Sara Beery

CompSust Open Graduate Seminar
April 3rd, 2020

# Improving Computer Vision for Camera Traps

## Leveraging Practitioner Insight to Build Solutions for Real-World Challenges

Big goal: monitoring biodiversity, globally and in real time.

# Big goal: monitoring biodiversity, globally and in real time.

# How can we contribute?

# Camera traps


# Camera traps

- 1,000s of organizations
- 10,000s of projects
- 1,000,000s of camera traps
- 100,000,000s of images

*estimates by Eric Fegraus, Conservation International

# Camera traps

- 1,000s of organizations
- 10,000s of projects
- 1,000,000s of camera traps
- 100,000,000s of images

For example: Idaho Department of Fish and Game alone has 5 years of unprocessed, unlabeled data, around 5 million images

*estimates by Eric Fegraus, Conservation International

# Camera trap data is challenging

(1) Illumination

(2) Blur

(3) ROI Size

(4) Occlusion

(5) Camouflage

(6) Perspective

# All these images have an animal in them

(1) Illumination

(2) Blur

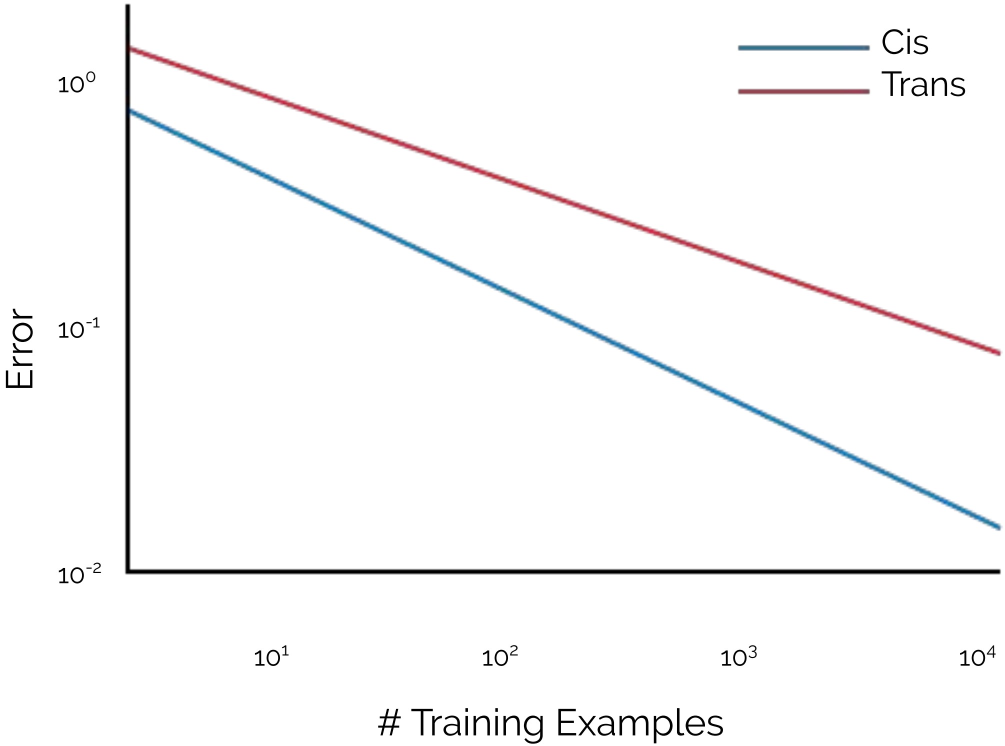
(3) ROI Size

(4) Occlusion

(5) Camouflage

(6) Perspective

# SOA models don’t generalize






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

# Class-agnostic detectors generalize best

## MegaDetector

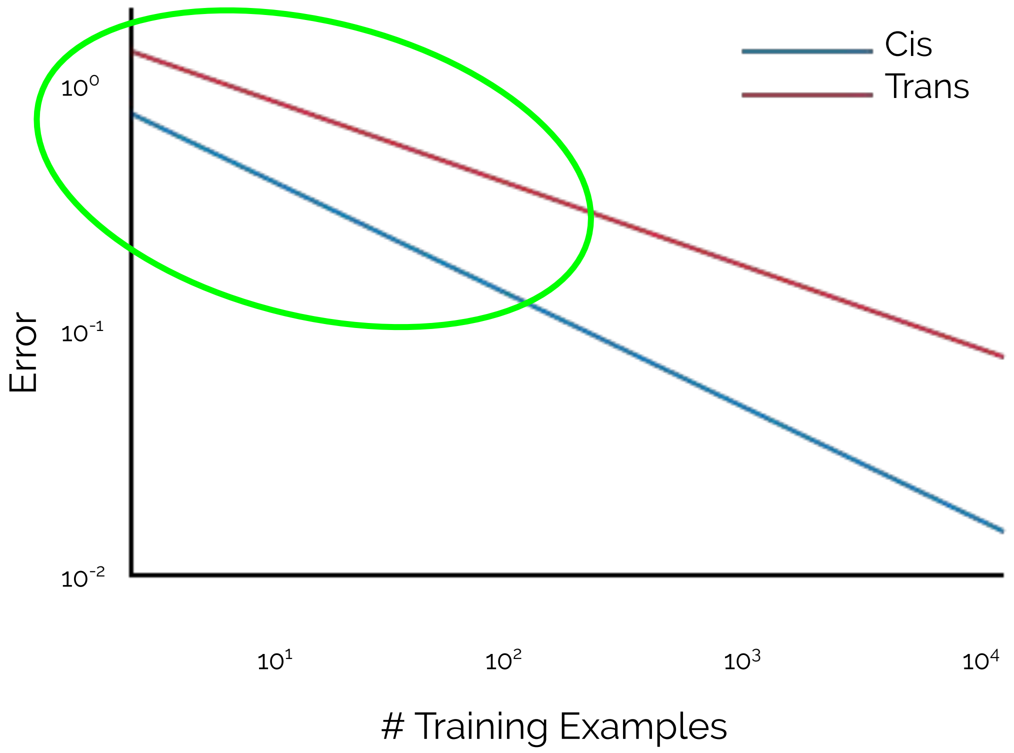
**Microsoft AI for Earth**


*Efficient Pipeline for Automating Species ID in new Camera Trap Projects,* Beery, et al., BiodiversityNext 2019
https://github.com/microsoft/CameraTraps/blob/master/megadetector.md


Sorted 4.8 million images in ~2.75 days

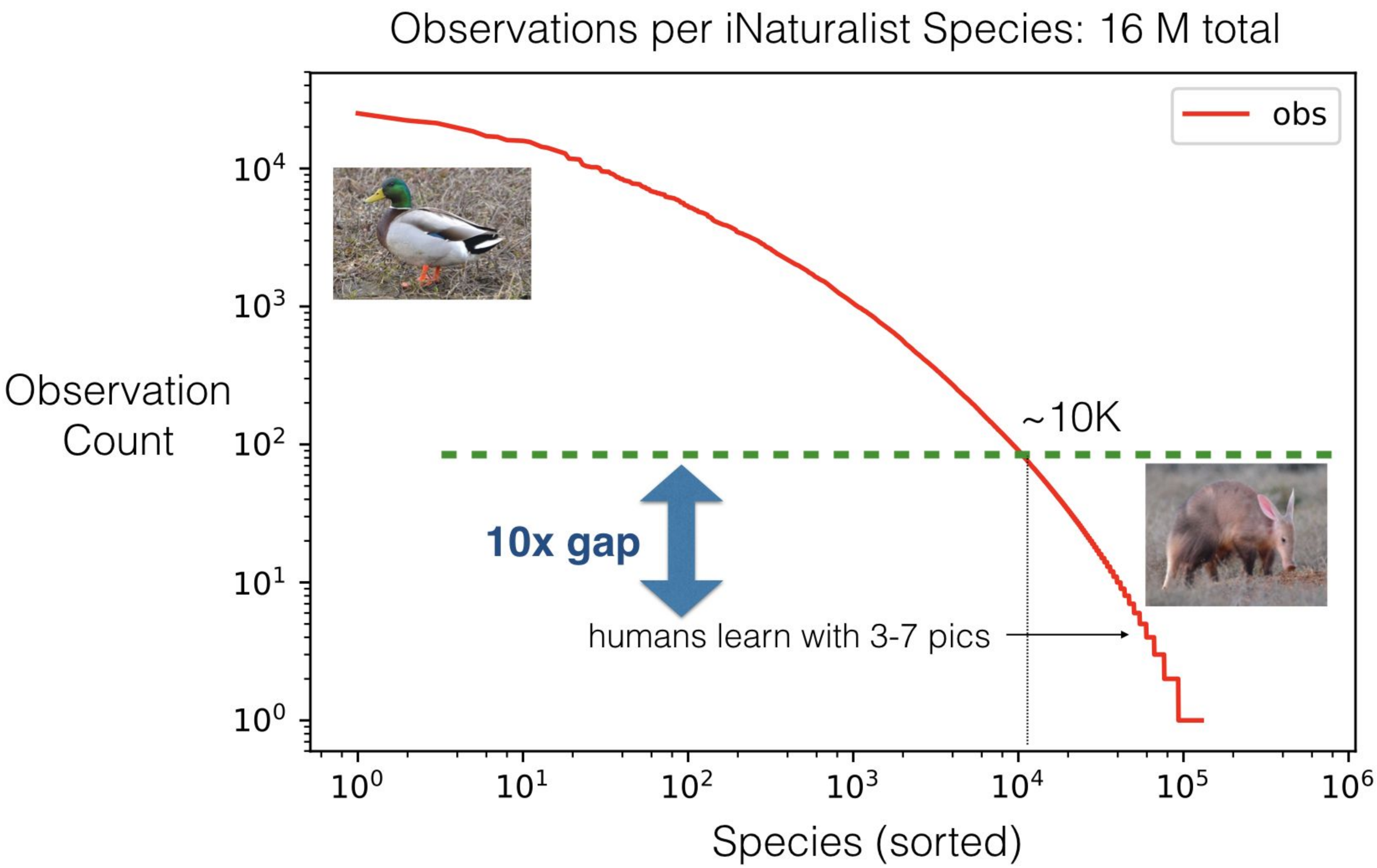
This would have taken 10 people working full-time 40 weeks to complete

