allowing for richer reasoning to support the final answer. Due to the limitations of multi-image input, this study focuses on single-round tool invocation scenarios. Since the question is related to the navigation goal, the retrieved memories further help the model understand the task progress and determine the next exploration action.

Formally, when the model retrieves memories, the first round generates an initial response $y' \sim \pi_\theta(\cdot \mid I, O, Q)$ containing the tool call and query text. Then, the memory retrieval tool $M_{re} = \mathcal{R}(y', M)$ is called to obtain the most relevant memories retrieved from long-term memory. In the

external Python environment, the text and observations in memory are encoded as text features f_t and observation features f_o by CLIP [28], respectively, while the query text is encoded as query features f_q. Cosine similarity is then calculated to obtain the $topk$ most similar memories. Finally, the retrieved memories m_i are obtained by combining the results from both. The retrieval process can be represented as:

$$\mathcal{R} = \{m_i \mid i \in \text{top-}k(\cos(f_q, f_i^{(t,o)}))\}. \tag{3}$$

Finally, the final response can be represented as $y \sim \pi_\theta(\cdot \mid I, O, M_{re})$. Additionally, if the memory retrieval tool call fails, the final response is the first-round response, *i.e.*, $y \sim \pi_\theta(\cdot \mid I, O, Q)$.

Reinforcement Fine-Tuning. MemoryExplorer utilizes RFT to learn active exploration and memory retrieval. Therefore, our training objective is:

$$\max_{\pi_\theta} \quad \mathbb{E}_{(I,O,Q)\sim D,\, y\sim\pi_\theta(\cdot|I,O,Q;M)} \left[r_\phi(I, O, Q, y)\right] - \beta\, D_{\text{KL}}(\pi_\theta(\cdot \mid I, O, Q; M) \,\|\, \pi_{\text{ref}}(\cdot \mid I, O, Q; M)) \,. \tag{4}$$

Note that our RFT training process does not optimize intermediate responses from memory retrieval tools, as our goal is to encourage the model to think and make decisions autonomously, using final reward feedback to evaluate the effectiveness of tool calls. For the specific training method, we employ Group Relative Policy Optimization (GRPO) [13], a well-established policy gradient method that estimates the baseline model by sampling response data, thus saving training resources.

Reward Modeling. Unlike single-question-answer training [9, 44, 45], we adopt a multi-task training approach that includes exploration actions, frontier selection, and memory-based question answering. This design enables the model to understand spatial and action relationships while actively invoking memory retrieval tools and generating retrieval content, thereby facilitating autonomous exploration.

To achieve multi-dimensional response optimization, we design a Multi-Task Reward function r_{total} that integrates four complementary components: action accuracy, frontier correctness, answer precision, and output format completeness. The total reward is defined as follows:

$$\begin{aligned} r_{\text{total}} =& w_{act} \cdot r_{\text{action}} \cdot c + w_{front} \cdot r_{\text{frontier}} \cdot c \\ &+ w_{ans} \cdot r_{\text{answer}} + w_{fmt} \cdot r_{\text{format}}, \end{aligned} \tag{5}$$

where $r_{\text{action}}, r_{\text{frontier}}, r_{\text{answer}}, r_{\text{format}} \in [0, 1]$ represent the sub-rewards for action accuracy, frontier correctness, answer precision, and output format completeness, respectively. Specifically, r_{answer} reflects the accuracy of the predicted answer, and r_{format} encourages structured and parsable responses, determined by whether the output includes complete `ACTION`, `FRONTIER`, and `ANSWER` segments. c is a consistency coefficient that penalizes logically inconsistent pairs between action and frontier. $w_{act}, w_{front}, w_{ans}, w_{fmt}$ denote the weighting coefficients for each reward component.

To further differentiate performance in scenarios involving tool assistance and tool invocation failure, a scaling factor α is applied to each sub-reward. This adjustment reduces all sub-scores when no external tool is employed, while amplifying them in tool-based reasoning conditions, thereby encouraging efficient tool utilization. The final reward r_{total} is clipped to the range $[0, 1]$ to ensure stability and comparability across tasks.

