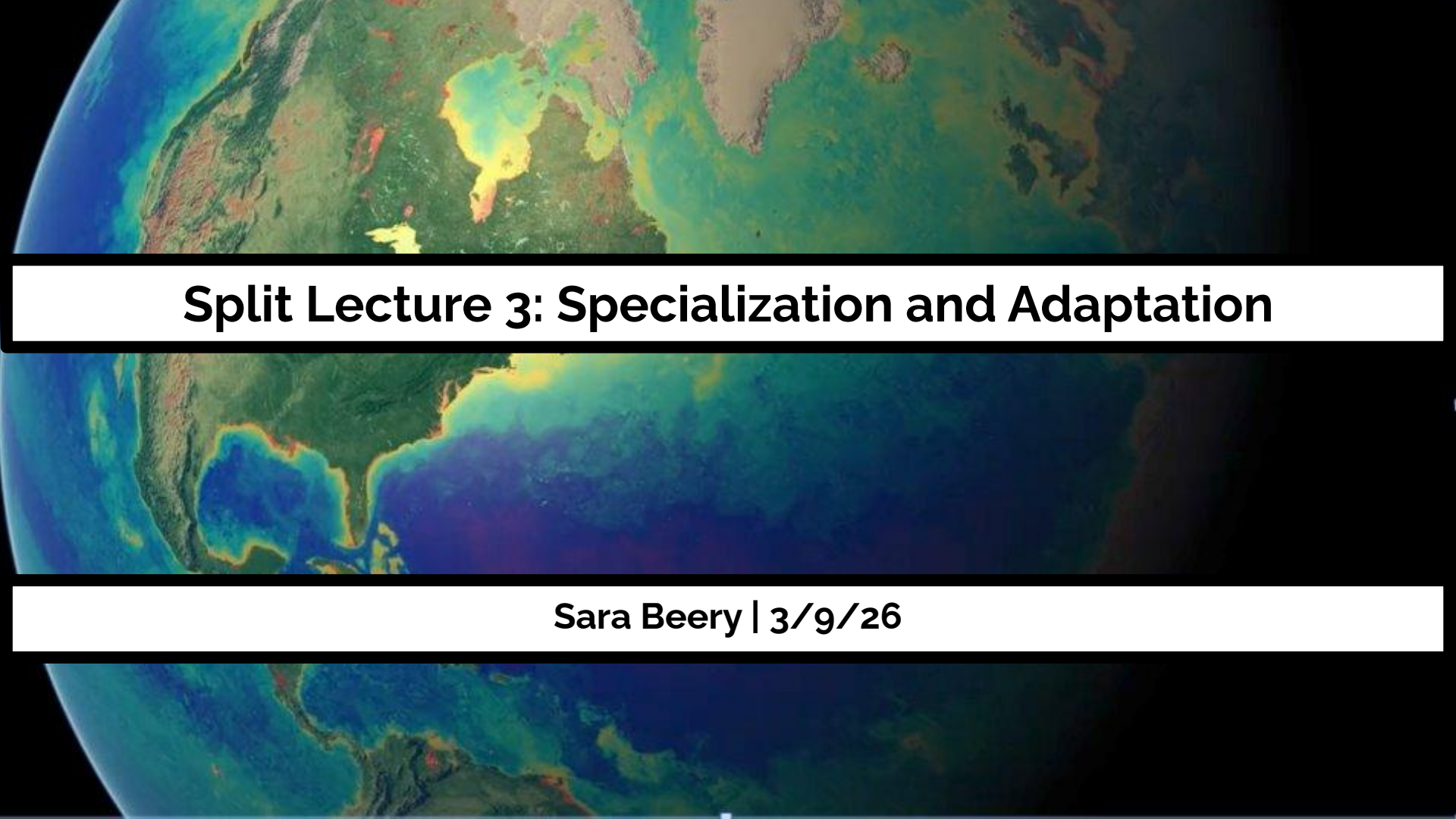


Split Lecture 3: Specialization and Adaptation

Sara Beery | 3/9/26

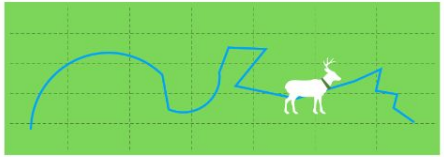
Incredible diversity and scale of data collected

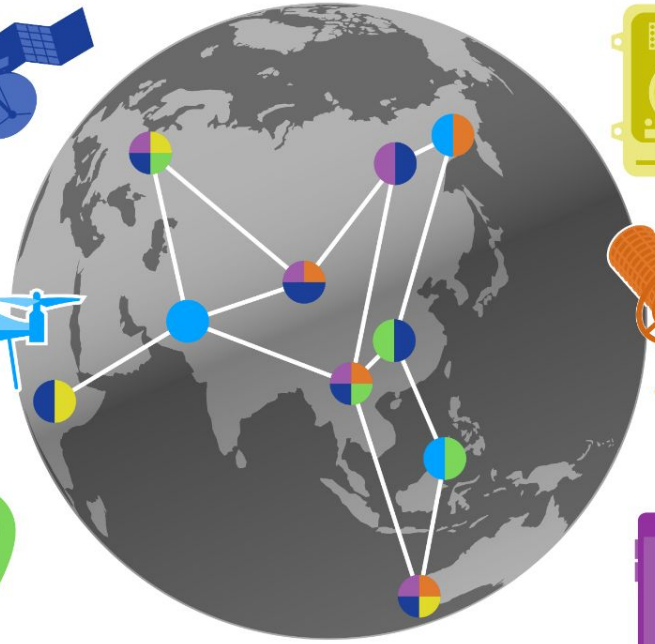
Mobile Sensors

Satellite (optical, SAR, LiDAR)

UAV (RGB, thermal, LiDAR)

On-Animal Sensors

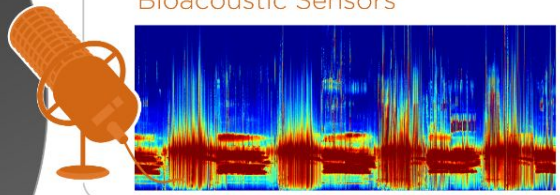




Stationary Sensors

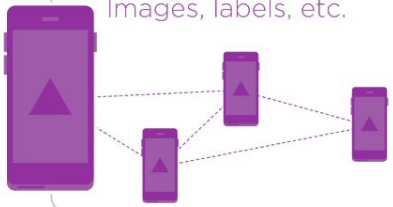
Camera Traps

Bioacoustic Sensors



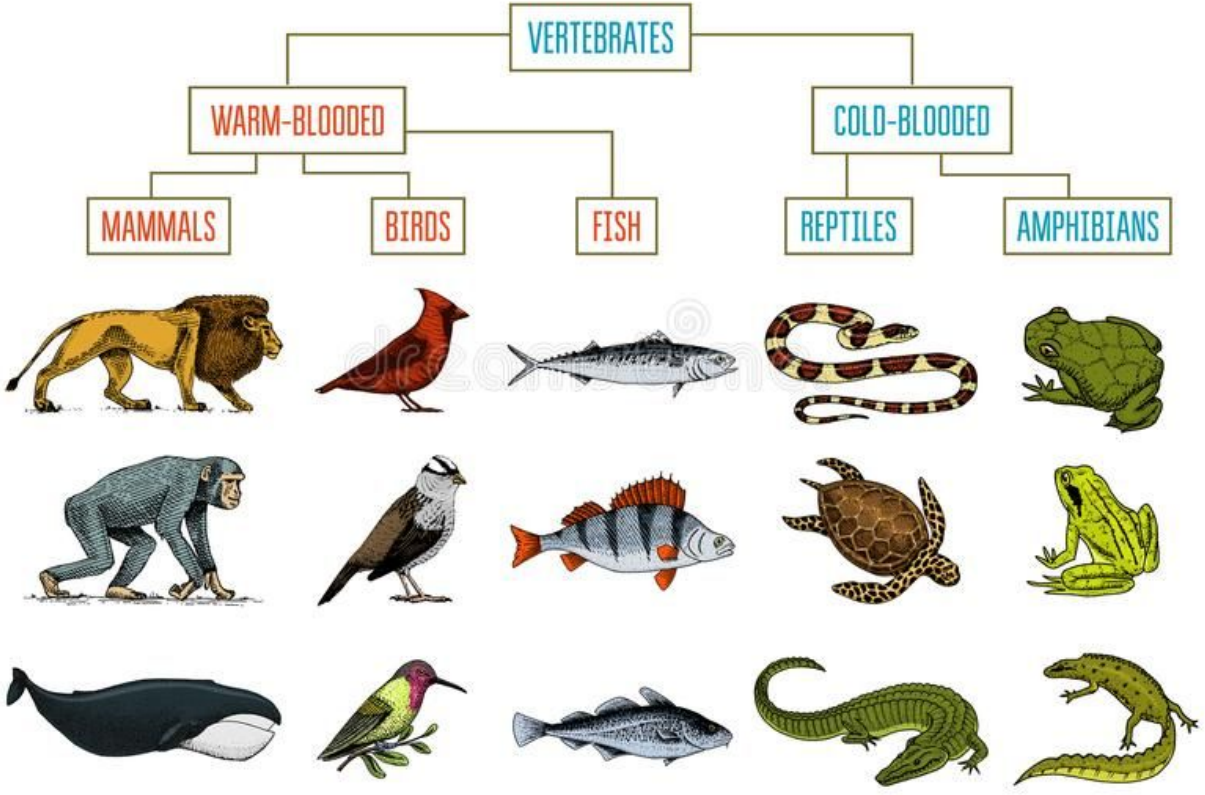
Community Science

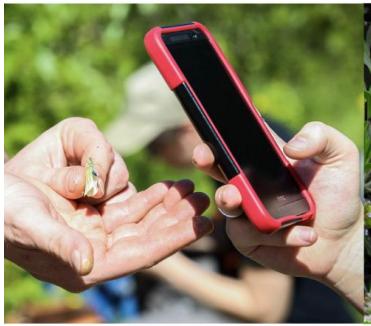
Images, labels, etc.

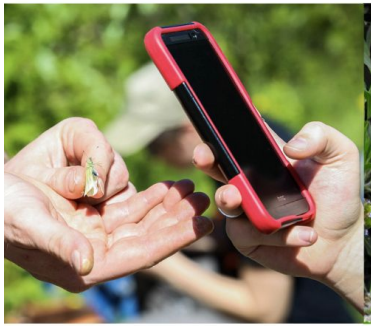


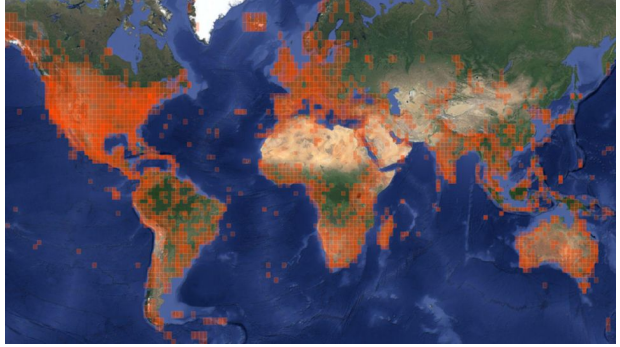

Sampling bias is not uniform for any modality

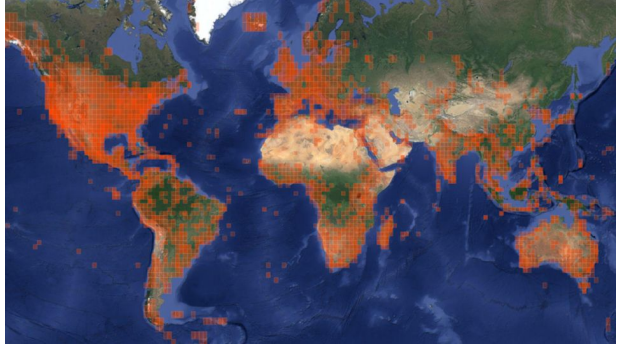
e.g. PIR detection rates vary per-species based on size and temperature





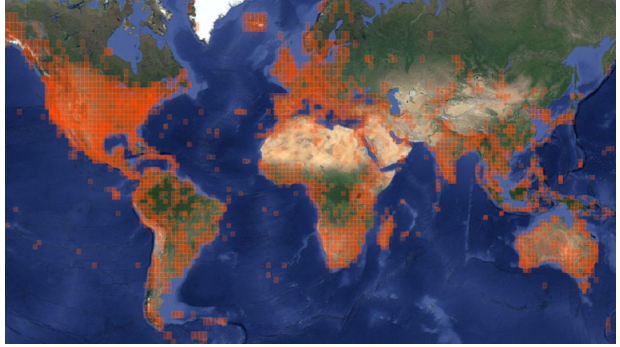


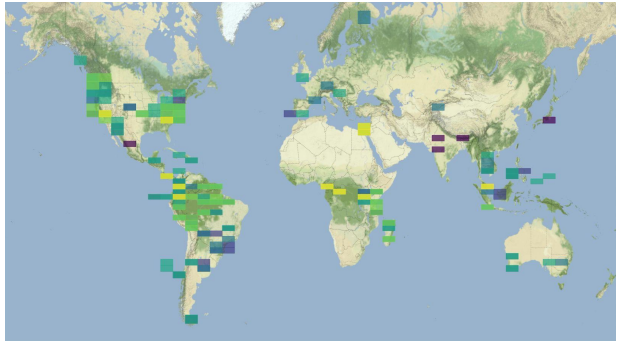


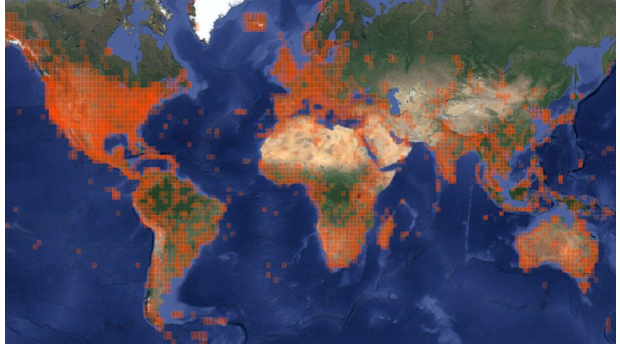

Ltl Acorn 0009 062°F 017°C


09.27.2012 Ltl Acorn 0009 082°F 028°C
10.19.2012 16:57:09

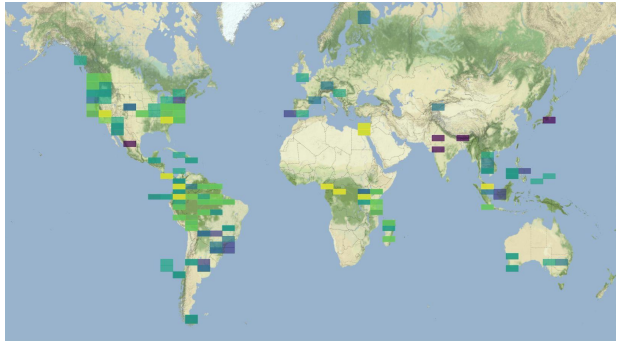


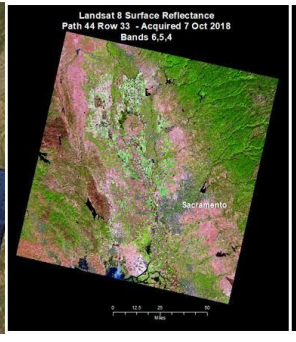

Ltl Acorn 0009 062°F 017°C
09.27.2012 Ltl Acorn 0009 082°F 028°C
10.19.2012 16:57:09

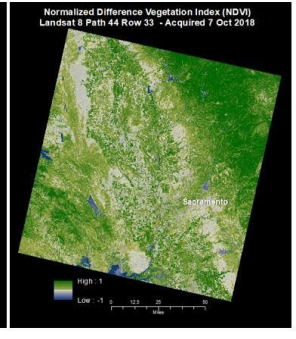


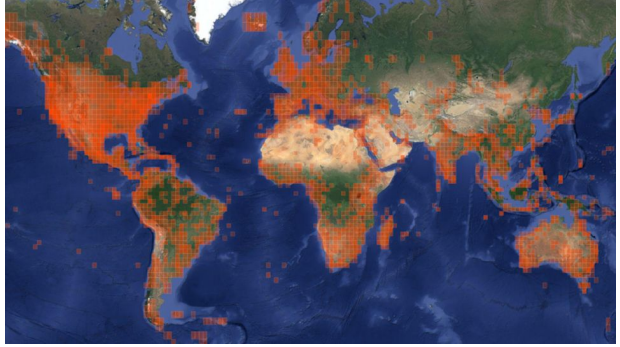



Ltl Acorn 0009 062°F 017°C
09.27.2012 Ltl Acorn 0009 082°F 028°C
10.19.2012 16:57:09