# Rare classes are hard





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


Observations per iNaturalist Species: 16 M total
obs
Observation Count
$10^0$
$10^1$
$10^2$
$10^3$
$10^4$
~10K
10x gap
humans learn with 3-7 pics
$10^0$ $10^1$ $10^2$ $10^3$ $10^4$ $10^5$ $10^6$
Species (sorted)

# E.g. learning pose variability

# Camera traps are static, and objects of interest are habitual


# Synthetic data improves rare-class performance

(f) Real Camera Traps (g) TrapCam-Unity (h) TrapCam-AirSim (i) Sim on Empty (j) Real on Empty

*Synthetic Examples Improve Generalization for Rare Classes,* Beery et al., WACV 2020

# Camera traps are static, and objects of interest are habitual


# Human labeling method

# Human labeling method


# Human labeling method

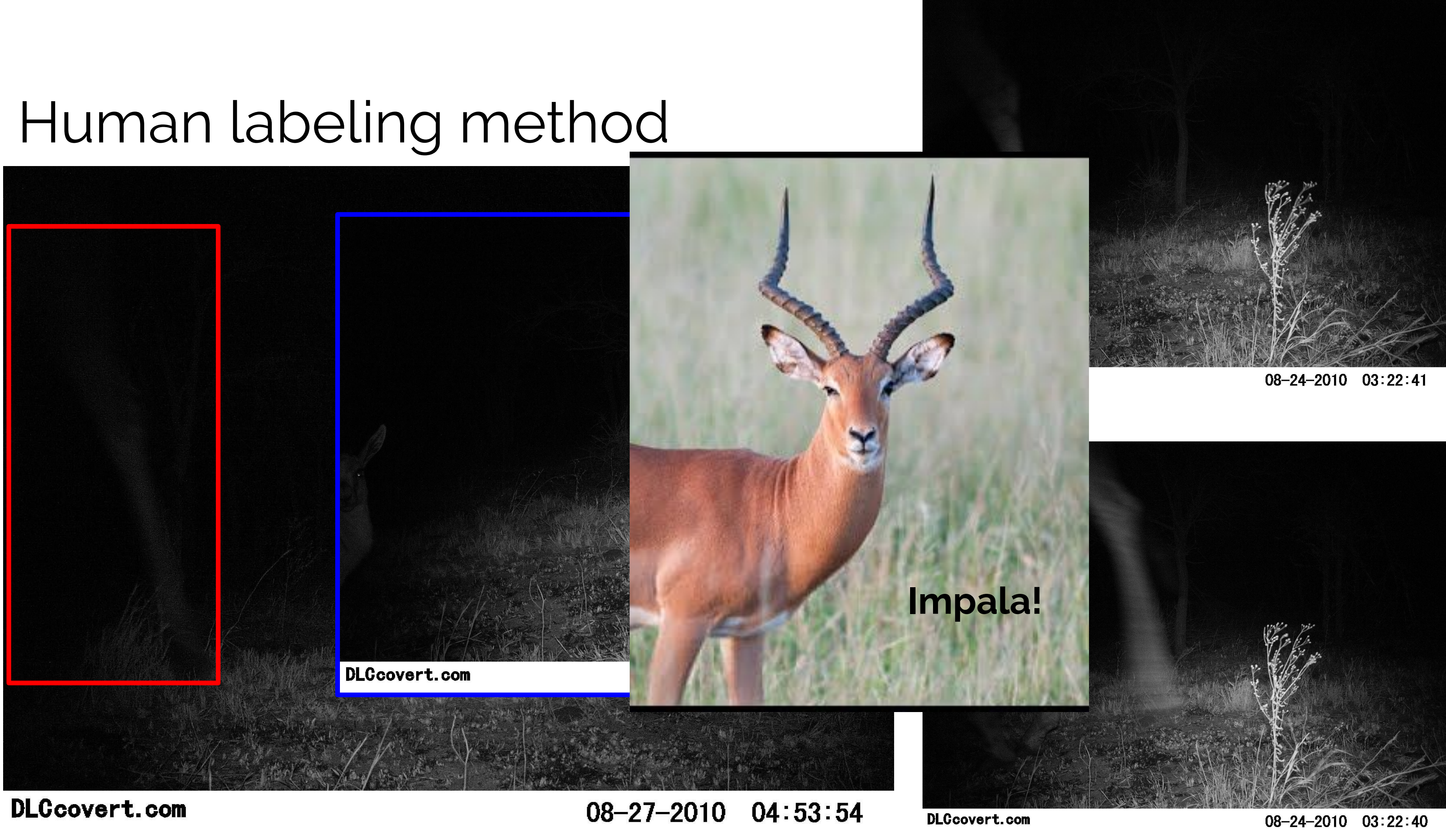


# Human labeling method



# Human labeling method

# Human labeling method



# Camera traps are static, and objects of interest are habitual

Human practitioners use this information, can we build a machine learning model that can do the same?

*Context R-CNN: Long Term Context for Per-Camera Object Detection,* Beery et al., CVPR 2020


# Camera traps are static, and objects of interest are habitual

1. Improve per-location object classification



These are probably the same species, and if we're confident about one, that should help us classify the other

# Camera traps are static, and objects of interest are habitual

1. Improve per-location object classification
2. Ignore salient false positives



These rocks have not moved in a month, they're probably not animals.

