PARALLELWORLD: TEST-TIME SCALING FOR AGENTIC EMBODIED REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Embodied AI has emerged as a key paradigm for autonomous systems operating in complex physical environments. While end-to-end models have established strong local capabilities for perception and control, scaling these capabilities to extended, multi-stage tasks remains difficult. Agentic embodied systems address this challenge by introducing high-level planners that organize perception, reasoning, and executable skills over longer interaction horizons. However, existing agentic embodied frameworks typically generate actions for direct execution without explicitly simulating how alternative action sequences shape future observations and task progress, leading agents to commit to suboptimal interactions before considering their long-term consequences. To address this limitation, we propose **ParallelWorld**, an adaptive-horizon test-time scaling framework that grounds agentic embodied reasoning in the prospective simulation of action outcomes. Specifically, we design an agent system to simulate and evaluate alternative trajectories in parallel before committing to an action. A verifier assesses intermediate state transitions with respect to the task objective, dynamically pruning unpromising branches and guiding subsequent expansion. Based on the evaluated long-horizon outcomes, the agent selects a trajectory, executes only its first action, and replans as new observations become available. Experiments on ESI-Bench and LIBERO-Pro demonstrate improvements in active spatial reasoning and language-conditioned manipulation, respectively, highlighting the effectiveness of prospective planning for both understanding and manipulation tasks.

1 INTRODUCTION

Embodied intelligence enables autonomous systems to integrate perception, reasoning, and action in complex physical environments, supporting tasks such as embodied understanding, navigation, and manipulation (Yang et al., 2025a; Majumdar et al., 2024). Recent end-to-end approaches have achieved strong performance on embodied understanding from provided observations and on short-horizon interaction tasks, including navigation and manipulation (Chen et al., 2024; 2021; Kim et al., 2024; Black et al., 2024; Intelligence et al., 2025). However, long-horizon tasks require sustained coordination across changing observations, intermediate task states, and successive decisions, which is difficult to achieve through local end-to-end predictions alone. Recent advances in large language and vision-language models have made such coordination increasingly tractable, giving rise to agentic embodied systems that decompose complex objectives and orchestrate executable skills over extended interactions (Liang et al., 2023; Hong et al., 2026; Zhang et al., 2026).

As shown in Figure 1, current agentic embodied systems typically employ a high-level planner to organize task-dependent interactions over perception and control modules. For embodied understanding, the planner selects exploratory actions to acquire task-relevant evidence, whereas in manipulation it composes executable skills or primitives to achieve a specified physical goal (Hong et al., 2026; Lei et al., 2026; Zhang et al., 2026; Fu et al., 2026; Liang et al., 2023; Lu et al., 2026). In both settings, planning is commonly interleaved with immediate execution, such that the next decision is made only after observing the outcome of the current action without explicitly simulating how alternative action sequences may alter future observations and task progress. As a result, agents can commit to locally plausible interactions before their downstream consequences are sufficiently assessed, causing early decisions to restrict later evidence acquisition or place the environment in states that are less favorable for subsequent execution.

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.

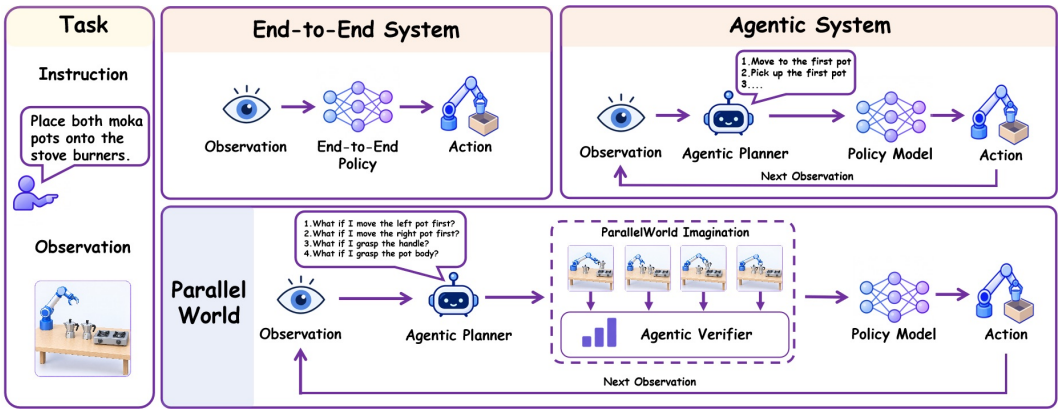

Figure 1: **Comparison of embodied paradigms.** End-to-end systems directly map observations and instructions to actions, while conventional agentic systems introduce a high-level planner to organize executable skills but still commit to actions without explicit prospective evaluation. ParallelWorld augments agentic planning with multi-step world imagination: the planner proposes alternative interaction trajectories, a verifier evaluates their prospective outcomes, and the selected action is executed before replanning from the next observation.

To tackle these challenges, we propose **ParallelWorld**, an adaptive-horizon test-time scaling framework that augments agentic embodied reasoning with explicit evaluation of future action consequences. Specifically, we construct a search tree through imagination, with each branch representing a candidate action sequence. By extending multiple branches in parallel over several interaction steps, the planner can assess an action in terms of the subsequent observations. A verifier guides this process by assessing the evolving branches against the task objective, retaining promising continuations and pruning less useful alternatives. We couple this prospective search with receding-horizon execution: the agent performs only the first action of the selected trajectory and initiates a new search after observing the resulting state. ParallelWorld directs additional test-time computation toward examining alternative courses of interaction before execution.

