LCA-on-the-Line: Benchmarking Out of Distribution Generalization with Class Taxonomies

Jia Shi[1], Gautam Gare[1], Jinjin Tian[1], Siqi Chai[1], Zhiqiu Lin[1], Arun Vasudevan[1], Di Feng[23], Francesco Ferroni[24], Shu Kong[56]

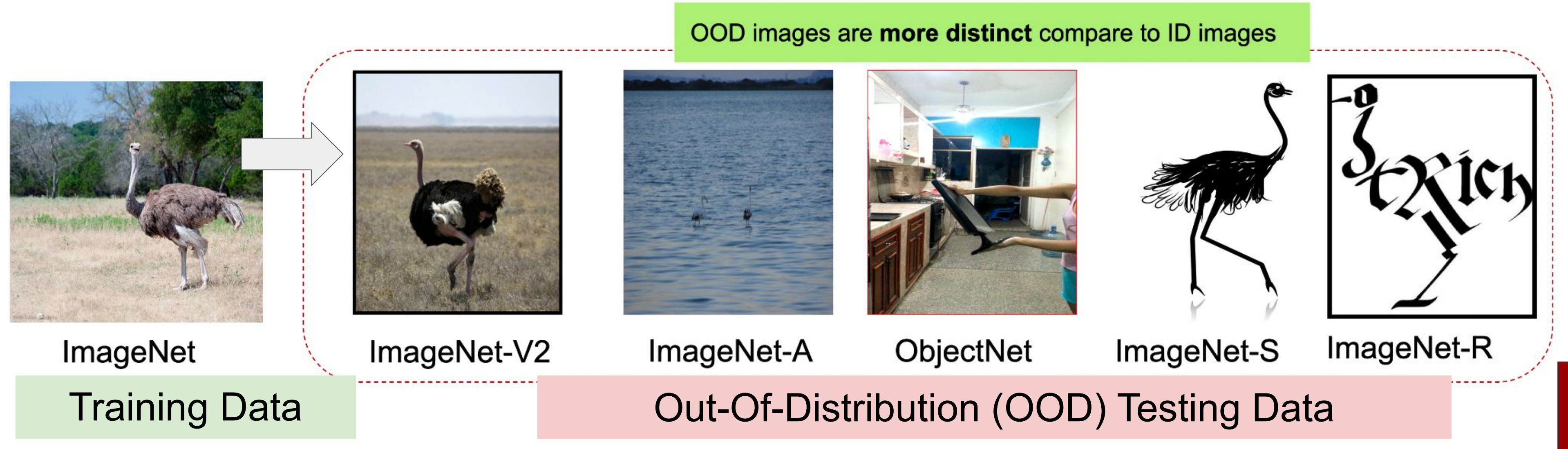
Robotics Institute, Carnegie Mellon University[1], Argo AI[2], Apple[3], Nvidia[4],Texas A&M University[5], University of Macau[6]

ICML
International Conference
On Machine Learning

Introduction

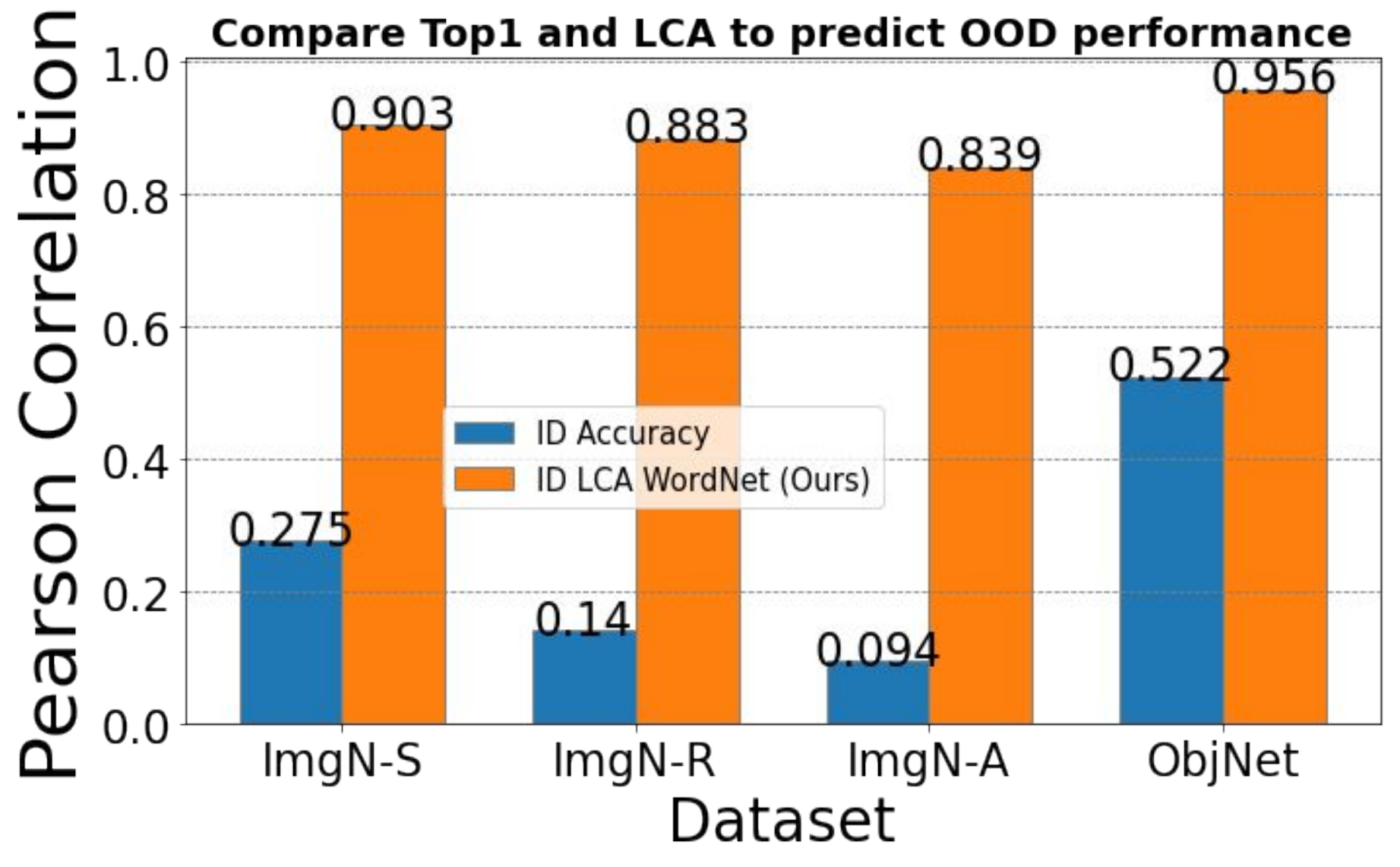
- LCA distance, a metric measuring prediction performance w.r.t an ontology/hierarchy on ID (in-distribution) data, can robustly predict the model's OOD (out of distribution) performance.
- It unifies Vision Models (VMs) and Vision-Language Models (VLMs) across different modalities and training data sources in terms of measuring model generalization, outperforming "accuracy- on-the-line".