5. Experiments

We propose the Long-term Memory Embodied Exploration benchmark. It consists of two components: multi-goal navigation and memory-based question answering. The agent is first required to perform multi-goal navigation in an unknown environment, storing and utilizing memories from the exploration process to complete the tasks. Subsequently, the agent must answer questions related to the navigation targets based on its memory. Additionally, during multi-goal navigation, the agent can leverage memory retrieval to locate specified targets. We also evaluate the agent's exploration and memory retrieval capabilities on the multimodal lifelong navigation benchmark platform, GOAT-Bench.

Experimental details. Our model is trained based on Qwen2.5-VL-7B-Instruct [1]. We use EasyR1, a simplified version of the VERL framework. The learning rate is set to 1e-6, and a KL penalty coefficient of 0.1 is applied to maintain training stability. Training is conducted on 8 NVIDIA H200 GPUs for 160 steps with a global batch size of 128. The $topk$ is set to 3. The consistency coefficient c is set to 0.5. The scaling factor α is set to 1.2 when tool assistance is involved, and to 0.5 for r_{answer} and r_{format}, and 0.6 for r_{action} and r_{frontier} when tool invocation fails. More experimental details are provided in the *Supplementary Material*.

Benchmark. LMEE-Bench consists of 166 tasks with 828 goals and 406 questions. Due to resource constraints, we randomly select a subset of 58 tasks for evaluation, covering 272 goals and 145 questions. We also provide results on the full test set in the *Supplementary Material*. LMEE consists of two types of tasks: multi-goal navigation and goal-oriented memory-based question answering. The questions cover five types: attributes, counting, location, relationships, and states, while the answers are provided in two formats: open-ended and multiple-choice. **GOAT-Bench** is a multimodal lifelong navigation benchmark. The task requires agents to navigate to multiple targets. Each target is described by category name, language description, and image. Due to the large scale and limited resources of GOAT-Bench, we followed 3D-Mem to evaluate a subset of "Val Unseen" set, including 36 scenarios, one exploration round each, and 278 navigation subtasks in total.

Table 2. Experiments on LMEE-Bench. Score represents the MLLM-Score for open-ended answers, and Acc represents the accuracy rate of the answer choices.

Method	Easy				Medium				Hard				Total			
	Nav		QA		Nav		QA		Nav		QA		Nav		QA	
	SR	SPL	Score	Acc	SR	SPL	Score	Acc	SR	SPL	Score	Acc	SR	SPL	Score	Acc
Retrieval-Augmented QA																
LLaVA-OneVision-7B [20]	-	-	35.71	57.14	-	-	42.02	62.77	-	-	25.00	43.75	-	-	38.62	59.31
Llama3.2-Vision-11B [23]	-	-	43.57	48.57	-	-	28.19	37.23	-	-	20.31	31.25	-	-	31.03	39.31
Qwen2.5-VL-7B [1]	-	-	36.43	65.71	-	-	43.62	53.19	-	-	42.19	43.75	-	-	41.72	55.86
InternVL3-8B [61]	-	-	42.14	65.71	-	-	41.22	54.26	-	-	26.56	43.75	-	-	39.83	55.17
Qwen3-VL-8B [37]	-	-	43.57	65.71	-	-	36.70	58.51	-	-	32.81	62.50	-	-	37.93	60.69
GPT-4o	-	-	43.57	65.71	-	-	37.50	51.06	-	-	40.62	56.25	-	-	39.31	55.17
Gemini-2.5-Flash	-	-	44.29	51.43	-	-	37.77	50.00	-	-	23.44	50.00	-	-	37.76	50.34
MLLM Exploration																
Explore-EQA [31]	4.26	4.26	-	-	15.17	8.67	-	-	14.89	7.25	-	-	13.24	7.66	-	-
3D-Mem [52]	21.28	9.68	30.71	42.86	15.73	6.09	34.57	44.68	17.02	6.97	25.00	18.75	16.91	6.86	32.59	41.38
RA-Mem	25.53	15.65	34.29	54.29	**21.35**	12.14	37.77	62.77	14.89	8.86	25.00	43.75	20.96	12.18	35.52	58.62
MemoryExplorer (Ours)	**31.91**	**21.11**	**35.71**	**68.57**	**21.35**	**14.03**	**48.14**	**63.83**	**23.40**	**12.51**	**34.38**	**68.75**	**23.53**	**14.99**	**43.62**	**65.52**

