

Lecture 7: Efficiency in Training, Eval, and Deployment

Sara Beery | 4/1/25

Challenge: AI models require expensive hardware



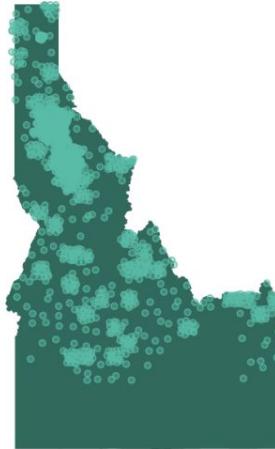
These are inaccessible to many potential AI users in biodiversity



Increased efficiency reduces cost

MegaDetector is used to process data for NGOs and conservation organizations globally

Idaho Dept. of Fish and Game



WOLF
pop. mgmt

2,000
cameras

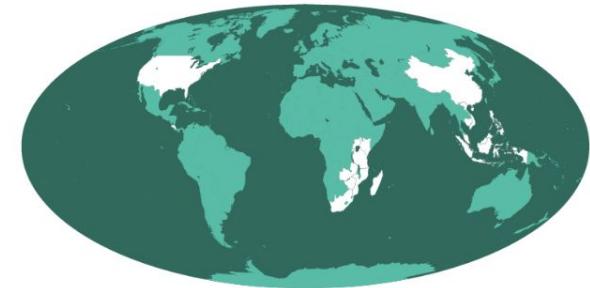
11M
images



Less than 15% of
images require
human review



Wildlife Protection Solutions



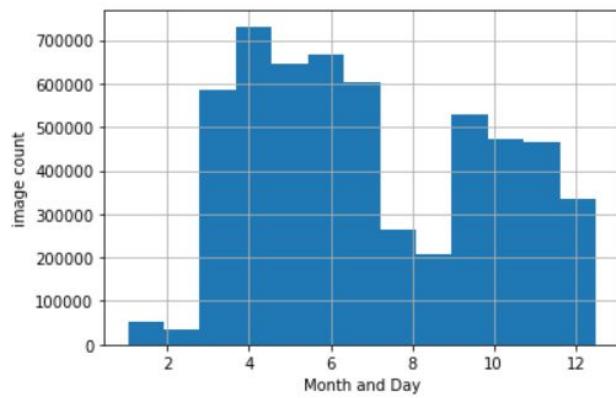
WILDLIFE CRIME PREVENTION

18 nations | 800 cameras | 900K images

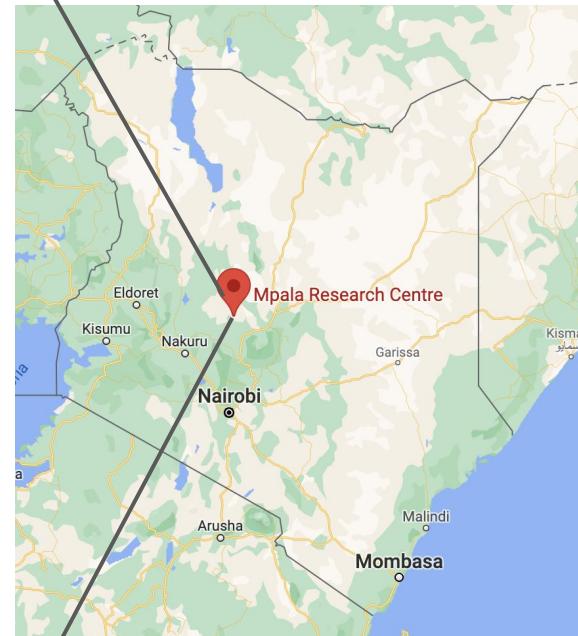
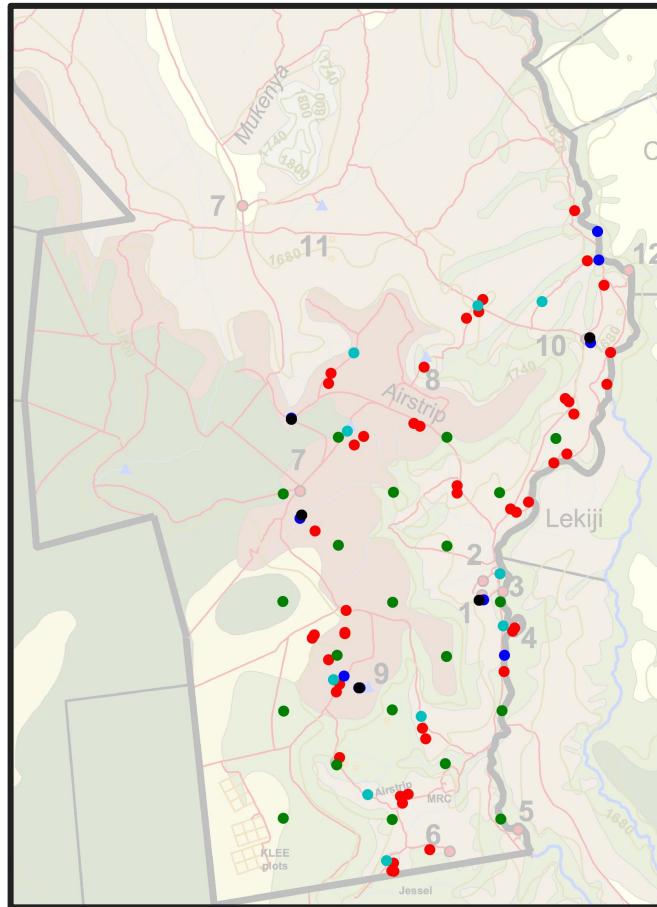
Real-time alerts
Detects one real wildlife threat per week on average