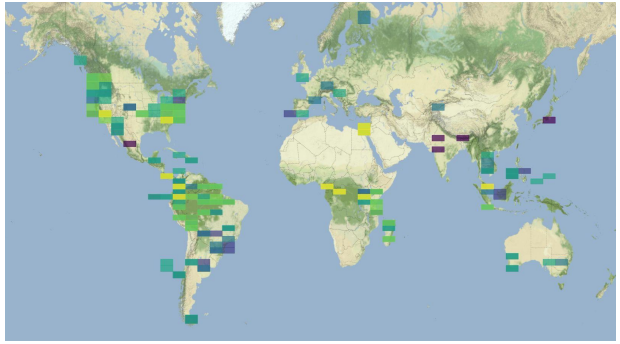


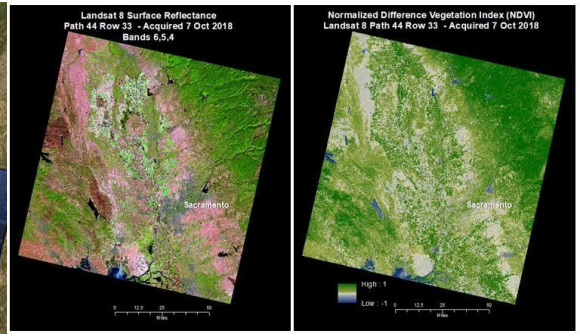

Landsat 8 Surface Reflectance
Path 44 Row 33 - Acquired 7 Oct 2018
Bands 6,5,4
Sacramento


Normalized Difference Vegetation Index (NDVI)
Landsat 8 Path 44 Row 33 - Acquired 7 Oct 2018
Sacramento
High : 1
Low : -1

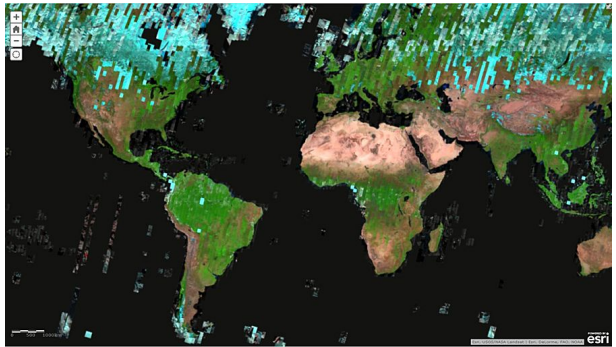



Ltl Acorn 0009 062°F 017°C
09.27.2012 Ltl Acorn 0009 082°F 028°C
10.19.2012 16:57:09

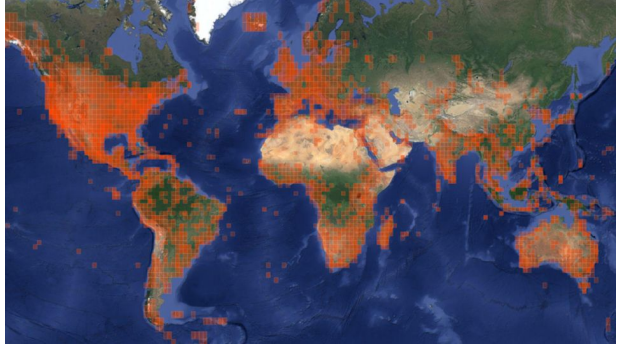


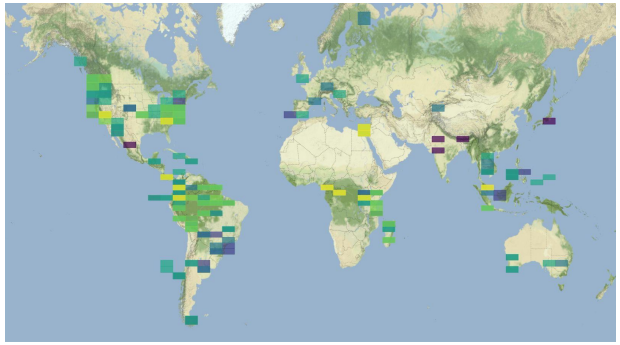

Landsat 8 Surface Reflectance
Path 44 Row 33 - Acquired 7 Oct 2018
Bands 6,5,4
Sacramento

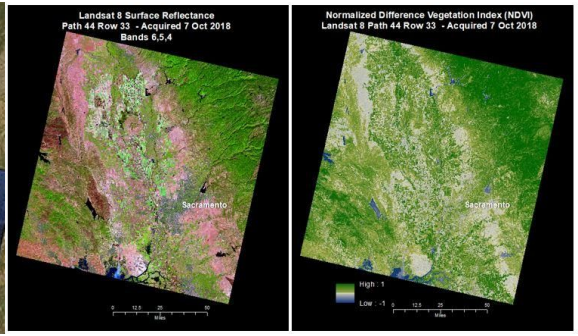
Normalized Difference Vegetation Index (NDVI)
Landsat 8 Path 44 Row 33 - Acquired 7 Oct 2018
Sacramento
High : 1
Low : -1

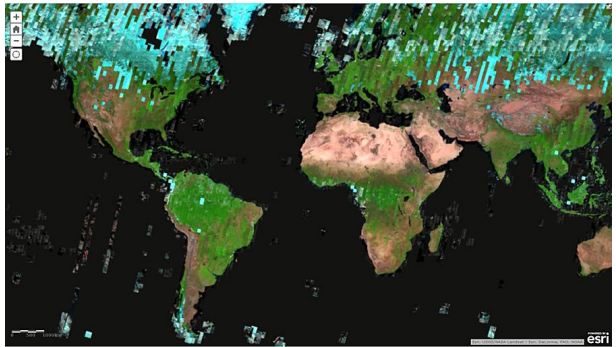


Heterogeneous Sampling

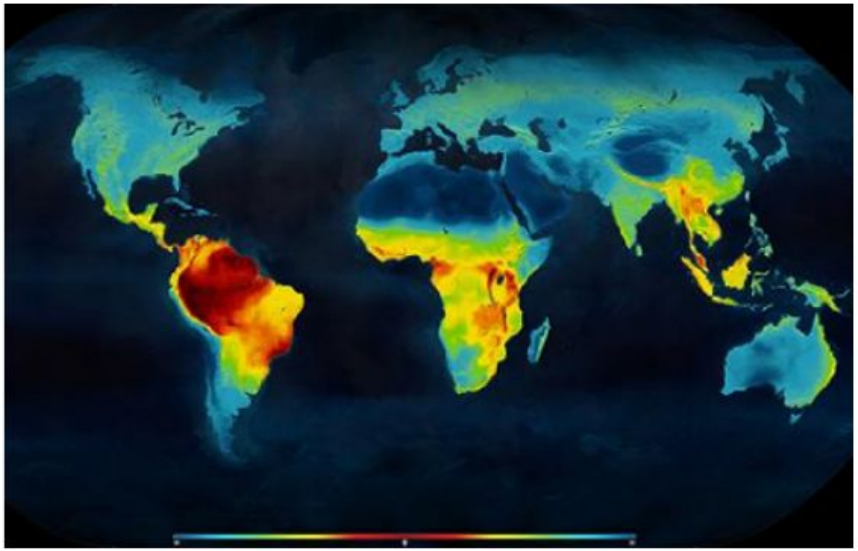




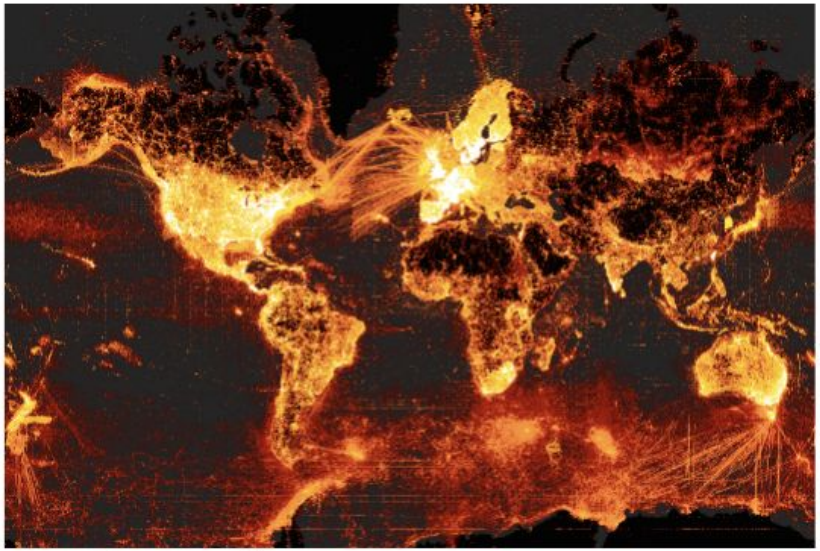




Biodiversity data is not IID - spatially, temporally, or taxonomically



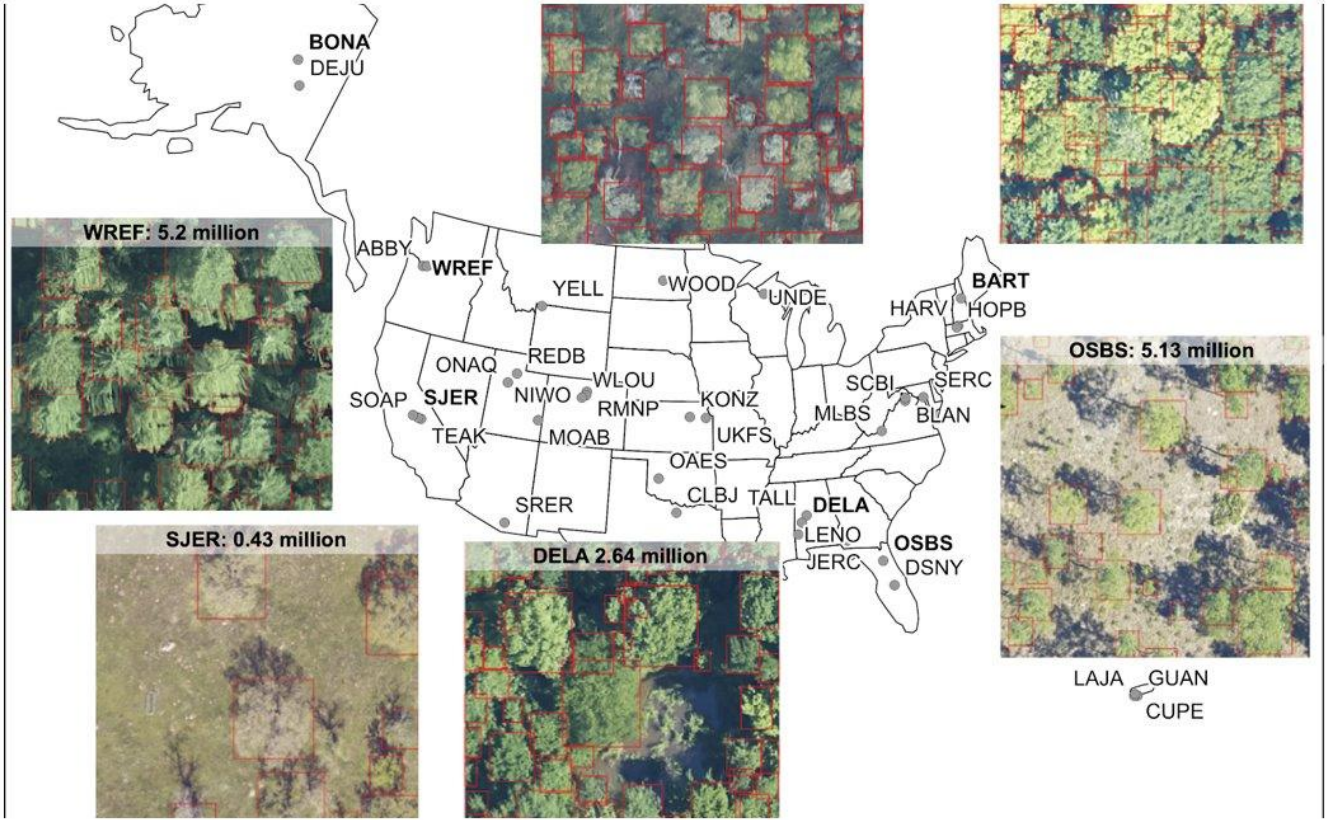
Map of global biodiversity



Species occurrence data in GBIF

Scaling Biodiversity Monitoring for the Data Age - ACM XRDS 2021

Detecting individual tree crowns



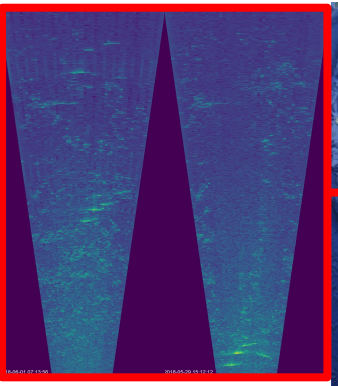
NEONCROWNS Dataset

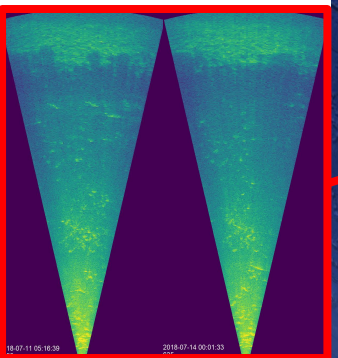
104,675,304 trees

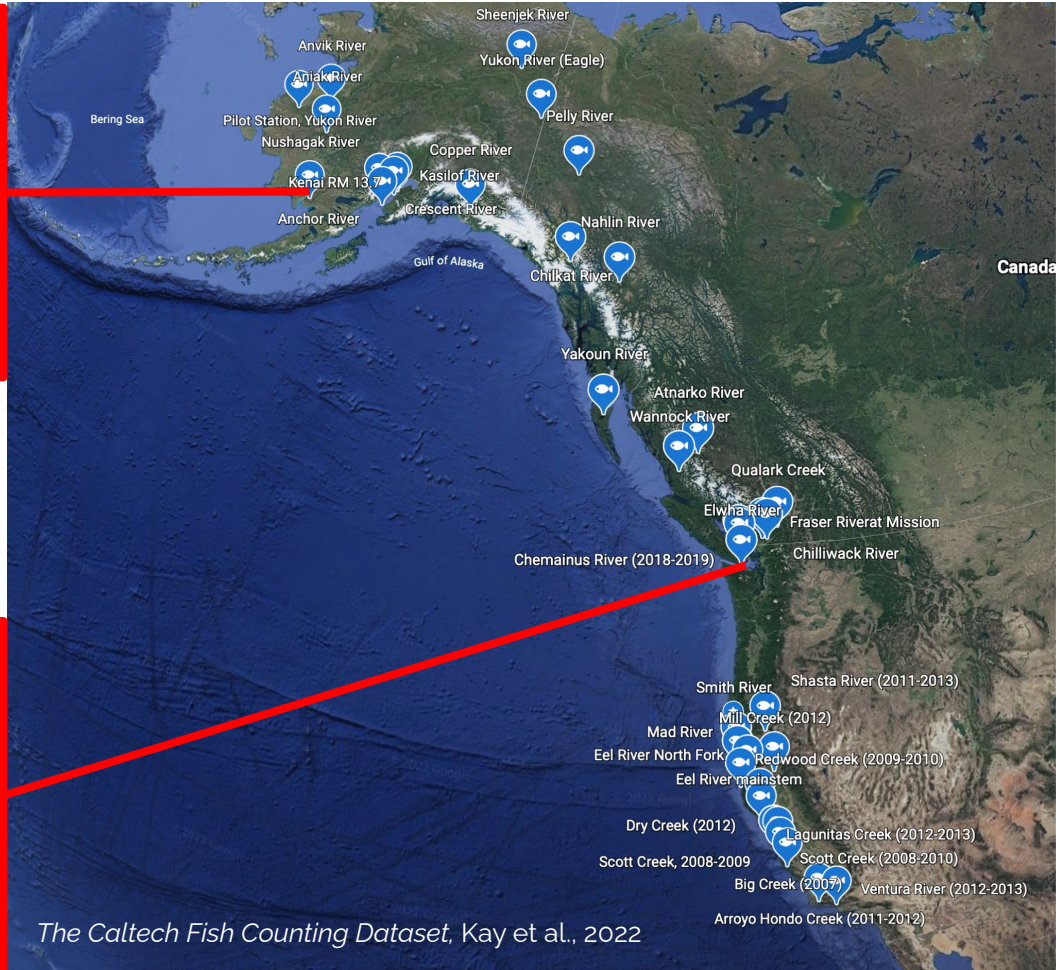
http://visualize.idtrees.org/

Weinstein et al., 2020

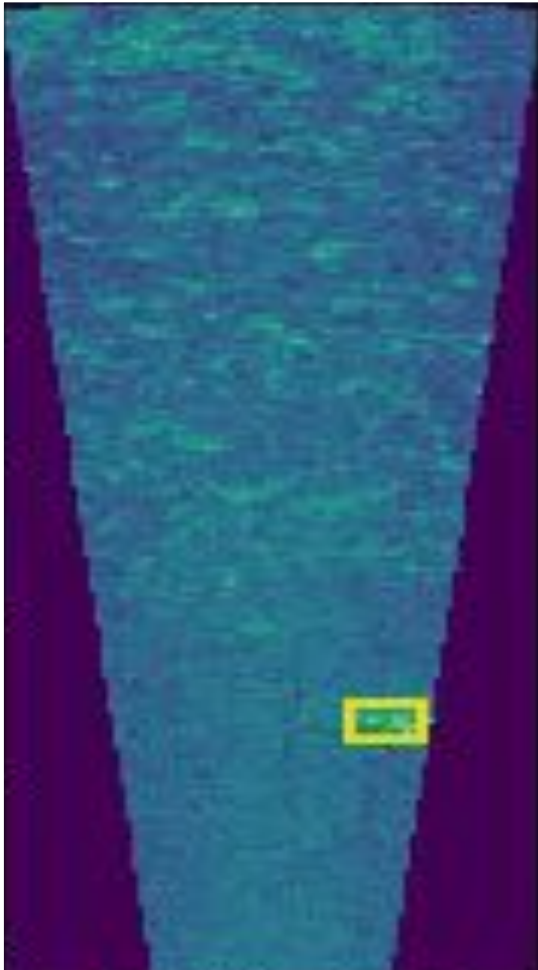
Detecting and counting salmon in static sonar