We conduct extensive experiments on ESI-Bench (Hong et al., 2026) and LIBERO-Pro (Zhou et al., 2025), covering embodied understanding and manipulation, respectively. These benchmarks examine two complementary aspects of agentic embodied reasoning: planning interactions to acquire task-relevant evidence and planning actions to achieve a physical goal. Experiments demonstrate improvements in both settings, supporting prospective world simulation as a mechanism for strengthening high-level embodied planning. Our main contributions are summarized as follows:

- We propose ParallelWorld, an adaptive-horizon test-time scaling framework that enables agentic embodied reasoning through the prospective simulation of alternative trajectories.
- We introduce a verifier-guided tree-search mechanism that evaluates action sequences and couples trajectory selection with receding-horizon execution.
- We demonstrate improvements on ESI-Bench and LIBERO-Pro, establishing the effectiveness of prospective planning for both understanding and manipulation embodied tasks.

2 Related Work

2.1 LLM-Based Agentic Systems

Early LLM-based agents developed different ways to connect reasoning with external actions. ReAct (Yao et al., 2022) interleaved reasoning and acting during task solving, while Toolformer (Schick et al., 2023) trained a language model to invoke external tools. AutoGen (Wu et al., 2023) supported more complex workflows through programmable coordination among agents. Executable code provided another action interface: CodeAct (Wang et al., 2024) represented actions as Python programs, enabling compositional tool use and refinement based on execution feedback. In software engineering, SWE-agent (Yang et al., 2024) introduced interfaces for repository navigation, editing, and

testing, while OpenHands (Wang et al., 2025) integrated code execution, command-line interaction, and web browsing. Meta-Harness (Lee et al., 2026) further searched over harness code, using evaluations and execution traces of previous candidates to guide subsequent revisions. These works established interfaces for planning, tool use, and action execution. In embodied tasks, agents must also ground their decisions in visual observations and executable actions within an environment.

2.2 Agentic Embodied Systems

Recent embodied systems increasingly rely on high-level agents to structure perception, reasoning, and physical interaction into goal-directed behaviors. For embodied understanding, earlier studies improve spatial reasoning from given observations through geometry-aware representations and supervision (Chen et al., 2024; Cheng et al., 2024; Wu et al., 2025; Fan et al., 2026), while active approaches acquire additional task-relevant evidence through exploration or physical interaction (Das et al., 2018; Majumdar et al., 2024; Ren et al., 2024; Zhou et al., 2026; Hong et al., 2026). In robotic manipulation, agentic systems translate language instructions into executable programs or high-level primitives, as exemplified by Code as Policies (Liang et al., 2023), CaP-X (Fu et al., 2026), ASPIRE (Lu et al., 2026), VLAs-as-Tools (Lei et al., 2026), and Harness VLA (Zhang et al., 2026). Although these two lines of work differ in their immediate objectives, both rely on high-level agents to organize sequences of interactions according to the task context: understanding-oriented agents act to acquire informative evidence, whereas manipulation-oriented agents act to transform the environment toward a desired goal. This common structure motivates prospective reasoning over alternative action outcomes before execution, which forms the basis of ParallelWorld.

2.3 Embodied Test-Time Scaling

Test-time scaling improves model inference by allocating additional computation at deployment through candidate generation, search, and verification. Early language-model approaches generate multiple candidate solutions and rank them with learned verifiers (Cobbe et al., 2021), while subsequent work extends this paradigm through structured search (Yao et al., 2023) and adaptive allocation of inference-time computation (Snell et al., 2025).

Recent studies have extended test-time scaling to embodied intelligence, where additional computation is used to improve action selection before execution. RoboMonkey (Kwok et al., 2025) samples multiple VLA actions and selects among them using a vision-language verifier, while RoVer (Dai et al., 2025) introduces a robot process reward model to score and iteratively refine candidate actions. TACO (Yang et al., 2025b) improves inference stability by selecting reliable action chunks from multiple VLA samples using a lightweight pseudo-count verifier, whereas E-TTS (Ye et al., 2026) jointly scales reasoning and action generation through history-aware verification and iterative refinement. Beyond direct action verification, ω-EVA (Sun et al., 2026b) incorporates proposal-conditioned future consequences into action refinement, and Gated GeoBoN (Zhao et al., 2026) selectively allocates additional rollout computation and ranks candidate rollouts according to action–future and geometric consistency. ParallelWorld advances this line of research by introducing prospective reasoning over alternative high-level embodied interactions before execution.

3 Methods

3.1 Problem Formulation

We consider agentic embodied reasoning as high-level planning in a partially observed environment. Given a language input ℓ and the current observation o_t, the planner selects executable high-level actions. An interaction episode yields a sequence:

$$\tau = (a_0, a_1, \ldots, a_{T-1}), \qquad a_t = (p_t, \theta_t) \in \mathcal{A}, \tag{1}$$

where p_t denotes an atomic action primitive and θ_t its associated arguments. Each primitive defines a bounded unit of interaction with a well-defined execution scope and termination condition. The planner therefore operates over semantic actions rather than directly predicting low-level trajectories.