Bandwidth is limited in the field



Up to 700K
hi-resolution
images per
month



Commercial edge-based AI camera traps in development



Instant Detect



Focused on
anti-poaching

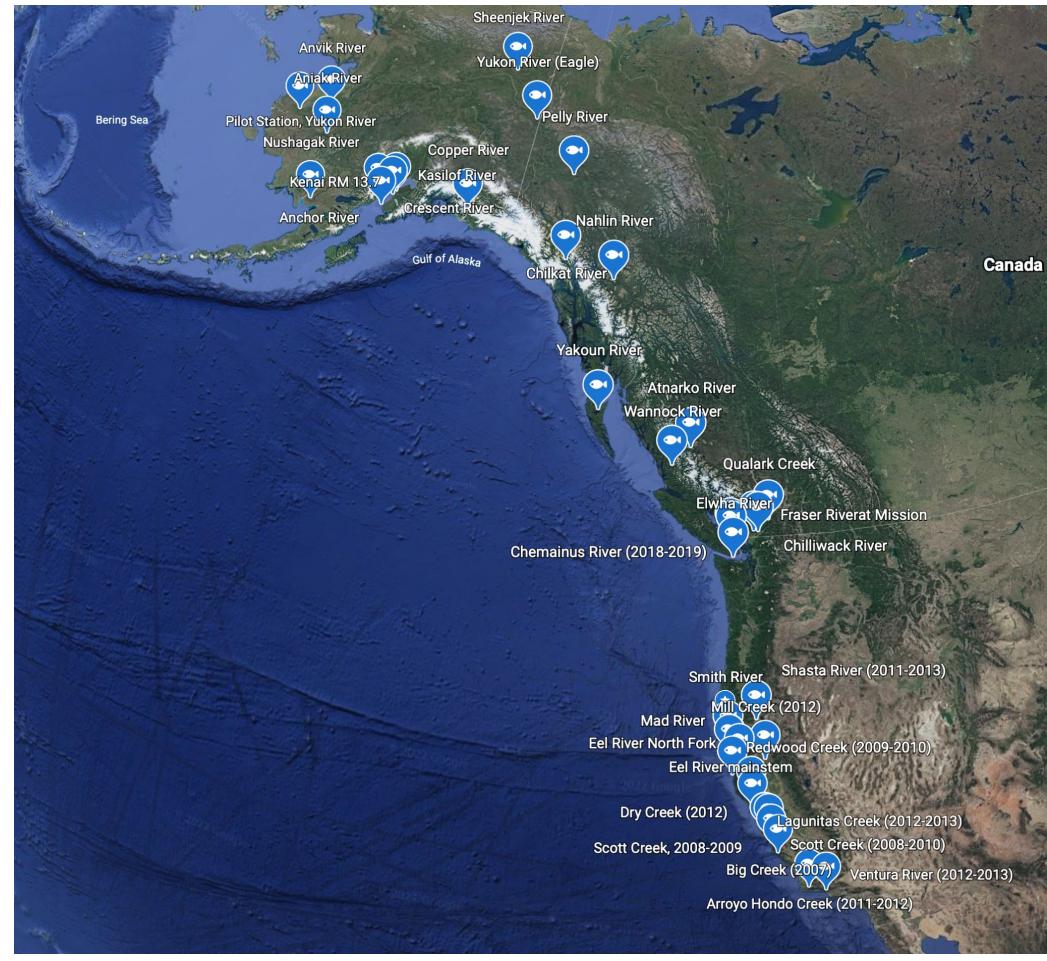


Focused on
anti-poaching &
human-wildlife
conflict



Focused on real-time
animal behavior
monitoring

Sonar deployment to monitor salmon returns

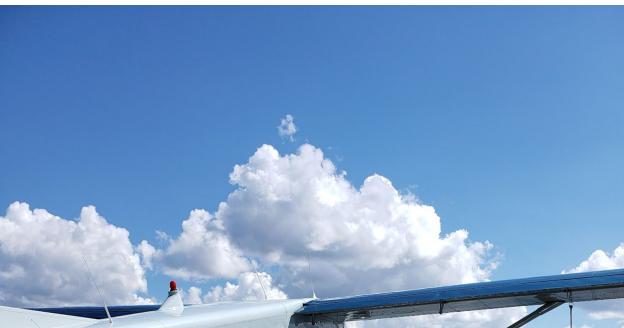


Photos courtesy of ADFG

We need near-real-time counts from remote field sites



We need near-real-time counts from remote field sites



This requires edge-based models that are robust and reliable even as environmental conditions change





www.inaturalist.org



iNaturalist is a joint initiative of the
California Academy of Sciences and the
National Geographic Society.