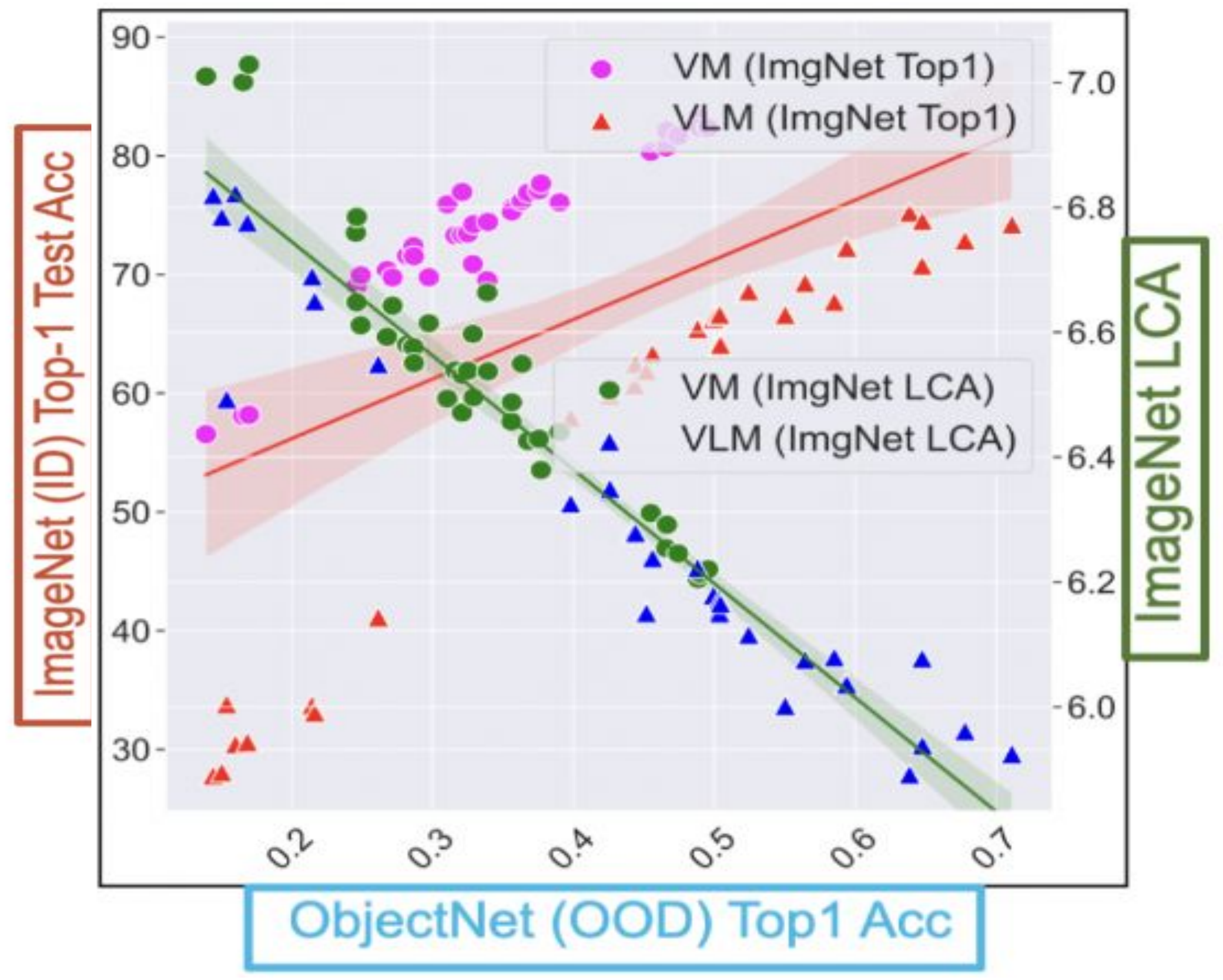
LCA-on-the-Line evaluates on severe visual shift datasets


