
Demonstration-Augmented Skill injection for Human-guided VLA

Minkyu Kang

Department of Computer Science and Engineering
Ulsan National Institute of Science and Technology
Ulsan, South Korea
perelman@unist.ac.kr

Abstract

Vision-Language-Action (VLA) models have demonstrated superior performance over conventional robot learning approaches, yet deploying them in real-world scenarios remains challenging due to data scarcity and limited generalization. To enable practical on-device execution, recent works have adopted efficiency-oriented techniques such as Mixture-of-Experts (MoE) and quantization to meet the latency and memory constraints of real-time systems. Despite these architectural advances, the core bottleneck remains: the scarcity of task-relevant training data leads to significant distribution shift when the model encounters novel environments or tasks at deployment time. Existing remedies such as data augmentation provide only marginal improvements, as synthetically diversified data still fails to capture the specific characteristics of the target setting. To address this problem, we propose DASH (Demonstration-Augmented Skill Injection for Human-guided VLA), a framework that injects task-specific skills into a pretrained VLA through a small number of human demonstrations. By formulating skill acquisition as few-shot learning conditioned on the target environment, DASH bridges the distribution gap between large-scale pretraining data and the downstream deployment setting without requiring costly data collection or full model retraining. Experimental results show that DASH enables robust policy execution in previously unseen environments and tasks with minimal human effort.

1 Introduction

References

A Appendix

You may include other additional sections here.