Executing a primitive induces a transition in the physical environment,

$$s_{t+1} = \mathcal{T}(s_t, a_t), \qquad o_{t+1} = \mathcal{O}(s_{t+1}), \tag{2}$$

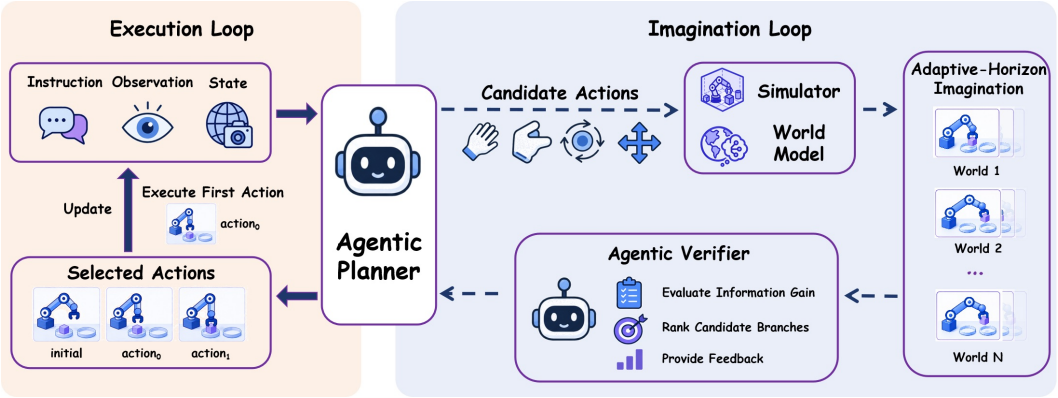

Figure 2: **Overview of ParallelWorld.** Given a task instruction and the current observation, an agentic planner explores alternative actions through simulator-based or world-model-based imagination. A verifier selects promising branches, while the planner determines whether to continue imagination or commit to execution. Only the first action of the selected trajectory is executed, followed by replanning from the resulting observation. Dashed and solid arrows denote imagination and environment interaction, respectively.

where s_t denotes the underlying environment state and o_t the resulting observation. The primitive space may combine perceptual and physical interactions. For example, camera motion and object interaction form the action space for embodied understanding, while robotic manipulation is expressed through a fixed set of analytic and VLA-backed primitives (Zhang et al., 2026).

Within this shared formulation, the two task settings differ primarily in their objectives. For understanding, the language input is a question q, and the agent interacts with the environment to acquire evidence $\mathcal{E}_{1:T}$ for predicting an answer:

$$\hat{y} = A\left(q, \mathcal{E}_{1:T}\right). \tag{3}$$

For manipulation, the language input is an instruction g. Repeated application of the single-step transition $\mathcal{T}$ should bring the environment to a terminal state satisfying the task goal:

$$s_T \in \mathcal{G}(g), \tag{4}$$

where $\mathcal{G}(g)$ denotes the state consistent with g. The overall framework is illustrated in Figure 2.

3.2 World Imagination

Before committing to an action, ParallelWorld constructs multiple prospective worlds to evaluate alternative courses of interaction. Starting from the current environment observation o_t, the imagination process proceeds recursively over prospective steps indexed by h, which are distinguished from actual environment interactions indexed by t.

We represent a prospective world as:

$$b_h = (o_h, \xi_h), \tag{5}$$

where o_h is its prospective observation and ξ_h is an optional restorable simulator state. We write $o(b_h) = o_h$ and $\xi(b_h) = \xi_h$ for its two components. For simulator-based imagination, ξ_h retains the state needed to resume that branch; for learned-model imagination, $\xi_h = \varnothing$ because candidate outcomes are generated without restoring a simulator state. The root branch is:

$$\mathcal{B}_0 = \{b_0\}, \qquad b_0 = (o_t, \xi_t), \tag{6}$$

where ξ_t is the current simulator snapshot when simulator-based imagination is used.

At imagination step h, the planner proposes a set of candidate primitive actions conditioned on the language objective and the current prospective worlds:

$$\mathcal{C}_h = \Pi_{\text{cand}}\left(\ell, \mathcal{B}_{h-1}\right), \qquad K_h = |\mathcal{C}_h|, \tag{7}$$

where $\mathcal{B}_{h-1}$ denotes the prospective worlds retained from the previous step, and K_h is the search width determined by the planner. It may consider several alternative actions when further comparison is necessary, while proposing fewer candidates when the next interaction is sufficiently clear. Each candidate

$$c_h^{(i)} = \left(b_{h-1}^{p(i)}, a_h^{(i)}\right), \qquad i = 1, \ldots, K_h, \tag{8}$$

associates an atomic primitive $a_h^{(i)}$ with one retained parent world $b_{h-1}^{p(i)}$. Here $p(i)$ identifies the parent of candidate i.

Each candidate is mapped to a prospective successor through a world-imagination operator:

$$\tilde{b}_h^{(i)} = \mathcal{W}\left(b_{h-1}^{p(i)}, a_h^{(i)}; \ell\right). \tag{9}$$

The implementation of $\mathcal{W}$ depends on the task setting. For understanding, the task often requires acquiring evidence that is not observable from the current view. Predicting such content directly from the available observation would require a learned world model to infer the very information being sought. We therefore obtain prospective outcomes from the simulator:

$$\xi_h^{(i)} = \mathcal{T}\left(\xi\big(b_{h-1}^{p(i)}\big), a_h^{(i)}\right), \qquad \tilde{b}_h^{(i)} = \left(\mathcal{O}\big(\xi_h^{(i)}\big), \xi_h^{(i)}\right). \tag{10}$$

Before evaluating a candidate, its parent simulator state is restored. This allows different actions to be evaluated from the same parent world. The successor retains both its resulting observation and the state needed for further expansion.