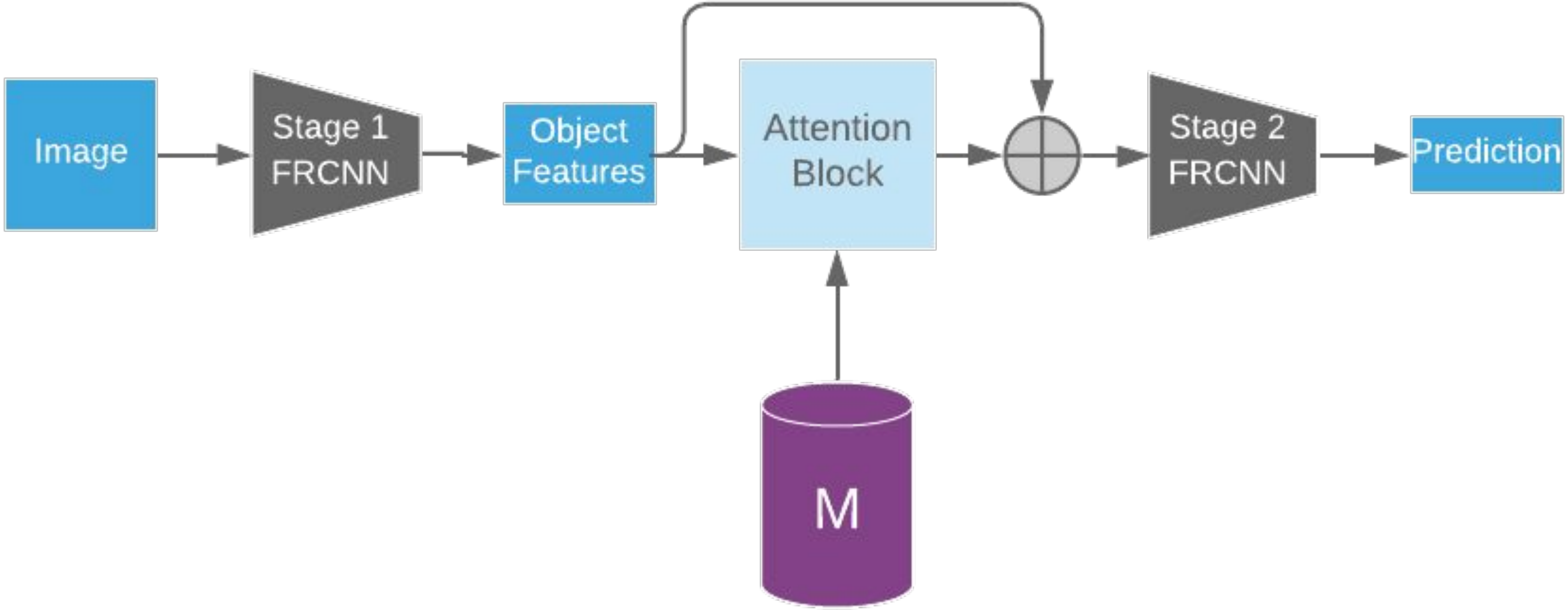
# Contextual memory strategy

- Extract features offline
- Reduce feature size
- Curate features
