

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

How Far Does Vision Go? Interpretable Visual Statistics from Word-Linked Web Image Distributions Predict Multisensory Norms for Chinese Words

Permalink

<https://escholarship.org/uc/item/5hp18350>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Author

WU, YUFENG

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

How Far Does Vision Go? Interpretable Visual Statistics from Word-Linked Web Image Distributions Predict Multisensory Norms for Chinese Words

Yufeng WU

City University of Hong Kong

Abstract

Recent work reports that large language models, including those with visual training, fail to recover the sensorimotor structure of human concepts. We argue this reflects how visual information has been operationalized, as opaque embeddings tied to single training instances, rather than a limit of vision itself. We test this by representing 593 Chinese disyllabic words through interpretable low-level visual statistics, pixel-level and mid-level features aggregated over word-linked, CLIP-filtered web images, and evaluating their predictive power for human sensorimotor norms across six modalities beyond lexical controls. We find a graded profile of cross-modal decodability. Visual norms are predicted best, followed by gustatory, olfactory, tactile, and auditory, while interoception sits at the predicted lower bound. The graded pattern, with substantial gains for non-visual modalities, indicates that prior negative results reflect a choice of representation rather than a failure of low-level visual statistics per se.