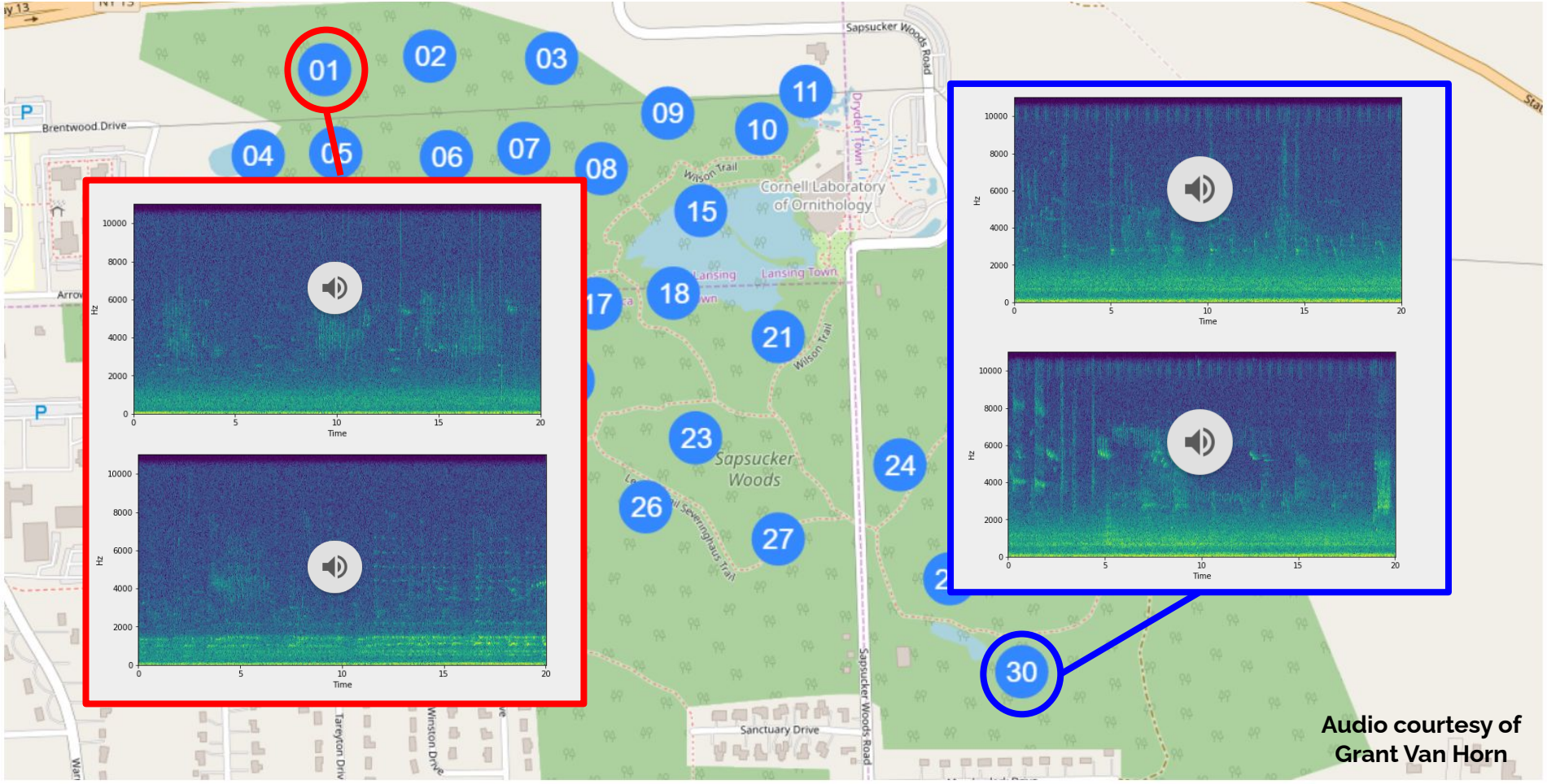




The Caltech Fish Counting Dataset, Kay et al., 2022



Detecting and categorizing birdsong in static bioacoustic sensors



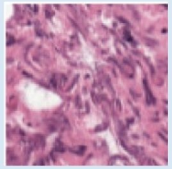
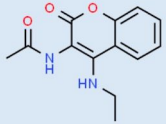
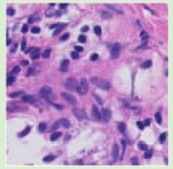
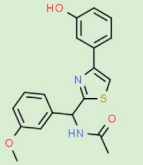
Audio courtesy of Grant Van Horn

Distribution shifts are ubiquitous in real-world scenarios

WILDS

https://wilds.stanford.edu/

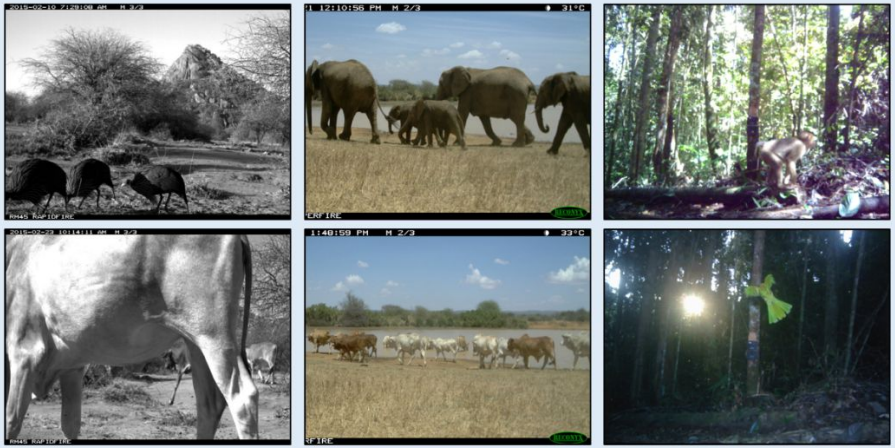
Pang Wei Koh*, Shiori Sagawa*, Henrik Marklund, Sang Michael Xie, Marvin Zhang,
Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Sara Beery,
Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang

	Camelyon17	iWildCam	PovertyMap	FMoW	Amazon	CivilComments	OGB-MolPCBA
Shift	Hospitals	Locations	Countries	Time	Users	Demographics	Scaffold
Train					Overall a solid package that has a good quality of construction for the price.	What do Black and LGBT people have to do with bicycle licensing?	
Test					I *loved* my French press, it's so perfect and came with all this fun stuff!	As a Christian, I will not be patronizing any of those businesses.	
Adapted from	Bandi et al. 2018	Beery et al. 2020	Yeh et al. 2020	Christie et al. 2018	Ni et al. 2019	Borkan et al. 2019	Hu et al. 2020

We can build evaluation frameworks that measure ID vs OOD performance

Training data

Camera 1 | Camera 2 | … | Camera 245



Out-of-distribution (OOD) test data

Camera 246

…

Control: In-distribution (ID) test data

Camera 1 | Camera 2 | … | Camera 245

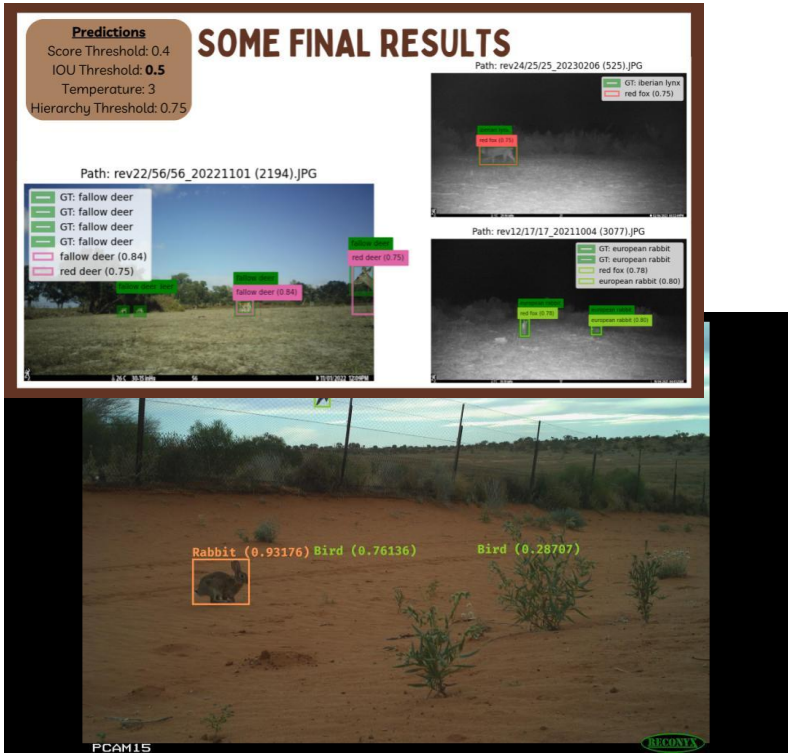
Macro F1

ID 47.0% → -16.0% → OOD 31.0%

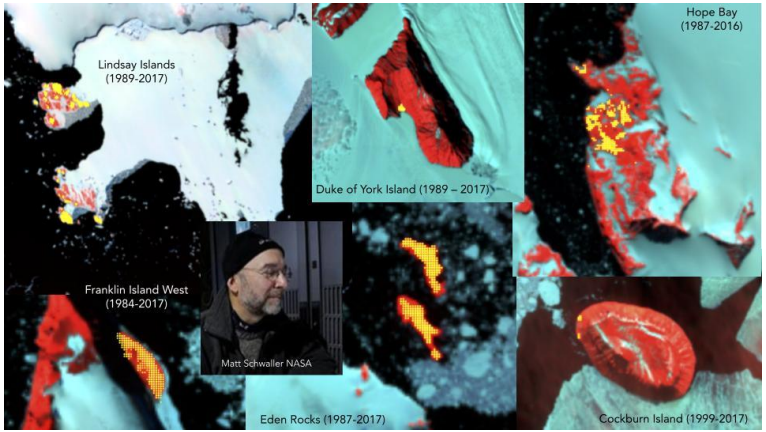
[Beery et al., 2020; Koh et al., 2021]

Slide from Shiori Sagawa

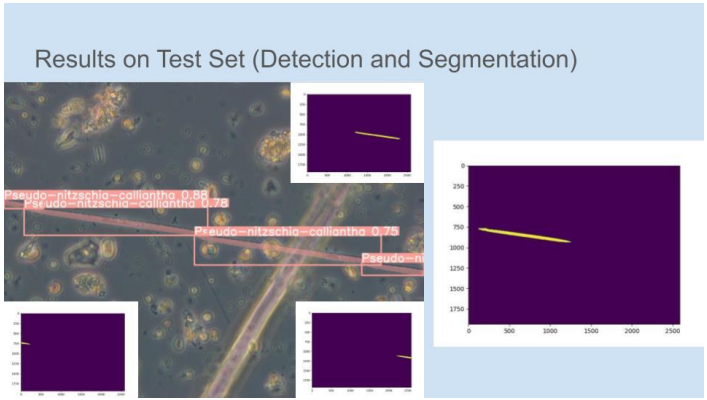
Typical model development / benchmarking



Animal occupancy/abundance, Marquez-Rodriguez, Tamm

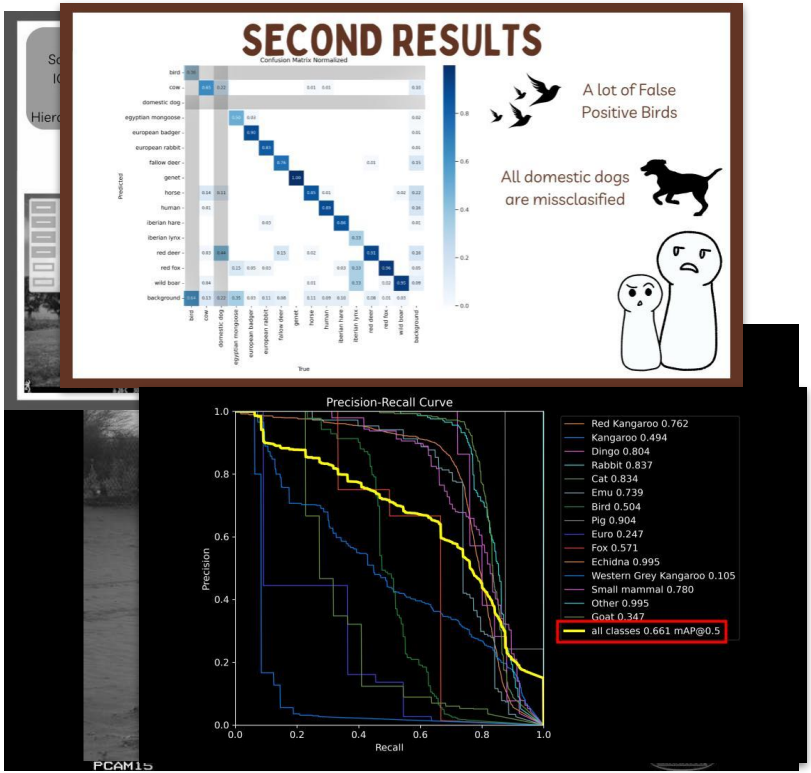


Guano surface area, Che-Castaldo

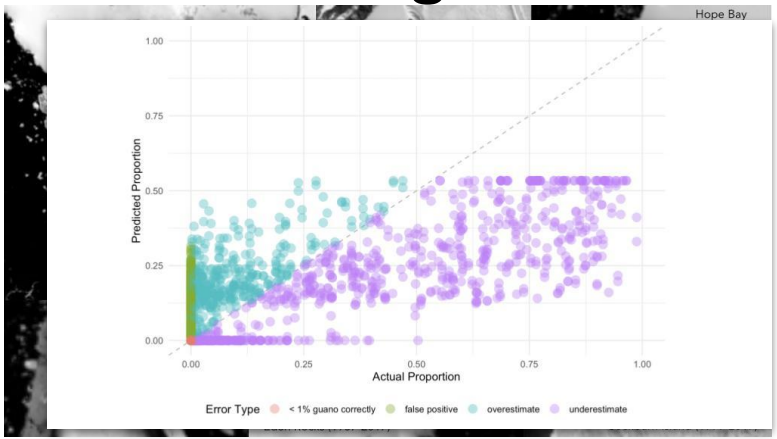


Phytoplankton biovolume, Marzidovšek

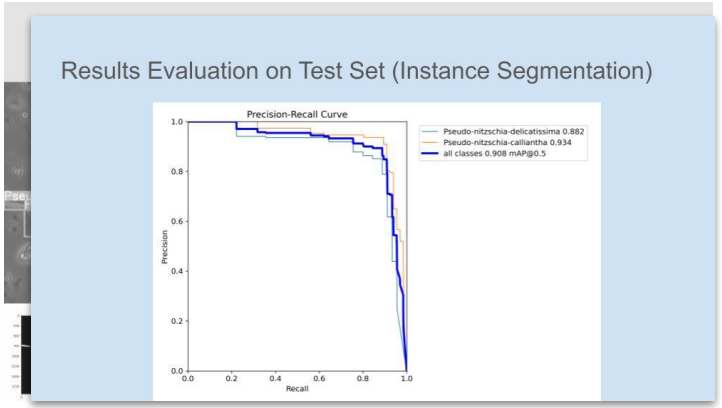
Typical model development / benchmarking



Animal occupancy/abundance, Marquez-Rodriguez, Tamm



Guano surface area, Che-Castaldo



Phytoplankton biovolume, Marzidovšek

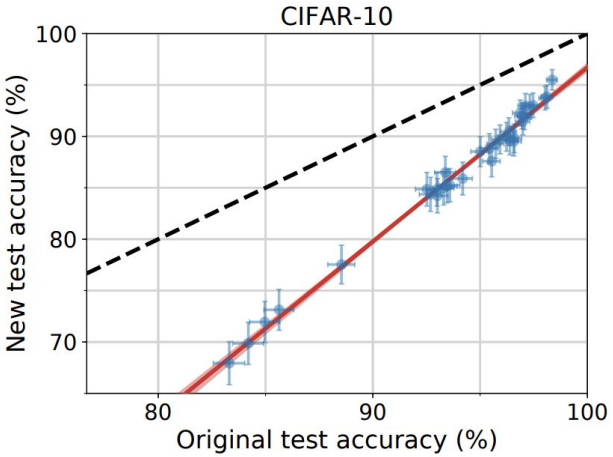
Performance on new data?

