ERACT: TOWARDS REASONING OVER MEMORIZATION VIA SEMANTIC EXPERIENCE RATIONALE INFORMED PLANNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Vision-Language Models (VLMs) have shown strong potential for embodied planning, yet their performance can degrade substantially under distribution shifts. While accumulated experience provides a scalable source of external knowledge, raw trajectory-based experience provides concrete behavioral references but leaves the decision logic underlying actions implicit, making it difficult to reuse flexibly when current executions diverge from past trajectories. Based on this insight, we propose Experience Rationale Informed Planning (ERAct), a rationale-driven embodied planning framework that transforms past experiences into explicit reasoning knowledge and transfers decision-relevant reasoning cues to novel situations. ERAct consists of a Semantic Experience Rationale Retriever (SERR), which performs coarse-grained retrieval of reasoning-relevant experience instances, and a Rationale-Informed Fusion Network (RIFN), which conducts fine-grained dual-path relevance modeling to extract useful reasoning cues and adaptively inject them through a confidence-aware residual mechanism. Extensive experiments on the EB-ALFRED benchmark within EmbodiedBench demonstrate that ERAct improves both instruction understanding and environmental reasoning, consistently outperforming prior agents. Further analyses show that ERAct exhibits stronger reasoning generalization and more effective experience utilization than direct trajectory reuse.

1 INTRODUCTION

Vision-Language Models (VLMs) have increasingly been extended to embodied planning tasks, where agents are required to integrate diverse reasoning capabilities to interact with the environment through generating appropriate sequences of atomic actions (Zeng et al., 2023). Despite strong performance on familiar situations, embodied planners trained on finite interaction distributions suffer substantial performance degradation under distribution shifts (Liu et al., 2025b), suggesting that their apparent reasoning ability relies heavily on memorized patterns from the training distribution rather than robust reasoning Zhang et al. (2024).

Expanding training to increasingly diverse multi-task and multi-environment data can mitigate this issue, but requires costly embodied data collection and continual model re-training to accommodate newly encountered situations. A more principled approach is to leverage accumulated past experiences as external grounding knowledge (Kuang et al., 2024). Previous approaches primarily reuse past experiences as behavioral references, conditioning current decisions on trajectories, action sequences, or perceptually similar observations from previous executions (Li et al., 2026; Wu et al., 2026). However, tasks sharing the same high-level goal structure unfold through substantially different execution trajectories due to variations in objects and environmental configurations (Shridhar et al., 2020b). Such behavior-centric experience reuse methods therefore become less effective when past trajectories do not align with the agent's current execution state, limiting their utility for ongoing reasoning, as illustrated in Figure 1. This limitation stems partly from the nature of raw experience trajectories. Interleaved visual observations and actions form a continuous execution stream, leaving implicit and entangled how the environment affects execution outcomes, how task progress evolves, and why a particular next decision is appropriate (Hu et al., 2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

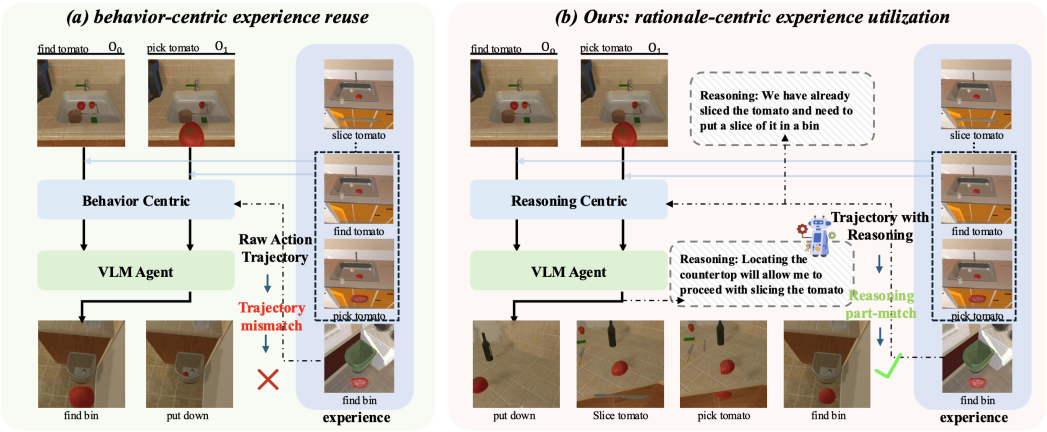


Figure 1: When retrieved experiences share similar execution structures (e.g., "Place a cooked tomato slice into the bin.") but unfold along different trajectories, (a) behavior-centric method misalign the agent with its actual task progress. (b) ERAct can extract the underlying shared decision logic and ground them in the current execution state.

Our key insight is that, beyond recording what was done, past experiences also contain the decision rationale underlying individual actions—why a particular action was appropriate given the task progress and environmental state at that moment (Zhang et al., 2026a). This makes reasoning relevance fundamentally distinct from trajectory-level similarity. Rather than reproducing past behavioral solutions, we ask whether embodied agents can leverage such fine-grained reasoning cues from relevant experiences to inform decisions under the current task and execution state. By explicitly exposing the relationship between task progress, environmental state, and action choice, experience rationales make the reasoning knowledge otherwise entangled within raw trajectories more accessible for selective reuse. However, effectively exploiting such rationale-based experiences introduces two key challenges. First, a retrieved rationale remains grounded in a specific past task and state, while only a subset of its reasoning cues can be useful for the present decision. Second, the utility of retrieved experiences varies substantially across situations. Statically and indiscriminately incorporating them interfere with the planner's own understanding of the current situation.