- Maintain spatiotemporal information



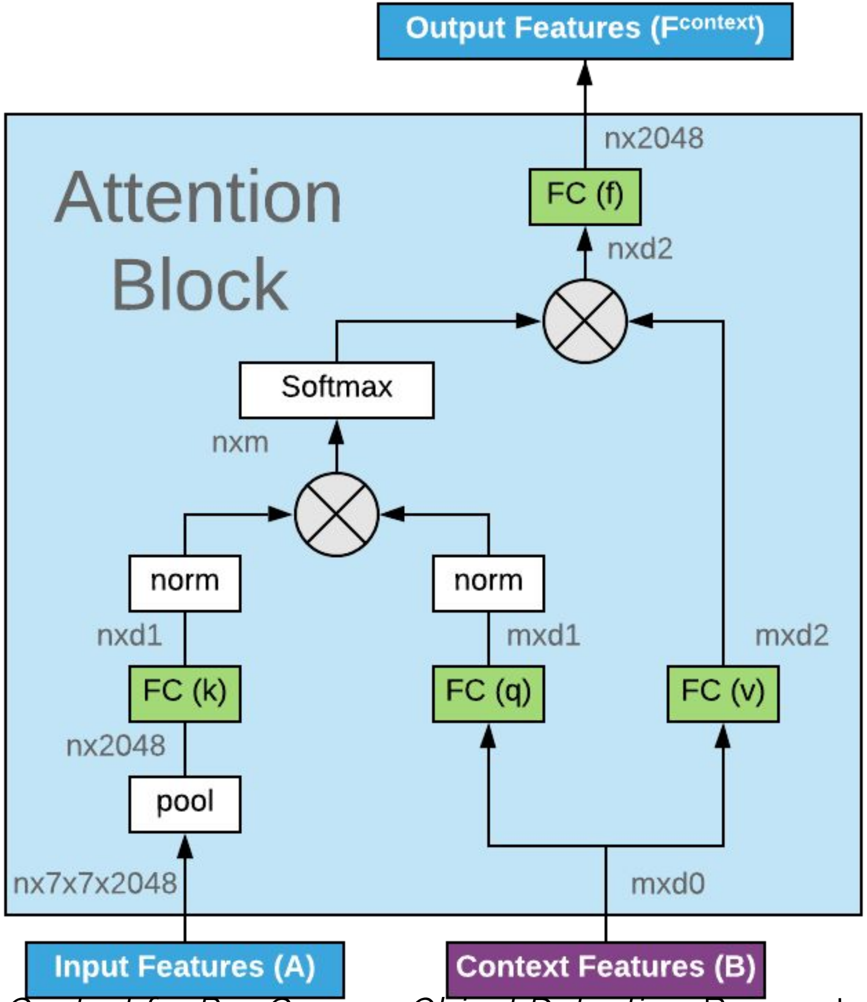

*Context R-CNN: Long Term Context for Per-Camera Object Detection*, Beery et al., CVPR 2020