For manipulation, we instead use an image generation model to predict the immediate visual outcome of each primitive. Given a proposed primitive $a_h^{(i)}$, we first convert it into a natural-language action prompt:

$$\sigma_h^{(i)} = \mathcal{F}\left(\ell, o\big(b_{h-1}^{p(i)}\big), a_h^{(i)}\right). \tag{11}$$

The image model then generates the corresponding post-action observation:

$$\tilde{b}_h^{(i)} = \left(G_\psi\left(o\big(b_{h-1}^{p(i)}\big), \sigma_h^{(i)}\right), \varnothing\right). \tag{12}$$

The generated observation serves as the starting point for subsequent prospective expansion.

The resulting candidate worlds

$$\widetilde{\mathcal{B}}_h = \left\{\tilde{b}_h^{(1)}, \ldots, \tilde{b}_h^{(K_h)}\right\} \tag{13}$$

are jointly evaluated by a verifier:

$$(\mathcal{B}_h, \rho_h^*, f_h) = V_\phi\left(\ell, \widetilde{\mathcal{B}}_h\right), \qquad M_h = |\mathcal{B}_h|, \tag{14}$$

where $\mathcal{B}_h \subseteq \widetilde{\mathcal{B}}_h$ denotes the prospective worlds retained for further expansion, ρ_h^* is the verifier's currently preferred root-to-leaf trajectory, and f_h denotes its comparative feedback. The verifier evaluates the observations associated with the candidates; simulator states are retained internally for branch restoration rather than presented as visual evidence. Parent links are preserved so that the preferred trajectory can be traced from its selected leaf to the root. The retention width M_h is determined by the verifier after observing the prospective outcomes.

Finally, the planner determines whether another imagination step is necessary:

$$z_h = \Pi_{\text{ctrl}}\left(\ell, \mathcal{B}_h, \rho_h^*, f_h\right), \qquad z_h \in \{\text{continue}, \text{stop}\}. \tag{15}$$

If $z_h = \text{continue}$, the retained worlds become the parents of the next prospective step and the planner generates a new candidate set $\mathcal{C}_{h+1}$; otherwise, imagination terminates and the selected trajectory ρ_h^* is passed to the interaction stage.

3.3 Foresight-Guided Interaction

When the planner terminates world imagination, the verifier-selected trajectory determines the next environment interaction. Let H denote the imagination depth at termination, and let

$$\rho_H^* = \left(a_{t,1}^*, a_{t,2}^*, \ldots, a_{t,H}^*\right) \tag{16}$$

Table 1: **Quantitative results on ESI-Bench (Hong et al., 2026).** We report accuracy for Passive, Active Exploration, and ParallelWorld. All active methods use the same answer-model. Row colors follow the ESI-Bench taxonomy: yellow denotes object representation, purple denotes layout and geometry, blue denotes number representation, and green denotes agents and goal-directed actions.

Category	Subcategory	Agent Evaluation		
		Passive	Active	ParallelWorld
Perceptual Grounding	*View Hallucination*	53.32%	66.82%	**71.40%**
	Partial Occlusion	36.84%	57.89%	**77.89%**
	Material Transparency	69.72%	71.10%	**73.85%**
Physical Structure	*Rigid Containment*	55.00%	**60.00%**	50.00%
	Deformable Objects	37.76%	40.82%	**42.86%**
Physical Dynamics	*Inclined Plane*	59.02%	83.61%	**85.25%**
	Stacking & Stability	1.12%	45.98%	**53.95%**
Specular Reflection	*Reflection Authoring*	56.57%	51.52%	**60.61%**
	Spatial Relations	38.94%	36.28%	**51.33%**
	Correspondence	42.05%	47.73%	**48.86%**
Spatial Relations	*Linear Alignment*	52.13%	68.09%	**70.21%**
	Geometric Configuration	16.55%	**17.96%**	16.90%
	Physical Contact	66.39%	68.07%	**73.95%**
Metric Comparison	*Dimensional Size*	39.52%	44.91%	**49.10%**
	Spatial Distance	61.18%	58.55%	**67.11%**
Cognitive Mapping	*Connectivity*	**66.67%**	55.00%	60.00%
	Traversable Passage	68.33%	65.00%	**75.00%**
	Regional Boundary	45.00%	57.50%	**63.75%**
	Long-Term Navigation	16.67%	16.67%	**18.33%**
Enumerative Perception	*Counting with Occlusion*	23.33%	26.67%	**30.00%**
	Spatial Segmentation	20.00%	20.00%	**23.33%**
	Merged Observation	0.00%	1.67%	**6.67%**
	Category Ambiguity	0.00%	8.33%	**13.33%**
	Structural Enclosure	2.50%	5.00%	**7.50%**
	Illumination Variability	0.00%	6.00%	**10.00%**
Temporal Understanding	*Unobserved Change*	38.62%	70.95%	**89.86%**
	Agent Observation	34.59%	48.87%	**62.41%**
Action Sequencing	*Action Order Inference*	37.66%	53.25%	**61.04%**

denote the preferred trajectory returned by the verifier, where t indexes the current environment interaction step and $a^*_{t,j}$ is the j-th imagined primitive along the trajectory.

