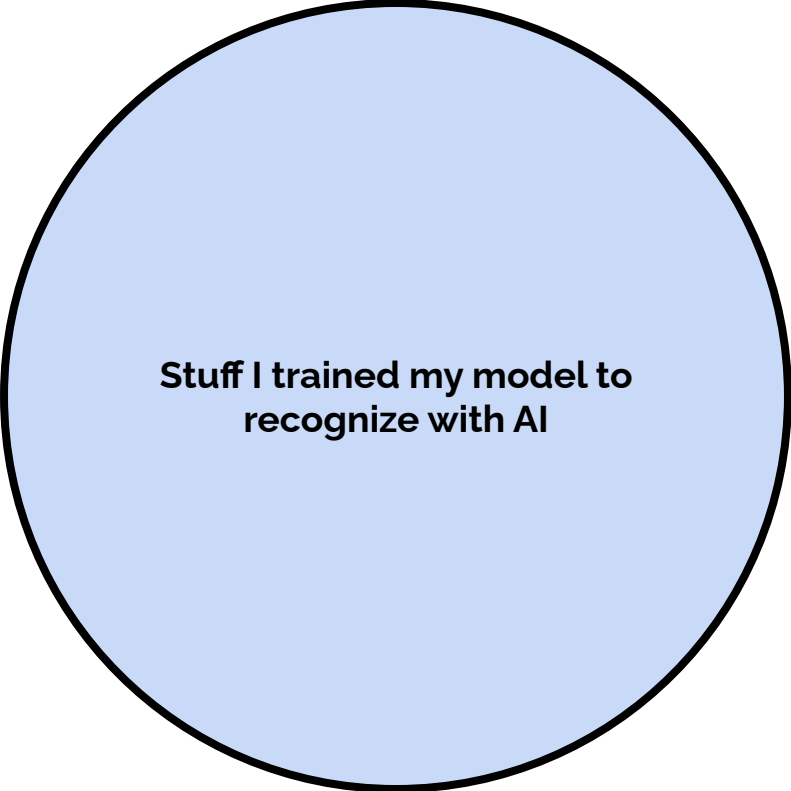


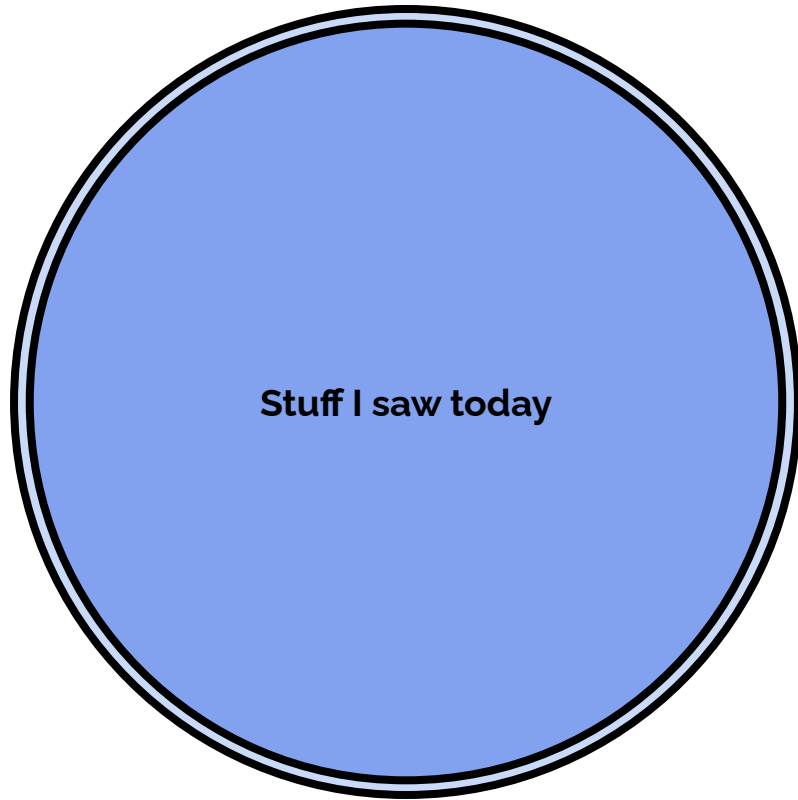
A satellite-style map of the Earth showing a color overlay. The colors range from dark blue (low values) to yellow and red (high values), with a concentration of high values in the North Atlantic and Arctic regions. The map is partially obscured by two white text boxes.