Instead of relying on powerful foundation models to implicitly interpret and utilize retrieved experiences through prompting, we propose **Experience Rationale Informed Planning (ERAct)**, a compact rationale-driven embodied agent that learns to identify and exploit useful reasoning cues from rationale-relevant experiences. ERAct leverages experience rationales to reduce its dependence on trajectory-level correspondence and support more robust planning in novel situations through two key components. The first component, the Semantic Experience Rationale Retriever (SERR), retrieves reasoning-relevant experiences according to their decision-dependent semantic rationale relevance, thereby providing explicit reasoning knowledge as complementary guidance. The second component, the Rationale-Informed Fusion Network (RIFN), adopts a residual formulation that selectively extracts situation-specific reasoning cues through dual-path alignment with the current task progress and visual environment, while dynamically modulating the strength of the injection with current tokens according to relevance confidence. Under the challenging yet practical closed-loop setting, SERR performs coarse-grained step-level experience retrieval, while RIFN further conducts fine-grained relevance modeling following a selection-to-grounding process to construct reasoning-enhanced representations. Finally, the VLM backbone takes these rationale-enhanced representations as input, leveraging the grounded reasoning cues to reason over the current execution state and generate the subsequent decision.

We conduct extensive evaluations of ERAct on the EB-ALFRED benchmark within EmbodiedBench, a challenging suite of high-level embodied planning tasks. ERAct consistently facilitate instruction and visual generalization by leveraging reasoning priors from offline experience memory. With the same amount of experience data, ERAct outperforms behavior-centric experience

reuse in both leveraging underlying reasoning cues and preserving the planner's own understanding of the current situation, validating the benefit of reasoning transfer over solution reuse.

The contributions of this work can be summarized as follows:

- We propose ERAct, a rationale-driven embodied agent that extracts transferable reasoning cues shared across task-specific experiences and adaptively integrates them through a residual framework for situation-aware planning, thereby improving generalization to unseen tasks and environments while enhancing adaptability to complex real-world scenarios.
- We design a live experience bank for coarse-grained retrieval of reasoning-relevant instances in an embedding space, and further develop RIFN for fine-grained relevance modeling, which selectively extracts reasoning cues pertinent to the current execution state and dynamically controls their injection strength to suppress mismatched information.
- Extensive experiments demonstrate that ERAct significantly outperforms prior state-of-the-art agents in both instruction understanding and environmental reasoning.

2 Related Work

2.1 Embodied Task Planning

Recent advances in large language and vision-language models have substantially improved embodied task planning, enabling agents to reason over natural-language goals, visual observations, and interaction histories to generate high-level actions. Existing methods can broadly be divided into training-free and training-based paradigms. Training-free approaches typically exploit prompting, task decomposition, replanning, or multi-agent coordination to elicit planning capabilities from powerful foundation models. LLM-Planner (Song et al., 2023) uses in-context example retrieval and replanning to improve task completion rate. ExpeL (Zhao et al., 2024) summarizes insights from past experience and uses those insights to guide future tasks. GiG (Li et al., 2026) encode state graph retrieve environments-similar priors to guide execution. In contrast, training-based approaches adapt open-source VLMs into task-specialized planners through supervised fine-tuning (Zhang et al., 2026b; Zhou et al., 2025), preference optimization (Jiao et al., 2025), or reinforcement learning (Yang et al., 2024) . Although these approaches substantially improve embodied planning, their acquired reasoning capabilities are primarily encoded in model parameters and are consequently shaped by the task and environment distributions observed during training.

2.2 Learning to Utilize Retrieved Experience

Recent studies have explored leveraging past interaction experience to improve embodied decision making. ExpRAG (Ferraz et al., 2026) performs retrieval-augmented fine-tuning by conditioning the policy on retrieved trajectories, while RAEA (Zhu et al., 2024) and Dejavu (Wu et al., 2026) incorporate retrieved behavioral experience into embodied policies for action generation or correction. Beyond directly reusing raw trajectories, some methods abstract past executions into higher-level behavioral representations, such as strategies, skills, or subgoals (Shi et al., 2026a). Despite different forms, these approaches remain behavior-centric, primarily transferring reusable behavioral patterns or solution structures while leaving the decision logic behind individual actions implicit. In contrast, ERAct uses fine-grained step-level rationales to explicitly capture why an action is appropriate under a specific execution state, enabling transfer of decision-relevant reasoning cues rather than behavioral solutions.

3 Method

3.1 Problem Formulation

3.1.1 VLM Agent

In embodied planning tasks, an agent is required to complete a task by performing a sequence of atomic actions in the environment. Specifically, the agent receives a natural language instruction I that describes the goal of the task. Then for each step t, it gets an egocentric RGB observation image

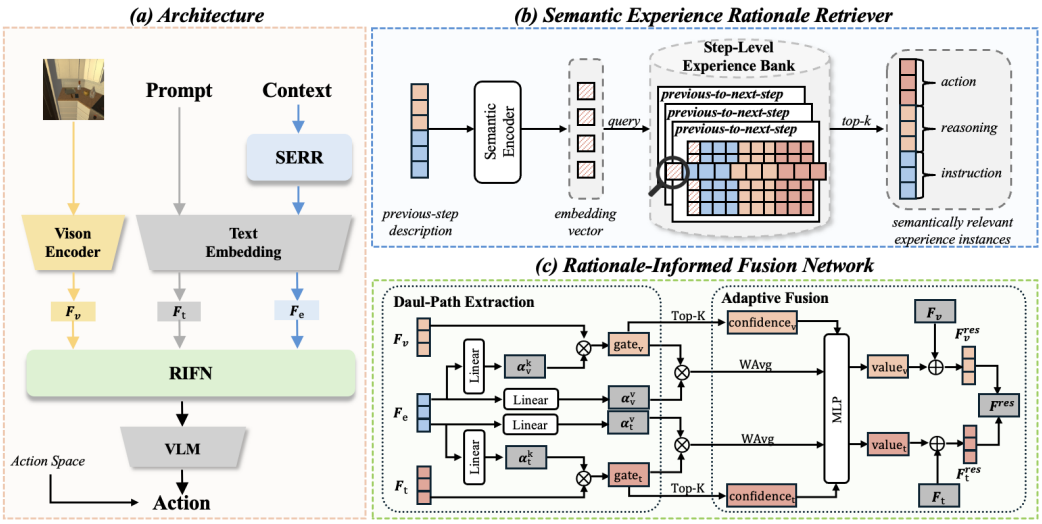


Figure 2: (a) Architecture of our proposed method. (b) In the Semantic Experience Rationale Retriever (SERR), we shift experience retrieval from trajectory matching toward step-level reasoning alignment. (c) In Rationale-Informed Fusion Network (RIFN) module, we propose a residual formulation that performs fine-grained relevance modeling over retrieved rationales and selectively injects decision-relevant reasoning cues into the planning.

