NEURAL INFORMATION PROCESSING SYSTEMS

Spectral Stratification of Semantic Abstraction in Vision-Language Models

NeurIPS 2026

Jeonghwan Cheon[1], Marin Vogelsang[2], Lukas Vogelsang[2], Pawan Sinha[2]

[1] Independent Researcher,
[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

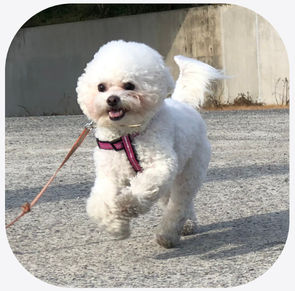
Introduction

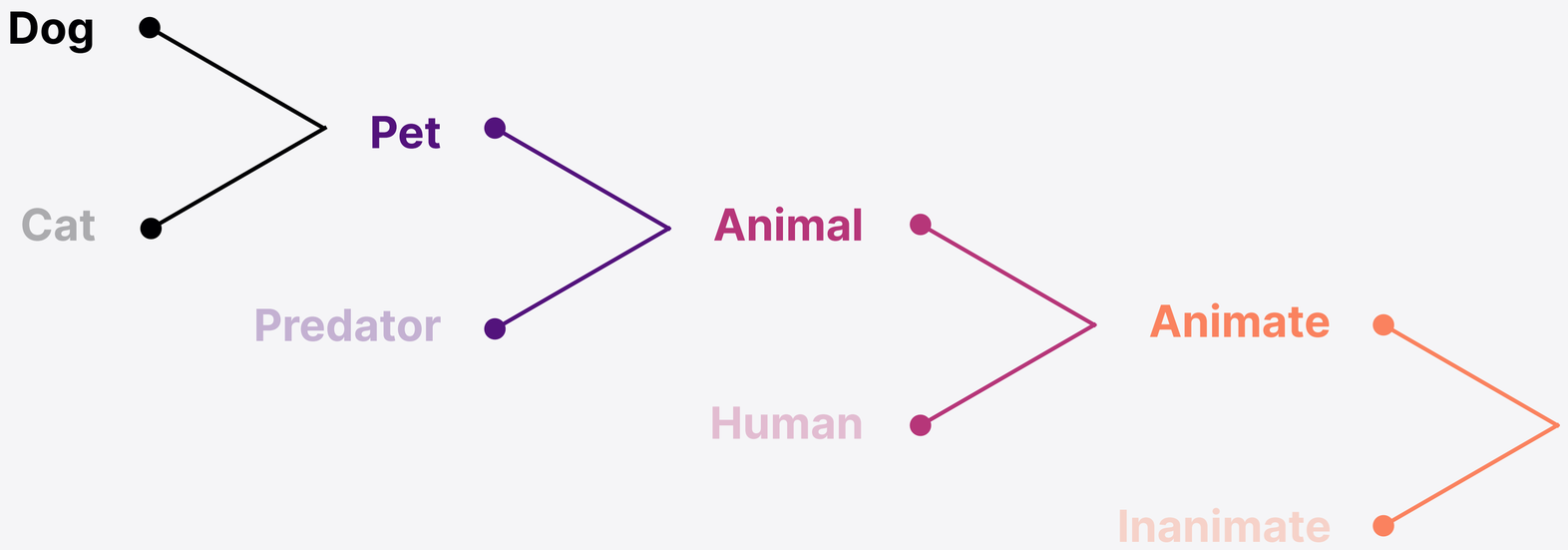
Semantic abstraction

01

Concepts are commonly described as organized in taxonomic hierarchies in which abstract categories subsume more specific ones.