# Use attention to incorporate context



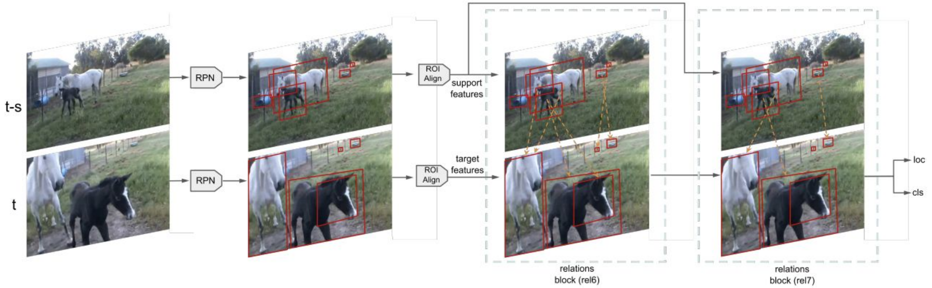

*Context R-CNN: Long Term Context for Per-Camera Object Detection*, Beery et al., CVPR 2020

Context is incorporated based on relevance

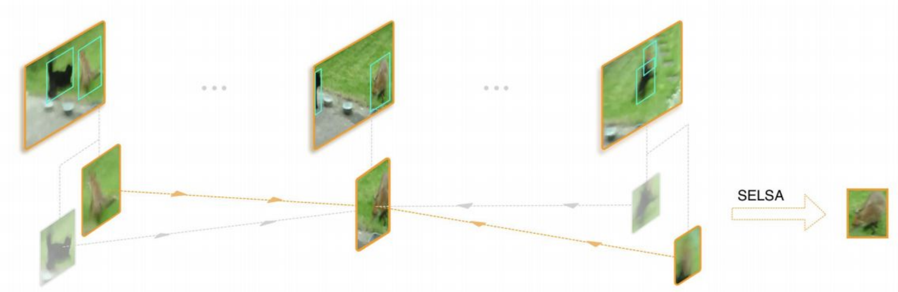



*Context R-CNN: Long Term Context for Per-Camera Object Detection*, Beery et al., CVPR 2020

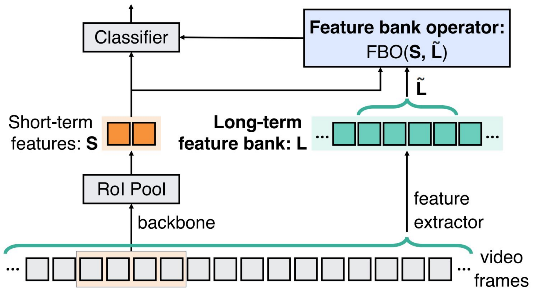
# Related Work: long-term temporal context in video



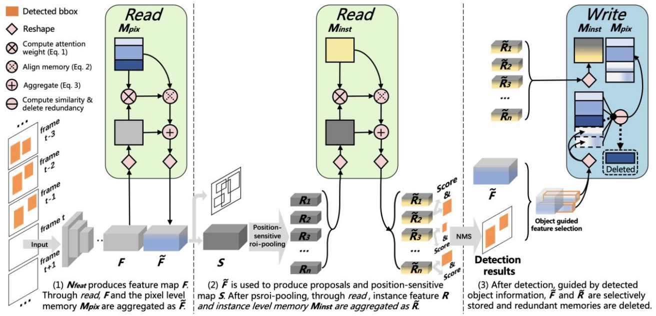

Shvets et al., *Leveraging Long-Range Temporal Relationships Between Proposals for Video Object Detection*