OOD images are **more distinct** compare to ID images
ImageNet
ImageNet-V2
ImageNet-A
ObjectNet
ImageNet-S
ImageNet-R
Training Data
Out-Of-Distribution (OOD) Testing Data

LCA is an indicator to OOD accuracy

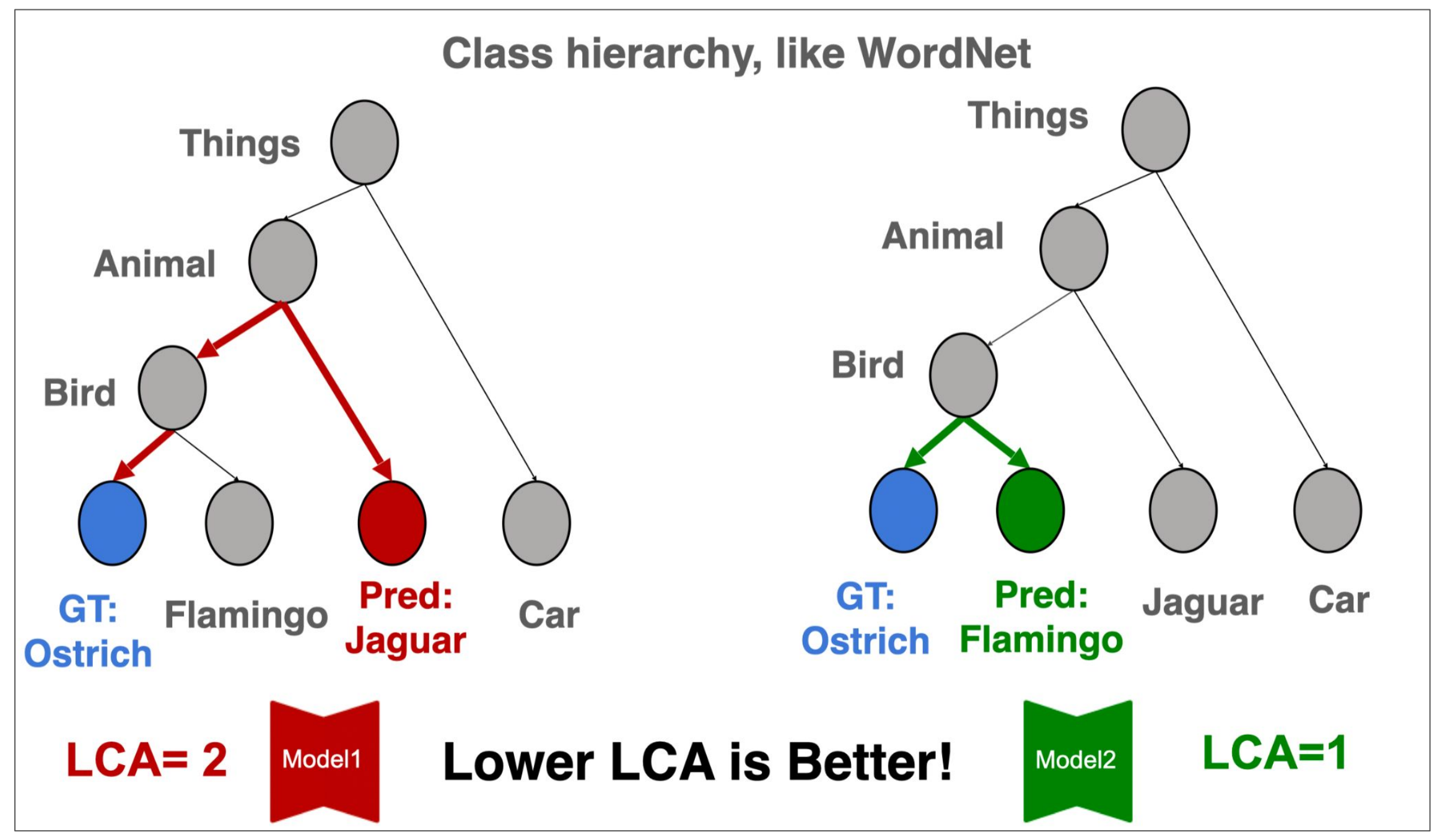
- LCA distance achieves strong linear correlation consistently on multiple ImageNet-OOD datasets.
- For VMs + VLMs, ID accuracy is *not* on the line, while ID LCA is on the line.