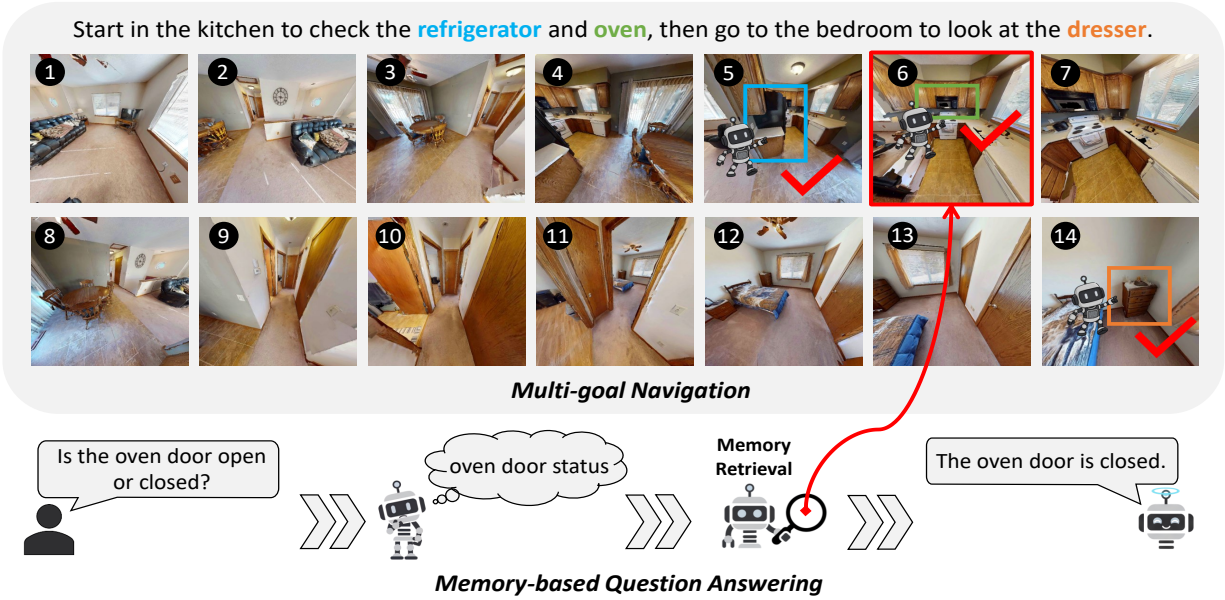

Figure 4. Qualitative example of LMEE-Bench.

Metric. We evaluate performance using Success Rate (SR) and Success weighted by Path Length (SPL). A navigation episode is considered successful if the agent's final position is within 1 meter of the target. For goal-oriented open-ended questions, we propose MLLM-Score, a quantitative metric leveraging an MLLM. For each question, the ground-truth answer, the observation of the target object, and the predicted answer are provided to the MLLM. Due to resource constraints, we use Qwen3-VL-30B-A3B-Instruct as our evaluation model. The MLLM assigns a score from 1 to 5 to each prediction to assess its quality. The MLLM-Score measures answer accuracy by averaging the scores across all questions and converting the result to a 0–100 scale. For multiple-choice questions, we evaluate predicted answers using standard accuracy.

Baselines. For the baseline evaluation of long-term memory exploration with MLLMs, we primarily compare our method with Explore-EQA [31] and 3D-Mem [52]. Since Explore-EQA lacks memory capabilities, we only evaluate its performance on multi-goal navigation tasks. 3D-Mem serves as an active exploration baseline leveraging long-term memory, enabling agent exploration and lifelong learning through memory snapshots and frontier snapshots. In addition, we develop a Retrieval-Augmented Memory (RA-Mem) variant based on 3D-Mem. While 3D-Mem uses an object-based memory filtering approach to limit the model's context window, it does not fully exploit the potential of MLLMs. In contrast, RA-Mem independently generates queries based on the task and current observations to retrieve memory and guide the agent in completing

Table 3. Experiments on GOAT-Bench. Evaluated on the "Val Unseen" split. Methods denoted by * are from GOAT-Bench, and those with † are evaluated on the subset. All MLLM-based exploration methods are implemented based on Qwen2.5-VL-7B.

