LUMOS: Language-Conditioned Imitation Learning with World Models

Iman Nematollahi[1] Branton DeMoss[2] Akshay L Chandra[1]
Nick Hawes[2] Wolfram Burgard[3] Ingmar Posner[2]
[1]University of Freiburg [2]University of Oxford [3]University of Technology Nuremberg
Accepted at ICRA 2025 lumos.cs.uni-freiburg.de

Abstract

We introduce LUMOS, a language-conditioned imitation learning framework that acquires multi-task, long-horizon skills by training in the latent space of a learned world model and transfers them zero-shot to real robots. By learning on-policy in latent space, LUMOS mitigates distribution shift common in offline imitation learning. Coherent long-horizon behavior is achieved through latent planning, multimodal hindsight relabeling, and intrinsic rewards defined over multi-step rollouts. On the CALVIN benchmark, LUMOS outperforms prior methods on chained multi-task evaluations and is, to our knowledge, the first to achieve real-world, language-conditioned visuomotor control using an offline world model. ***Full paper available at:*** *https://arxiv.org/abs/2503.10370*

1. Introduction

Imitation learning suffers from covariate shift due to mismatches between expert and agent behavior, while reinforcement learning mitigates this via trial-and-error correction, but requires extensive real-world interaction. World model approaches improve sample efficiency but typically require structured environments or task-specific rewards. LUMOS learns a latent world model from teleoperated play data and trains language-conditioned policies offline through imagined rollouts. It requires under 1% language supervision, supports long-horizon tasks, and eliminates the need for further real-world interaction.

2. Approach

LUMOS learns policies through latent imagination in two stages. First, a DreamerV2-style world model [2] is trained on real-world RGB data to model latent dynamics. Then, a language-conditioned policy is trained offline in this space using a DITTO-style intrinsic reward [1] that matches expert trajectories. A CVAE-based planner samples language-conditioned latent segments, and a contrastive loss aligns language with rollouts. The policy is optimized via actor-critic learning framework, enabling long-horizon skills without further real-world data.

Figure 1. LUMOS learns language-conditioned policies in latent space, enabling error recovery and long-horizon performance.

3. Relevance to the Learning to Simulate Robot Worlds Workshop

LUMOS enables offline, language-conditioned imitation learning by training policies in the latent space of a world model learned from real-world teleoperated play data. It achieves state-of-the-art results on the CALVIN benchmark, outperforming prior methods across single-stage and multi-stage tasks. LUMOS excels particularly in long-horizon settings, reducing compounding errors and recovering from intermediate failures. In real-world experiments, policies trained entirely in imagination transfer zero-shot to physical robots, succeeding in complex tasks such as drawer opening, object pushing, and placement. These results demonstrate that compact, learned world models can function as high-fidelity, task-specific simulators, eliminating the need for manual simulation or reward design. This directly supports the workshop's goal of replacing handcrafted or photorealistic simulations with scalable, data-driven models for realistic and efficient robot learning.

References

[1] Branton DeMoss, Paul Duckworth, Nick Hawes, and Ingmar Posner. Ditto: Offline imitation learning with world models. *arXiv preprint arXiv:2302.03086*, 2023. 1

[2] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*, 2020. 1