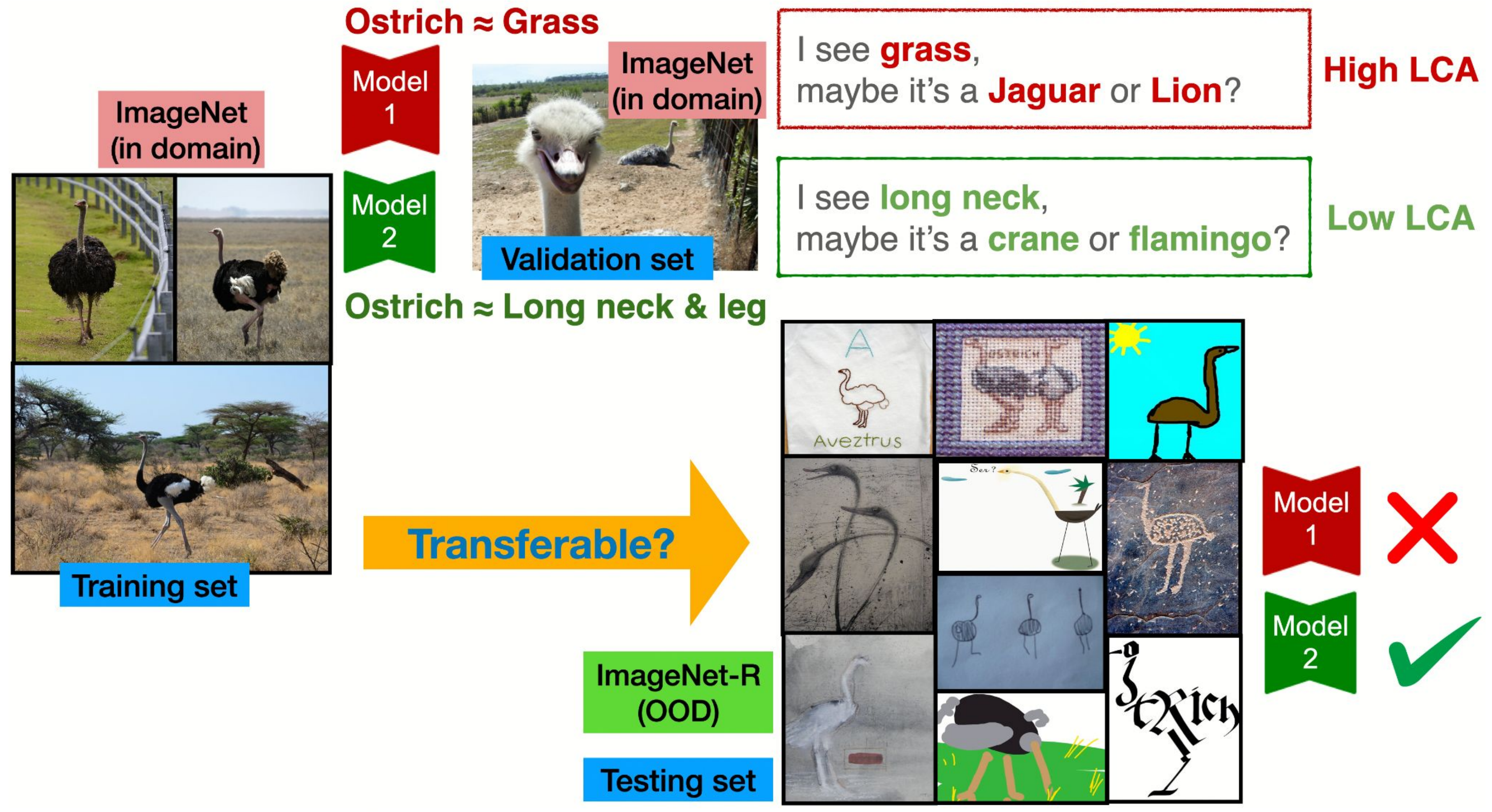
How to calculate LCA distance?

- LCA distance measures class-pairwise distance within a given **ontology/hierarchy**.
- Here, we measure the distance between ground-truth and prediction in a hierarchy, like **WordNet.**



Why does LCA distance work?

- LCA distance reflects model's learned predictive feature, measuring its transferability to OOD.
- Low LCA distance reflects small spurious correlation.