Lecture 4: Open-Set Learning

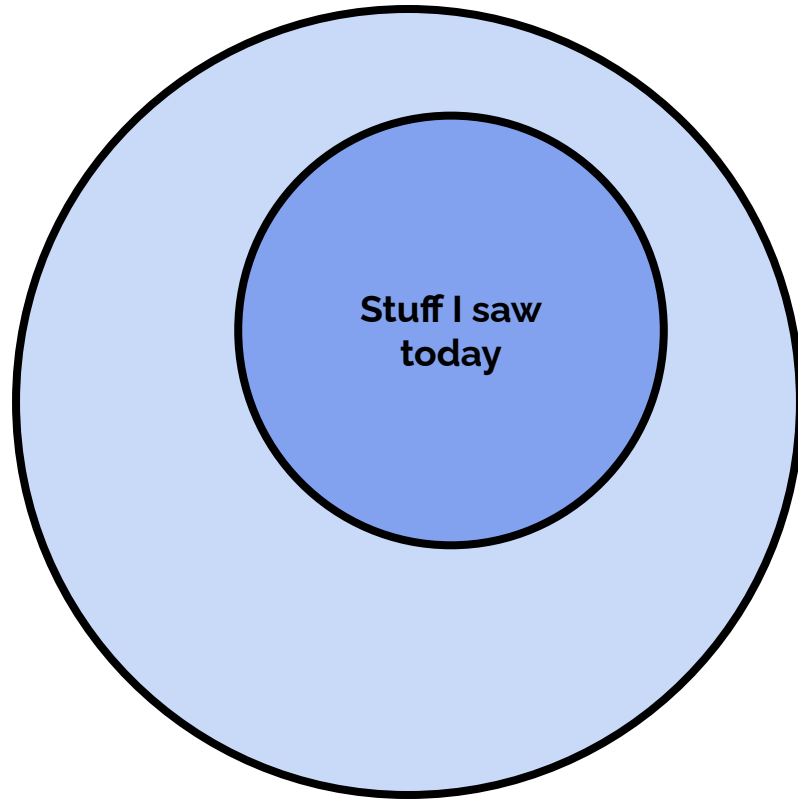
Sara Beery | 3/4/25



**Stuff I trained my model to
recognize with AI**



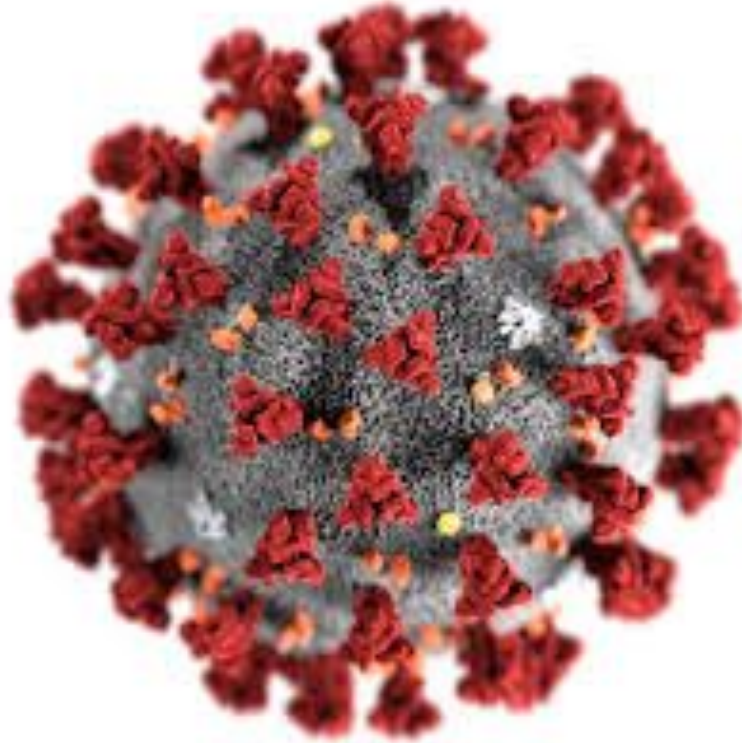
Stuff I saw today