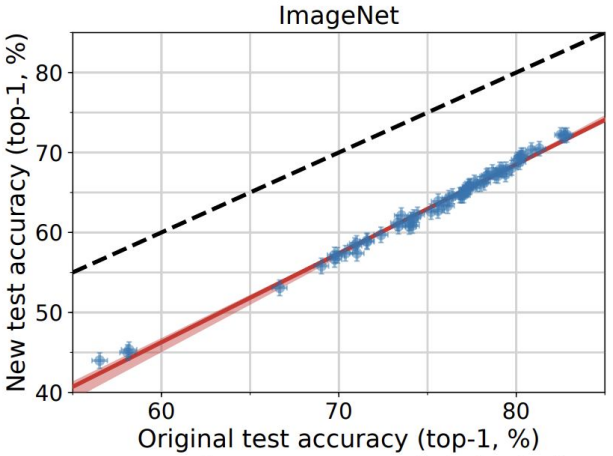
Do ImageNet Classifiers Generalize to ImageNet?

Benjamin Recht* UC Berkeley
Rebecca Roelofs UC Berkeley
Ludwig Schmidt UC Berkeley
Vaishaal Shankar UC Berkeley

Abstract

We build new test sets for the CIFAR-10 and ImageNet datasets. Both benchmarks have been the focus of intense research for almost a decade, raising the danger of overfitting to excessively re-used test sets. By closely following the original dataset creation processes, we test to what extent current classification models generalize to new data. We evaluate a broad range of models and find accuracy drops of 3% – 15% on CIFAR-10 and 11% – 14% on ImageNet. However, accuracy gains on the original test sets translate to larger gains on the new test sets. Our results suggest that the accuracy drops are not caused by adaptivity, but by the models' inability to generalize to slightly "harder" images than those found in the original test sets.



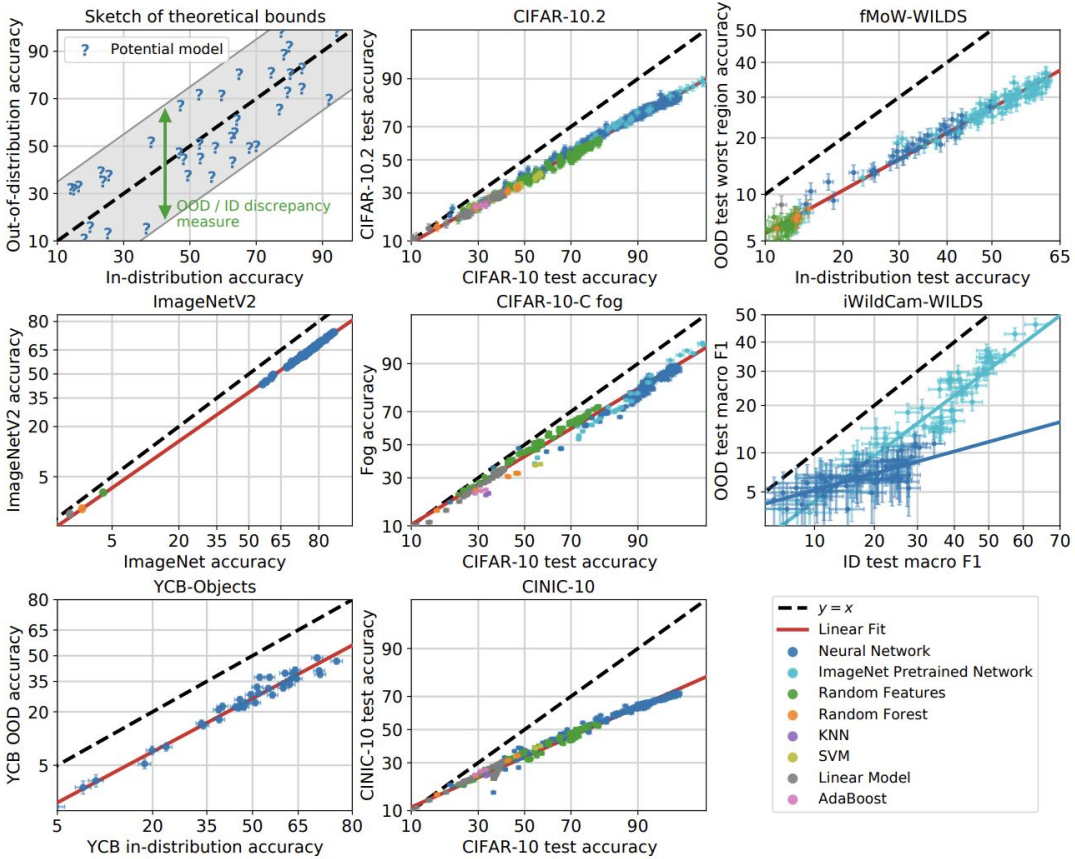


https://arxiv.org/pdf/1902.10811

Ideal reproducibility Model accuracy Linear fit

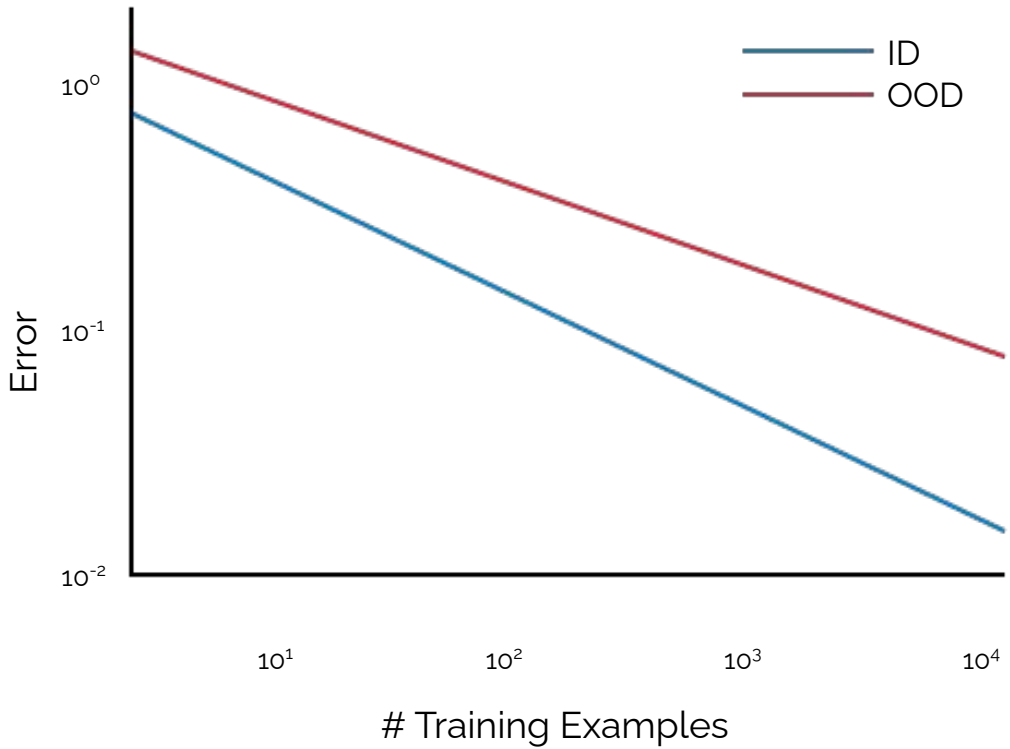
Performance on new data?

In-distribution and
out-of-distribution accuracy
often correlated, but
differently correlated on
different test data



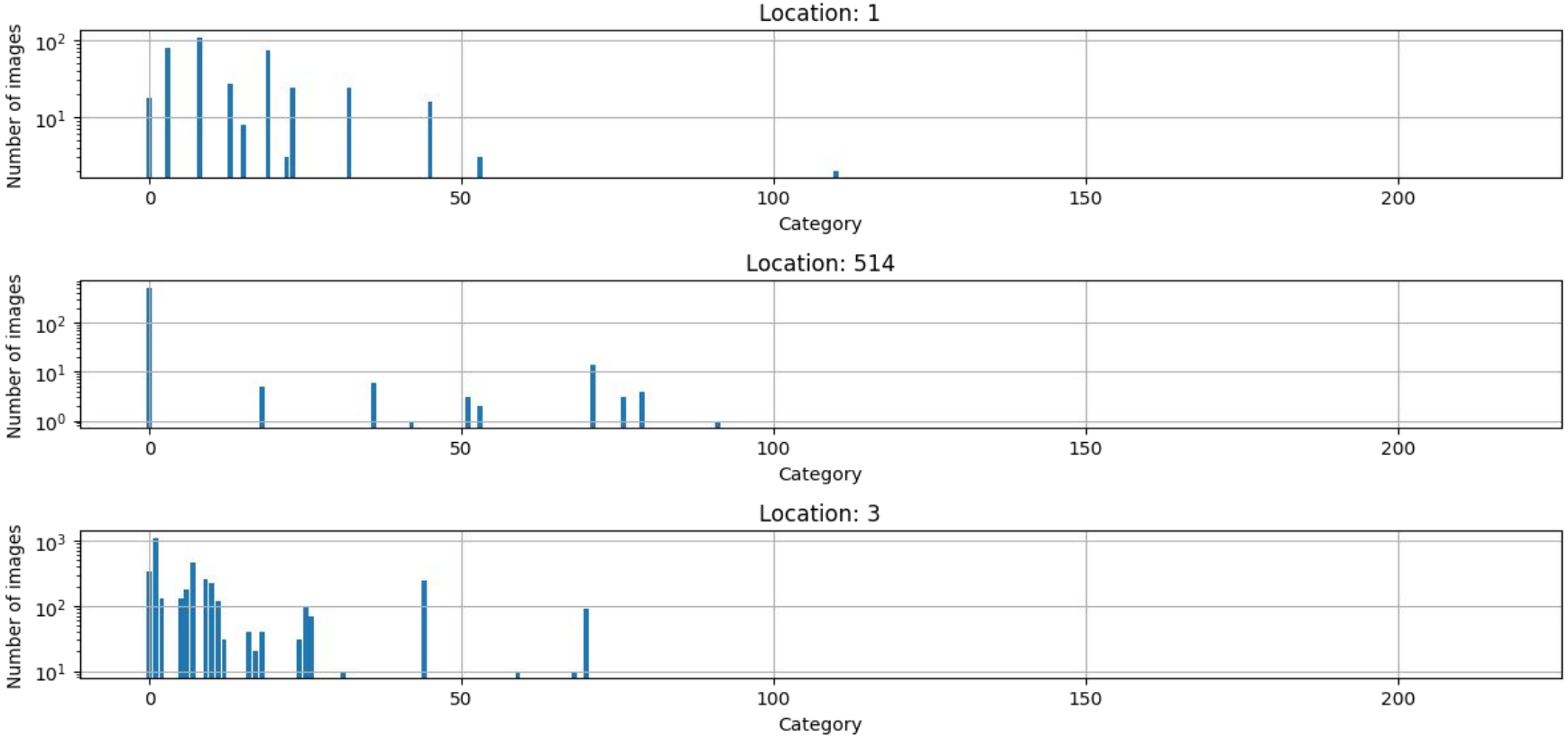
https://proceedings.mlr.press/v139/miller21b/miller21b.pdf

Performance degrades OOD even for common species



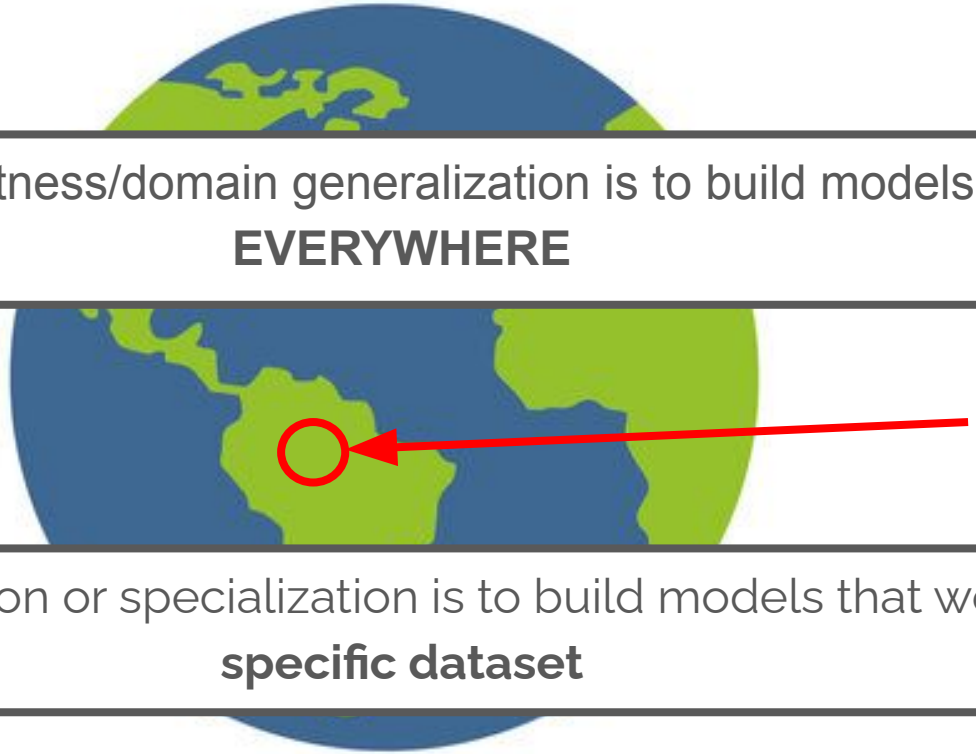
Recognition in Terra Incognita, Beery et al., ECCV 2018

Class distribution is different for each static sensor location



Robustness vs Specialization

The goal of robustness/domain generalization is to build models that work **EVERYWHERE**



The goal of adaptation or specialization is to build models that work well on a **specific dataset**

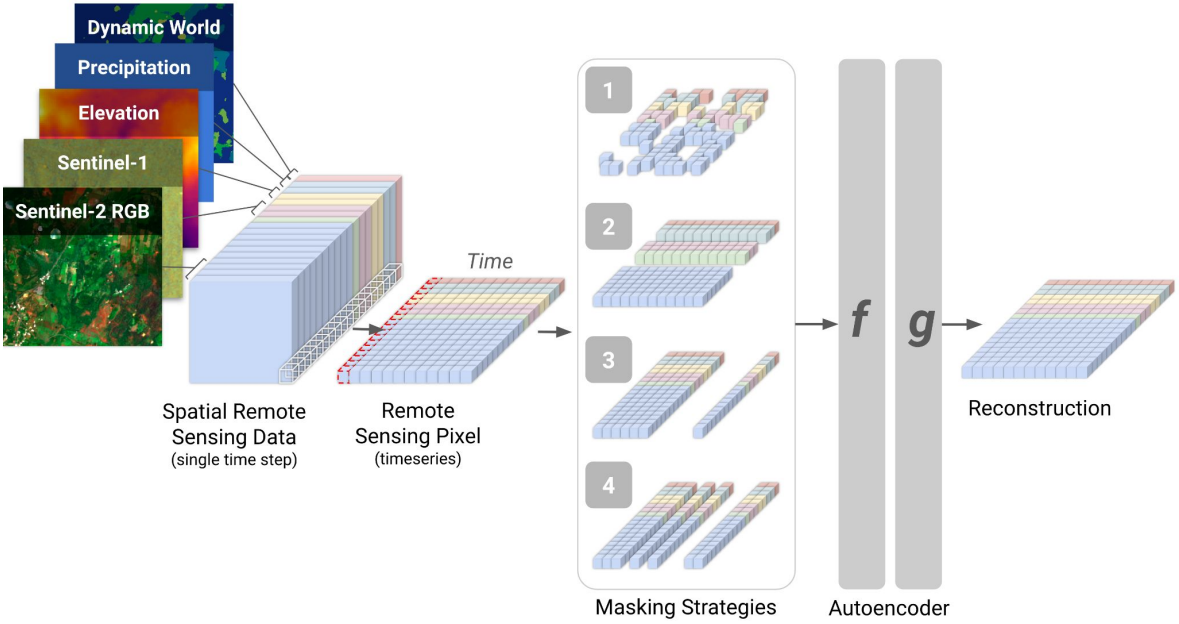
Task granularity vs generalizeability

https://github.com/ecologize/CameraTraps

Efficient Pipeline for Camera Trap Image Review, Beery, et al., DMAIC @ KDD 2019

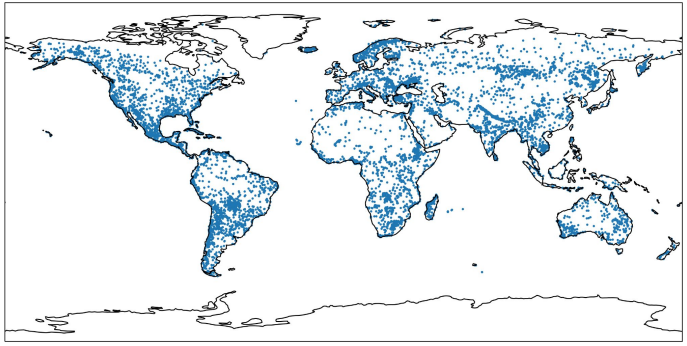
Sarah Bassing @S_Bassing · May 19

Thank goodness for the **#MegaDetector** helping me find the ONE animal image mixed in with 170,787 pictures of blowing grass and clouds from this #CameraTrap! Image recognition software is a game changer. #painless #tech4wildlife #WAPredatorPreyProject