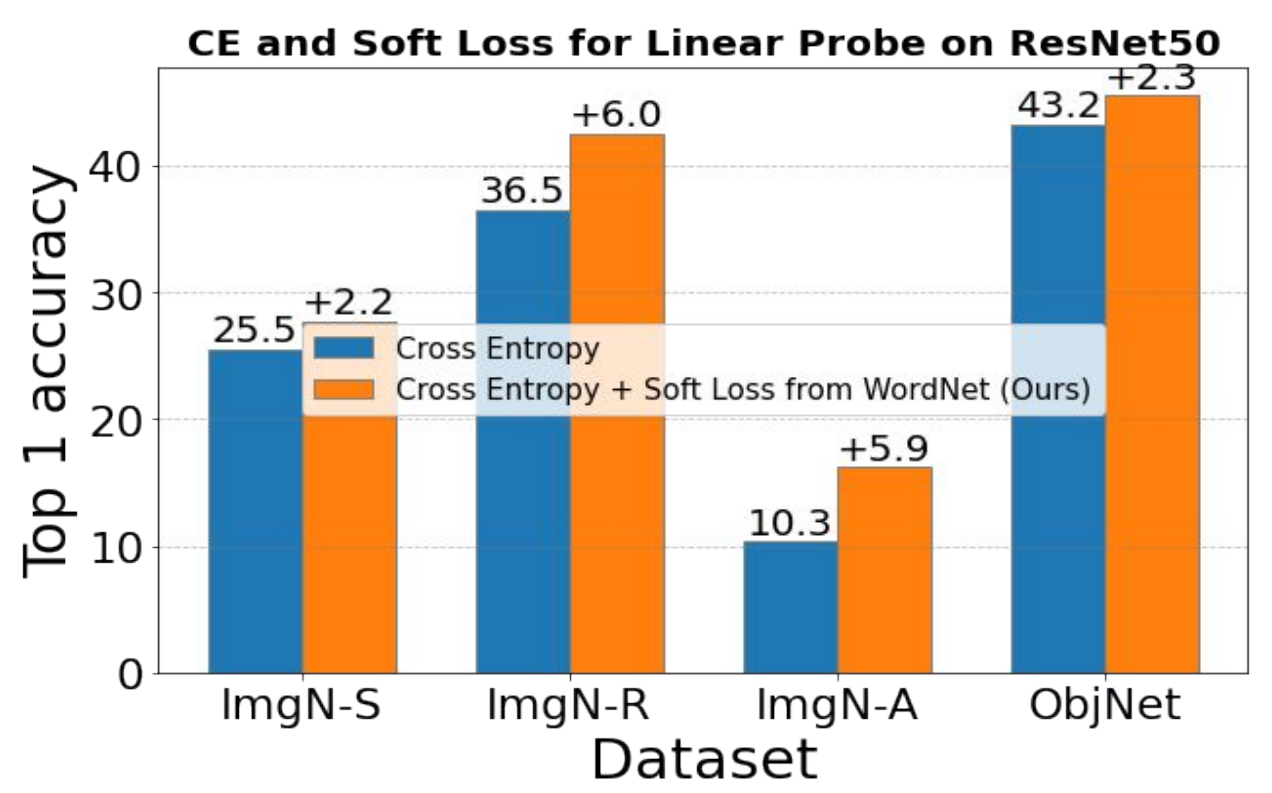


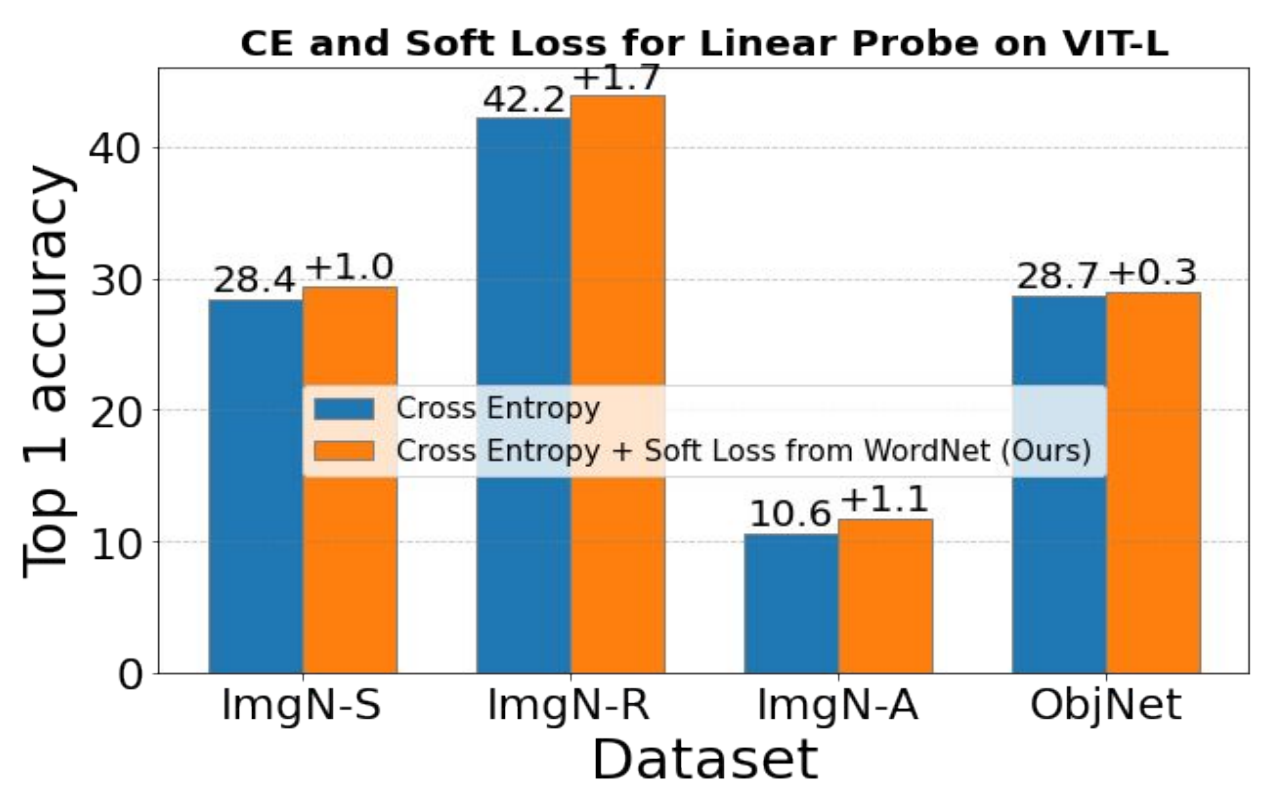
I see **grass**, maybe it's a **Jaguar** or **Lion**?