Collins 1969, Smith 1981, Murphy 2002



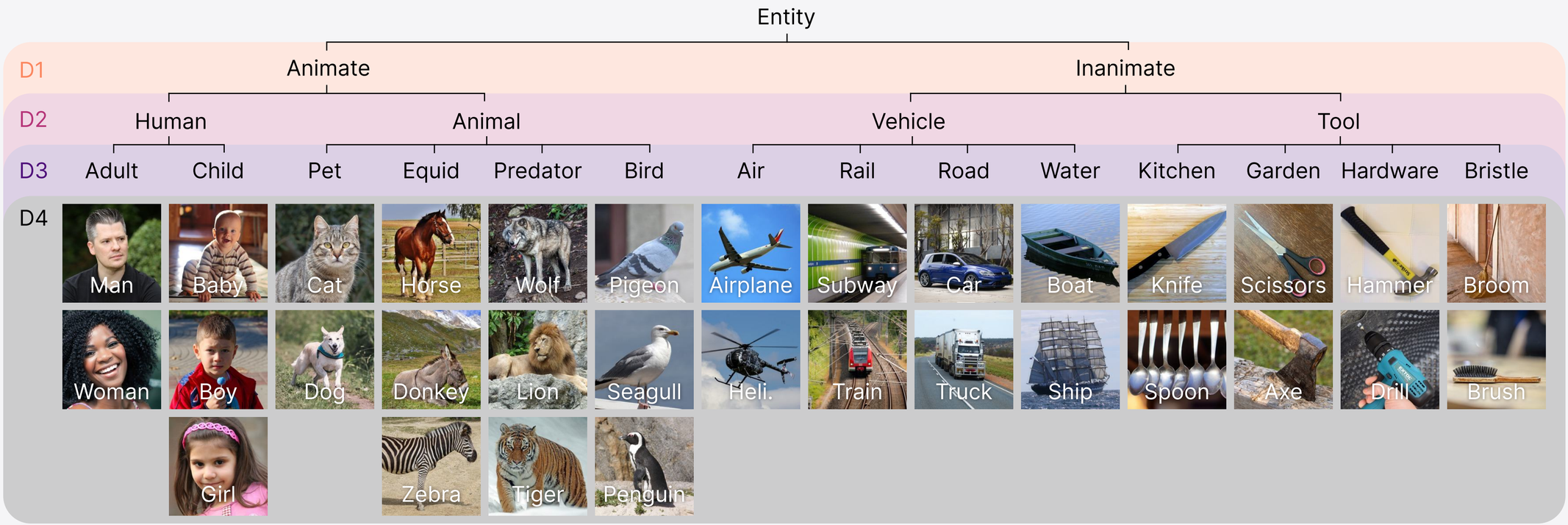


Setup

Semantic taxonomy

02

We measure abstraction with a four-level (D1–D4) probing taxonomy,
ranging from the coarse animate–inanimate distinction to individual object concepts.