o_t as environmental state feedback, and outputs an action a_t chosen from a pre-defined action set A. The environment state will evolve in response to the action of the agent: $s_{t+1} = \mathcal{E}(s_t, a_t)$, where $\mathcal{E}$ is implemented by a simulator. The agent is required to continuously adapt to environmental changes until it determines that the task has been completed and selects the terminal action. Standard VLM Agents directly model:

$$\pi(a_t \mid I, H_t, o_t) \tag{1}$$

where H_t denotes the interaction context accumulated up to timestep t. Such task-specific trained agents, however, often suffer substantial performance degradation under distribution shifts. We operate within a closed-loop framework, where each action decision is made based solely on the robot's available information, including the current camera observation and historical interaction context, without access to auxiliary privileged feedback, such as action-success signals. This setting mirrors real-world robotic deployment scenarios, where agents must rely on robust and generalizable reasoning capabilities to perform a range of intermediate processes, including instruction understanding, commonsense reasoning, action modeling, environment analysis, and progress monitoring.

3.1.2 VLM Agent with Experience Rationale

Our key insight is to introduce an experience rationale informed planner, which leveraging more general reasoning cues underlying past experiences to enhance the VLM model. With ERAct, the conditional distribution is reformulated as follows:

$$\pi(a_t, r_t \mid I, H_t, o_t) = \sum_{e_t} \pi_l(a_t, r_t \mid I, H_t, o_t, e_t)\pi_h(e_t \mid I, H_t) \tag{2}$$

Here, π_h denotes our rationale-relevant experience retriever, which performs coarse-grained retrieval to obtain rationales associated with experience instances, conditioned on the overall task instruction and the action reasoning process from the previous step. Conditioned on the retrieved instances, π_l is a decision-level fusion planner, which further conducts fine-grained relevance modeling to dynamically align transferable reasoning cues with the current state for action decision-making. Unlike trajectory-based experience reuse, where retrieved trajectories primarily serve as references for reproducing past solutions, ERAct treats retrieved rationales as reasoning knowledge from which decision-relevant reasoning cues are further distilled and integrated into the current planning context.

3.2 SEMANTIC EXPERIENCE RATIONALE RETRIEVER

Trajectory-based retrieval primarily focuses on surface-level similarities of textual or visual cues in trajectories, potentially overlooking the underlying decision logic that may be useful for the current unseen task. The SERR is designed to address this by functioning as a step-level semantic retrieval over experience rationales. Its goal is to prioritize whether an experience contains relevant reasoning knowledge, rather than whether its task or observation closely matches the current situation. As shown in Figure 3, the rationale relevance provides a more utility aligned signal for experience retrieval. The SERR π_h consists of two components: a semantic encoder π_e and an experience retriever π_b. Given the interaction context, π_e encodes the semantic description of the agent's inferred task progress and decision context into a dense representation d_t, while π_b grounds it into concrete experience instances containing relevant reasoning knowledge from the experience bank D:

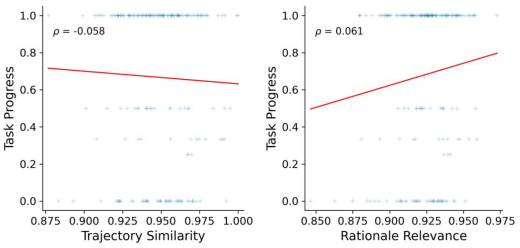


Figure 3: Scatter plots illustrating the relationship between success rate and trajectory similarity (left) or rationale relevance score (right). The red line indicates a linear fit to the data.

$$\pi_h(e_t \mid I, H_t; D) = \pi_b(e_t \mid D, d_t), \qquad d_t = \pi_e(I, H_t). \tag{3}$$

As a lightweight coarse-grained selector, SERR operates at the granularity of individual decision steps, shifting experience retrieval from trajectory matching toward step-level reasoning alignment.

3.2.1 EXPERIENCE BANK DESIGN

We design a step-level experience rationale bank that organizes past trajectories into decision-level reasoning instances. A conventional experience trajectory is represented as $\tau = (o_1, a_1, \ldots, o_T, a_T)$. We augment each action with its corresponding rationale, yielding $\tau^R = (o_1, r_1, a_1, \ldots, o_T, r_T, a_T)$, where r_t describes the task progress and environmental context and explains why action a_t is appropriate, thereby making the underlying decision logic explicit rather than leaving it implicitly entangled with task-specific behaviors, see Appendix for further discussion. To ensure that the experience bank provides reliable reasoning references, we retain only successful trajectories that accomplish the task goal. For each experience τ, we also store it additionally with a fixed task instruction l_τ. For every valid timestep t, we store an experience instance

$$m_t = (l_\tau, k_t, r_t, a_t), \tag{4}$$

where k_t is a compact semantic key derived from the task instruction and the preceding action reasoning for retrieval, r_t provides the reasoning knowledge associated with the current decision, and a_t is the corresponding executed action. The complete experience bank is therefore defined as $D = \{m_t\}_{\tau,t}$, containing all valid decision-level instances across all successful experience trajectories encountered during data collection.

3.2.2 DECISION-DEPENDENT KEY CONSTRUCTION AND TOP-K RETRIEVAL

SERR constructs retrieval representations from the semantic decision context rather than directly matching raw instructions or trajectories. For each stored experience instance m_i, we concatenate its task instruction l_{τ_i} with the rationale from the preceding decision step r_{i-1} and encode them using the semantic encoder:

$$k_i = \pi_e([l_{\tau_i}; r_{i-1}]), \tag{5}$$

The task instruction provides the overall task objective, while the preceding rationale summarizes the task progress and reasoning context leading to the next decision. At timestep t, we construct the query in the same manner $q_t = \pi_e([I; r_{t-1}])$. The explicit reasoning context makes retrieval progress-aware, allowing experiences to become relevant when they involve similar underlying decision logic, even when their trajectories differ substantially. We then compute the cosine similarity between the query and each stored key $s_i = \cos(q_t, k_i)$, and retrieve the top-K experience instances

$\mathcal{N}_K(q_t) = \mathrm{TopK}_i(s_i)$. The corresponding rationales $\{r_i\}_{i\in\mathcal{N}_K(q_t)}$ provide coarse-grained reasoning knowledge for the current decision, which is subsequently refined by RIFN through fine-grained relevance modeling against the current task and visual state.