Method	Success Rate	SPL
GOAT-Bench Baselines [18]		
Modular GOAT*	24.9	17.2
Modular CLIP on Wheels*	16.1	10.4
SenseAct-NN Skill Chain*	29.5	11.3
SenseAct-NN Monolithic*	12.3	6.8
MLLM Exploration		
Explore-EQA [31]†	23.02	14.43
3D-Mem [52]†	37.05	20.26
RA-Mem†	42.81	21.95
MemoryExplorer (Ours)†	**46.40**	**28.03**

tasks. However, these methods still rely solely on MLLM reasoning. Our proposed MemoryExplorer extends RA-Mem by incorporating reinforcement learning, improving the model's active memory retrieval and exploration capabilities. Beyond active exploration, we also evaluate a post-exploration setting, where MemoryExplorer collects observations during navigation to construct a memory bank, and we compare the performance of different MLLM models using Retrieval-Augmented Question Answering.

Quantitative Comparison. As shown in Tab. 2, our model demonstrates higher robustness and efficiency compared to existing embodied exploration methods on LMEE-Bench. In addition, we evaluate current MLLMs using retrieval-augmented question answering. We observe that Qwen2.5-VL-7B performs better on open-ended questions, while Qwen3-VL-8B and LLAVA-OV-7B excel at multiple-choice questions. This suggests that misalignment between cognitive understanding and action decision-making may lead to hallucinations in the models.

The results on GOAT-Bench are shown in Tab. 3, RA-Mem is more flexible than the memory pre-filtering mechanism used in 3D-Mem, as it actively generates memory query texts to more effectively leverage long-term memory, thereby significantly improving the model's success rate. Moreover, MemoryExplorer further enhances the agent's success rate and efficiency in multi-goal long-horizon navigation, demonstrating the performance gains brought by improved utilization of exploratory memories.

Qualitative results on LMEE-Bench. To intuitively demonstrate the effectiveness of long-term memory–based embodied exploration, we present a test case on LMEE-Bench in Fig. 4. Through multi-goal exploration and memory accumulation, the agent progressively understands the environment. When a question is presented, the agent retrieves relevant memories and answers the question.

Ablation Study. Our baseline is RA-Mem, and we apply RFT to enhance the agent's ability to actively retrieve memories and exploration. We begin with a relatively simple task-progress question: "*Which objects in the task have already been found or completed?*" The answer corresponds to the discovered target objects, such as "*tv, couch, oven, sink.*" We then increase the difficulty by extending it to multiple-choice questions covering all five dataset question types. Finally, we combine the two settings above. As shown in Tab. 4, training the model using memory retrieval tools significantly improved model performance, while there is a non-linear positive correlation between the question types and model performance. Using only a single question type cannot achieve the best results, whereas incorporating a richer variety of question types enables the model to achieve strong performance.

Table 4. Ablation study on question type.

Setting	LMEE-Bench			GOAT-Bench	
	SR	SPL	Score	SR	SPL
Baseline	20.96	12.18	35.52	42.81	21.95
Simple	20.80	12.49	41.33	44.24	27.29
Multiple-choice	**23.53**	14.99	**43.62**	46.40	28.03
All	22.06	**15.13**	43.28	**48.20**	**29.36**

Visualization. We visualize the training reward curve and tool usage percentage in Fig. 5.The model gradually learns to invoke the memory-retrieval tool more accurately, which in turn leads to improved answer accuracy.

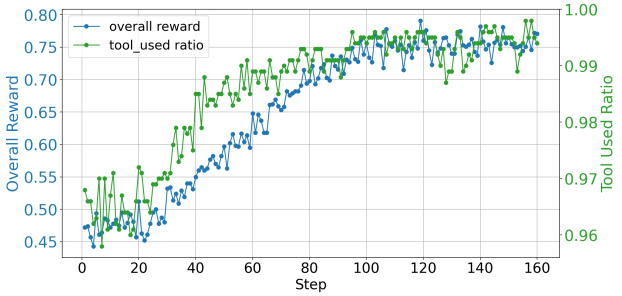

Figure 5. Training reward curve and tool usage percentage.

6. Conclusion

We propose Long-term Memory Embodied Exploration (LMEE), which constructs a episodic memory bank through Multi-goal Navigation and leverages Memory-based Question Answering to promote the integration of cognitive and decision-making. We further introduce MemoryExplorer, a reinforcement learning framework that trains the model to actively retrieve memory. A Multi-Task reward function combining action prediction, frontier selection, and question answering enables autonomous exploration and proactive memory use. Our approach achieves strong performance on both the LMEE-Bench and GOAT-Bench.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62302167, 62222602, 62502159, U23A20343, W2521174), the Shanghai Committee of Science and Technology (Grant Nos. 25511103300, 25511104302, 25511102700), the Natural Science Foundation of Chongqing (CSTB2023NSCQ-JQX0007), and the Young Elite Scientists Sponsorship Program by CAST YESS20240780.