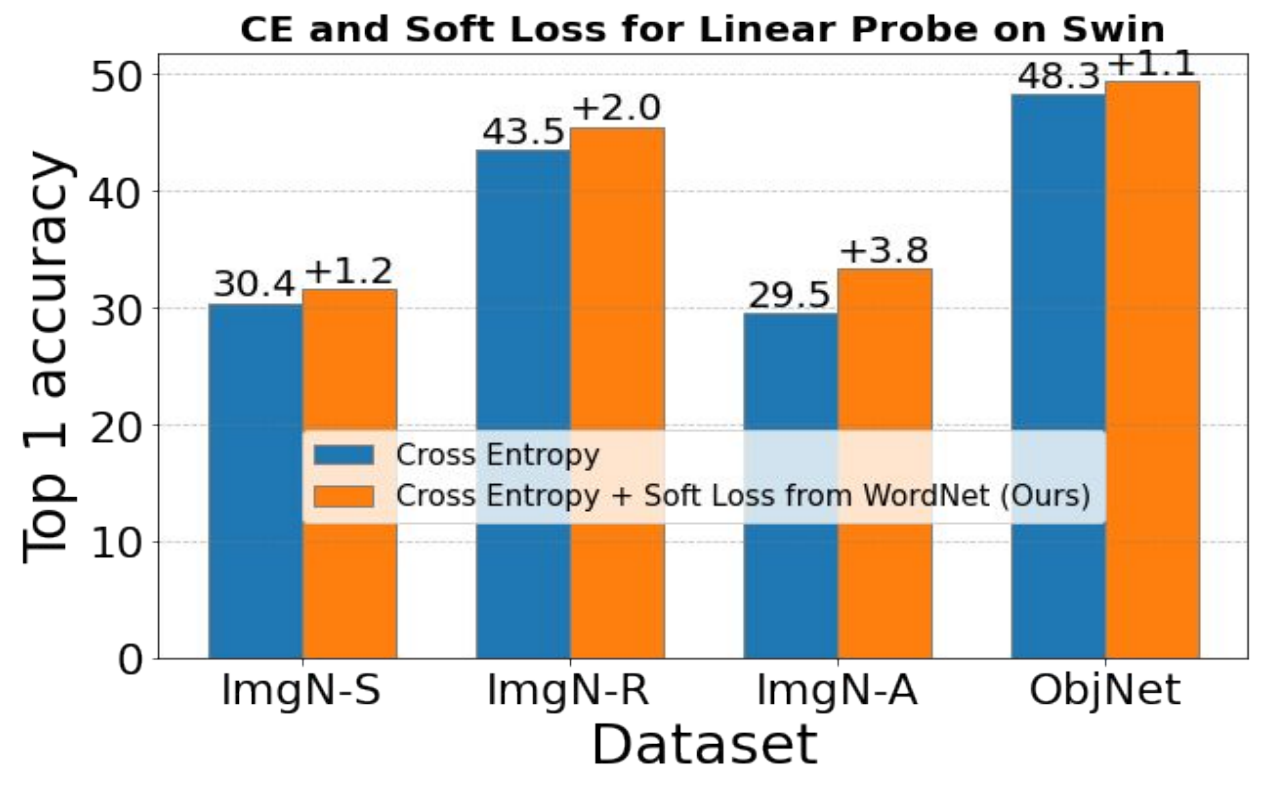
I see **long neck**, maybe it's a **crane** or **flamingo**?

LCA as soft labels improves OOD accuracy

- Hierarchy allows exploiting class-pairwise distances in model training!
- We train with cross-entropy + soft labels loss.

