18 March 2025 · CC-BY 4.0

Qeios

Peer Review

Review of: "What's the Move? Hybrid Imitation Learning via Salient Points"

Nikhil Potu Surya Prakash[1]

[1]. University of California, Berkeley, United States

The paper presents SPHINX (Salient Point-Based Hybrid ImitatioN and eXecution), a hybrid imitation learning (IL) approach for robot manipulation tasks. SPHINX effectively combines multimodal observations—point clouds for long-range waypoints and wrist-camera images for precise manipulation—into a unified policy that switches between two action modes: waypoint-based coarse movements and dense fine-grained control. By learning to identify salient points within a scene, SPHINX significantly enhances spatial generalization and task execution efficiency. The method is evaluated across six robotic tasks, demonstrating superior performance over state-of-the-art IL baselines in both real-world and simulated environments.

Merits

Adaptive Hybrid Control: The core innovation of SPHINX lies in its ability to dynamically switch between waypoint-based and dense action policies using salient points. This adaptive approach addresses the limitations of prior IL methods that either rely exclusively on coarse waypoints or dense per-step control.

Efficient Data Collection & Training: The authors introduce a novel teleoperation system that enables efficient annotation of salient points and hybrid action collection, reducing the burden on human demonstrators.

Strong Empirical Validation: The method is benchmarked rigorously across multiple real-world tasks, with a significant performance boost (up to 1.7× speedup) over baselines such as Diffusion Policy, HYDRA, and OpenVLA. Notably, SPHINX shows strong generalization to new viewpoints and scene variations.

Thoughtful Baseline Comparisons: The authors carefully compare SPHINX against both image-based and waypoint-only policies, providing insights into the contributions of salient point detection and hybrid action switching.

doi.org/10.32388/AOM9SZ

Suggestions:

Computational Complexity & Training Overhead: While the hybrid policy structure is innovative, a more detailed analysis of its training time, computational cost, and memory footprint would benefit readers interested in scaling this approach.

Practical Deployment Considerations: The discussion would be strengthened by elaborating on real-world deployment scenarios. Specifically, how well does SPHINX perform under varying levels of sensor noise, occlusions, or deformable object interactions?

Hyper-parameter Sensitivity: The choice of hyperparameters, particularly for salient point detection and waypoint interpolation, is not deeply explored. A sensitivity analysis would provide a better understanding of how robust the method is to tuning variations.

Comparison with Model-Free Reinforcement Learning (RL): While IL is the primary focus, an additional discussion on how SPHINX compares to model-free RL methods in terms of sample efficiency and policy generalization could provide a more comprehensive perspective.

I look forward to seeing further refinements and applications of this approach in dynamic and unstructured real-world environments.

Declarations

Potential competing interests: No potential competing interests to declare.