References

[1] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 6, 7

[2] Yihan Cao, Jiazhao Zhang, Zhinan Yu, Shuzhen Liu, Zheng Qin, Qin Zou, Bo Du, and Kai Xu. Cognav: Cognitive process modeling for object goal navigation with llms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9550–9560, 2025. 3

[3] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158*, 2017. 2, 3

[4] Matthew Chang, Theophile Gervet, Mukul Khanna, Sriram Yenamandra, Dhruv Shah, So Yeon Min, Kavit Shah, Chris Paxton, Saurabh Gupta, Dhruv Batra, et al. Goat: Go to any thing. *arXiv preprint arXiv:2311.06430*, 2023. 5

[5] Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Russ R Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. *Advances in Neural Information Processing Systems*, 33:4247–4258, 2020. 3

[6] Zhenfang Chen, Delin Chen, Rui Sun, Wenjun Liu, and Chuang Gan. Scaling autonomous agents via automatic reward modeling and planning. *arXiv preprint arXiv:2502.12130*, 2025. 3

[7] Zhili Cheng, Yuge Tu, Ran Li, Shiqi Dai, Jinyi Hu, Shengding Hu, Jiahao Li, Yang Shi, Tianyu Yu, Weize Chen, et al. Embodiedeval: Evaluate multimodal llms as embodied agents. *arXiv preprint arXiv:2501.11858*, 2025. 3

[8] Abhishek Das, Samyak Datta, Georgia Gkioxari, Stefan Lee, Devi Parikh, and Dhruv Batra. Embodied question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–10, 2018. 3

[9] Xingyu Fu, Minqian Liu, Zhengyuan Yang, John Corring, Yijuan Lu, Jianwei Yang, Dan Roth, Dinei Florencio, and Cha Zhang. Refocus: Visual editing as a chain of thought for structured image understanding. *arXiv preprint arXiv:2501.05452*, 2025. 6

[10] Chen Gao, Liankai Jin, Xingyu Peng, Jiazhao Zhang, Yue Deng, Annan Li, He Wang, and Si Liu. Octonav: Towards generalist embodied navigation. *arXiv preprint arXiv:2506.09839*, 2025. 3

[11] Junyao Gao, Yanan Sun, Yanchen Liu, Yinhao Tang, Yanhong Zeng, Ding Qi, Kai Chen, and Cairong Zhao. Styleshot: A snapshot on any style. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 3

[12] Tianjun Gu, Linfeng Li, Xuhong Wang, Chenghua Gong, Jingyu Gong, Zhizhong Zhang, Yuan Xie, Lizhuang Ma, and Xin Tan. Doraemon: Decentralized ontology-aware reliable agent with enhanced memory oriented navigation. *arXiv preprint arXiv:2505.21969*, 2025. 3

[13] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025. 6

[14] Meera Hahn, Devendra Singh Chaplot, Shubham Tulsiani, Mustafa Mukadam, James M Rehg, and Abhinav Gupta. No rl, no simulation: Learning to navigate without navigating. *Advances in Neural Information Processing Systems*, 34:26661–26673, 2021. 3

[15] Wenbo Hu, Yining Hong, Yanjun Wang, Leison Gao, Zibu Wei, Xingcheng Yao, Nanyun Peng, Yonatan Bitton, Idan Szpektor, and Kai-Wei Chang. 3dllm-mem: Long-term spatial-temporal memory for embodied 3d large language model. *arXiv preprint arXiv:2505.22657*, 2025. 2, 3

[16] Kaixuan Jiang, Yang Liu, Weixing Chen, Jingzhou Luo, Ziliang Chen, Ling Pan, Guanbin Li, and Liang Lin. Beyond the destination: A novel benchmark for exploration-aware embodied question answering. *arXiv preprint arXiv:2503.11117*, 2025. 2, 3

