Unveiling the Feature Geometry of Robot Foundation Models via Nyström NCut

Ningze Zhong, Kyle Zhang, Ily Rafaeli

University of Pennsylvania

{zhong666, kyle100, ilyr}@seas.upenn.edu

Abstract

Vision-Language-Action (VLA) models have emerged as powerful robot foundation models, yet their internal representations remain largely opaque. We apply Nyström Normalized Cuts to visualize and analyze feature geometry across VLA hidden layers, revealing how these models process multimodal inputs during manipulation tasks. Our analysis uncovers strong object-centric semantic clustering and sensitivity to failure indicators, but also exposes critical limitations in spatial reasoning and language grounding. Through controlled experiments on OpenVLA, we demonstrate that visual features dominate over language by a factor of $10\times$, and that spatial information fails to generalize despite perfect training-set memorization. These findings provide actionable insights for debugging and improving robot foundation models.