3.3 Rationale-Informed Fusion Network

The retrieved rationales remain grounded in specific past tasks, and therefore contain mismatched information. To bridge this gap, we formulate RIFN as a hierarchical reasoning transfer framework governed by two levels of latent reasoning representations: it first extracts transferable reasoning cues underlying retrieved experience, then selectively aggregates their decision-relevant components, before adaptively integrating it through a confidence-aware residual pathway.

3.3.1 Dual-Path Reasoning Extraction

To extract the reasoning structure that can be reused beyond its original task, we introduce a dual-path architecture comprising textual context and visual environment, which interleaves reasoning prior modeling with multimodal visual-textual fusion. Let $E_t \in \mathbb{R}^{L_e\times d}$ denote the representation of the retrieved experience rationale, rather than using a shared projection, we respectively construct key-value representations for the textual and visual pathways:

$$K_t^{\text{text}} = W_k^{\text{text}} E_t, \qquad V_t^{\text{text}} = W_v^{\text{text}} E_t, \tag{6}$$

$$K_t^{\text{img}} = W_k^{\text{img}} E_t, \qquad V_t^{\text{img}} = W_v^{\text{img}} E_t. \tag{7}$$

The textual pathway emphasizes reasoning cues associated with the task instruction and execution progress, while the visual pathway transforms the rationale into a representation suitable for alignment with the current environmental observation. In this way, the task-specific experience is not treated as fixed textual content, but is transformed into reasoning cues that can support cross-task transfer and be subsequently instantiated according to the current decision.

3.3.2 Decision-Conditioned Reasoning Instantiation

After independent dual-path extraction, RIFN further instantiates the transferable components that are relevant to the current situation. RIFN acts as a fine-grained selector, dynamically suppressing irrelevant knowledge while preserving critical reasoning components in the latent state. Let X_t^m denote the input representation for modality $m \in \{\text{text}, \text{img}\}$. We perform semantic alignment between the current representation and the corresponding reasoning cues by computing token-wise relevance scores against the rationale keys:

$$S_t^m = X_t^m \left(K_t^m\right)^\top, \qquad m \in \{\text{text}, \text{img}\}. \tag{8}$$

Conditioned on these interactions, current-context representation queries the retrieved rationale to identify the critical components that are relevant to the current decision. We then distill them into situation-specific reasoning representation through normalized cross-token aggregation:

$$C_t^m = \left\{ \frac{\sum_j \sigma\left(S_{t,ij}^m/\tau\right) V_{t,j}^m}{\sum_j \sigma\left(S_{t,ij}^m/\tau\right) + \epsilon} \right\}_i. \tag{9}$$

In this way, the retrieved rationale is no longer incorporated as a complete task-specific experience. Instead, its reasoning cues are selectively recomposed, yielding situation-specific guidance for subsequent planning.

3.3.3 Confidence-Aware Residual Fusion

Although decision-conditioned instantiation identifies the relative contribution of different reasoning components, normalized aggregation inevitably produces a reasoning cue regardless of its overall relevance to the current situation. Directly incorporating these cues may therefore interfere with the planner when the experience branch is unreliable. To preserve the original planning capability while selectively exploiting useful experience, we introduce a confidence-aware residual fusion mechanism, where experience-derived reasoning is incorporated as a dynamically controlled residual

rather than directly serving as the original representation. Specifically, we estimate the confidence of the experience branch by aggregating the Top-k relevance scores:

$$\gamma_{t,i}^{m} = \mathrm{Mean}\left(\mathrm{TopK}\left(G_{t,i:}^{m}, k\right)\right). \tag{10}$$

The instantiated reasoning cues are transformed by a shared MLP and modulated by the estimated confidence to form an experience-induced residual, which is subsequently injected into the original representation:

$$\Delta_t = \gamma_t \odot F_{\mathrm{fusion}}\left(C_t\right), \tag{11}$$

$$\widetilde{X}_t = X_t + \Delta_t. \tag{12}$$

This residual formulation preserves an identity pathway for the original planner, allowing experience-derived reasoning to serve as complementary guidance rather than a replacement for its inherent decision capability. When the retrieved reasoning is strongly supported by the current context, a larger residual contribution informs the current decision; when the experience branch is unreliable, the confidence suppresses Δ_t^m, causing $\widetilde{X}_t^m$ to approach the original representation X_t^m. In this way, RIFN can exploit useful experience while reducing the risk of degrading the planner's original capability under mismatched or weakly relevant retrieval.

4 Experiment

4.1 Experimental setups

Evaluation Benchmarks. We evaluate our method on EB-ALFRED (Yang et al., 2025), a benchmark built upon the AI2-THOR simulator (Kolve et al., 2017) and the validation split of the ALFRED dataset (Shridhar et al., 2020a). Notably, all training data are derived exclusively from the ALFRED training split, with no access to any evaluation tasks from EB-ALFRED during training. Accordingly, the experience bank used for both training and evaluation is also constructed solely from trajectories in the ALFRED training split. This strict separation not only requires the agent to generalize to novel instructions and unseen environments, but also prevents it from relying on experience trajectories collected from the evaluation tasks themselves. Consequently, effective experience utilization requires going beyond direct trajectory reuse, the agent must identify transferable reasoning cues from retrieved rationales and ground them in the current execution state to support decision-making. We utilize success rate (SR) as the metric for EB-ALFRED.