[17] Yizhang Jin, Jian Li, Tianjun Gu, Yexin Liu, Bo Zhao, Jinxiang Lai, Zhenye Gan, Yabiao Wang, Chengjie Wang, Xin Tan, et al. Efficient multimodal large language models: A survey. *Visual Intelligence*, 3(1):27, 2025. 2

[18] Mukul Khanna, Ram Ramrakhya, Gunjan Chhablani, Sriram Yenamandra, Theophile Gervet, Matthew Chang, Zsolt Kira, Devendra Singh Chaplot, Dhruv Batra, and Roozbeh Mottaghi. Goat-bench: A benchmark for multi-modal lifelong navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16373–16383, 2024. 2, 3, 4, 5, 8

[19] Jacob Krantz, Stefan Lee, Jitendra Malik, Dhruv Batra, and Devendra Singh Chaplot. Instance-specific image goal navigation: Training embodied agents to find object instances. *arXiv preprint arXiv:2211.15876*, 2022. 2, 3

[20] Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, et al. Llava-onevision: Easy visual task transfer. *arXiv preprint arXiv:2408.03326*, 2024. 2, 7

[21] LinFeng Li, Jian Zhao, Yuan Xie, Xin Tan, and Xuelong Li. Compassnav: Steering from path imitation to decision understanding in navigation. *arXiv preprint arXiv:2510.10154*, 2025. 3

[22] Arjun Majumdar, Anurag Ajay, Xiaohan Zhang, Pranav Putta, Sriram Yenamandra, Mikael Henaff, Sneha Silwal, Paul Mcvay, Oleksandr Maksymets, Sergio Arnaud, et al. Openeqa: Embodied question answering in the era of foundation models. In *Proceedings of the IEEE/CVF conference*

on computer vision and pattern recognition, pages 16488–16498, 2024. 2, 3

[23] AI Meta. Llama 3.2: Revolutionizing edge ai and vision with open, customizable models. *Meta AI Blog. Retrieved December*, 20:2024, 2024. 2, 7

[24] Siru Ouyang, Jun Yan, I Hsu, Yanfei Chen, Ke Jiang, Zifeng Wang, Rujun Han, Long T Le, Samira Daruki, Xiangru Tang, et al. Reasoningbank: Scaling agent self-evolving with reasoning memory. *arXiv preprint arXiv:2509.25140*, 2025. 3

[25] Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G Patil, Ion Stoica, and Joseph E Gonzalez. Memgpt: Towards llms as operating systems. *arXiv preprint arXiv:2310.08560*, 2023. 3

[26] Zhangyang Qi, Zhixiong Zhang, Yizhou Yu, Jiaqi Wang, and Hengshuang Zhao. Vln-r1: Vision-language navigation via reinforcement fine-tuning. *arXiv preprint arXiv:2506.17221*, 2025. 3

[27] Yanyuan Qiao, Haodong Hong, Wenqi Lyu, Dong An, Siqi Zhang, Yutong Xie, Xinyu Wang, and Qi Wu. Navbench: Probing multimodal large language models for embodied navigation. *arXiv preprint arXiv:2506.01031*, 2025. 3

[28] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 4, 6

[29] Ram Ramrakhya, Eric Undersander, Dhruv Batra, and Abhishek Das. Habitat-web: Learning embodied object-search strategies from human demonstrations at scale. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5173–5183, 2022. 3

[30] Ram Ramrakhya, Matthew Chang, Xavier Puig, Ruta Desai, Zsolt Kira, and Roozbeh Mottaghi. Grounding multimodal llms to embodied agents that ask for help with reinforcement learning. *arXiv preprint arXiv:2504.00907*, 2025. 3

[31] Allen Z Ren, Jaden Clark, Anushri Dixit, Masha Itkina, Anirudha Majumdar, and Dorsa Sadigh. Explore until confident: Efficient exploration for embodied question answering. *arXiv preprint arXiv:2403.15941*, 2024. 2, 3, 7, 8

[32] Vishnu Sarukkai, Brennan Shacklett, Zander Majercik, Kush Bhatia, Christopher Ré, and Kayvon Fatahalian. Automated rewards via llm-generated progress functions. *arXiv preprint arXiv:2410.09187*, 2024. 3

[33] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9339–9347, 2019. 4

[34] Kunal Pratap Singh, Luca Weihs, Alvaro Herrasti, Jonghyun Choi, Aniruddha Kembhavi, and Roozbeh Mottaghi. Ask4help: Learning to leverage an expert for embodied tasks. *Advances in Neural Information Processing Systems*, 35:16221–16232, 2022. 3