Rather than executing the entire imagined trajectory, ParallelWorld commits only its first action:

$$\hat{a}_t = a^*_{t,1}. \tag{17}$$

The selected primitive is executed in the environment, yielding the next state and observation:

$$s_{t+1} = \mathcal{T}(s_t, \hat{a}_t), \qquad o_{t+1} = \mathcal{O}(s_{t+1}). \tag{18}$$

By executing only the first action, ParallelWorld avoids committing to an entire imagined trajectory whose later states may become increasingly inaccurate, while preserving the benefit of multi-step lookahead for current action. The newly obtained observation then initiates the next planning cycle.

4 Experiments

We evaluate ParallelWorld on embodied understanding and manipulation. We first evaluate the effectiveness of prospective world imagination for embodied understanding in Section 4.1. In Section 4.2, we further assess ParallelWorld on robotic manipulation, examining whether imagined action outcomes improve high-level planning and task execution. In Section 4.4, we conduct an ablation study on the search width K_h in ParallelWorld.

Table 2: **Quantitative results on LIBERO-Pro (Zhou et al., 2025).** We report success rates (%) under instruction-redirection (T) and position-swap (S) perturbations across the SPATIAL, OBJECT, GOAL, and LIBERO-10 suites. Each setting contains 100 trials (10 tasks $\times$ 10 seeds).

Method	Spatial		Object		Goal		LIBERO-10	
	T	S	T	S	T	S	T	S
OpenVLA (Kim et al., 2024)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
π_0 (Black et al., 2024)	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0
$\pi_{0.5}$ (Intelligence et al., 2025)	1.0	20.0	1.0	17.0	2.0	38.0	1.0	8.0
MolmoAct (Lee et al., 2025)	0.0	0.0	0.0	6.0	0.0	0.0	6.0	0.0
NORA (Hung et al., 2025)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
X-VLA (Zheng et al., 2026)	0.0	0.0	8.0	2.0	9.0	1.0	10.0	0.0
AtomVLA (Sun et al., 2026a)	1.0	16.0	0.0	10.0	11.0	2.0	9.0	1.0
HarnessVLA (Zhang et al., 2026)	79.0	56.0	79.0	85.0	62.0	57.0	42.0	29.0
ParallelWorld (Ours)	**88.0**	**83.0**	**94.0**	**88.0**	**65.0**	**66.0**	**50.0**	**30.0**

4.1 EMBODIED UNDERSTANDING

4.1.1 EXPERIMENTAL SETUP

Benchmark. We conduct experiments on ESI-Bench (Hong et al., 2026), a comprehensive benchmark for embodied spatial intelligence. The benchmark contains 10 task categories and 29 subcategories. Due to an environment issue that prevents the execution of pouring actions, we exclude the Liquid Volume subcategory and evaluate the remaining 28 subcategories.

Baselines. We selected two paradigms from ESI-Bench (Hong et al., 2026) as baselines, including Passive Single-View, which predicts the answer from a single observation at the initial pose, and Active Exploration, which sequentially selects camera or physical actions to gather task-relevant evidence before answering.

Implementation Details. We build our implementation upon the active exploration framework provided by ESI-Bench (Hong et al., 2026). The planner iteratively decides whether to continue exploration based on the confidence of the current prediction. Specifically, exploration terminates when the answer confidence exceeds 0.8 or when the maximum exploration depth of 15 is reached.

4.1.2 MAIN RESULTS

As shown in Table 1, ParallelWorld consistently improves over Active Exploration on the majority of ESI-Bench subcategories, demonstrating the effectiveness of prospective world imagination for embodied understanding. The gains are particularly evident on tasks that require additional interaction to reveal task-relevant evidence or resolve uncertainty, where the accuracy on Unobserved Change increases from 70.95% to 89.86% and Partial Occlusion improves from 57.89% to 77.89%. Compared with sequential action selection, ParallelWorld benefits from evaluating multiple possible future trajectories before execution, leading to more effective decisions across diverse spatial reasoning tasks. We also observe several subcategories where the improvement is limited or performance slightly decreases compared with the baselines. These cases are further analyzed in Section 4.3.

4.2 ROBOTIC MANIPULATION

4.2.1 EXPERIMENTAL SETUP

Benchmark. We evaluate manipulation performance on LIBERO-Pro (Zhou et al., 2025), an extension of LIBERO designed to assess robotic generalization under perturbations rather than the standard in-distribution setting. We consider the SPATIAL, OBJECT, GOAL and LIBERO-10 suites under instruction- redirection (**T**) and position swap (**S**) perturbation. Each evaluation cell contains 10 tasks evaluated over 10 seeds, resulting in 100 episodes per setting and 800 episodes in total.