Implementation Details. We employ Qwen2.5-VL-7B-Instruct (Wu et al., 2025a) as the base VLM and bge-large-en-v1.5 (Xiao et al., 2024) as the semantic encoder for SERR. For the first two training stages, we perform full-parameter supervised fine-tuning for 3 epochs with a learning rate of 1.0×10^{-5} and a global batch of 64 on 4 NVIDIA RTX PRO 6000 GPUs. For the third training stage, we jointly optimize the RIFN modules with the VLM backbone for 1 epoch. All parameters of RIFN are fully trainable, while the VLM backbone is adapted using LoRA. We use a learning rate of 2.0×10^{-4} and a global batch of 16 on an NVIDIA RTX PRO 6000 GPU. All evaluations in the simulated environment are conducted on an NVIDIA RTX 5090 GPU.

Baselines and Variants. For a fair comparison, we re-evaluate RoboAgent (Xu et al., 2026) under the same closed-loop setting without access to auxiliary privileged feedback. Following Ferraz et al. (2026), we further construct ExpRAG, a retrieval-augmented fine-tuning baseline in which experiences are directly incorporated into the training context. For a controlled comparison, ExpRAG is trained with the same backbone, training data, and experience bank as ERAct, allowing us to isolate the effect of the proposed RIFN module.

4.2 Main Results

As shown in Table 1, our framework of experience rationale informed planning achieves leading performance while operating under strictly realistic observation constraints. ERAct surpasses all existing fine-tuning-based methods in terms of average SR. It also outperforms powerful VLM models on the base and common splits. These results suggest that ERAct learns to identify and exploit reasoning-relevant experiences, allowing transferable reasoning cues to support robust generalization to novel tasks and environments without continual retraining. Compared with ExpRAG, the

Table 1: Performance comparison on EB-ALFRED.

	Loop	Avg	Base	Common	Complex	Visual	Spatial	Long
General VLM Models								
Gemini-1.5-Pro	open	62.3	70	64	72	58	52	58
Qwen-VL-Max	open	41.3	44	48	44	42	38	32
Qwen2.5-VL-72B	open	39.7	50	42	42	36	34	34
GPT-4o	open	56.3	64	54	68	46	52	54
GPT-4o	closed	26.0	30	26	28	22	26	24
Claude-3.5-Sonnet	open	64.0	72	66	**76**	60	**58**	52
Claude-3.5-Sonnet	closed	57.3	62	62	64	40	42	**74**
Embodied VLM Models								
RoboBrain2.0 (Team et al., 2025)	–	14.0	–	–	–	–	–	–
Vlaser (Yang et al., 2026)	–	50.0	–	–	–	–	–	–
LoHo-Manip-4B (Liu et al., 2026)	–	37.8	48	36	54	28	30	31
Fine-Tuning Methods								
REBP (Wu et al., 2025b)	–	35.6	54	42	46	28	38	6
RoboGPT-R1 (Liu et al., 2025a)	–	55.3	62	56	64	50	50	50
WAP (Shi et al., 2026b)	closed	62.7	66	62	70	56	**52**	70
RoboAgent (Xu et al., 2026)	closed	63.0	76	48	70	**70**	42	72
Experience-Enhanced Methods								
ExpRAG (Ferraz et al., 2026)	closed	61.7	74	68	**76**	40	50	62
ERAct (Ours)	closed	**65.0**	**78**	**72**	74	50	46	70

Table 2: Ablation study of ERAct on EB-ALFRED.

Method	Avg	Base	Common	Complex	Visual	Spatial	Long
ERAct (Ours)	**65.0**	**78**	**72**	74	50	46	**70**
w/o Dual Path	63.0	68	62	76	**54**	**52**	66
w/o Selector	61.0	**78**	64	66	**54**	42	62
w/o RIFN	63.3	74	68	**78**	46	48	66

results demonstrate that RIFN can dynamically select and aggregate reasoning cues aligned with the current execution state, while preserving the model's understanding of the ongoing task and solution progress. This reduces the interference introduced by mismatched retrieved information and allows experience to inform, rather than override, the model's current decision process.

4.3 Ablation Studies

The ablation results presented in Table 2 demonstrate the rationality of our proposed method. The degradation observed without the dual-path design further highlights the importance of modality-specific relevance modeling. By separately aligning experience rationales with textual task semantics and visual observations, the dual-path design enables reasoning cues to be grounded according to both task progress and the current perceptual state, reducing interference from cues that are semantically relevant but visually mismatched. The qualitative attention maps in Fig. 4 show that the model without the dual-path design often mislocalizes or overlooks task-relevant objects. This indicates that collapsing textual and visual relevance modeling into a shared pathway can entangle heterogeneous reasoning cues, leading to cross-modal interference and less precise grounding of experience information in the current observation. Without the selector, retrieved reasoning features are incorporated indiscriminately, which can overwhelm the model's original task-state representation and disrupt its ongoing reasoning process. Consistent with this, the qualitative attention maps show diffuse responses over multiple irrelevant objects, suggesting that the model fails to suppress mismatched experience cues. The w/o RIFN variant is trained on the same data using only LoRA adaptation of the VLM backbone, without access to the experience-fusion mechanism. The performance gap indicates that ERAct primarily learns how to leverage external experience to support reasoning, rather than merely internalizing additional patterns from the training data.

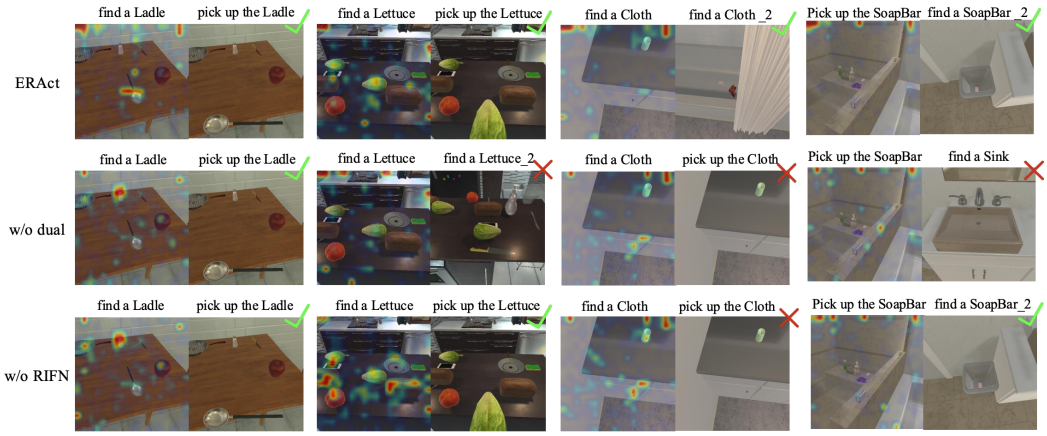


Figure 4: Qualitative attention maps for the ablation study on EB-ALFRED.