How It Works



1

Record your observations



2

Share with fellow naturalists



3

Discuss your findings



Observations



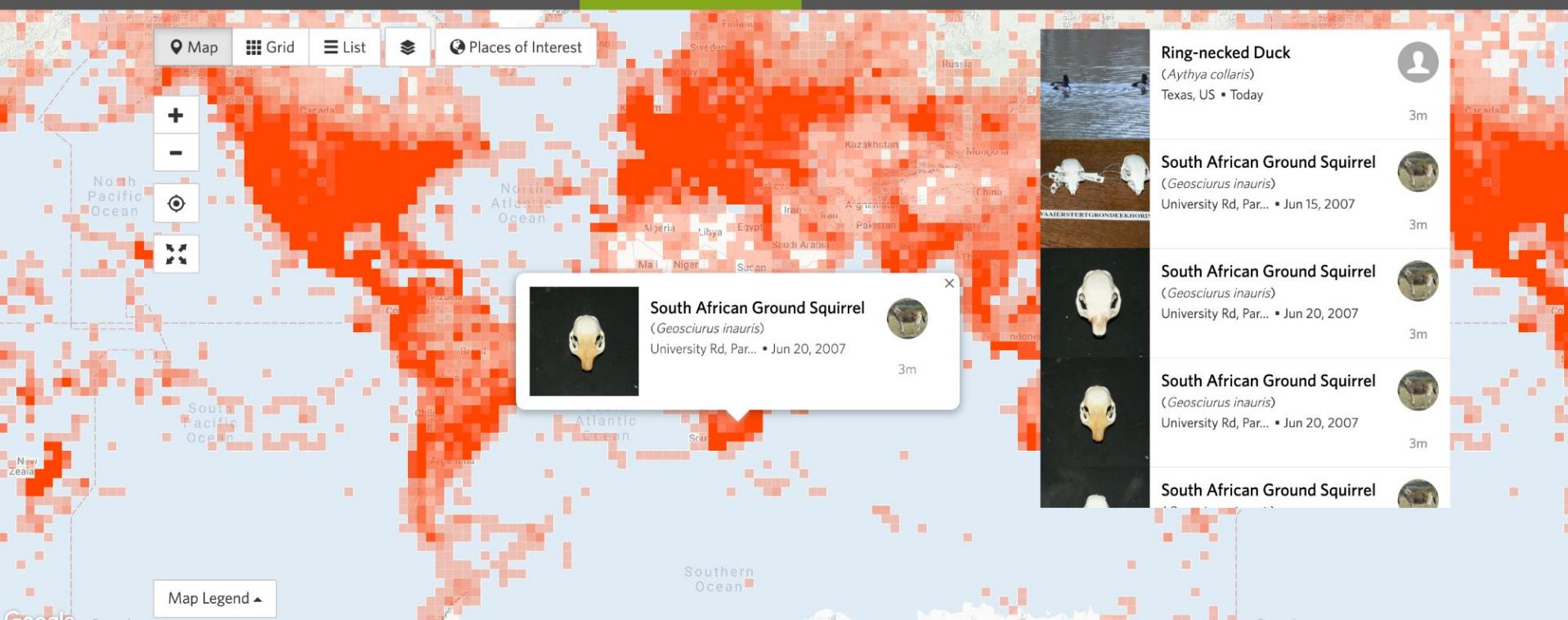
Species

Location

Go

Filters

The World

90,060,114
OBSERVATIONS344,629
SPECIES234,007
IDENTIFIERS2,015,371
OBSERVERS

Real-time, on-device fine grained categorization

The image shows the Seek app interface on three mobile devices. The left device displays the app's main landing page with a green background featuring leaf patterns, the "seek" logo, and the text "Get outside, explore, and learn about the nature all around you!". Logos for the California Academy of Sciences and National Geographic are also present. The middle device shows the "SPECIES" screen for a Monarch butterfly, with a camera icon at the bottom. The right device shows the "SPECIES NEARBY" screen for San Francisco, listing "California Poppy", "Western Blue-eyed Grass", and "Miner's Lettuce". It also features "CHALLENGES" and "ACHIEVEMENTS" sections. The achievements section highlights the "SURVEYOR" level, which requires observing 50 species.