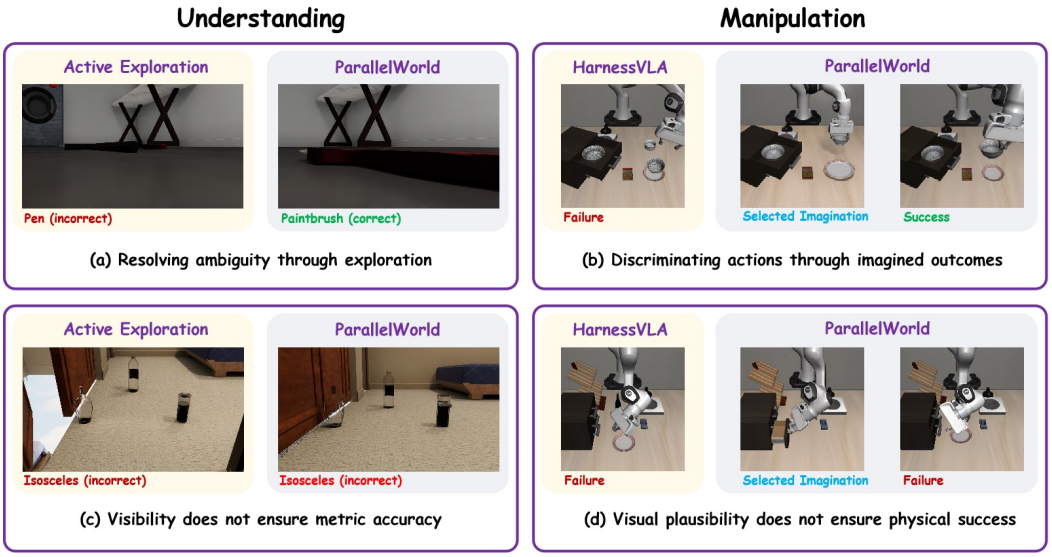

Figure 3: **Qualitative analysis of ParallelWorld.** (a) **Prospective exploration.** Active Exploration selects an uninformative view and fails, while ParallelWorld discovers the discriminative observation through future interaction. (b) **Prospective manipulation.** HarnessVLA (Zhang et al., 2026) fails with direct execution, whereas ParallelWorld selects an imagined outcome that guides the robot toward a better grasping configuration and achieves success. (c) **Limitation in geometric reasoning.** ParallelWorld obtains sufficient observations, but the model still fails to infer precise metric relationships between objects. (d) **Limitation in physical prediction.** The imagined outcome appears plausible, but visual imagination cannot fully capture contact dynamics required for successful physical execution.

Baselines. We compare ParallelWorld with representative end-to-end VLA models and agentic manipulation methods. The VLA baselines include OpenVLA (Kim et al., 2024), π_0 (Black et al., 2024), $\pi_{0.5}$ (Intelligence et al., 2025), MolmoAct (Lee et al., 2025), NORA (Hung et al., 2025), X-VLA (Zheng et al., 2026), and AtomVLA (Sun et al., 2026a). We use HarnessVLA (Zhang et al., 2026) as the direct agentic baseline and reproduce it with Codex as the agentic planner, whereas the results of the remaining frozen VLA baselines are taken from the original HarnessVLA paper.

Implementation Details. Our manipulation implementation is built upon the HarnessVLA (Zhang et al., 2026) framework. For visual imagination, we use GPT-image-2.5 as the image generation model with low image quality settings to predict the prospective outcomes of candidate primitives.

4.2.2 Main Results

As shown in Table 2, ParallelWorld surpasses HarnessVLA (Zhang et al., 2026) across all LIBERO-Pro evaluation settings, validating the effectiveness of prospective world imagination for robotic manipulation. The largest gains are observed on the Spatial and Object suites, where ParallelWorld substantially improves the success rates over the agentic baseline, including 56.0% to 83.0% on Spatial position-swap and 79.0% to 94.0% on Object instruction-redirection. Improvements on the Goal and LIBERO-10 suites are comparatively more moderate, although ParallelWorld remains consistently stronger across all corresponding settings. We further analyze the variation in performance gains across different task suites in Section 4.3.

4.3 Discussions

When does ParallelWorld help? ParallelWorld is most effective when task success depends on discriminating among alternative action outcomes and when the value of an action emerges through its downstream consequences. This is particularly evident in spatial and object-centric manipulation, where candidate actions differ in target identity, relative configuration, or placement geometry, as

Table 3: **Ablation of adaptive search width on Spatial-S.** Results are averaged over three tasks with 10 seeds each. Imagination Time and Total Time represent the average time cost of imagination and the entire episode, respectively.

Method	Success Rate (%) ↑	Total Time (s) ↓
Direct Execution (Zhang et al., 2026)	23.3	417.4
Fixed Search Width ($K_h = 2$)	10.0	2930.0
Fixed Search Width ($K_h = 4$)	36.7	4889.5
Adaptive Search Width (Ours)	**73.3**	**2402.2**

well as in embodied understanding tasks that require additional interaction to acquire otherwise unavailable evidence. By explicitly comparing prospective worlds before execution, ParallelWorld enables the planner to select interactions with higher expected information gain or task progress rather than relying on the most plausible immediate action. As illustrated in Figure 3, ParallelWorld can identify informative viewpoints in understanding tasks and select more favorable manipulation outcomes before execution, reducing errors caused by incorrect exploration or action selection.

Limitations. The effectiveness of ParallelWorld is limited when the decisive factors are not fully reflected in visual outcomes. As shown in Figure 3, the method may still fail when the underlying challenge requires precise metric reasoning or accurate prediction of physical interactions. For example, although sufficient visual evidence is acquired, current agentic models may struggle with geometric relationships, while image-based imagination may not faithfully capture contact dynamics and execution uncertainty. These limitations are also reflected in Goal and LIBERO-10 tasks, where articulated states, contact-dependent transitions, and long-horizon error accumulation play a larger role. Overall, the effectiveness of prospective imagination depends on whether task-relevant consequences can be reliably represented in the imagined observations.