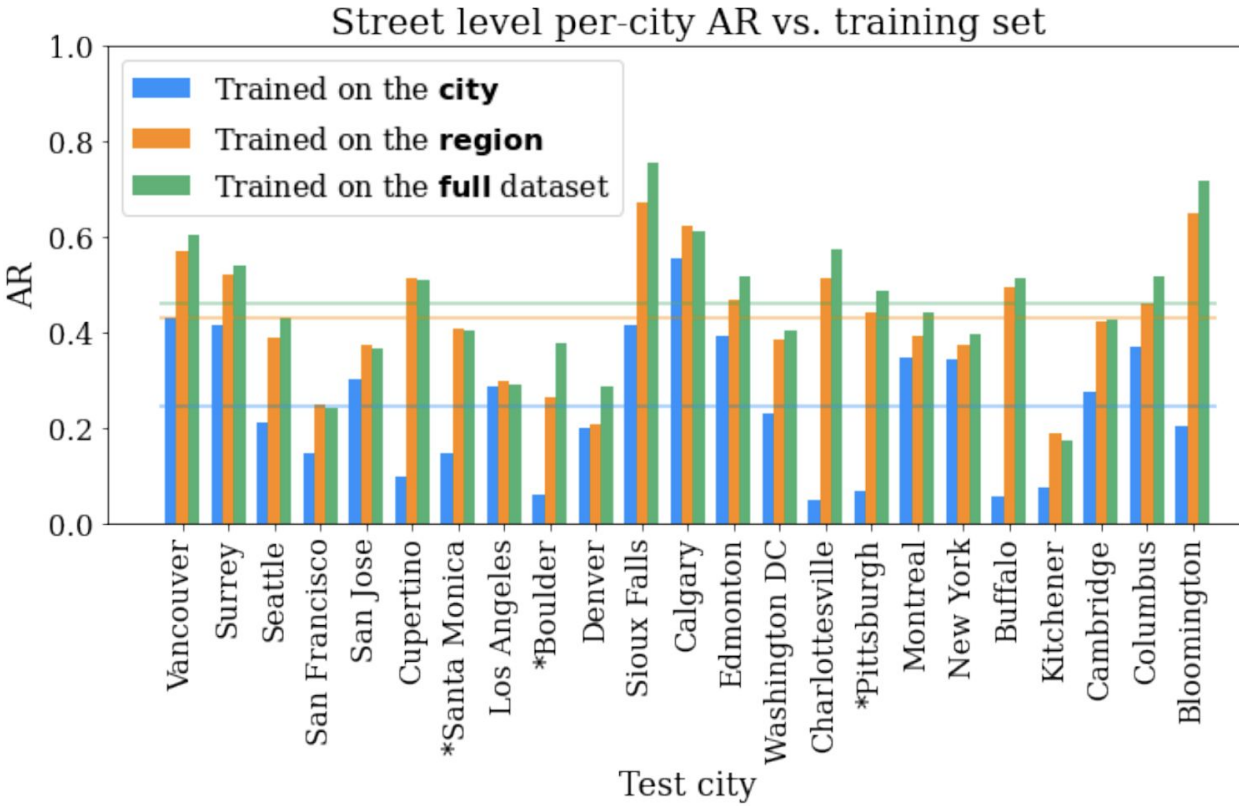


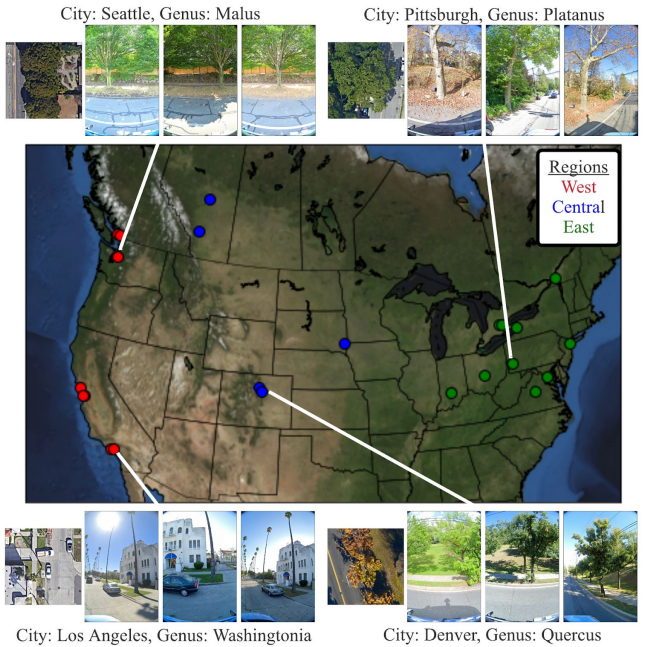
Self-supervision can learn global representations without expensive labels, *if you have global data in hand*

Presto: Lightweight, Pretrained Transformers for Remote Sensing Timeseries, Tseng*, Zvonkov*, Purohit, Rolnick, Kerner



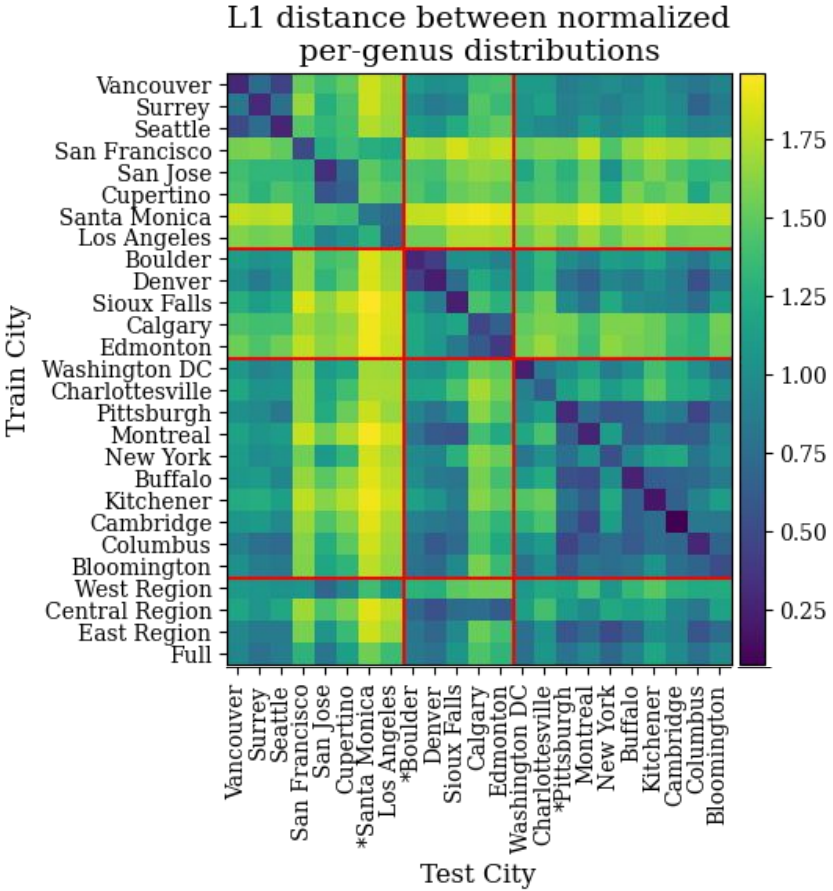
Large-scale cross-region models often outperform smaller city- or region-specific models, but not always

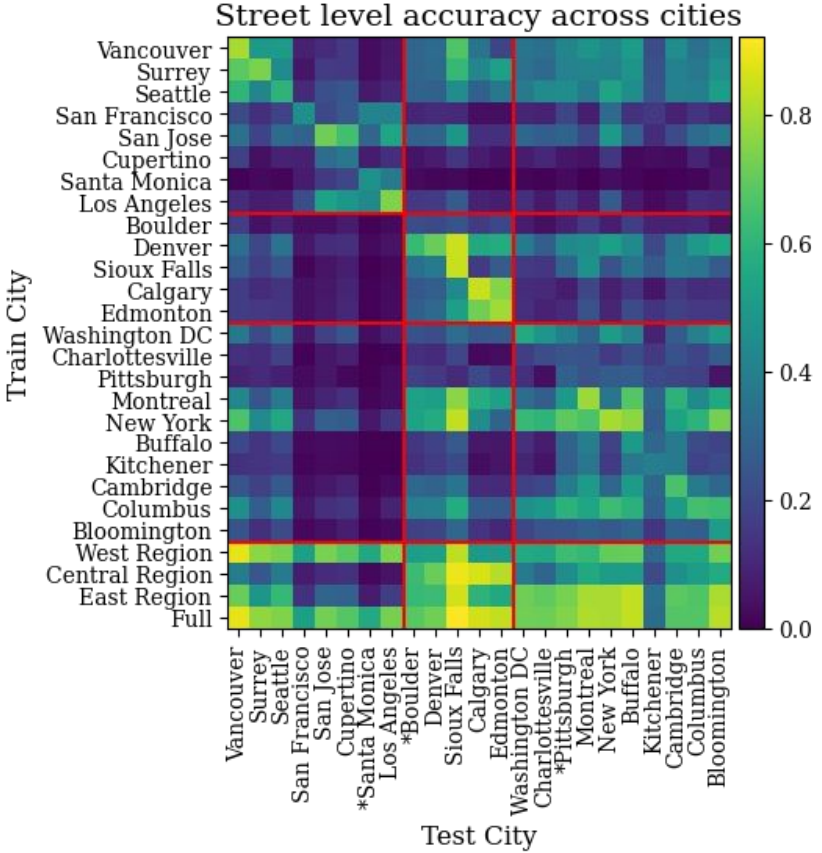




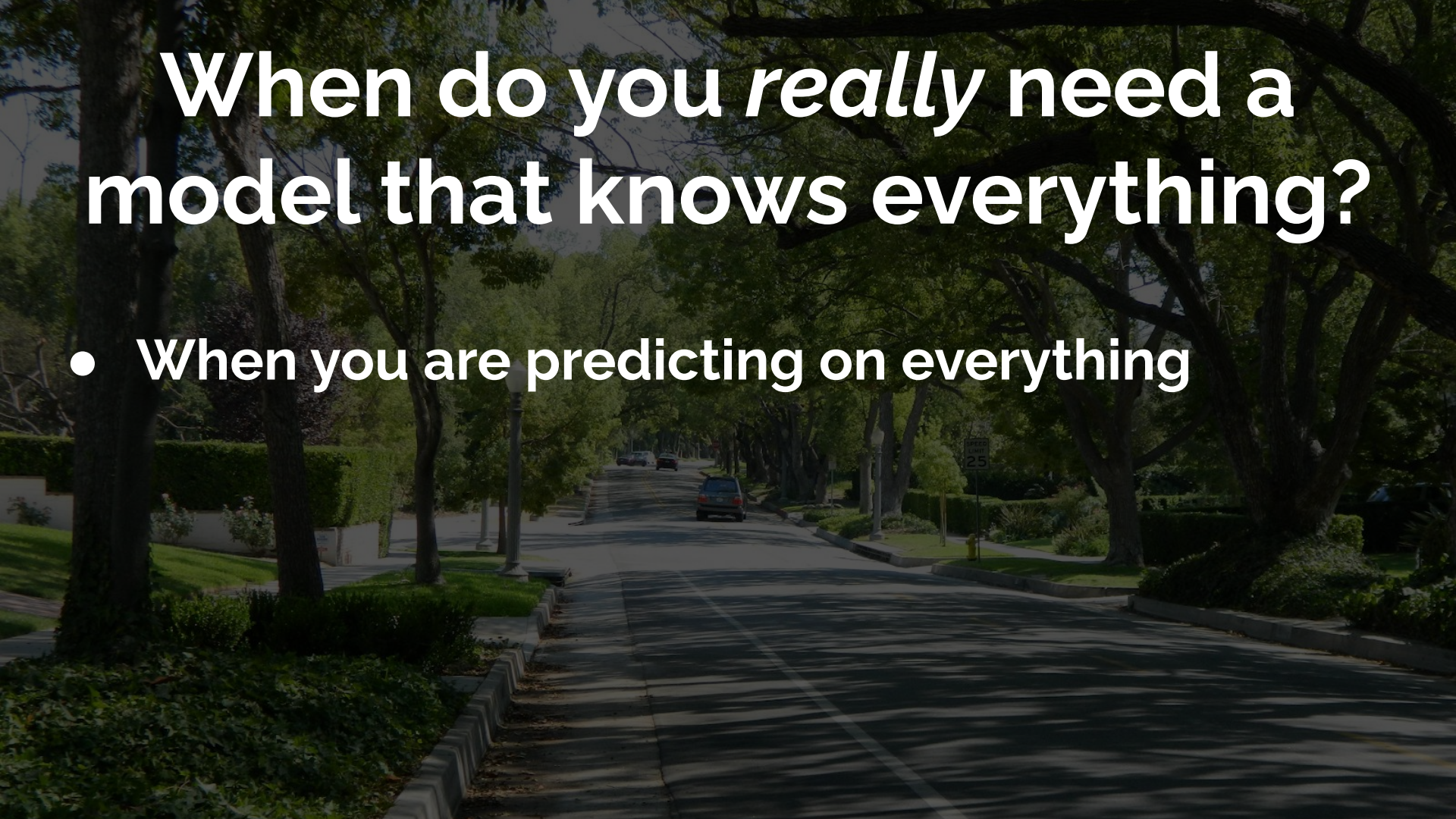
The Auto Arborist Dataset: A Large-Scale Benchmark for Multiview Urban Forest Monitoring Under Domain Shift, Beery et al., CVPR 2022

Distribution Shifts Across Cities



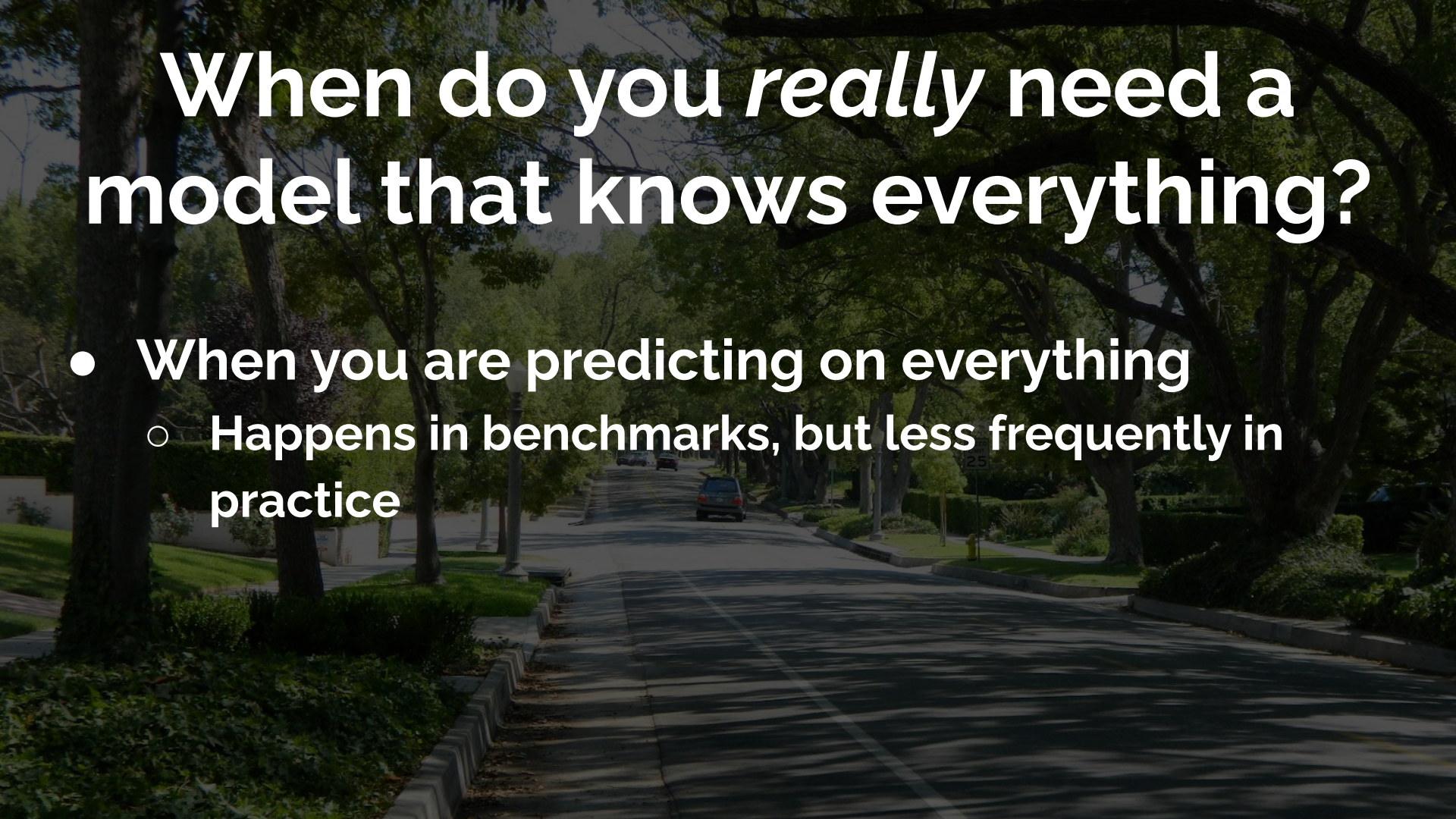



When do you *really* need a model that knows everything?