Table 3: Effect of rationale-based experience on EB-ALFRED.

Train Exp.	Test Exp.	Avg	Base	Common	Complex	Visual	Spatial	Long
Rationale Trajectory	w/ Exp.	**65.0**	**78**	72	**74**	50	**46**	**70**
	w/o Exp.	60.7	74	70	66	44	46	64
Action Trajectory	w/ Exp.	60.7	74	68	70	50	42	60
	w/o Exp.	62.7	74	**74**	68	**52**	44	64

4.4 Effect of Experience Rationales

We further investigate whether the benefit of ERAct arises from the explicit reasoning knowledge encoded in experience rationales or merely from access to past action trajectories. As shown in Table 3, incorporating rationale-based experience improves the average SR from 60.7 to 65.0, with particularly pronounced gains on Complex (+8), Visual (+6), and Long-Horizon (+6) tasks. In contrast, incorporating action-only experience decreases the average SR from 62.7 to 60.7. This contrast suggests that action trajectories, which primarily describe *what* was executed, provide limited information about the conditions under which past decisions remain applicable and may therefore introduce mismatched solution patterns under task and environment shifts. Experience rationales instead expose the underlying decision logic, allowing ERAct to extract reasoning cues that can be re-grounded to the current execution state. Notably, the rationale-trained model does not outperform its action-trained counterpart when experience is removed, whereas a clear advantage emerges once experience is provided. This indicates that rationale supervision primarily enhances the agent's capability to utilize external experience, rather than merely injecting additional reasoning knowledge into the model parameters. These results therefore support our central hypothesis that the transferable value of past experience lies less in reproducing previously executed solutions than in reusing the reasoning knowledge underlying those solutions.

4.5 Conclusion

In this work, we investigated how past embodied experiences can be reused as explicit reasoning knowledge rather than merely as behavioral references. We introduced ERAct, a rationale-driven embodied planning framework that exposes the decision logic underlying past actions and transfers decision-relevant reasoning cues to the current execution state. Through coarse-grained rationale-aware retrieval with SERR and fine-grained situation-dependent grounding with RIFN, ERAct selectively extracts useful experience information while adaptively controlling its influence.

AI USE STATEMENT

In this work, we used generative AI tools to generate step-level rationales from embodied task execution trajectories, and to assist with research design, result interpretation, translation, code debugging. All AI-assisted outputs were reviewed by the authors. Generated code was manually inspected and tested, and suggested factual claims were verified against original sources. We take responsibility for the final content of this work, including text, claims, code, data, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies embodied planning in simulated environments and does not involve human subjects or private user data. Our experiments are conducted on established embodied benchmarks and are intended for research purposes. While improved embodied planning may eventually contribute to real-world robotic systems, deploying such systems in safety-critical environments would require additional safeguards, validation, and consideration of biases inherited from pretrained models and training data.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of ERAct, including the construction of the experience bank, rationale generation and retrieval, the RIFN architecture, training settings, and evaluation protocol in the main paper and appendix. Implementation details, hyperparameters, baseline configurations, and additional experimental analyses are also documented to facilitate reproduction of our results.

REFERENCES

Thomas Palmeira Ferraz, Romain Deffayet, Vassilina Nikoulina, Hervé Déjean, and Stéphane Clinchant. Retrieval-augmented llm agents: Learning to learn from experience. *arXiv preprint arXiv:2603.18272*, 2026.

Bing Hu, Zaijing Li, Rui Shao, Junda Chen, April Hua Liu, Wei-Shi Zheng, and Liqiang Nie. From abstraction to instantiation: Learning behavioral representation for vision-language-action model. *arXiv preprint arXiv:2605.22671*, 2026.

Kechen Jiao, Zhirui Fang, Jiahao Liu, Bei Li, Qifan Wang, Xinyu Liu, Junhao Ruan, Zhongjian Qiao, Yifan Zhu, Yaxin Xu, et al. Tcpo: Thought-centric preference optimization for effective embodied decision-making. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 9585–9599, 2025.

Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Matt Deitke, Kiana Ehsani, Daniel Gordon, Yuke Zhu, et al. Ai2-thor: An interactive 3d environment for visual ai. *arXiv preprint arXiv:1712.05474*, 2017.

Yuxuan Kuang, Junjie Ye, Haoran Geng, Jiageng Mao, Congyue Deng, Leonidas Guibas, He Wang, and Yue Wang. Ram: Retrieval-based affordance transfer for generalizable zero-shot robotic manipulation. *arXiv preprint arXiv:2407.04689*, 2024.

Xiang Li, Ning Yan, and Masood Mortazavi. Embodied task planning via graph-informed action generation with large language models. *arXiv preprint arXiv:2601.21841*, 2026.

Isabella Liu, An-Chieh Cheng, Rui Yan, Geng Chen, Ri-Zhao Qiu, Xueyan Zou, Sha Yi, Hongxu Yin, Xiaolong Wang, and Sifei Liu. Long-horizon manipulation via trace-conditioned vla planning. *arXiv preprint arXiv:2604.21924*, 2026.

Jinrui Liu, Bingyan Nie, Boyu Li, Yaran Chen, Yuze Wang, Shunsen He, and Haoran Li. Robogpt-r1: Enhancing robot task planning with reinforcement learning. *arXiv preprint arXiv:2510.14828*, 2025a.

Yang Liu, Weixing Chen, Yongjie Bai, Xiaodan Liang, Guanbin Li, Wen Gao, and Liang Lin. Aligning cyber space with physical world: A comprehensive survey on embodied ai. *IEEE/ASME Transactions on Mechatronics*, 2025b.