4.4 Ablation Studies

We investigate the effect of adaptive search width by comparing ParallelWorld with fixed search width strategies on three Spatial-S tasks with 30 episodes in total. As shown in Table 3, direct execution achieves only 23.3% success rate, demonstrating the difficulty of solving these tasks without prospective evaluation. Fixed search width strategies provide limited improvements, where $K_h = 2$ achieves 10.0% success rate and $K_h = 4$ improves to 36.7% with higher computational cost.

The inferior performance of fixed search widths stems from their inability to adapt the number of candidate worlds to the uncertainty of the current decision. A small fixed width may discard informative candidates when multiple alternatives need to be explored, while a large fixed width introduces unnecessary candidates in relatively deterministic states, increasing both computation and the difficulty of trajectory selection. In contrast, Adaptive Search Width dynamically determines K_h according to the current task state and planning context, allocating more search capacity when alternative futures are ambiguous and reducing redundant exploration otherwise. It achieves a substantially higher success rate of 73.3% while maintaining lower runtime than the fixed $K_h = 4$ strategy, demonstrating the importance of adaptive search width for effective test-time planning.

5 Conclusion

We presented ParallelWorld, an adaptive-horizon test-time scaling framework for agentic embodied reasoning. Rather than directly committing to locally plausible actions, ParallelWorld explicitly explores alternative prospective worlds before execution, allowing future observations and task consequences to inform high-level decisions. We instantiate ParallelWorld for both embodied understanding and robotic manipulation, using simulator-based transitions and image-based outcome prediction, respectively. Experiments demonstrate consistent improvements over existing active and agentic baselines, while ablation studies further validate the effectiveness of adaptive search width. These results establish prospective world imagination as an effective mechanism for extending agentic embodied systems from reactive action selection toward foresight-driven decision making.

REFERENCES

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. π_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

Boyuan Chen, Zhuo Xu, Sean Kirmani, Brian Ichter, Dorsa Sadigh, Leonidas Guibas, and Fei Xia. SpatialVLM: Endowing vision-language models with spatial reasoning capabilities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14455–14465, 2024.

Shizhe Chen, Pierre-Louis Guhur, Cordelia Schmid, and Ivan Laptev. History aware multimodal transformer for vision-and-language navigation. *Advances in neural information processing systems*, 34:5834–5847, 2021.

An-Chieh Cheng, Hongxu Yin, Yang Fu, Qiushan Guo, Ruihan Yang, Jan Kautz, Xiaolong Wang, and Sifei Liu. SpatialRGPT: Grounded spatial reasoning in vision-language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 135062–135093, 2024.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Mingtong Dai, Lingbo Liu, Yongjie Bai, Yang Liu, Zhouxia Wang, Rui Su, Chunjie Chen, Liang Lin, and Xinyu Wu. Rover: Robot reward model as test-time verifier for vision-language-action model. *arXiv preprint arXiv:2510.10975*, 2025.

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*, pp. 1–10, 2018.

Zhiwen Fan, Jian Zhang, Renjie Li, Junge Zhang, Runjin Chen, Hezhen Hu, Kevin Wang, Peihao Wang, Huaizhi Qu, Shijie Zhou, et al. VLM-3R: Vision-language models augmented with instruction-aligned 3D reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 31054–31065, 2026.

Letian Fu, Justin Yu, Karim El-Refai, Ethan Kou, Haoru Xue, Huang Huang, Wenli Xiao, Guanzhi Wang, Dantong Niu, Fei-Fei Li, et al. Cap-x: A framework for benchmarking and improving coding agents for robot manipulation. *arXiv preprint arXiv:2603.22435*, 2026.

Yining Hong, Jiageng Liu, Han Yin, Manling Li, Leonidas Guibas, Li Fei-Fei, Jiajun Wu, and Yejin Choi. Esi-bench: Towards embodied spatial intelligence that closes the perception-action loop. *arXiv preprint arXiv:2605.18746*, 2026.

Chia-Yu Hung, Qi Sun, Pengfei Hong, Amir Zadeh, Chuan Li, U Tan, Navonil Majumder, Soujanya Poria, et al. Nora: A small open-sourced generalist vision language action model for embodied tasks. *arXiv preprint arXiv:2504.19854*, 2025.

Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, et al. $\pi_{0.5}$: a vision-language-action model with open-world generalization. *arXiv preprint arXiv:2504.16054*, 2025.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

Jacky Kwok, Christopher Agia, Rohan Sinha, Matt Foutter, Shulu Li, Ion Stoica, Azalia Mirhoseini, and Marco Pavone. Robomonkey: Scaling test-time sampling and verification for vision-language-action models. *arXiv preprint arXiv:2506.17811*, 2025.

Jason Lee, Jiafei Duan, Haoquan Fang, Yuquan Deng, Shuo Liu, Boyang Li, Bohan Fang, Jieyu Zhang, Yi Ru Wang, Sangho Lee, et al. Molmoact: Action reasoning models that can reason in space. *arXiv preprint arXiv:2508.07917*, 2025.

Yoonho Lee, Roshen Nair, Qizheng Zhang, Kangwook Lee, Omar Khattab, and Chelsea Finn. Meta-harness: End-to-end optimization of model harnesses. *arXiv preprint arXiv:2603.28052*, 2026.