This isn't the case in the real world

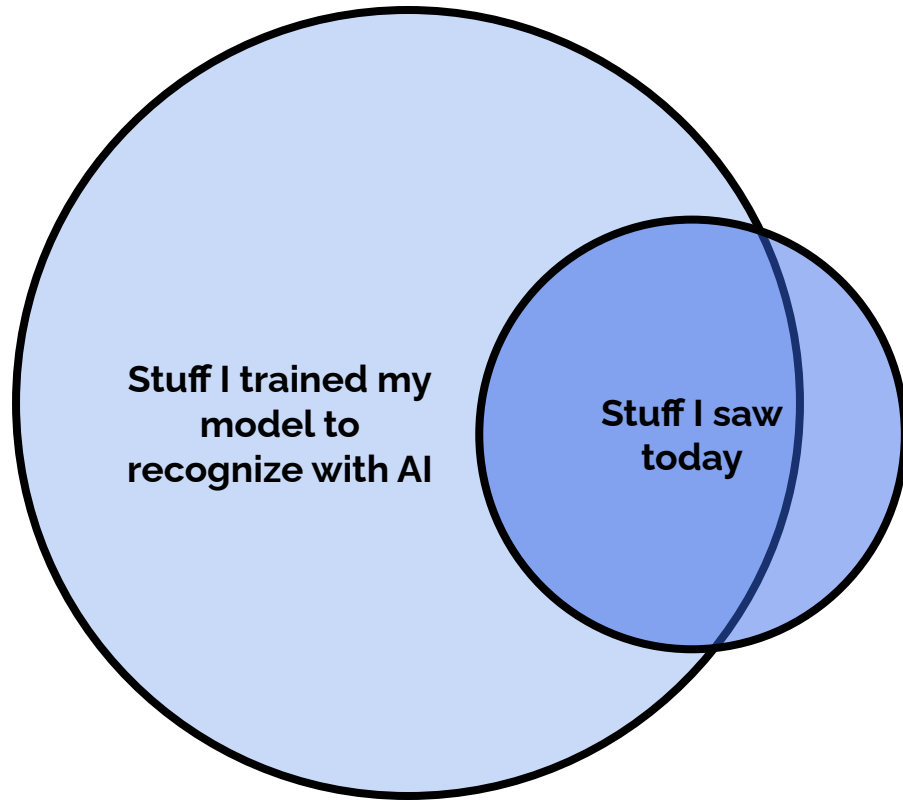


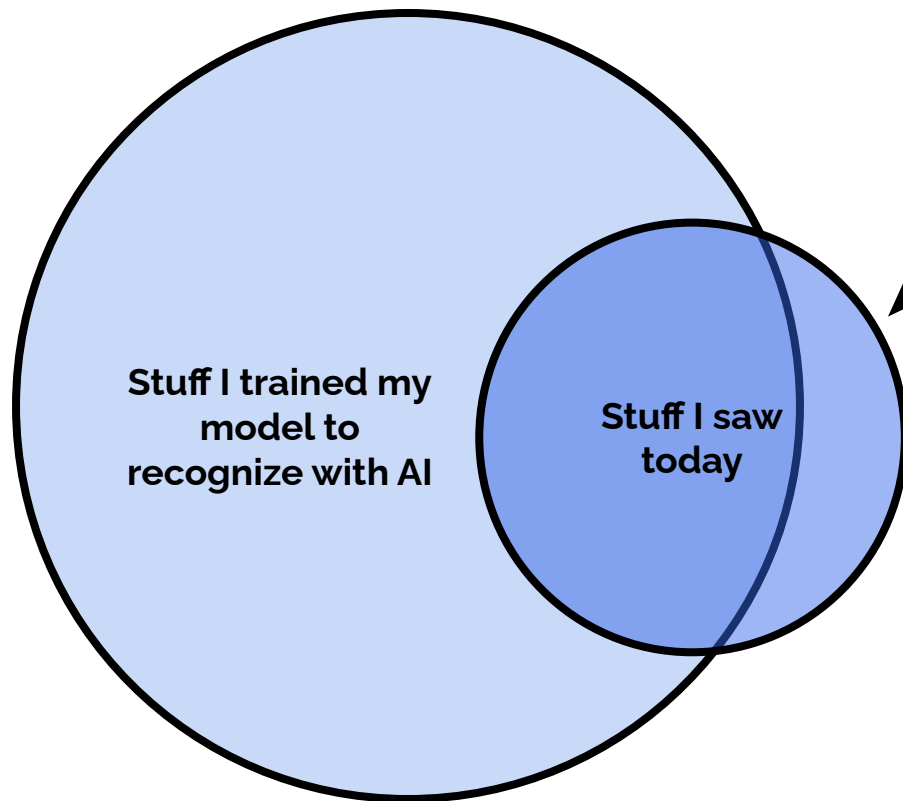
This isn't the case in the real world



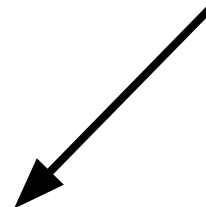
This isn't the case in the real world