seek by iNaturalist

Get outside, explore, and learn about the nature all around you!

CALIFORNIA ACADEMY OF SCIENCES

NATIONAL GEOGRAPHIC

SPECIES

Monarch

That's it! Now take a picture to add this species to your observations.

CAMERA PHOTOS

SPECIES NEARBY

SAN FRANCISCO

ALL SPECIES

California Poppy

Western Blue-eyed Grass

Miner's Lettuce

CHALLENGES

APRIL 2019 CONNECTIVITY CHALLENGE

OUR PLANET seek Camera to earn

ACHIEVEMENTS

YOUR LEVEL SURVEYOR

Observe 50 species to get to the next level!

SPECIES BADGES

Training efficiency

- Fine tuning often decreases training time, and thus costs
- Reducing sample overlap or removing noisy training data can reduce training costs without impacting performance, or sometimes improving performance (Coresets, DataComp)
- Smaller models train faster

Training on the edge

- Power
- Hardware
- Bandwidth
- Verification

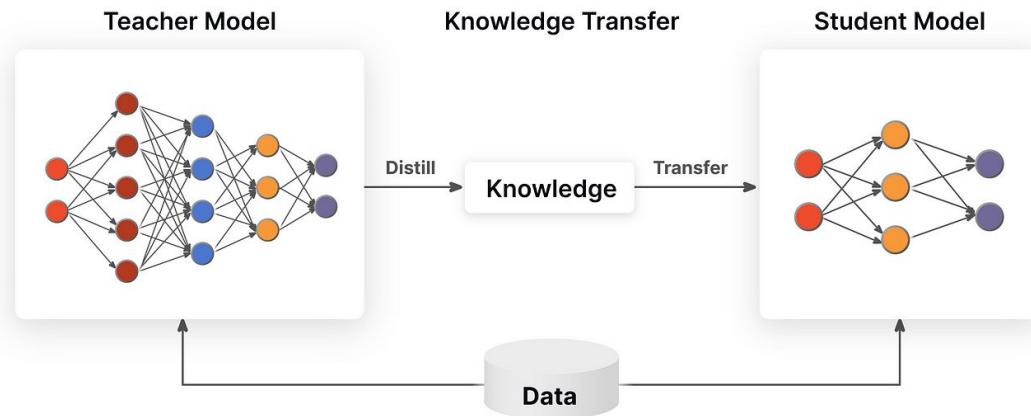
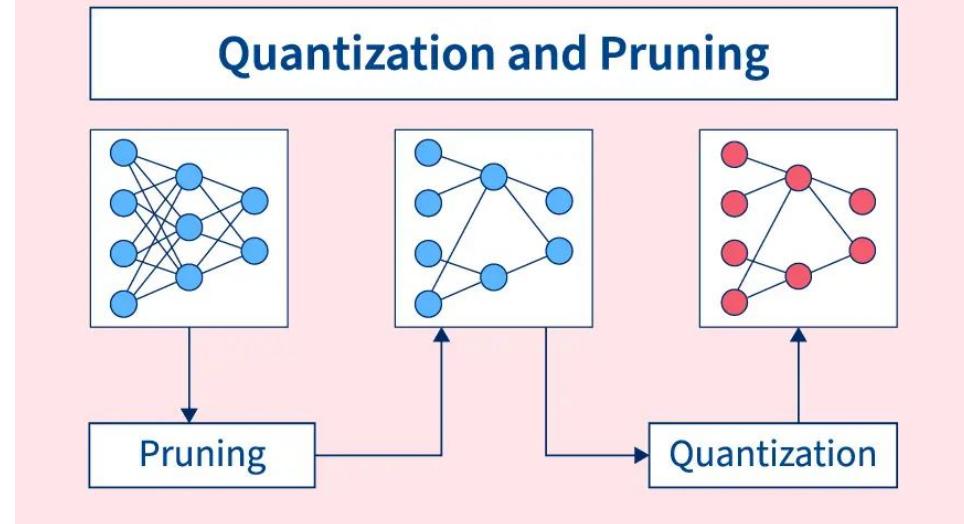