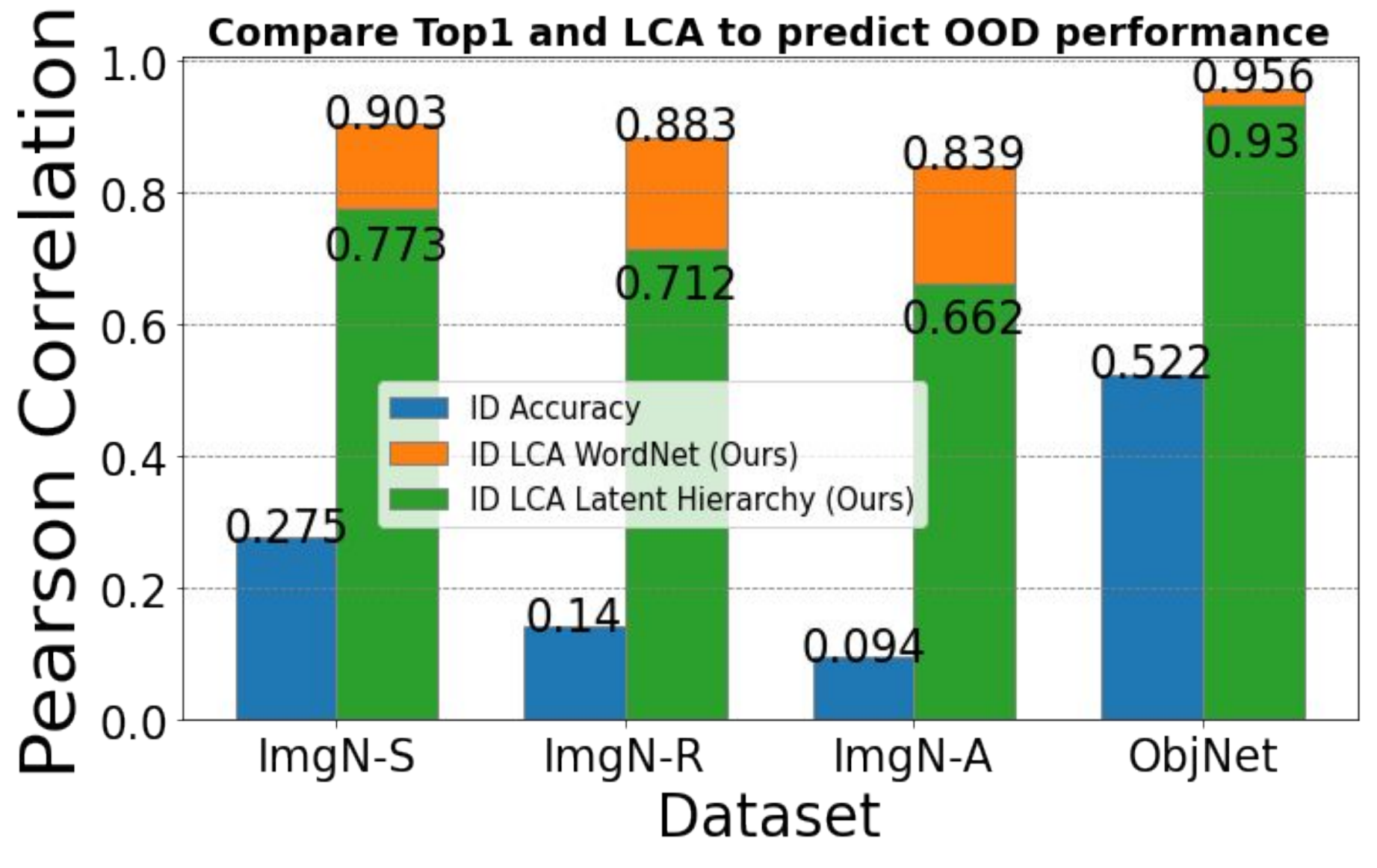


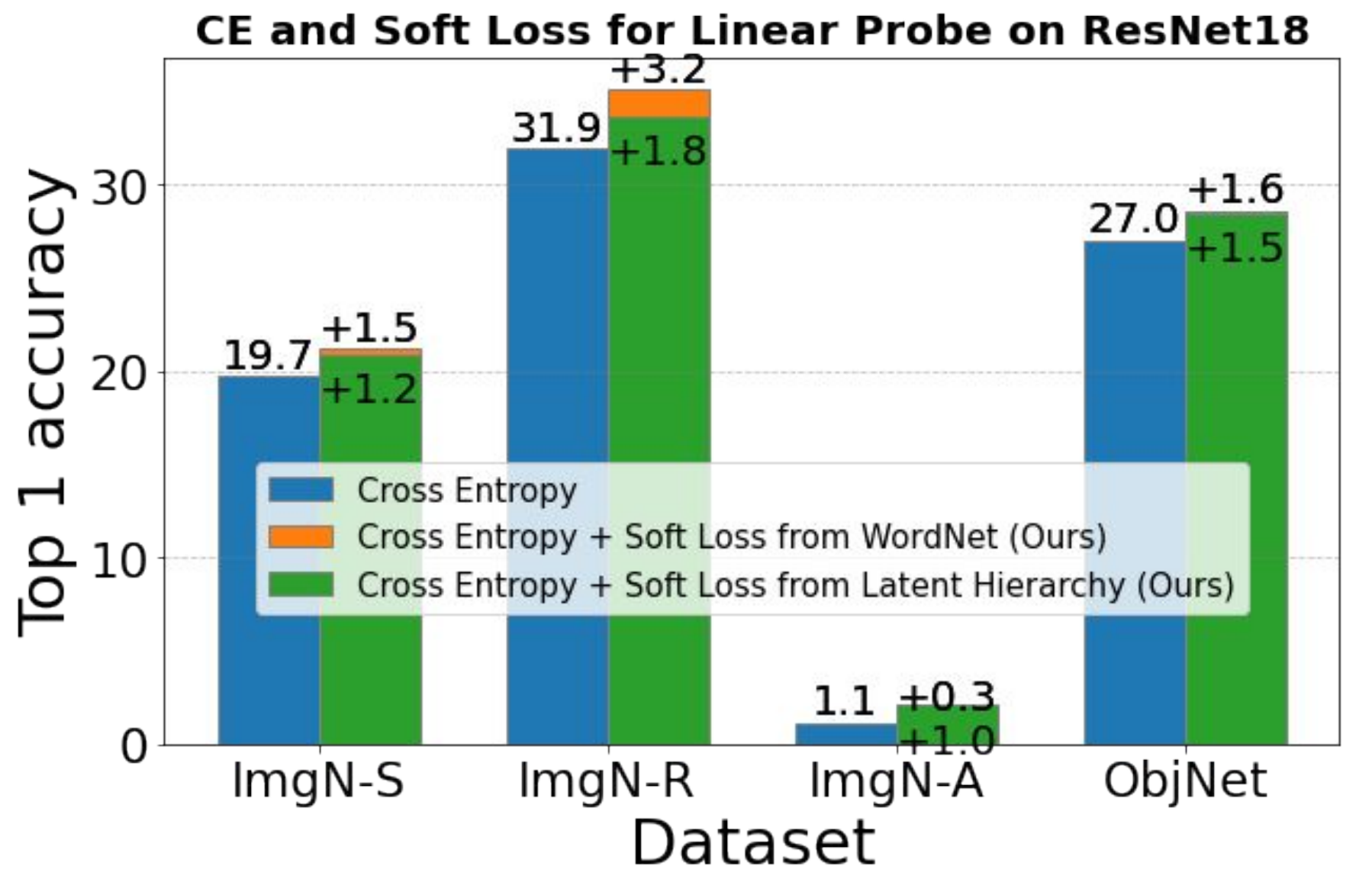




Constructing a latent hierarchy on any dataset

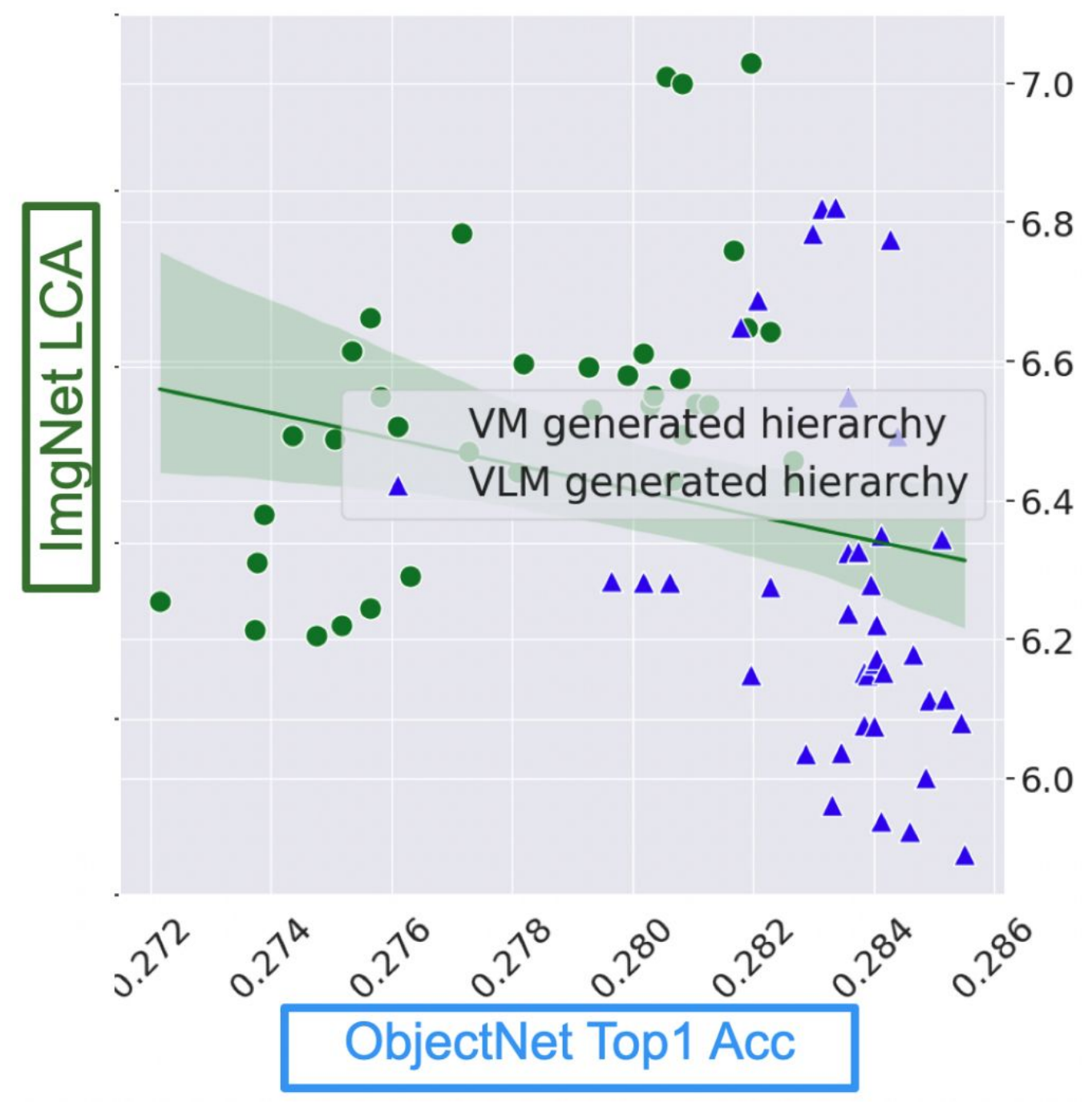
- We can construct a hierarchy by clustering per-class mean features using a foundation model (like CLIP).
- Using a latent hierarchy performs as well as WordNet.





New insights to VLM generalization

- Using soft labels from latent hierarchies generated by VLMs yields better OOD results than VMs.
- That said, VLMs have a better human-aligned feature distribution, i.e., their generated labels better align with human-world ontology (WordNet).


ImgNet LCA
VM generated hierarchy
VLM generated hierarchy
ObjectNet Top1 Acc