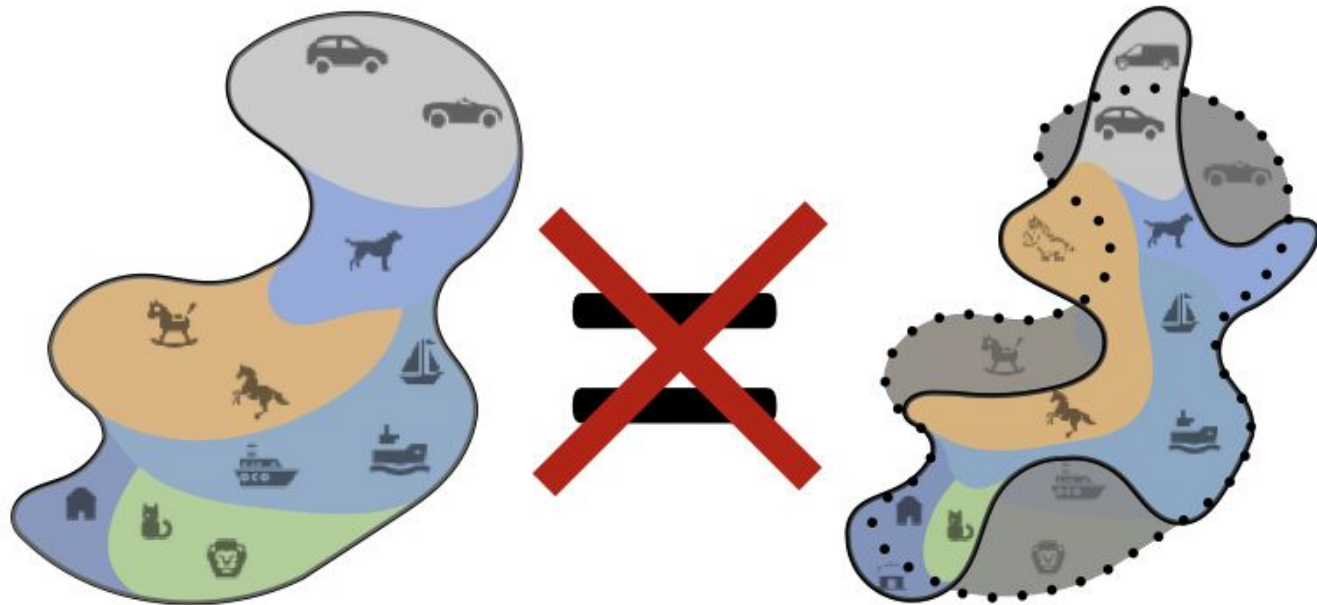






What's going to happen?





Training



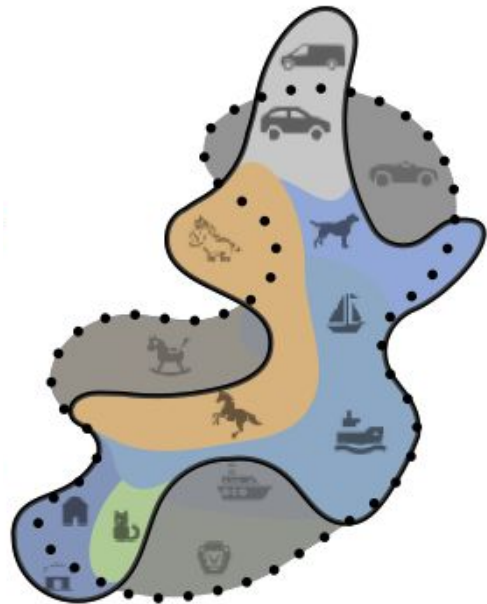
Inference



Open-set learning

What do you do when you aren't sure about the test distribution?