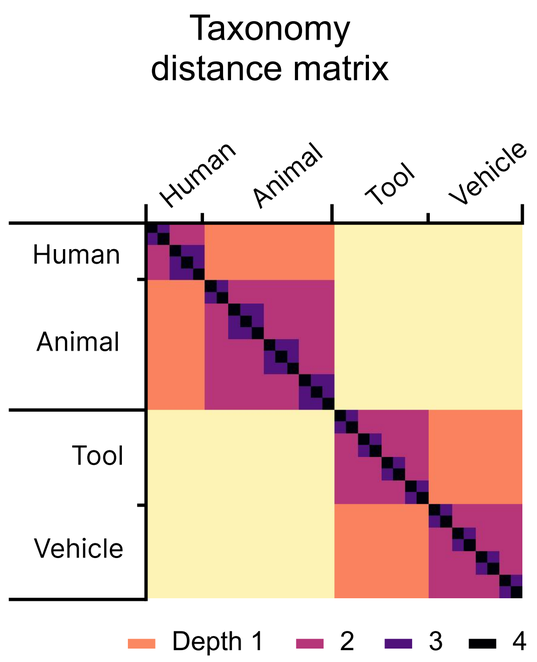


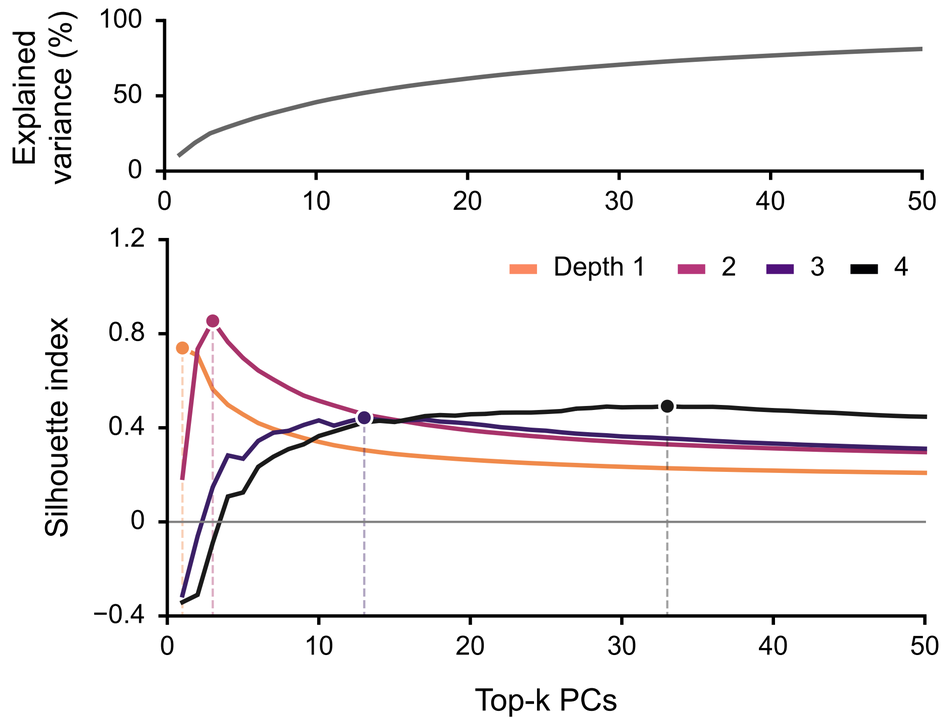
Results

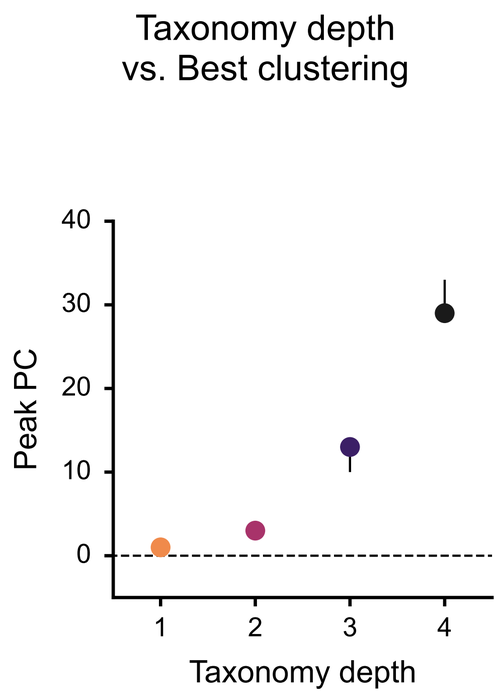
Spectrally stratified semantic abstraction

03

Leading principal components encode broad conceptual structure, while later components carry progressively finer-grained distinctions.