Zixing Lei, Changxing Liu, Yichen Xiong, Minhao Xiong, Yuanzhuo Ding, Zhipeng Zhang, Weixin Li, and Siheng Chen. Towards long-horizon embodied agents with tool-aligned vision-language-action models. *arXiv preprint arXiv:2605.13119*, 2026.

Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control. In *2023 IEEE International conference on robotics and automation (ICRA)*, pp. 9493–9500. IEEE, 2023.

Runyu Lu, Yubo Wu, Ethan Kou, Letian Fu, Wenli Xiao, Ajay Mandlekar, Yinzhen Xu, Guanya Shi, Ken Goldberg, Ang Chen, et al. Aspire: Agentic/skills discovery for robotics. *arXiv preprint arXiv:2607.00272*, 2026.

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 *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16488–16498. IEEE, 2024.

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.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *Advances in neural information processing systems*, 36:68539–68551, 2023.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling parameters for reasoning. In *International Conference on Learning Representations*, volume 2025, pp. 10131–10165, 2025.

Xiaoquan Sun, Zetian Xu, Chen Cao, Zonghe Liu, Yihan Sun, Jingrui Pang, Ruijian Zhang, Zhen Yang, Kang Pang, Dingxin He, et al. Atomvla: Scalable post-training for robotic manipulation via predictive latent world models. *arXiv preprint arXiv:2603.08519*, 2026a.

Zhenguo Sun, Yu Sun, Hande Huang, and Alois Knoll. ω-eva: Envision, verify, and act with latent interactive world models. *arXiv preprint arXiv:2606.09457*, 2026b.

Xingyao Wang, Yangyi Chen, Lifan Yuan, Yizhe Zhang, Yunzhu Li, Hao Peng, and Heng Ji. Executable code actions elicit better llm agents. *arXiv preprint arXiv:2402.01030*, 2024.

Xingyao Wang, Boxuan Li, Yufan Song, Frank F Xu, Xiangru Tang, Mingchen Zhuge, Jiayi Pan, Yueqi Song, Bowen Li, Jaskirat Singh, et al. Openhands: An open platform for ai software developers as generalist agents. In *International Conference on Learning Representations*, volume 2025, pp. 65882–65919, 2025.

Diankun Wu, Fangfu Liu, Yi-Hsin Hung, and Yueqi Duan. Spatial-MLLM: Boosting MLLM capabilities in visual-based spatial intelligence. In *Advances in Neural Information Processing Systems*, volume 38, pp. 13569–13597, 2025.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, et al. Autogen: Enabling next-gen llm applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*, 2023.

John Yang, Carlos Jimenez, Alexander Wettig, Kilian Lieret, Shunyu Yao, Karthik Narasimhan, and Ofir Press. Swe-agent: Agent-computer interfaces enable automated software engineering. *Advances in Neural Information Processing Systems*, 37:50528–50652, 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*, 2025a.

Siyuan Yang, Yang Zhang, Haoran He, Ling Pan, Xiu Li, Chenjia Bai, and Xuelong Li. Steering vision-language-action models as anti-exploration: A test-time scaling approach. *arXiv preprint arXiv:2512.02834*, 2025b.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822, 2023.

Wen Ye, Peiyan Li, Tingyu Yuan, Yuan Xu, Xiangnan Wu, Chaoyang Zhao, Jing Liu, Nianfeng Liu, Yan Huang, and Liang Wang. E-tts: A new embodied test-time scaling framework for robotic manipulation. In *European Conference on Computer Vision*, pp. 395–412. Springer, 2026.

Yixian Zhang, Huanming Zhang, Feng Gao, Xiao Li, Zhihao Liu, Chunyang Zhu, Jiaxing Qiu, Yuchen Yan, Jiyuan Liu, Wenhao Tang, et al. Harness vla: Steering frozen vlas into reliable manipulation primitives via memory-guided agents. *arXiv preprint arXiv:2607.08448*, 2026.

Zesen Zhao, Minkyoung Cho, Boyuan Zheng, Kunxiao Gao, Yulong Cao, Z Morley Mao, et al. Test-time scaling for world action models via zero-shot geometric evaluation. *arXiv preprint arXiv:2607.17454*, 2026.

Jinliang Zheng, Jianxiong Li, Zhihao Wang, Dongxiu Liu, Xirui Kang, Yuchun Feng, Yinan Zheng, Jiayin Zou, Yilun Chen, Jia Zeng, et al. X-vla: Soft-prompted transformer as scalable cross-embodiment vision-language-action model. In *International Conference on Learning Representations*, volume 2026, pp. 60580–60606, 2026.

Weijie Zhou, Xuantang Xiong, Yi Peng, Manli Tao, Chaoyang Zhao, Honghui Dong, Ming Tang, and Jinqiao Wang. Physvlm-avr: Active visual reasoning for multimodal large language models in physical environments. *Advances in Neural Information Processing Systems*, 38:3873–3906, 2026.

Xueyang Zhou, Yangming Xu, Guiyao Tie, Yongchao Chen, Guowen Zhang, Duanfeng Chu, Pan Zhou, and Lichao Sun. Libero-pro: Towards robust and fair evaluation of vision-language-action models beyond memorization. *arXiv preprint arXiv:2510.03827*, 2025.

A Appendix

A.1 AI Use Statement

We used generative AI tools to assist with language editing and polishing of the manuscript, including improving grammar, clarity, and readability. All AI-assisted edits were reviewed by the authors to ensure that they accurately reflected the intended meaning. The authors take full responsibility for the final content of the paper.