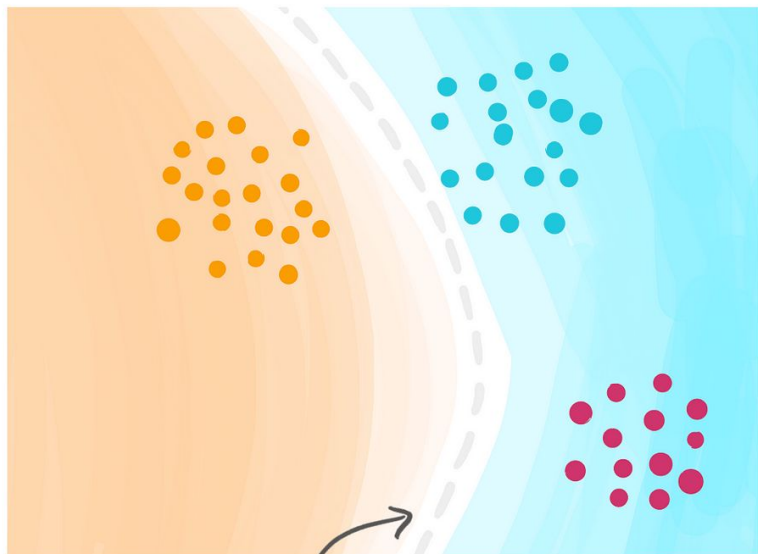
- Flag new things (OOD Detection)
- Build systems that can cluster new categories without supervision (Novel Category Discovery)
- Build systems that can recognize what you trained on AND cluster new stuff into categories (Generalized Category Discovery)



OOD detection

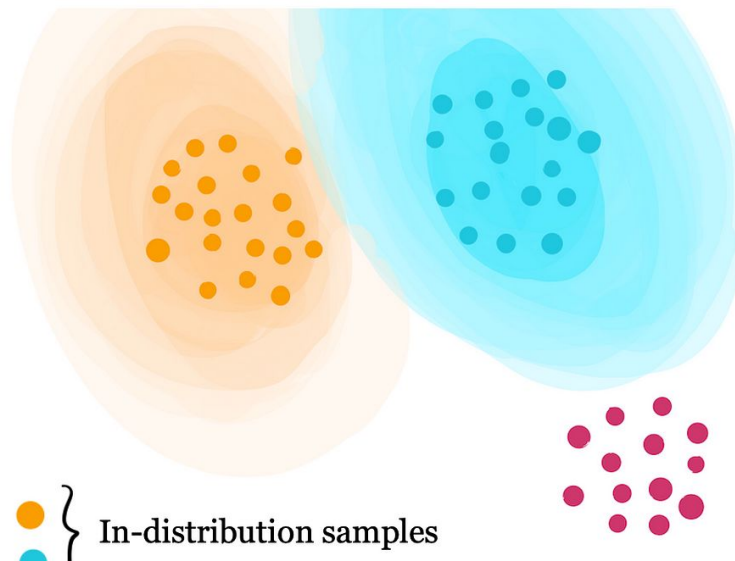
Design models that understand what they have and haven't seen




