ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Rapid Robot Manipulation Policy Learning via Hierarchical Foundation-Model Prior Distillation

***Abstract*—In robotic skill acquisition, rapid policy learning remains challenging due to high-dimensional state-action spaces and inefficient exploration in the early stage of training [1]. Although the pre-trained OpenVLA model exhibits cross-task generalization and can generate goal-directed actions for unseen tasks under suitable prompts, its direct application to novel manipulation tasks remains limited, while full fine-tuning is computationally expensive. To address this issue, we propose a hierarchical framework that combines OpenVLA with reinforcement learning for efficient skill acquisition. Specifically, OpenVLA is used to generate diverse task-related prior trajectories through prompt engineering, and reinforcement learning leverages these priors to fit local dynamics and constrain policy exploration. In this way, the proposed method improves adaptation efficiency and accelerates policy learning on new tasks. We evaluate the framework on multiple manipulation tasks in the LIBERO environment.**

I. Introduction

Robotic manipulation in industrial settings remains challenging due to complex and time-varying contact interactions. Although reinforcement learning methods such as SAC, GPS, and PPO avoid explicit dynamics modeling, they often converge slowly in high-dimensional state–action spaces, especially during early exploration [2]. Existing approaches that improve learning with expert priors usually depend on high-quality task-specific demonstrations, which are costly to collect and generalize poorly beyond the training distribution.

We observe that the pre-trained OpenVLA model shows cross-task generalization and can generate goal-related trajectories for unseen manipulation tasks under suitable prompts. Although suboptimal, these trajectories provide useful priors for policy initialization, while directly adapting OpenVLA to new tasks remains computationally expensive [3] [4].

Motivated by the brain–spinal hierarchical motor control mechanism, we propose a two-level framework for efficient robotic skill acquisition. At the high level, OpenVLA generates candidate prior trajectories through prompt guidance. At the low level, a lightweight task-specific controller fits local dynamics from the selected priors and is further refined through environment interaction. In this way, the proposed framework reduces the cost of large-model adaptation and improves exploration efficiency in reinforcement learning.

Our main contributions are: (1) a prompt-based mechanism for generating and selecting useful prior trajectories from OpenVLA for unseen tasks; (2) a hierarchical prior distillation optimization framework that combines foundation-model priors with local dynamics fitting and controller refinement; and (3) validation on multiple LIBERO tasks, showing improved sample efficiency and final performance over baselines.

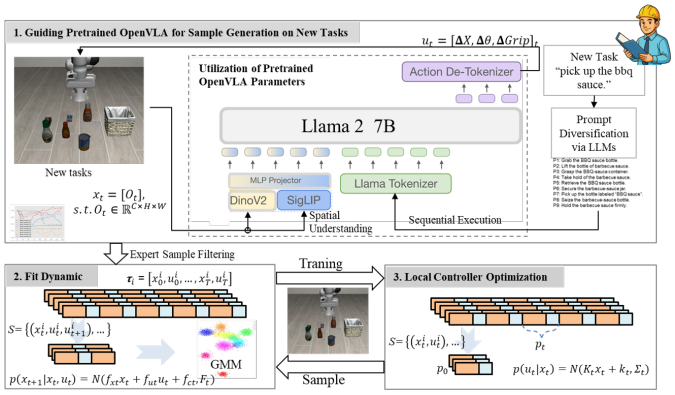

Fig. 1. The flowchart of the proposed Hierarchical Prior Distillation Optimization (HPDO) method.

II. METHOD

The proposed method has three stages. First, LLM-generated prompts guide OpenVLA to produce diverse candidate trajectories, from which informative priors are selected. Second, these priors are used to fit a local dynamics model under a continuous-control MDP. Third, the learned model initializes a local controller that is iteratively refined through interaction, enabling rapid policy acquisition for new tasks.

III. EXPERIMENTS

A. Experimental Setup

Experiments are conducted on four LIBERO manipulation tasks. We use the OpenVLA checkpoint fine-tuned on demonstrations from the same environment. The state includes 7-D joint positions and velocities, and the action includes 3-D end-effector displacement plus a gripper command, exchanged with the simulator at 10 Hz.

B. Generalization of the Pretrained OpenVLA Model

a) Evaluation of Prompt Generalization in Pretrained OpenVLA: We use ChatGPT-o3 to generate synonymous prompts for each task and guide OpenVLA to produce diverse trajectories. For Task A, examples include *Grab the BBQ sauce bottle*, *Lift the barbecue sauce bottle*, and *Pick up the bottle labeled BBQ sauce*. Fig. 3 shows that prompt variations can elicit task-relevant trajectories, providing useful priors for policy learning.

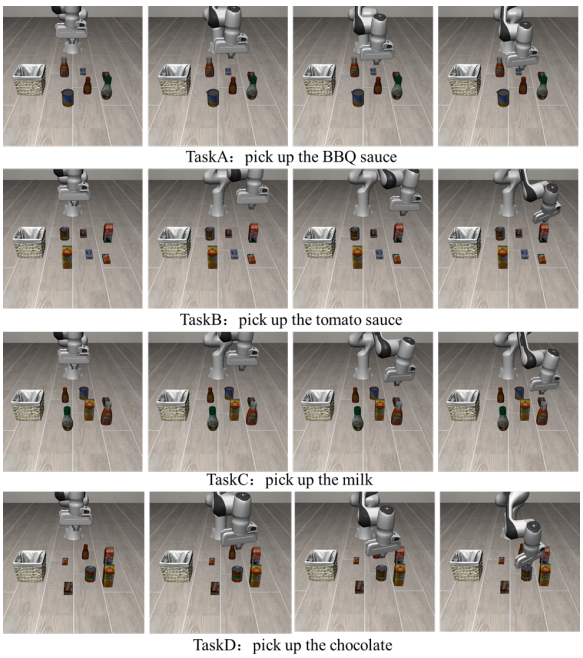

Fig. 2. Examples of experimental tasks are presented. The left image depicts the initial state of the task, the middle image shows the intermediate state, and the right image illustrates the final state.