Evaluation efficiency

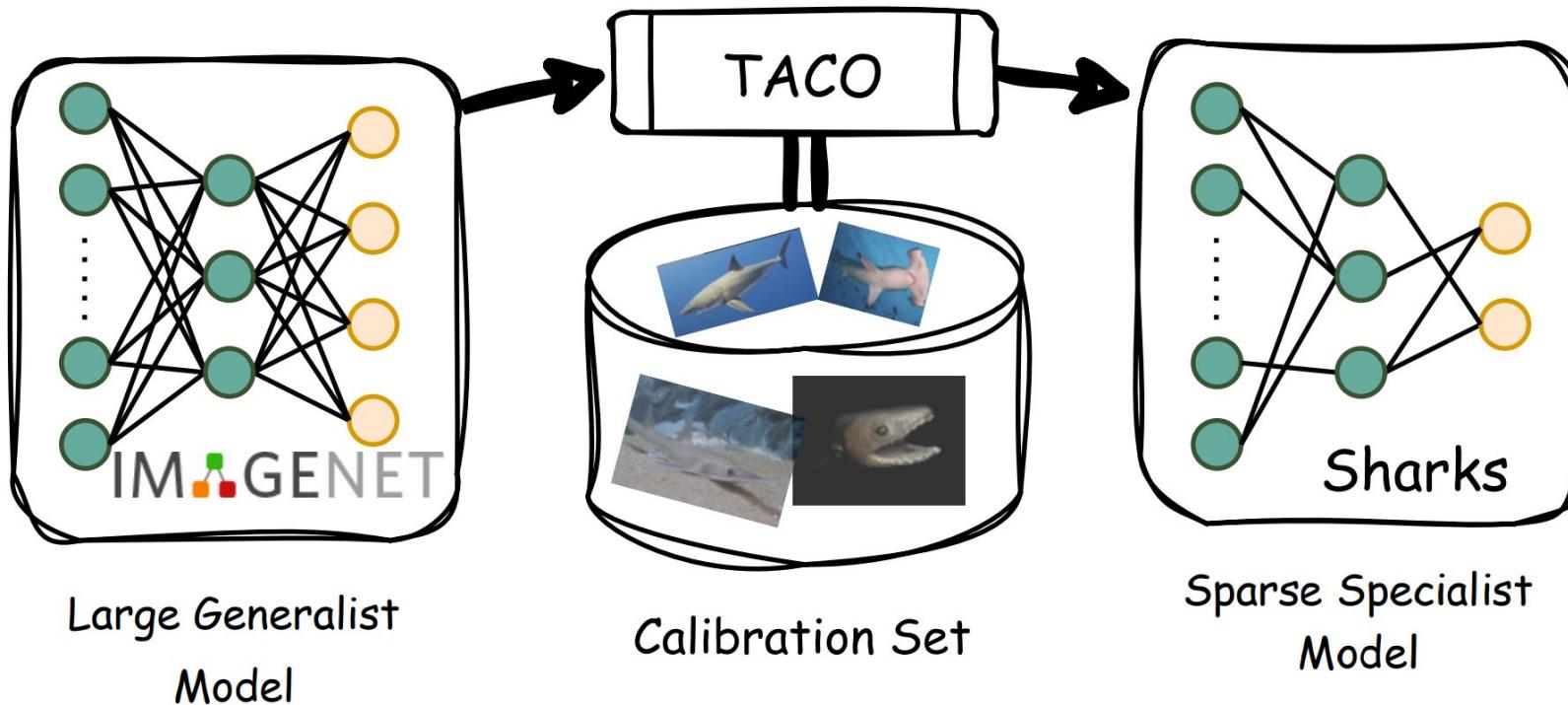
- Inexpensive proxy tests (ie brightness)
- Small (but representative) test sets
- Striation and multiple metrics
 - Make the most of your inference calls
- Active testing (soon!)

Inference efficiency

- Quantization
- Pruning
- Distillation
- Routing



We can quickly compress large generalist models into accurate and efficient specialists



Federated learning

- Maintains data privacy
- Can be efficient at the edge
- Requires bandwidth and synchronization