When do you *really* need a model that knows everything?

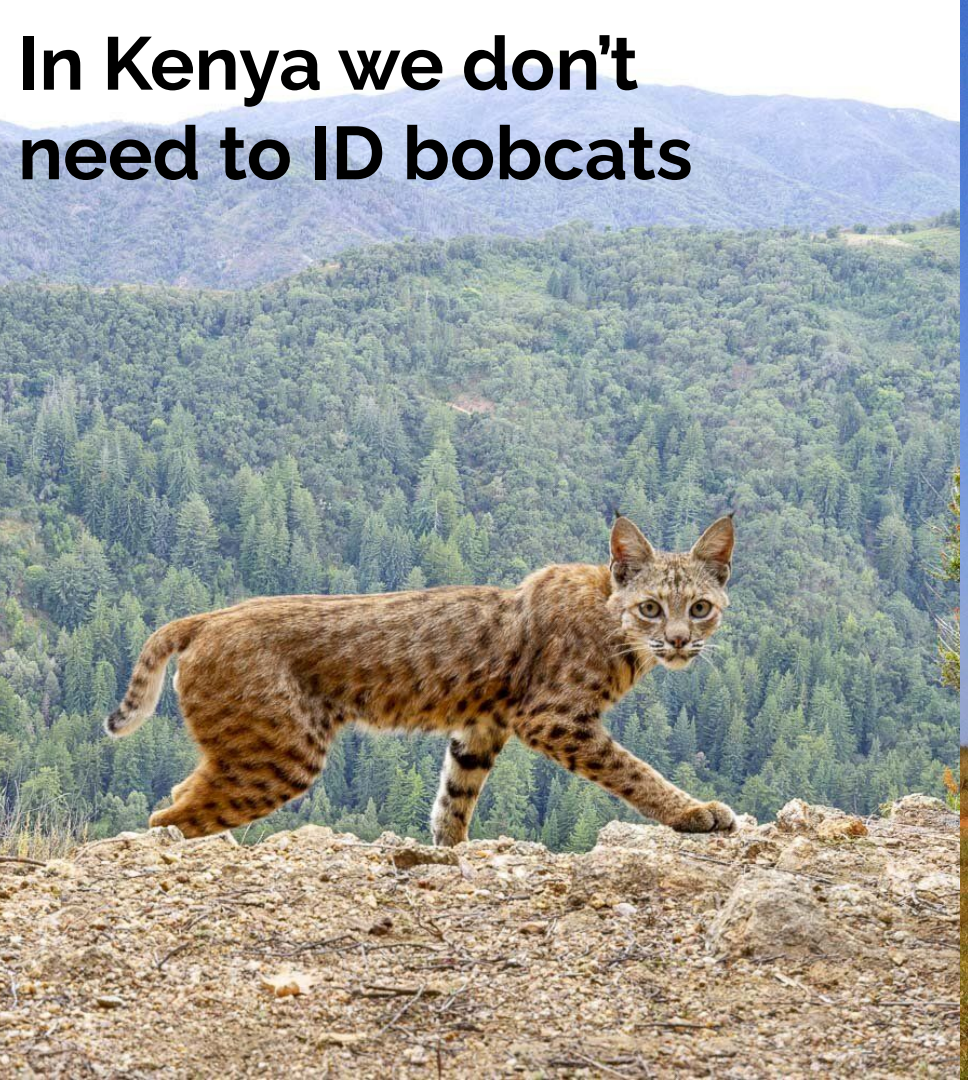
- When you are predicting on everything



When do you *really* need a model that knows everything?

- When you are predicting on everything
  - Happens in benchmarks, but less frequently in practice

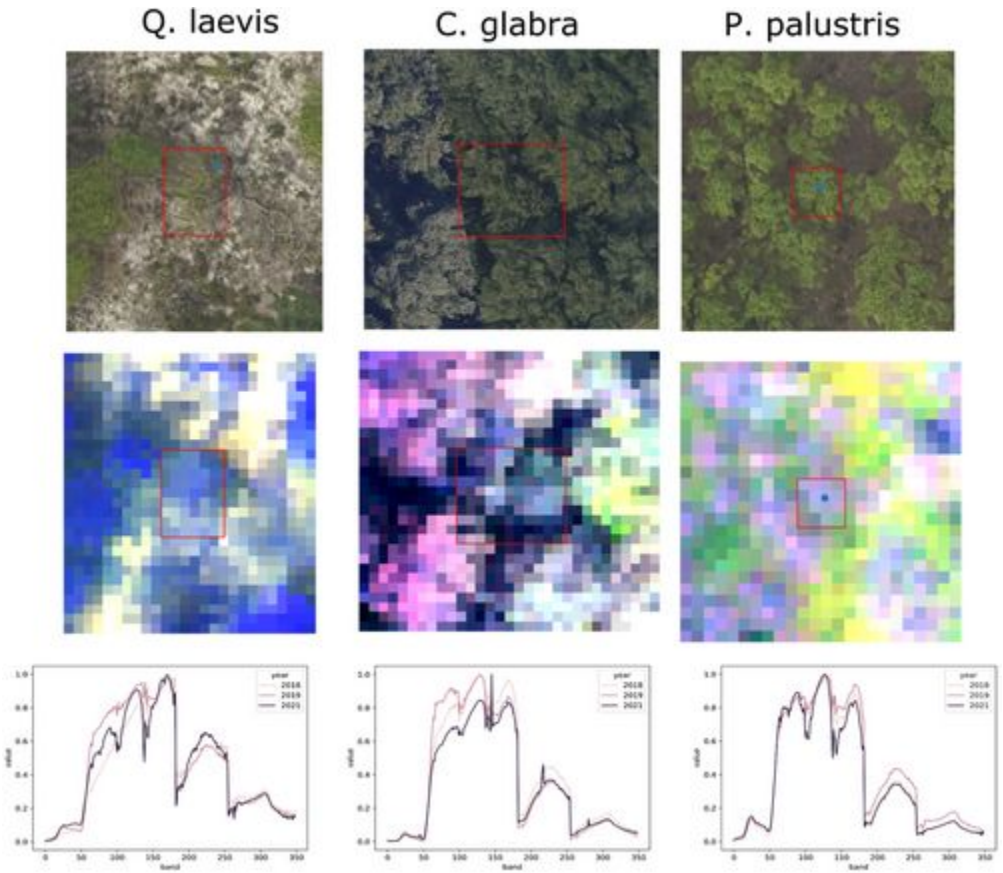
In Kenya we don't need to ID bobcats



In Idaho we don't need to ID giraffes


Lepidopterist
doesn't need to
ID sharks


Shark biologist
doesn't need to
ID moths



Individual hyperspectral tree species ID at NEON sites

- Best results come from an ensemble of specialized models
  - **per year**
  - **per taxonomic group**

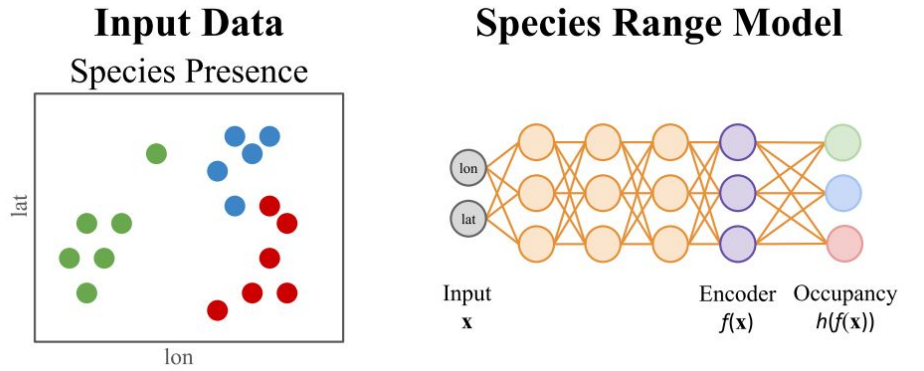
- **No expectation of generalization to future years**

Capturing long-tailed individual tree diversity using an airborne imaging and a multi-temporal hierarchical model, Weinstein et al., RSEC 2023

When do you *really* need a model that knows everything?

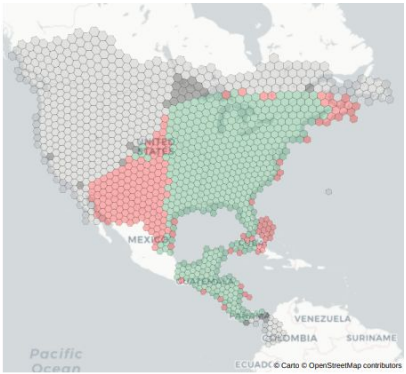
- When you are predicting on everything
  - Happens in benchmarks, but less frequently in practice
- Caveat: it is easier to maintain one model than many - so it depends on who is training models

Spatial representations of species help “specialize” in a single model

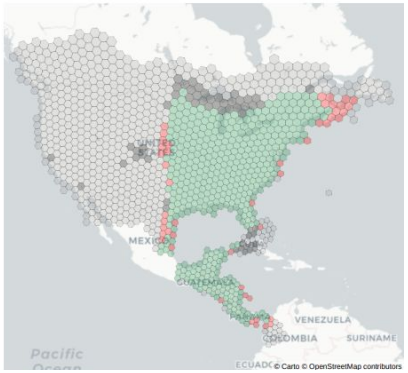


Wood thrush

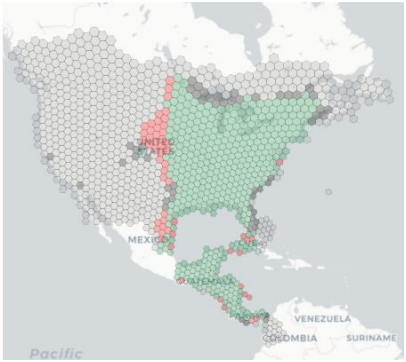
10 positives/class



100 positives/class



1000 positives/class



Spatial Implicit Neural Representations for Global-Scale Species Mapping, **Cole et al**., ICML 2023

Efficient ways to specialize

- Fine-tuning

Finetuning and few-shot learning

Which of these is an example of the same concept as the item in the box?

[Lake, Salakhutdinov, Tenenbaum, 2015]

“Deep learning”

Human learning

Representations

Models

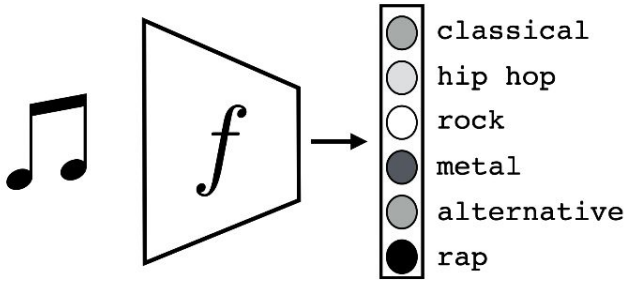
Skills

How can we give deep nets prior knowledge?