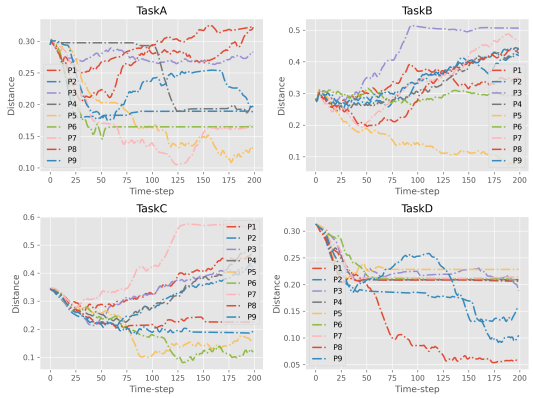

Fig. 3. Under the same pretrained model, robot manipulation trajectories are generated by OpenVLA using semantically equivalent prompts produced by an LLM. The horizontal axis represents the time steps, while the vertical axis denotes the distance to the target.

b) Generalization Capability of Different Pretrained Checkpoints: We evaluate four OpenVLA checkpoints: *spatial*, *object*, *goal*, and *10*. Since Tasks A–D match the scene distribution of *openvla-7b-finetuned-libero-object*, this checkpoint performs best and is used in subsequent experiments. The results also indicate limited cross-scene generalization of OpenVLA priors.

C. Comparison with Other Methods

To evaluate policy search efficiency, we compare the proposed method with the Linear Gaussian Controller (LGC) derived from Guided Policy Search (GPS), a representative model-based RL method with fast policy optimization. Both methods are tested on four manipulation tasks, with three runs per task; shaded areas indicate standard deviation. As shown in Fig. 5, the proposed method converges fastest on all four tasks. For example, in the lower-right plot, it reaches

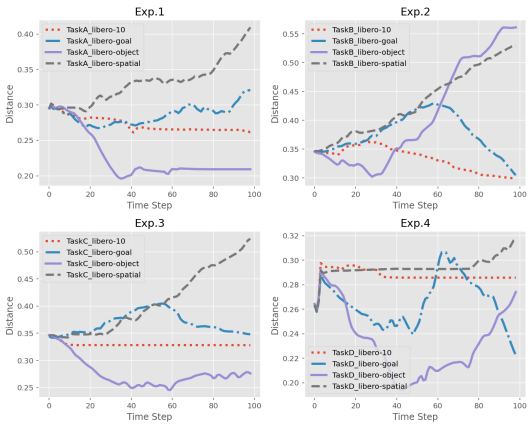

Fig. 4. Task performance of different OpenVLA checkpoints on novel tasks.

the target more quickly. These results verify the effectiveness of our method and match our design goal.

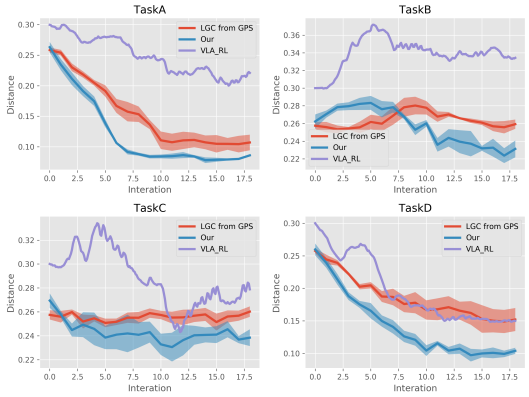

Fig. 5. Performance of Three Methods on Robotic Manipulation Tasks.

IV. CONCLUSIONS

Although multimodal foundation models show strong reasoning, scene understanding, and cross-task generalization, they are still insufficient for direct low-level robot control on unseen tasks without expert data or further adaptation. We show that guiding OpenVLA with LLM-generated prompt variations to produce trajectory priors, and combining these priors with reinforcement learning, can effectively constrain policy search and improve skill learning efficiency. Results in the LIBERO simulation environment verify the effectiveness of the proposed method.

REFERENCES

[1] K. Chatzilygeroudis, V. Vassiliades, F. Stulp *et al.*, "A survey on policy search algorithms for learning robot controllers in a handful of trials," *IEEE Trans. Robot.*, vol. 36, no. 2, pp. 328–347, 2019.

[2] S. Levine, N. Wagener, and P. Abbeel, "Learning contact-rich manipulation skills with guided policy search," *in Proc. IEEE Int. Conf. Robot. Autom.*, 2015.

[3] T. Kwon, N. Di Palo, and E. Johns, "Language models as zero-shot trajectory generators," *IEEE Robot. Autom. Lett.*, 2024.

[4] D. Qingwei, W. Tingting, Z. Peng *et al.*, "Enhancing robot learning through cognitive reasoning trajectory optimization under unknown dynamics," *IEEE Robot. Autom. Lett.*, 2025.