So Yeon Min, Devendra Singh Chaplot, Pradeep Ravikumar, Yonatan Bisk, and Ruslan Salakhutdinov. Film: Following instructions in language with modular methods. *arXiv preprint arXiv:2110.07342*, 2021.

Hao Shi, Bin Xie, Yingfei Liu, Lin Sun, Fengrong Liu, Tiancai Wang, Erjin Zhou, Haoqiang Fan, Xiangyu Zhang, and Gao Huang. Memoryvla: Perceptual-cognitive memory in vision-language-action models for robotic manipulation. In *International Conference on Learning Representations*, volume 2026, pp. 18567–18602, 2026a.

Junhao Shi, Zhaoye Fei, Siyin Wang, Qipeng Guo, Jingjing Gong, and Xipeng QIu. World-aware planning narratives enhance large vision-language model planner. *Advances in Neural Information Processing Systems*, 38:50557–50583, 2026b.

Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer, and Dieter Fox. Alfred: A benchmark for interpreting grounded instructions for everyday tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10740–10749, 2020a.

Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. Alfworld: Aligning text and embodied environments for interactive learning. *arXiv preprint arXiv:2010.03768*, 2020b.

Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M Sadler, Wei-Lun Chao, and Yu Su. Llm-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2998–3009, 2023.

BAAI RoboBrain Team, Mingyu Cao, Huajie Tan, Yuheng Ji, Xiansheng Chen, Minglan Lin, Zhiyu Li, Zhou Cao, Pengwei Wang, Enshen Zhou, et al. Robobrain 2.0 technical report. *arXiv preprint arXiv:2507.02029*, 2025.

Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng-ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, et al. Qwen-image technical report. *arXiv preprint arXiv:2508.02324*, 2025a.

Di Wu, Jiaxin Fan, Junzhe Zang, Guanbo Wang, Wei Yin, Wenhao Li, and Bo Jin. Reinforced reasoning for embodied planning. *arXiv preprint arXiv:2505.22050*, 2025b.

Shaokai Wu, Yanbiao Ji, Qiuchang Li, Zhiyi Zhang, Qichen He, Wenyuan Xie, Guodong Zhang, Bayram Bayramli, Yue Ding, and Hongtao Lu. Dejavu: Towards experience feedback learning for embodied intelligence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 29578–29587, 2026.

Shitao Xiao, Zheng Liu, Peitian Zhang, Niklas Muennighoff, Defu Lian, and Jian-Yun Nie. C-pack: Packed resources for general chinese embeddings. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pp. 641–649, 2024.

Peiran Xu, Jiaqi Zheng, and Yadong Mu. Roboagent: Chaining basic capabilities for embodied task planning. *arXiv preprint arXiv:2604.07774*, 2026.

Ganlin Yang, Tianyi Zhang, Haoran Hao, Weiyun Wang, Yibin Liu, Dehui Wang, Guanzhou Chen, Zijian Cai, Junting Chen, Weijie Su, et al. Vlaser: Vision-language-action model with synergistic embodied reasoning. In *International Conference on Learning Representations*, volume 2026, pp. 65736–65765, 2026.

Jingkang Yang, Yuhao Dong, Shuai Liu, Bo Li, Ziyue Wang, Haoran Tan, Chencheng Jiang, Jiamu Kang, Yuanhan Zhang, Kaiyang Zhou, et al. Octopus: Embodied vision-language programmer from environmental feedback. In *European conference on computer vision*, pp. 20–38. Springer, 2024.

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.

Fanlong Zeng, Wensheng Gan, Yongheng Wang, Ning Liu, and Philip S Yu. Large language models for robotics: A survey. *arXiv preprint arXiv:2311.07226*, 2023.

Kaidong Zhang, Pengzhen Ren, Bingqian Lin, Junfan Lin, Shikui Ma, Hang Xu, and Xiaodan Liang. Pivot-r: Primitive-driven waypoint-aware world model for robotic manipulation. *Advances in Neural Information Processing Systems*, 37:54105–54136, 2024.

Wenqi Zhang, Mengna Wang, Gangao Liu, Huixin Xu, Yiwei Jiang, Yongliang Shen, Guiyang Hou, Zhe Zheng, Hang Zhang, Xin Li, Jiajun Liu, Weiming Lu, Peng Li, and Yueting Zhuang. Embodied-reasoner: Synergizing visual search, reasoning, and action for embodied interactive tasks. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 41178–41207, San Diego, California, United States, July 2026a. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.1910. URL https://aclanthology.org/2026.acl-long.1910/.

Wenqi Zhang, Mengna Wang, Gangao Liu, Huixin Xu, 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. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 41178–41207, 2026b.

Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 19632–19642, 2024.

Weijie Zhou, Manli Tao, Chaoyang Zhao, Honghui Dong, Ming Tang, and Jinqiao Wang. Lightplanner: unleashing the reasoning capabilities of lightweight large language models in task planning. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 14813–14820. IEEE, 2025.

Yichen Zhu, Zhicai Ou, Xiaofeng Mou, and Jian Tang. Retrieval-augmented embodied agents. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17985–17995. IEEE, 2024.

A APPENDIX

A.1 TRAINING STRATEGY

A.1.1 PROGRESSIVE EMBODIED REASONING TRAINING

Our training procedure consists of three stages that gradually enhance reasoning capabilities.

Stage 1: Training with Expert Trajectories We collect the expert trajectories paired with direct instructions (e.g., the instruction "Place a knife in a sink.") from the training split of ALFRED Shridhar et al. (2020a). These trajectories are used for full-parameter supervised fine-tuning, enabling the model to establish foundational action mapping capabilities.

Stage 2: Training with Instruction-Augmented Trajectories We perform data augmentation on the training set to enhance its diversity. Specifically, following WAP Shi et al. (2026b), we employ an LLM to generate reasoning-intensive instructions (e.g., the instruction "Place an object typically used for slicing in a sink.") across four dimensions: visual, spatial, functional, and syntactic. Fine-tuning on these instruction-augmented trajectories helps the model develop basic instruction understanding capabilities.