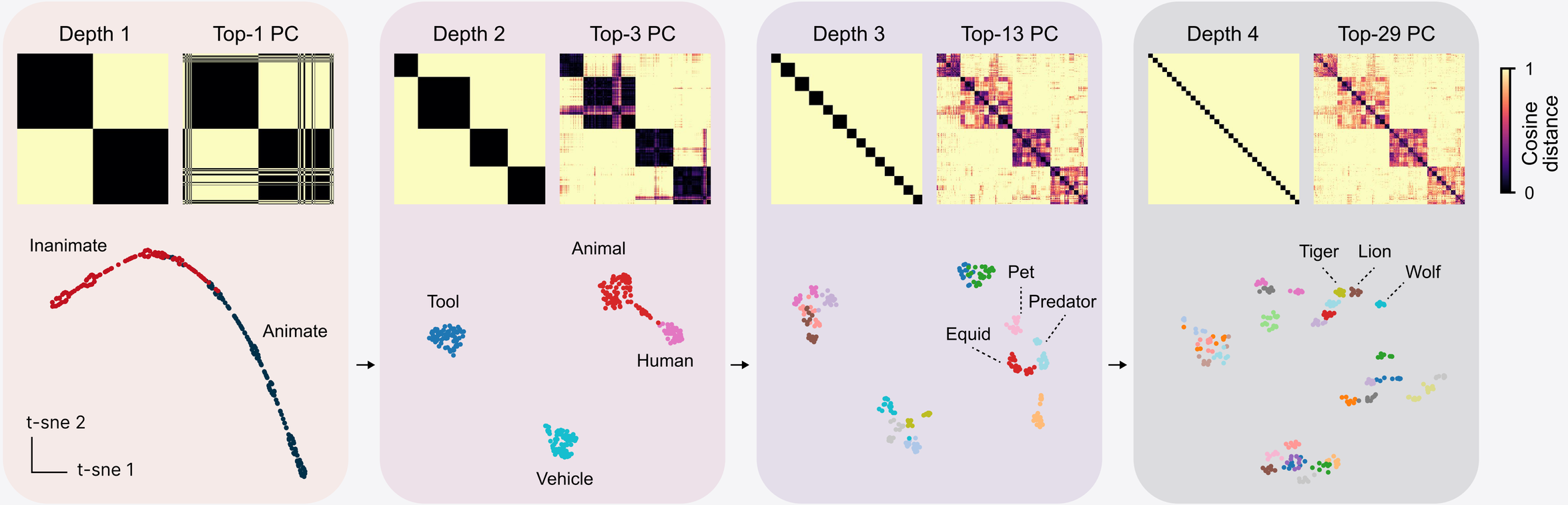


Results

Hierarchical structure emerges across spectral ranks

04

Increasing the number of retained principal components progressively reveals finer-grained category structure.