[35] Andrew Szot, Max Schwarzer, Harsh Agrawal, Bogdan Mazoure, Rin Metcalf, Walter Talbott, Natalie Mackraz, R Devon Hjelm, and Alexander T Toshev. Large language models as generalizable policies for embodied tasks. In *The Twelfth International Conference on Learning Representations*, 2023. 3

[36] Xin Tan, Bangwei Liu, Yicheng Bao, Qijian Tian, Zhenkun Gao, Xiongbin Wu, Zhihao Luo, Sen Wang, Yuqi Zhang, Xuhong Wang, et al. Towards safe and trustworthy embodied ai: foundations, status, and prospects. 2025. 3

[37] Qwen Team. Qwen3-vl: Sharper vision, deeper thought, broader action. *Qwen Blog. Accessed*, pages 10–04, 2025. 2, 7

[38] Wanxin Tian, Shijie Zhang, Kevin Zhang, Xiaowei Chi, Yulin Luo, Junyu Lu, Chunkai Fan, Qiang Zhou, Yiming Zhao, Ning Liu Siyu Lin, et al. Seea-r1: Tree-structured reinforcement fine-tuning for self-evolving embodied agents. *arXiv preprint arXiv:2506.21669*, 2025. 3

[39] Sen Wang, Shao Zeng, Tianjun Gu, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, et al. From enhancement to understanding: Build a generalized bridge for low-light vision via semantically consistent unsupervised fine-tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13804–13814, 2025. 2

[40] Sen Wang, Jiangning Zhang, Xin Tan, Zhifeng Xie, Chengjie Wang, and Lizhuang Ma. Mmofusion: Multi-modal co-speech motion generation with diffusion model. *Pattern Recognition*, 169:111774, 2026. 2

[41] Yu Wang, Ryuichi Takanobu, Zhiqi Liang, Yuzhen Mao, Yuanzhe Hu, Julian McAuley, and Xiaojian Wu. Mem-{\alpha}: Learning memory construction via reinforcement learning. *arXiv preprint arXiv:2509.25911*, 2025. 3

[42] Erik Wijmans, Samyak Datta, Oleksandr Maksymets, Abhishek Das, Georgia Gkioxari, Stefan Lee, Irfan Essa, Devi Parikh, and Dhruv Batra. Embodied question answering in photorealistic environments with point cloud perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6659–6668, 2019. 3

[43] Erik Wijmans, Abhishek Kadian, Ari Morcos, Stefan Lee, Irfan Essa, Devi Parikh, Manolis Savva, and Dhruv Batra. Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames. *arXiv preprint arXiv:1911.00357*, 2019. 3

[44] Jinming Wu, Zihao Deng, Wei Li, Yiding Liu, Bo You, Bo Li, Zejun Ma, and Ziwei Liu. Mmsearch-r1: Incentivizing lmms to search. *arXiv preprint arXiv:2506.20670*, 2025. 6

[45] Mingyuan Wu, Jingcheng Yang, Jize Jiang, Meitang Li, Kaizhuo Yan, Hanchao Yu, Minjia Zhang, Chengxiang Zhai, and Klara Nahrstedt. Vtool-r1: Vlms learn to think with images via reinforcement learning on multimodal tool use. *arXiv preprint arXiv:2505.19255*, 2025. 6

[46] Yuefeng Xi, Chenyang Zhu, Yao Duan, Renjiao Yi, Lintao Zheng, Hongjun He, and Kai Xu. Thp: Tensor-field-driven hierarchical path planning for autonomous scene exploration with depth sensors. *Computational Visual Media*, 10(6): 1121–1135, 2024. 3

[47] Zhifeng Xie, Sen Wang, Ke Xu, Zhizhong Zhang, Xin Tan, Yuan Xie, and Lizhuang Ma. Boosting night-time scene parsing with learnable frequency. *IEEE Transactions on Image Processing*, 32:2386–2398, 2023. 2

[48] Zhifeng Xie, Sen Wang, Qiucheng Yu, Xin Tan, and Yuan Xie. Csfwinformer: Cross-space-frequency window transformer for mirror detection. *IEEE Transactions on Image Processing*, 33:1853–1867, 2024. 2