Discriminators



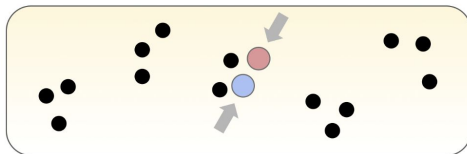
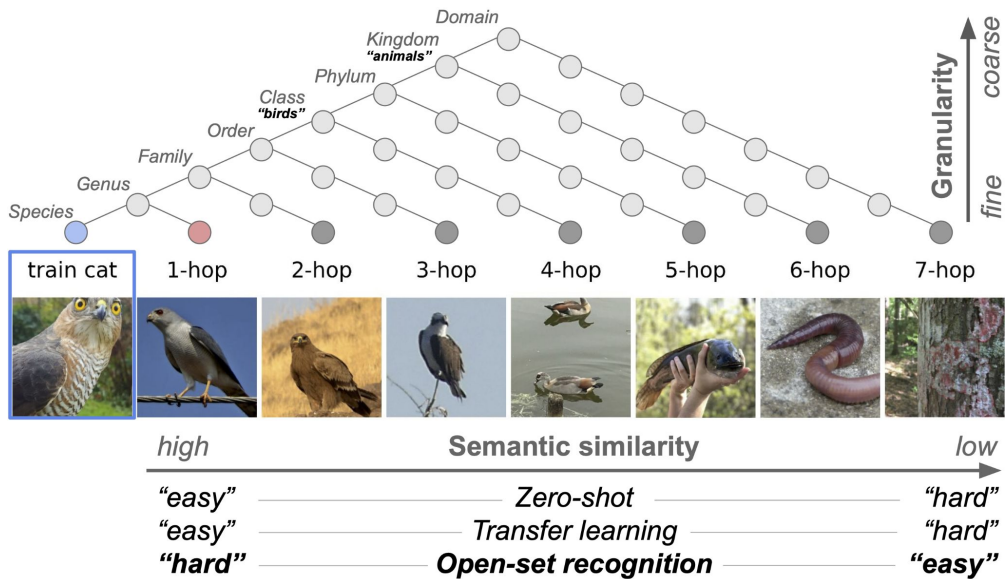
decision boundary

Density Estimators

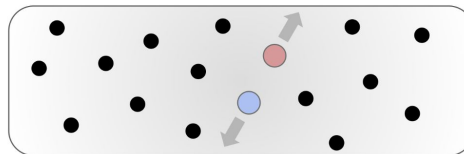


-   } In-distribution samples
-  } Out-of-distribution samples

How does this interact with fine-grained?