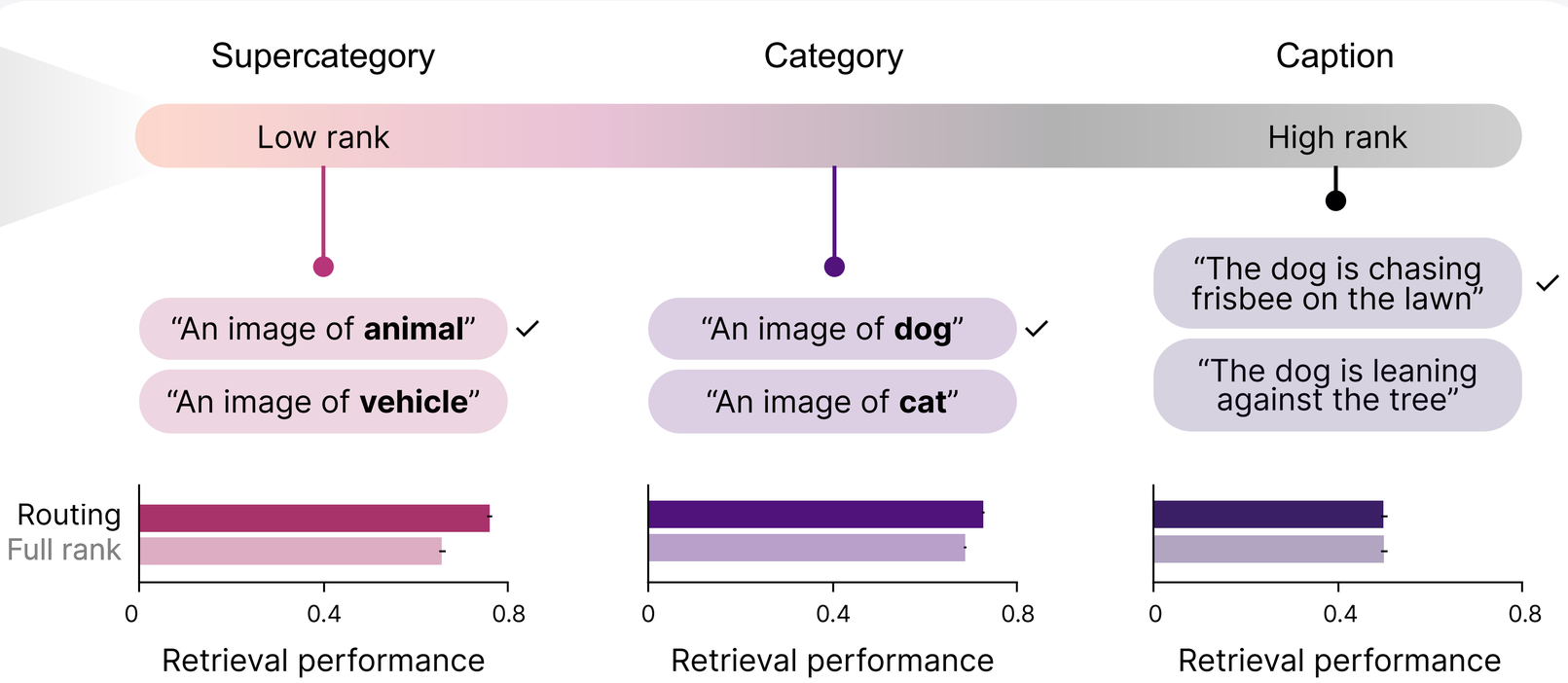
Results

Rank-based routing changes retrieval behavior

05

Selecting spectral subspaces matched to the target abstraction level improves retrieval performance without retraining.



