Mental Rotation or Pattern Matching? Representational Structure and Angle-Dependent Behavior in Vision–Language Models

Junzhe Ren
University of Glasgow, Glasgow, Scotland, UK

Jiayuan Zhang
University of Glasgow, Glasgow, Scotland, UK

Abstract

Vision-language models (VLMs) often perform well on spatial reasoning tasks, but it remains unclear whether their performance is supported by internal representations that track rotation angle. We study this question in a controlled same or mirror 3D object task using LLaVA-1.5-7B as an intervenable computational system. First, the model's decision margin varies systematically with rotation angle, mainly for rotated rather than mirrored stimuli. Second, we identify an angle-related direction in intermediate hidden representations whose activity covaries with behavior. Third, targeted projection ablations produce progressive flattening of the angle-margin relationship as intervention strength increases. Together, these findings provide mechanistic evidence that angle-sensitive internal geometry contributes to the model's behavior on this task. The results support a cautious interpretation of structured, intervenable spatial processing rather than a strong claim of human-like mental rotation.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*