

Probing Perceptual Constancy in Large Vision Language Models

Haoran Sun

Johns Hopkins University, Baltimore, Maryland, United States

Suyang Yu

University of Washington, Seattle, Washington, United States

Yijiang Li

Electrical and computer engineering, La Jolla, California, United States

Qingying Gao

Computer Science, Baltimore, Maryland, United States

Haiyun Lyu

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

Hokin Deng

Harvard University, Boston, Massachusetts, United States

Dezhi Luo

University of Michigan, Ann Arbor, Michigan, United States

Abstract

Perceptual constancy is the ability to maintain stable perceptions of objects despite changes in sensory input, such as variations in distance, angle, or lighting. This ability is crucial for recognizing visual information in a dynamic world, making it essential for Vision-Language Models (VLMs). However, whether VLMs are currently and theoretically capable of mastering this ability remains underexplored. In this study, we evaluate 33 VLMs using 253 experiments across three domains: color, size, and shape constancy. The experiments include single-image and video adaptations of classic cognitive tasks, along with novel tasks in in-the-wild conditions, to evaluate the models' recognition of object properties under varying conditions. We find significant variability in VLM performance, with models excelling in shape constancy but struggling with color and size constancy. These results suggest that while VLMs are proficient in object recognition, they do not fully replicate the robustness of human perceptual constancy.