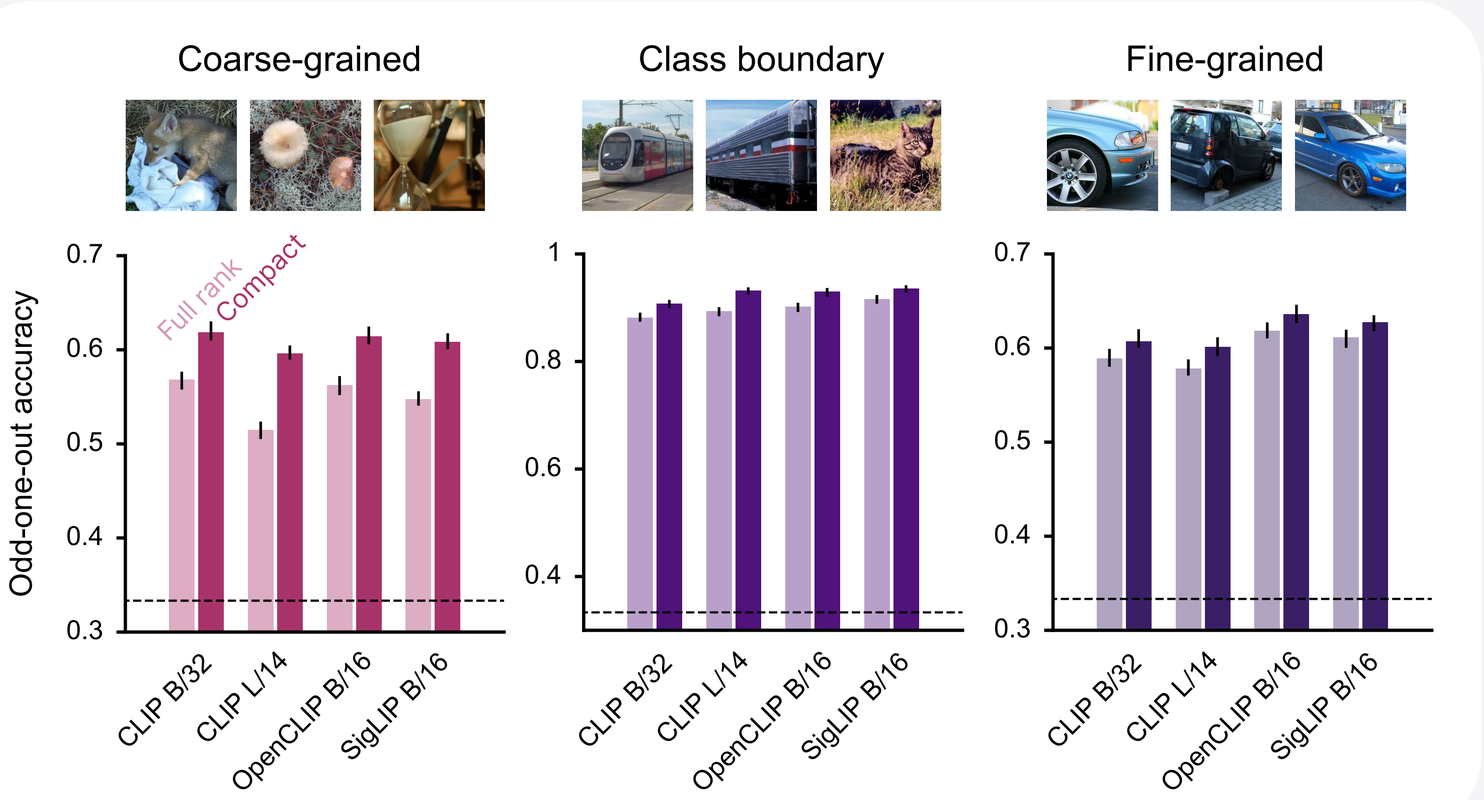


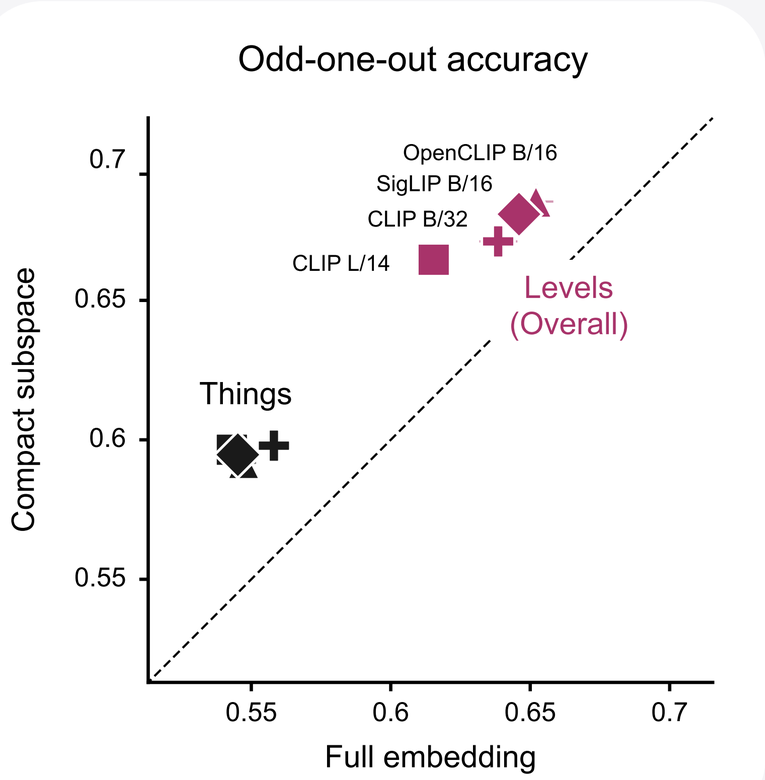
Results

Compact subspaces better match human judgments

06

Compact spectral subspaces that retain only leading components improve agreement with human odd-one-out judgments.





Takeaway

07

Semantic abstraction in VLMs is hierarchically organized
along the principal component spectrum,
with compact spectral subspaces improving human alignment.