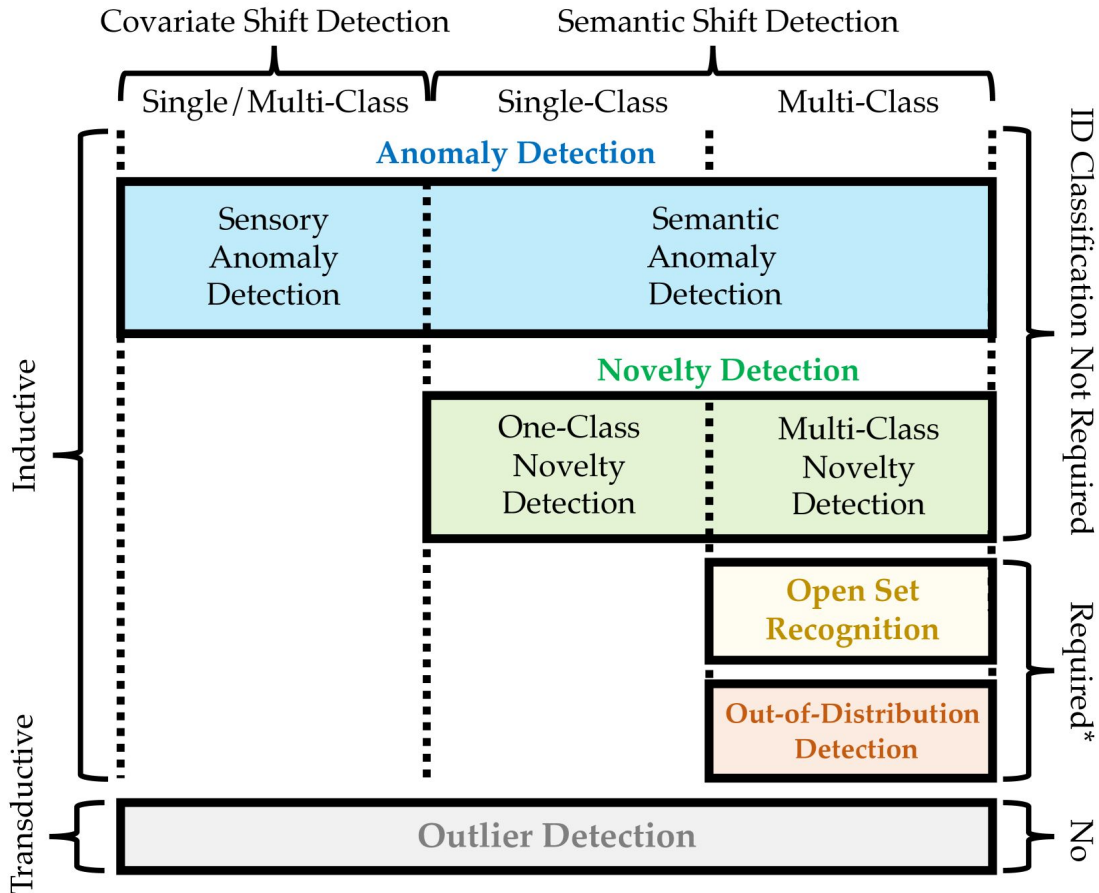


Hierarchy-supportive
representation space



Hierarchy-adversarial
representation space

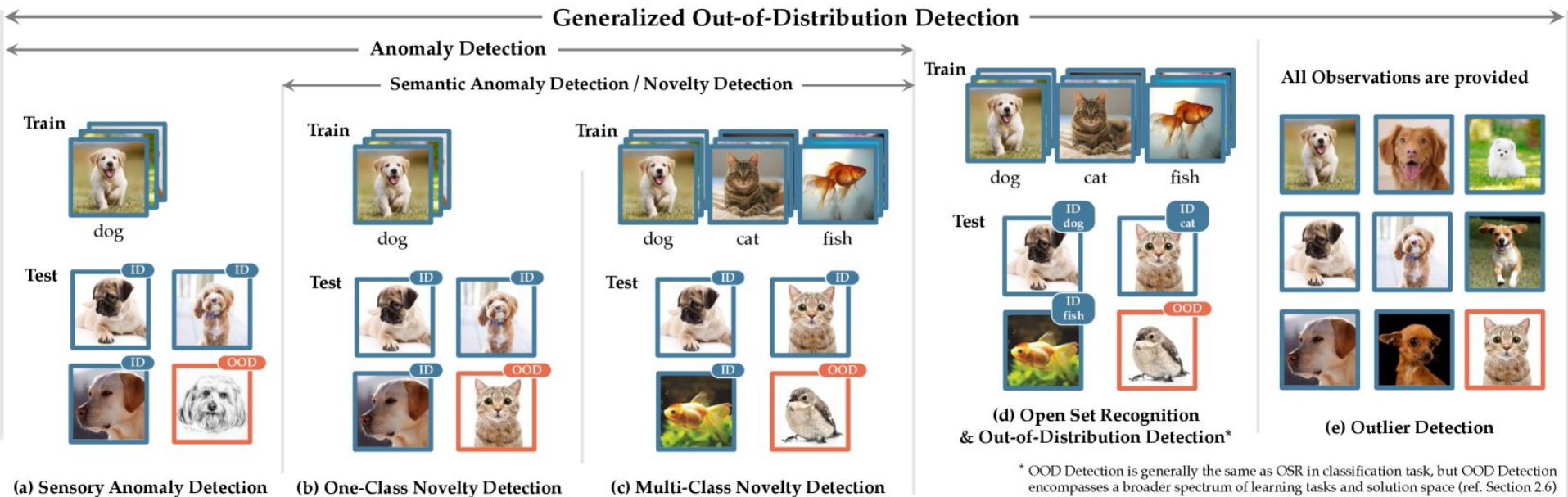
Is this different from anomaly detection?