Fine-tuning

Pretraining

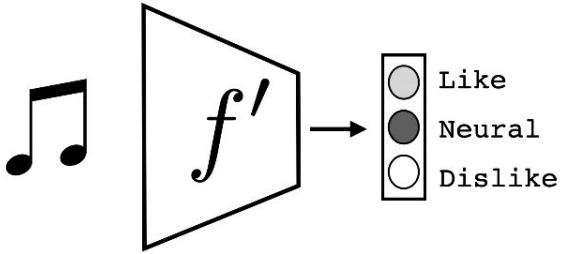
Genre recognition



A lot of data

Adapting

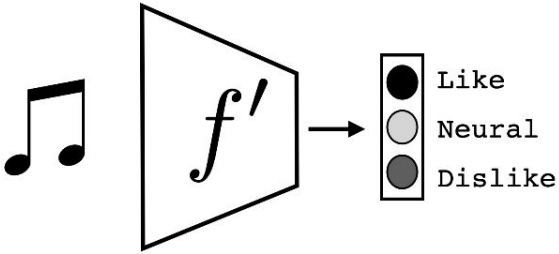
Preference prediction



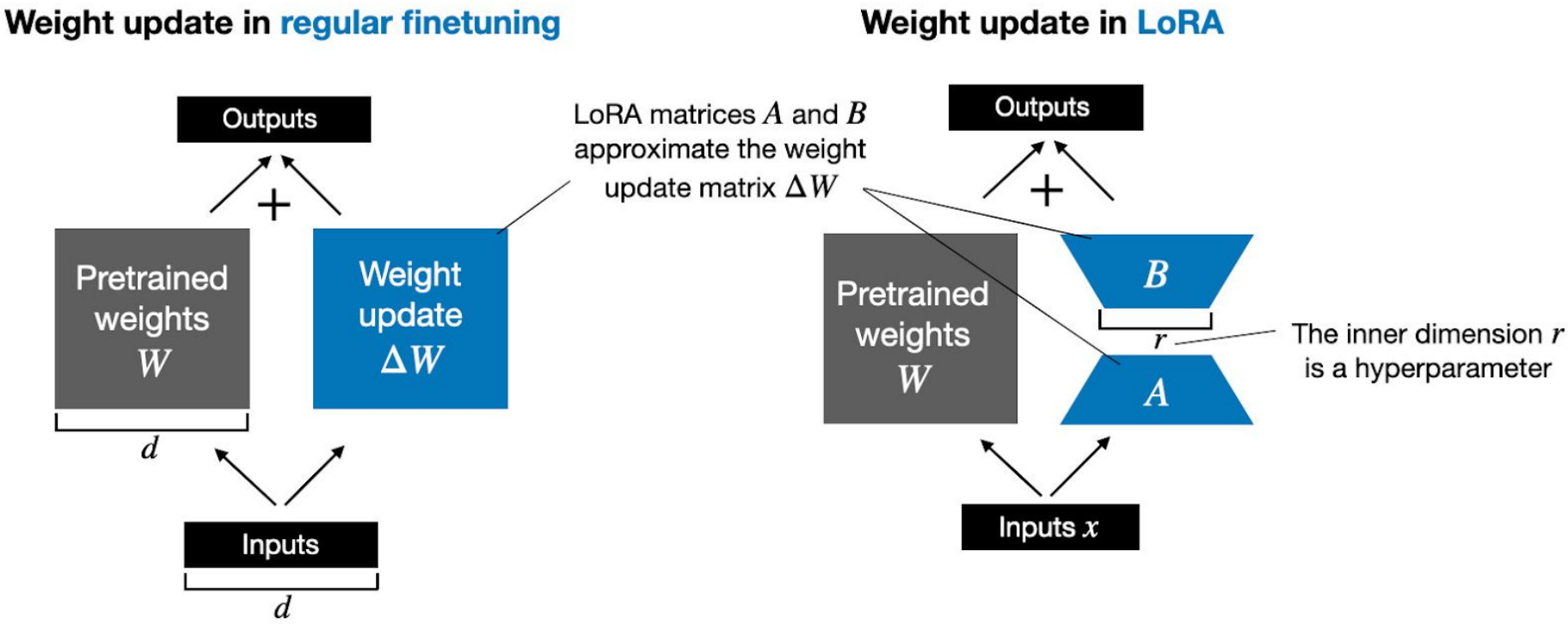
A little data

Testing

Preference prediction



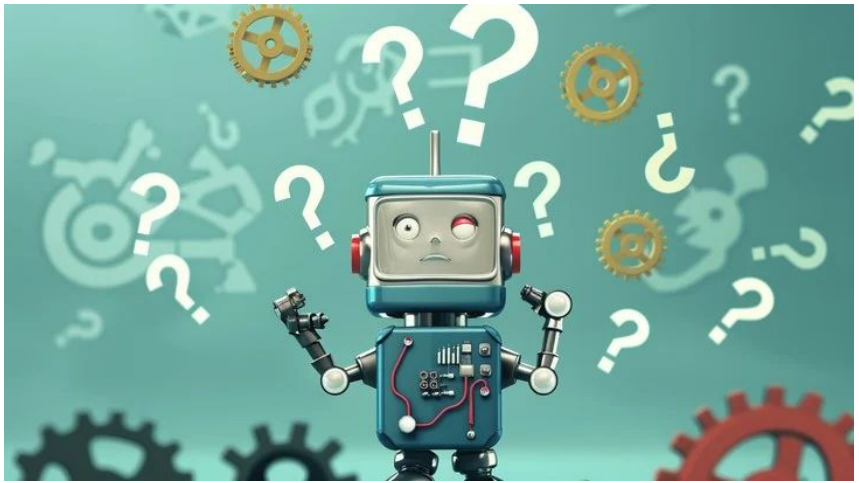
LORA: Low-rank finetuning



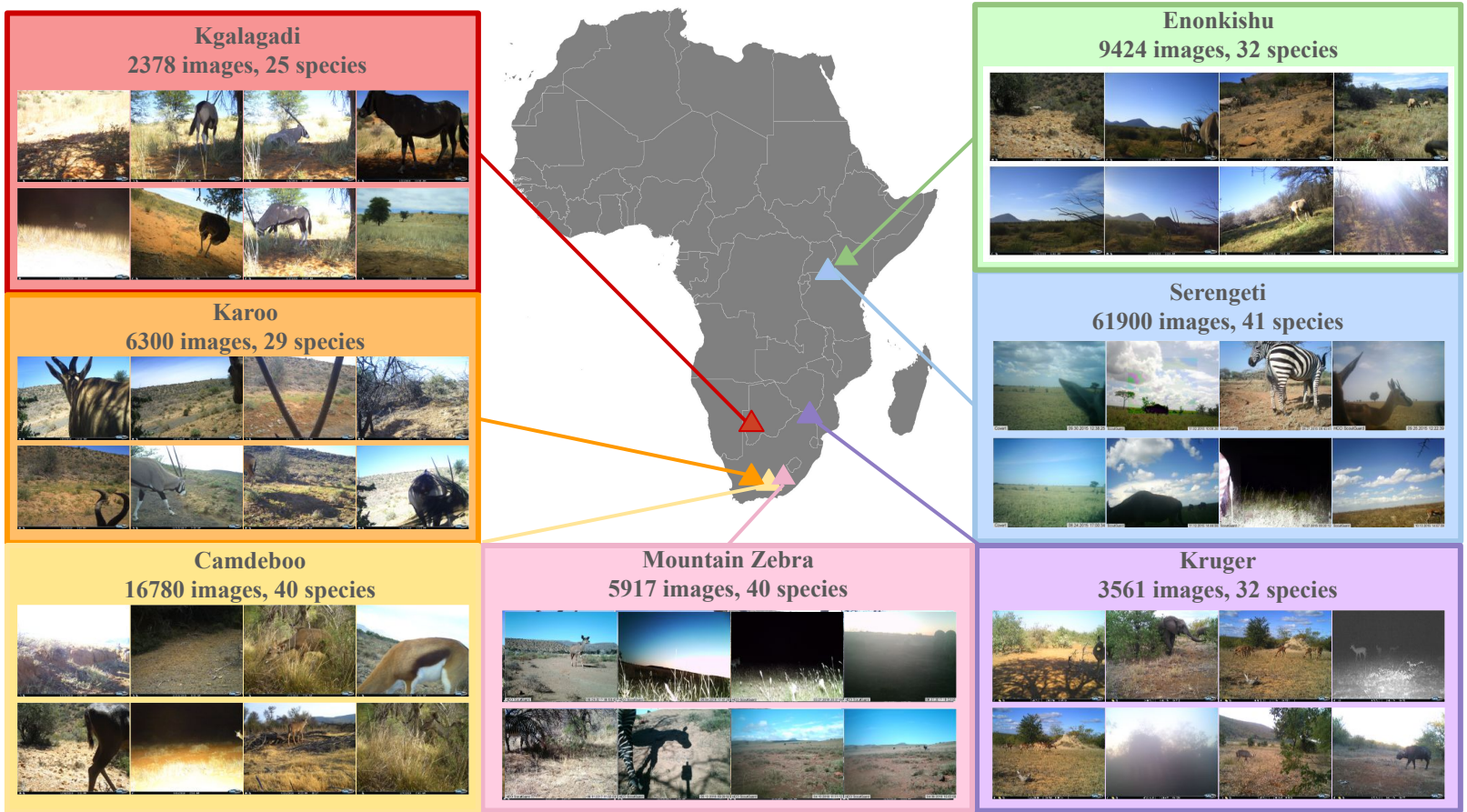
https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms

When does vanilla fine-tuning “fail”?

- Dataset isn’t representative -> bad performance
- Pretraining isn’t useful -> not so bad, maybe inefficient?
- Dataset is too small -> overfitting, lost capacity from the original model



Data Sharing, Resources, and AI - *how, when, and for whom* does sharing data improve ML accuracy and accessibility?



Efficient ways to specialize

- Fine-tuning
- Active learning

Deep active learning to adapt species ID to new projects

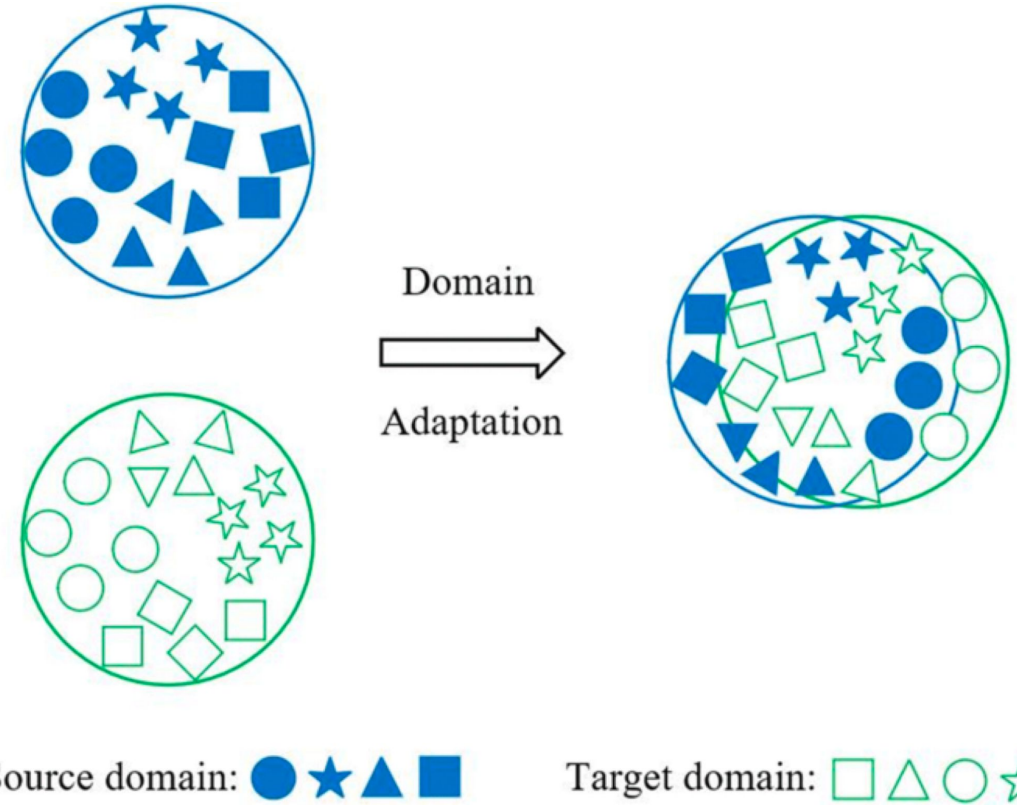
- Uses the MegaDetector to crop
- Cluster animals based on visual similarity in new cameras
- Humans ID examples from each cluster (active learning criteria)
- Gets same accuracy with **99.5% fewer labels**

A deep active learning system for species identification and counting in camera trap images, Norouzzadeh, Morris, Beery, Joshi, Jojic, Clune, Methods in Ecology & Evolution, 2021

Efficient ways to specialize

- Fine-tuning
- Active learning
- Domain adaptation

Supervised domain adaptation



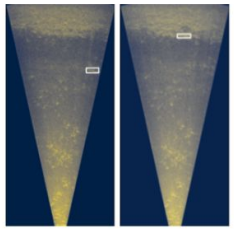
Examples:

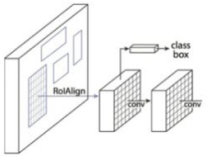
- Co-training
- DANN (adversarial)
- CORAL (correlation alignment)
- ...can be semi-supervised

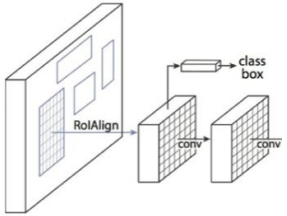
https://link.springer.com/article/10.1007/s10489-022-03709-8