Stage 3: Training with Experience-Augmented Trajectories While each action of the original trajectories is successfully executed, they provide limited supervision for monitoring task progress based on environmental observations. To address this limitation, we synthesize additional trajectories that require the model to analyze the current observation and determine whether an action has succeeded. We then augment each action step in the collected trajectories with relevant experience ratioanles retrieved from a live experience bank, which is constructed using the training data from Stage 1 and Stage 2. Fine-tuning on these experience-augmented trajectories teaches the model identify and leverage useful reasoning cues to support deeper reasoning for current state, thereby integrating and strengthening the planner's multiple reasoning capabilities required for embodied task planning.

A.1.2 TRAINING DATASET

ALFRED (Shridhar et al., 2020a) is originally proposed as an embodied instruction following simulator (Min et al., 2021). Our training procedure consists of three stages that progressively equip the agent with the capabilities required for experience-informed planning, with the detailed training procedure provided in the Appendix. Following WAP(Shi et al., 2026b), we construct a training corpus of 85K instruction–trajectory pairs derived from the original 16K ALFRED trajectories. Among them, 69K pairs are used in the first two stages to establish the fundamental capabilities of the VLM. These trajectories are further organized into the experience bank used during both the third-stage training and evaluation, with the construction of experience rationales detailed in the Appendix. The remaining 16K pairs are reserved for the third stage, where ERAct is trained to effectively leverage retrieved experience rationales to enhance reasoning and decision-making.

Table 4: Performance across different training stages on EB-ALFRED.

Training Stage	Avg	Base	Common	Complex	Visual	Spatial	Long
Stage 1	44.3	60	42	4	26	36	48
Stage 2	49.0	62	54	60	28	40	50
Stage 3	**65.0**	**78**	**72**	**74**	**50**	**46**	**70**

A.2 STEPWISE RATIONALE GENERATION

For each action a_t in a trajectory τ, we generate an explicit rationale r_t that explains why the action is appropriate given the current task progress and environmental context:

$$r_t = \mathcal{M}\left(\widetilde{I}, o_t, \{(\cdot, a_t)\} \cup \{(r_i, a_i)\}_{i=1}^{t-1}\right), \tag{13}$$

where $\mathcal{M}$ denotes the rationale generation model and $\widetilde{I}$ represents the task instruction. The current action a_t is provided while its rationale is left unspecified, prompting $\mathcal{M}$ to retrospectively explain the decision based on the current observation and preceding execution history. Specifically, r_t describes the current task progress, decision-relevant environmental state, and the reasoning that makes a_t appropriate at the current step. By applying this procedure throughout the trajectory, we transform the original observation–action stream into a sequence of rationale–action instances,

$$\tau^r = \{(o_t, r_t, a_t)\}_{t=1}^{T}, \tag{14}$$

which explicitly exposes the decision-level reasoning knowledge implicit in the original execution trajectory. These rationale-augmented instances form the experience bank used by ERAct for subsequent retrieval and reasoning-informed augmentation.

A.3 Prompt Templates

Rationale Generation Prompt

You are tasked with generating the reasoning process for an embodied agent executing a specific action.

Instruction: {enhanced_instruction}
Current observation: [IMAGE]
Current action: {action}
Previous actions and reasoning: {previous_actions_and_reasoning}
Please generate detailed reasoning that explains WHY the agent should take the current action. Include:
1. What the agent observes in the environment
2. How this relates to the instruction
3. How this action contributes to the overall task goal
4. Reasoning:

Figure 5: Prompt Templates for Rationale Generation.

Planning Prompt

You are a robot operating in a home. Given a task, you must accomplish the task using a defined set of actions to achieve the desired outcome.

The available action id (0 ~ {action count}) and action names are: {action space}.\n

\n\n## Now the human instruction is: {instrcution} \n\n The interaction history: {interaction history} \n\n

You need to describe current visual state from the image, summarize interaction history and reason why the last action or plan failed and did not finish the task, output your new plan to achieve the goal from current state. At the end, output the excutable action with action id(0 ~ {action count}) from the available actions.

Figure 6: Prompt Templates for Planning.

A.4 Effect of Experience Rationales

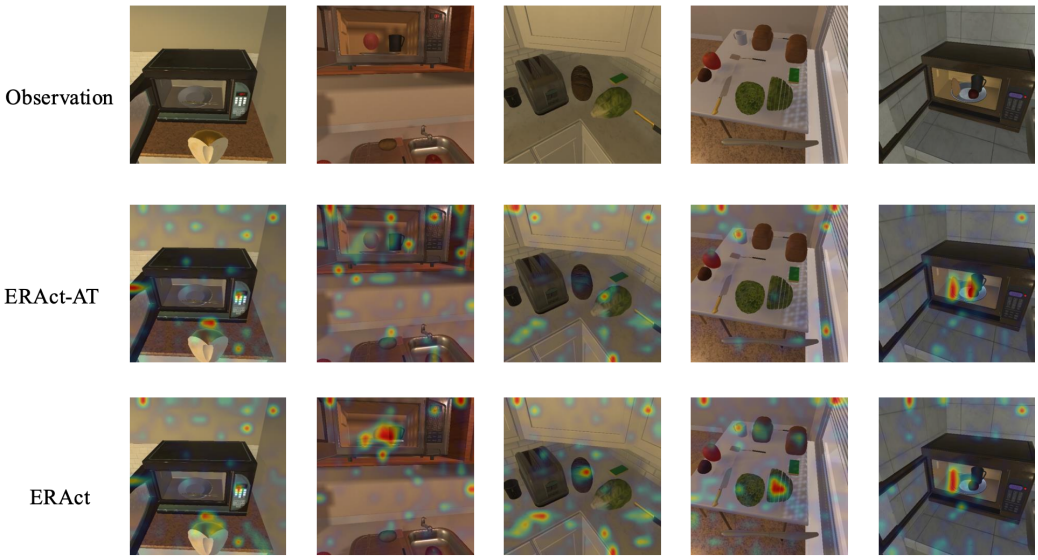


Figure 7: Qualitative attention maps for effect of experience rationales.