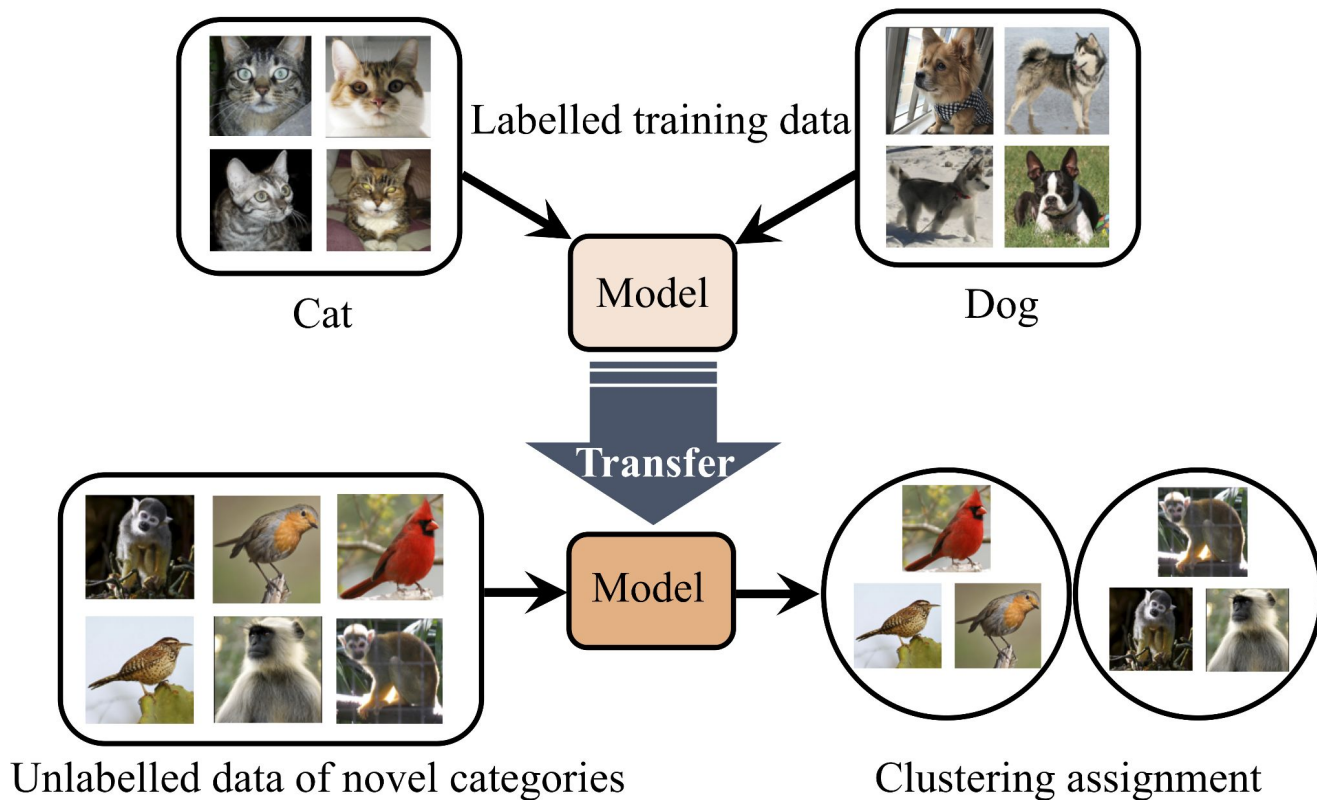
It's complicated...

*Exception: In OOD Detection, density-based methods do not require ID classification

The boundaries between OOD tasks can be difficult to parse



Beyond OOD: novel category discovery



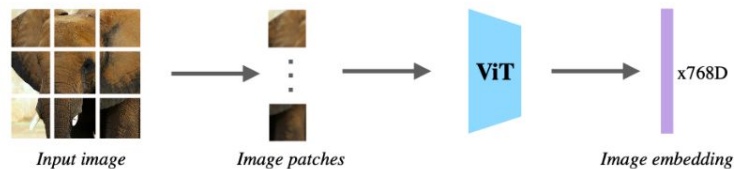
Generalized category discovery

Setting: Generalized Category Discovery

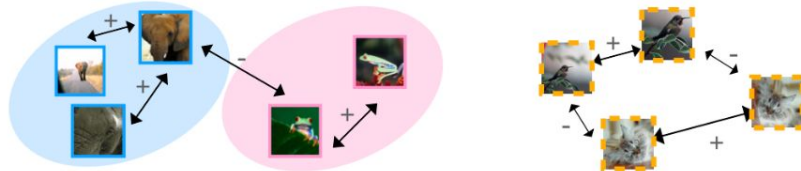


Method

(1) Feature extraction with vision transformer



(2) Supervised Contrastive (left) & Self-supervised Contrastive (right)



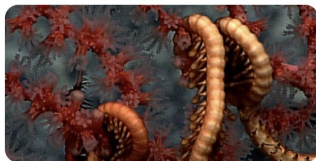
(3) Semi-supervised K-Means Clustering



Open-set challenges in ecology

FathomNet 2023

Shifting seas, shifting species: Out-of-sample detection in the deep ocean



<https://www.kaggle.com/competitions/fathomnet-out-of-sample-detection/overview>