Unsupervised Domain Adaptation

Unlabeled target

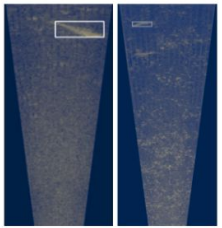


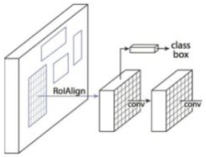




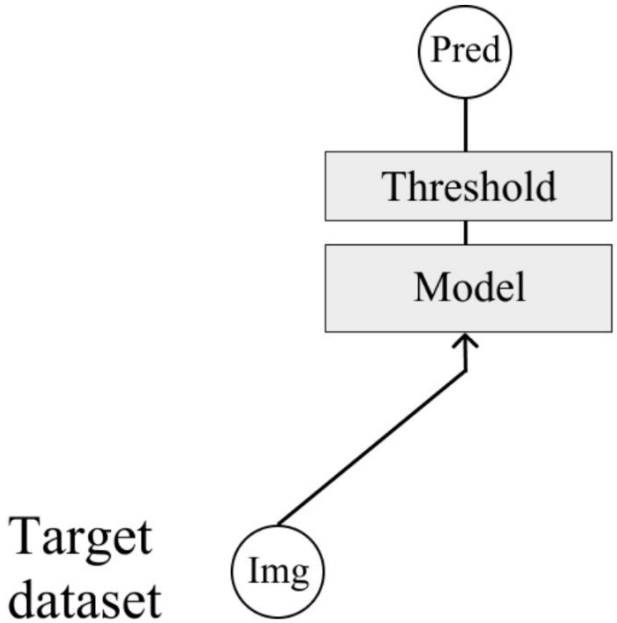
UDA model

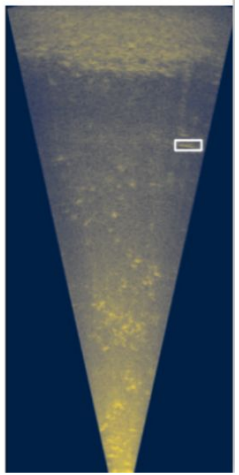
Labeled source



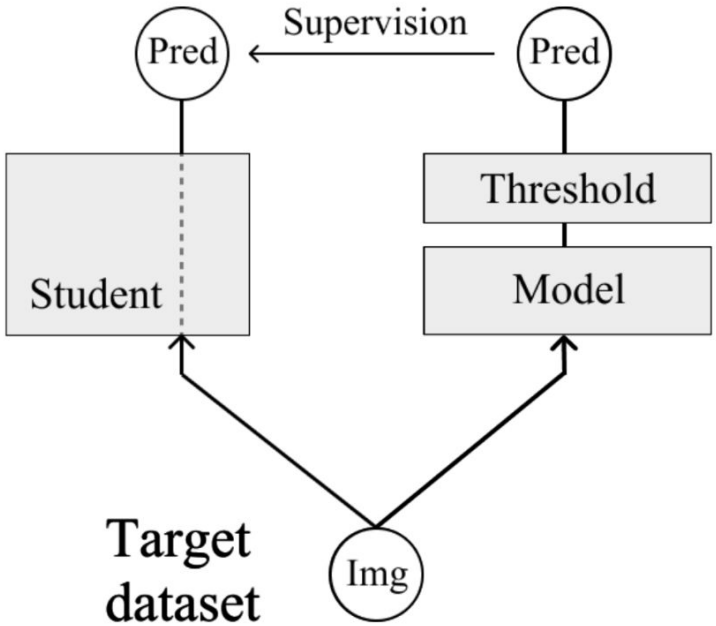


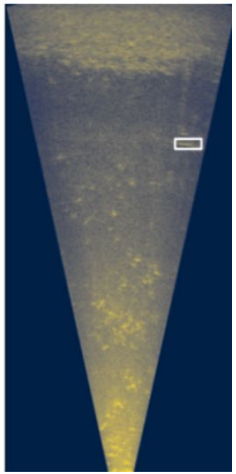
https://aldi-daod.github.io/



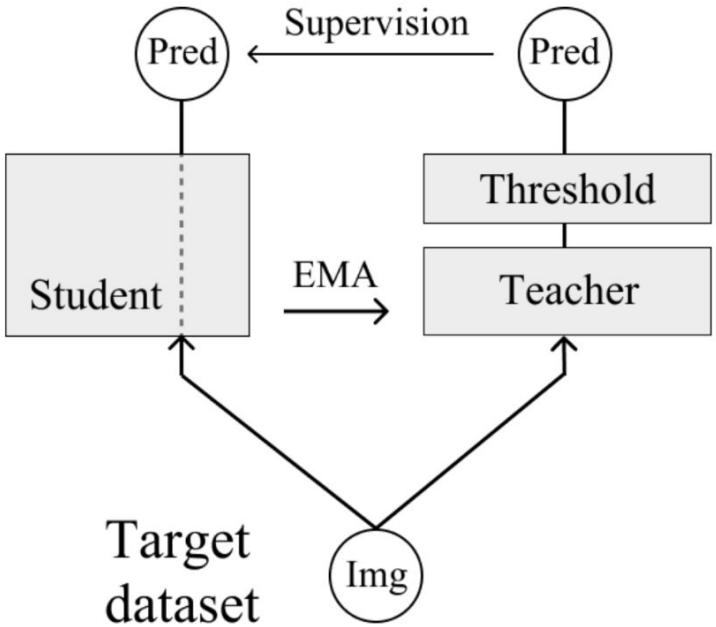


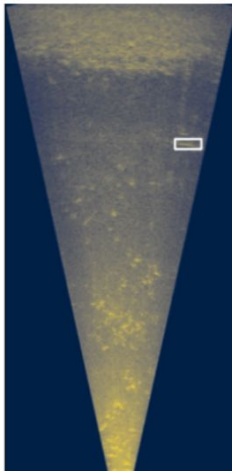
https://aldi-daod.github.io/



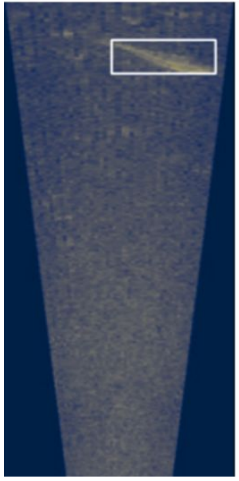


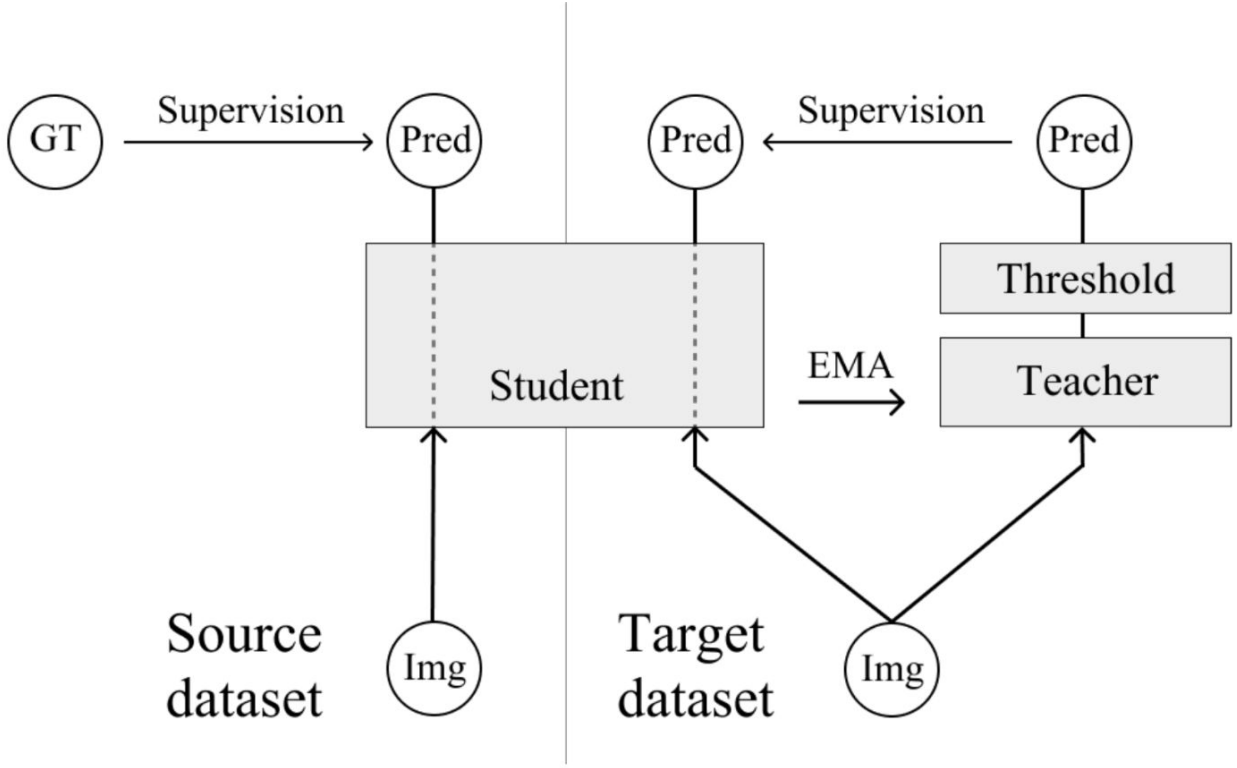
https://aldi-daod.github.io/

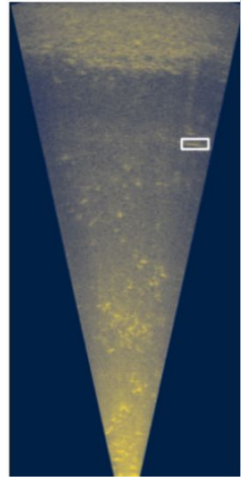




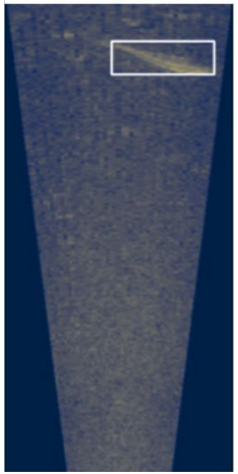
https://aldi-daod.github.io/

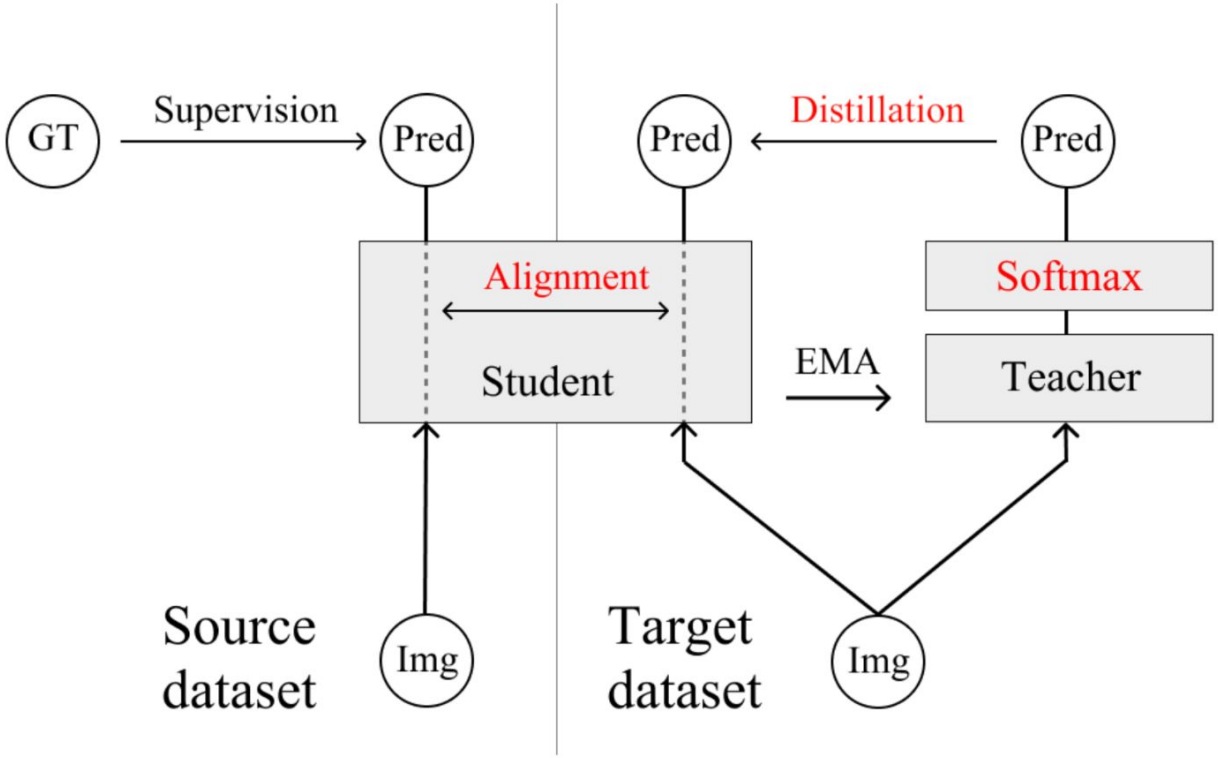


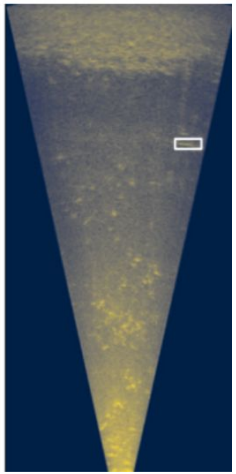




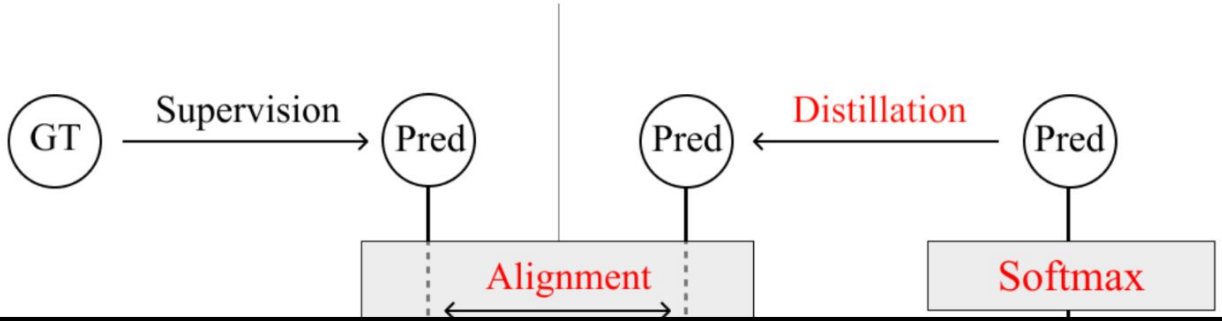
https://aldi-daod.github.io/



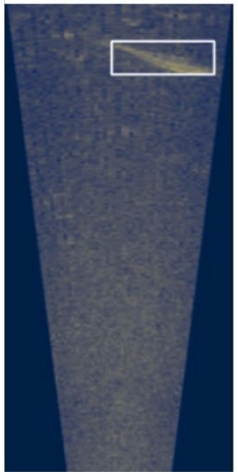


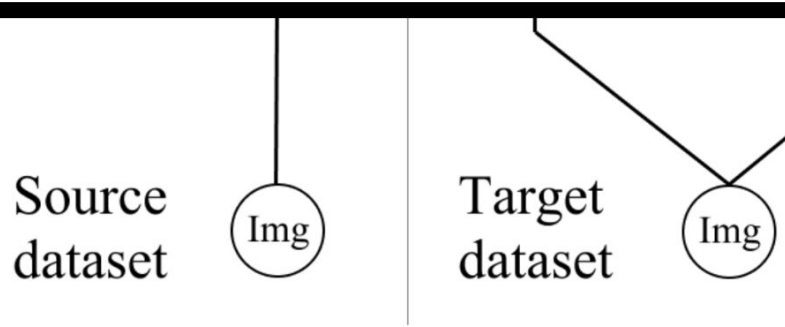


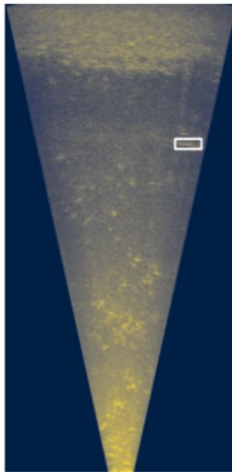
https://aldi-daod.github.io/



How do we know when to stop training?







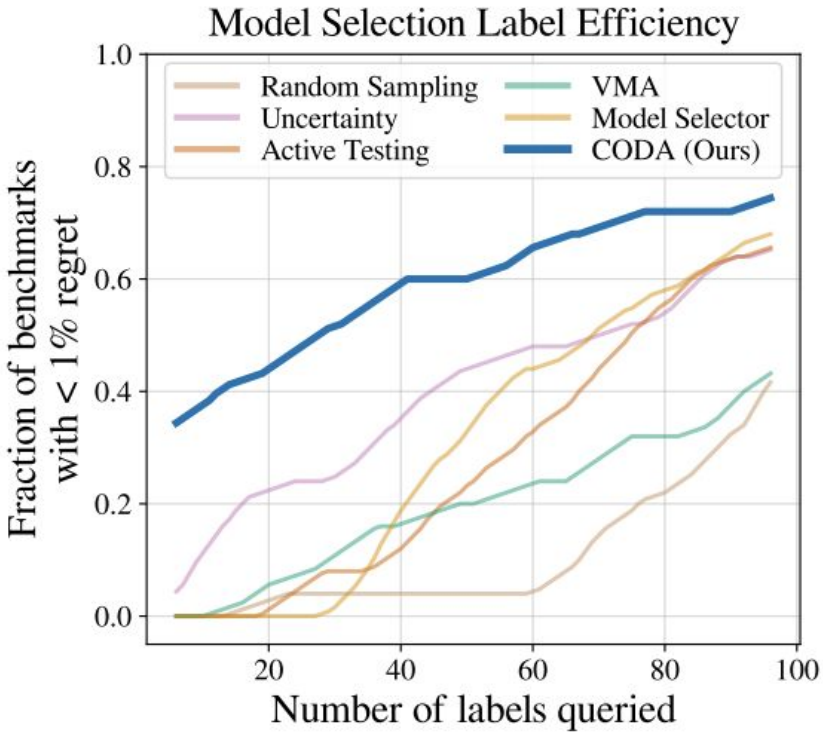
https://aldi-daod.github.io/

Active model selection

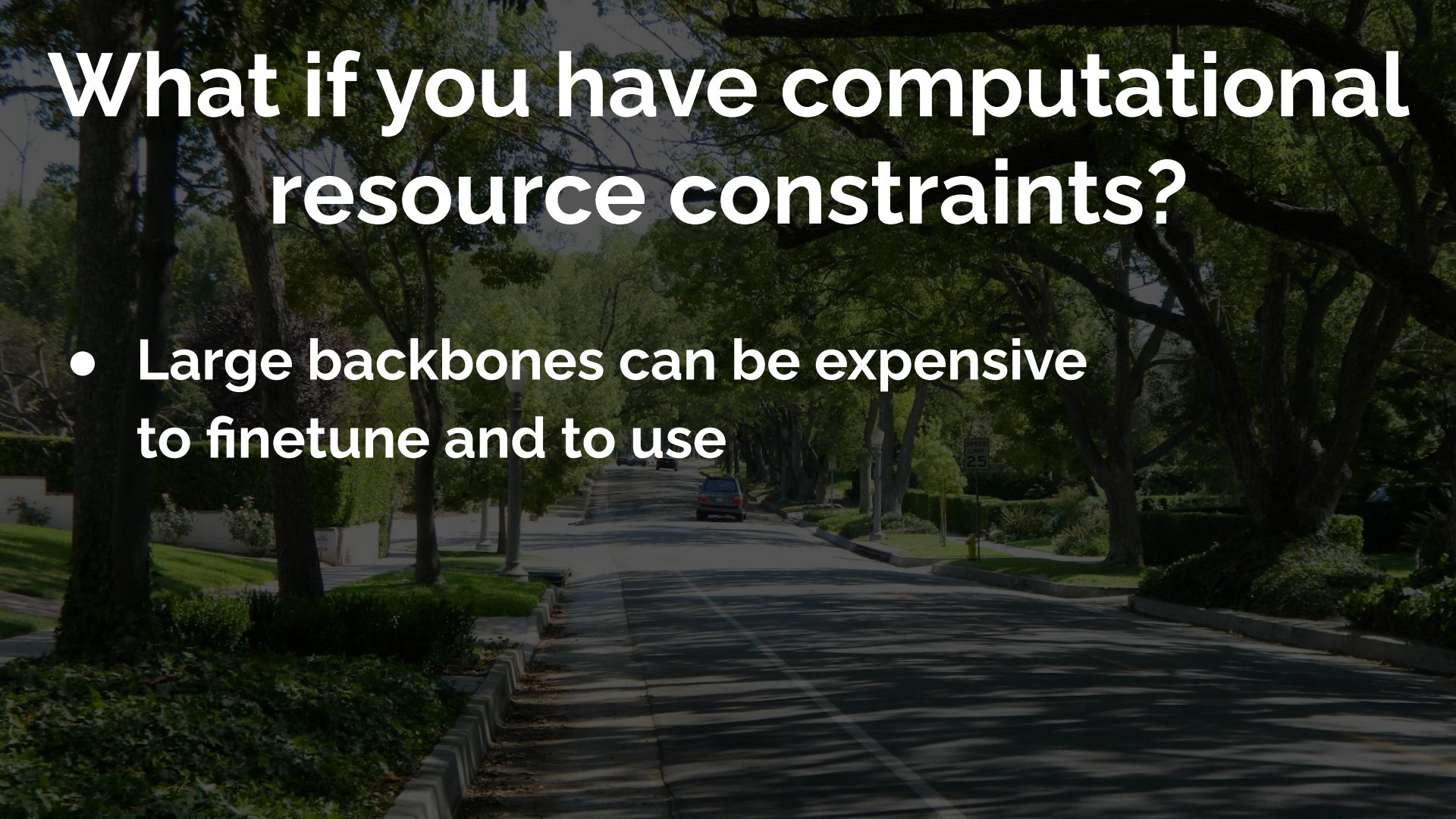
You need to choose a model to use on your data, but don't have labels

- Model zoos
- Across checkpoint runs
- Domain adaptation

How to figure out what model you should use?



Consensus-Driven Active Model Selection, Kay et al. 2025



What if you have computational resource constraints?

- Large backbones can be expensive to finetune and to use

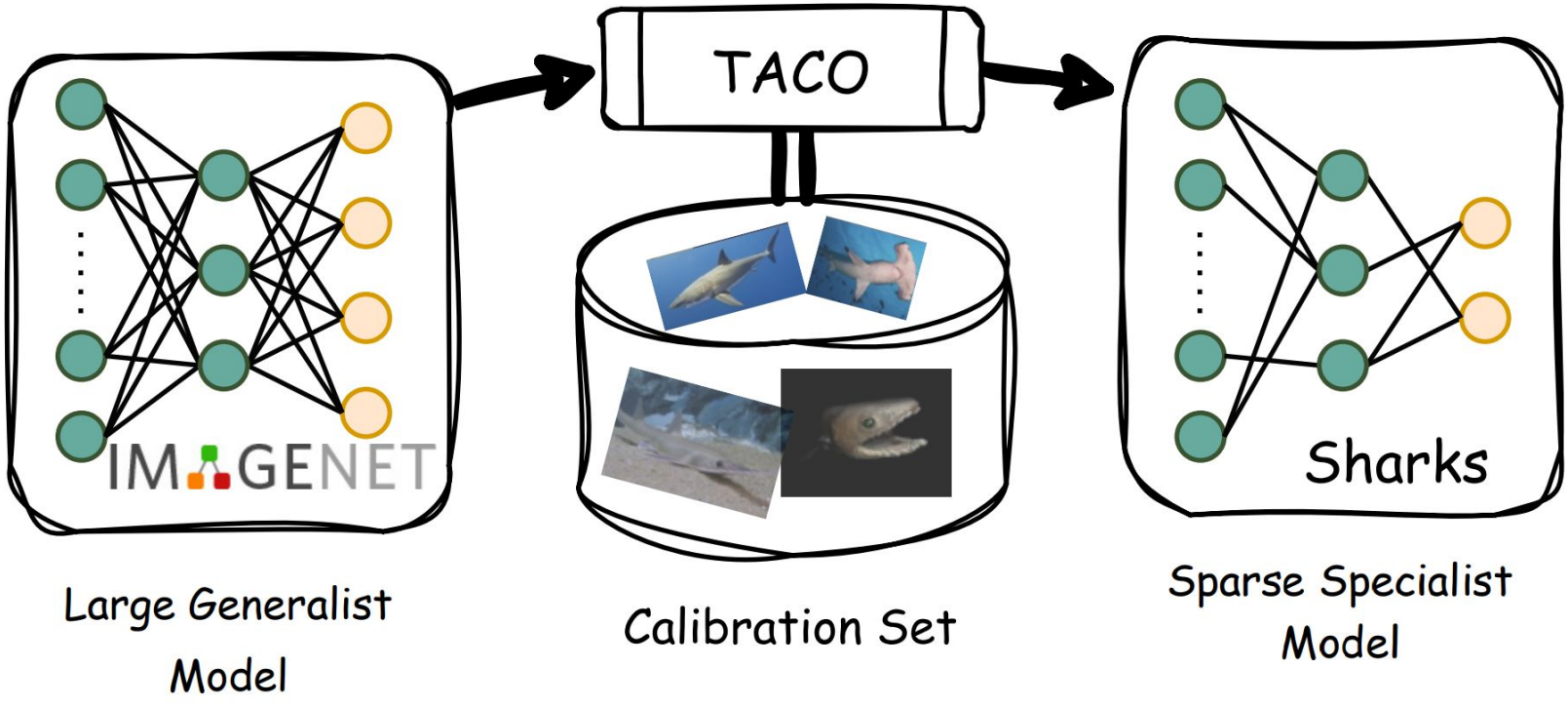
What if you have computational resource constraints?

- Large backbones can be expensive to finetune and to use
- Not all stakeholders have access to GPUs or bandwidth to move data

What if you have computational resource constraints?

- Large backbones can be expensive to finetune and to use
- Not all stakeholders have access to GPUs or bandwidth to move data

Can we quickly compress large generalist models into accurate and efficient specialists?



Institute of Science and Technology Austria

Skoltech
Skolkovo Institute of Science and Technology

Vision Models Can Be Efficiently Specialized via Few-Shot Task-Aware Compression, Kuznedelev,..., Beery, Kurtic, Alistarh et al. 2023

Discussion: what are the tradeoffs between focusing on methods for generalization and specialization in *research*