[49] Karmesh Yadav, Jacob Krantz, Ram Ramrakhya, Santhosh Kumar Ramakrishnan, Jimmy Yang, Austin Wang, John Turner, Aaron Gokaslan, Vincent-Pierre Berges, Roozbeh Mootaghi, Oleksandr Maksymets, Angel X Chang, Manolis Savva, Alexander Clegg, Devendra Singh Chaplot, and Dhruv Batra. Habitat challenge 2023. https://aihabitat.org/challenge/2023/, 2023. 2, 3

[50] Karmesh Yadav, Ram Ramrakhya, Santhosh Kumar Ramakrishnan, Theo Gervet, John Turner, Aaron Gokaslan, Noah Maestre, Angel Xuan Chang, Dhruv Batra, Manolis Savva, et al. Habitat-matterport 3d semantics dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4927–4936, 2023. 4

[51] Rui Yang, Hanyang Chen, Junyu Zhang, Mark Zhao, Cheng Qian, Kangrui Wang, Qineng Wang, Teja Venkat Koripella, Marziyeh Movahedi, Manling Li, et al. Embodiedbench: Comprehensive benchmarking multi-modal large language models for vision-driven embodied agents. *arXiv preprint arXiv:2502.09560*, 2025. 3

[52] Yuncong Yang, Han Yang, Jiachen Zhou, Peihao Chen, Hongxin Zhang, Yilun Du, and Chuang Gan. 3d-mem: 3d scene memory for embodied exploration and reasoning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17294–17303, 2025. 2, 3, 7, 8

[53] Hang Yin, Xiuwei Xu, Linqing Zhao, Ziwei Wang, Jie Zhou, and Jiwen Lu. Unigoal: Towards universal zero-shot goal-oriented navigation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 19057–19066, 2025. 3

[54] Naoki Yokoyama, Ram Ramrakhya, Abhishek Das, Dhruv Batra, and Sehoon Ha. Hm3d-ovon: A dataset and benchmark for open-vocabulary object goal navigation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5543–5550. IEEE, 2024. 2, 3

[55] Hongli Yu, Tinghong Chen, Jiangtao Feng, Jiangjie Chen, Weinan Dai, Qiying Yu, Ya-Qin Zhang, Wei-Ying Ma, Jingjing Liu, Mingxuan Wang, et al. Memagent: Reshaping long-context llm with multi-conv rl-based memory agent. *arXiv preprint arXiv:2507.02259*, 2025. 3

[56] Licheng Yu, Xinlei Chen, Georgia Gkioxari, Mohit Bansal, Tamara L Berg, and Dhruv Batra. Multi-target embodied question answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6309–6318, 2019. 3

[57] Mingliang Zhai, Zhi Gao, Yuwei Wu, and Yunde Jia. Memory-centric embodied question answer. *arXiv preprint arXiv:2505.13948*, 2025. 2, 3

[58] Wenqi Zhang, Mengna Wang, Gangao Liu, Xu Huixin, Yiwei Jiang, Yongliang Shen, Guiyang Hou, Zhe Zheng, Hang Zhang, Xin Li, et al. Embodied-reasoner: Synergizing visual search, reasoning, and action for embodied interactive tasks. *arXiv preprint arXiv:2503.21696*, 2025. 2, 3

[59] Youcai Zhang, Xinyu Huang, Jinyu Ma, Zhaoyang Li, Zhaochuan Luo, Yanchun Xie, Yuzhuo Qin, Tong Luo, Yaqian Li, Shilong Liu, et al. Recognize anything: A strong image tagging model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1724–1732, 2024. 4

[60] Cairong Zhao, Yubin Wang, Xinyang Jiang, Yifei Shen, Kaitao Song, Dongsheng Li, and Duoqian Miao. Learning domain invariant prompt for vision-language models. *IEEE Transactions on Image Processing*, 33:1348–1360, 2024. 3

[61] Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, et al. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*, 2025. 7

[62] Ziyu Zhu, Xilin Wang, Yixuan Li, Zhuofan Zhang, Xiaojian Ma, Yixin Chen, Baoxiong Jia, Wei Liang, Qian Yu, Zhidong Deng, et al. Move to understand a 3d scene: Bridging visual grounding and exploration for efficient and versatile embodied navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8120–8132, 2025. 2, 3, 5