Wu et al., *Sequence Level Semantics Aggregation for Video Object Detection*




Wu et al., *Long-Term Feature Banks for Detailed Video Understanding*




Deng et al., *Object Guided External Memory Network for Video Object Detection*

# Datasets

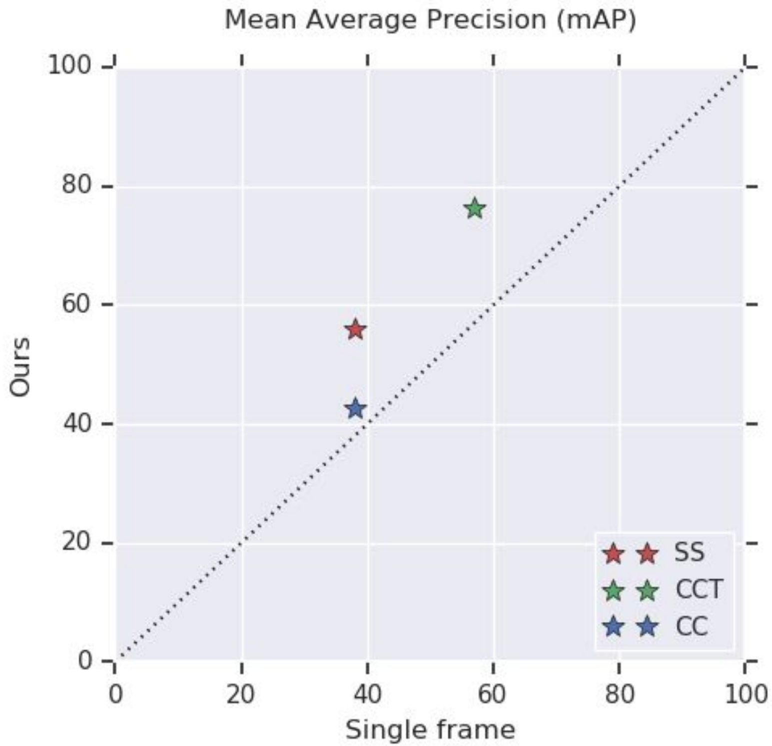
- **Snapshot Serengeti (SS):** 225 cameras, 3.4M images, 48 classes, Eastern African game preserve
- **Caltech Camera Traps (CCT):** 140 cameras, 243K images, 18 classes, American Southwestern urban wildlife
- **CityCam (CC):** 17 cameras, 60K images, 10 vehicle classes, traffic cameras from NYC

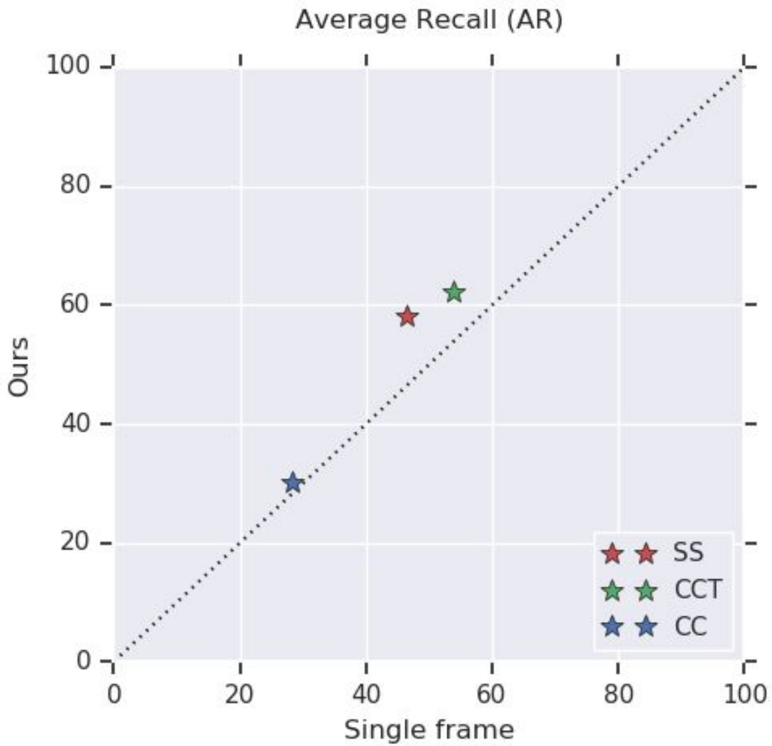

*Context R-CNN: Long Term Context for Per-Camera Object Detection*, Beery et al., CVPR 2020

# Results

**SS:** Snapshot Serengeti
**CCT:** Caltech Camera Traps
**CC:** CityCam

| | SS | | CCT | | CC | |
|---|---|---|---|---|---|---|
| Model | mAP | AR | mAP | AR | mAP | AR |
| Single Frame | 37.9 | 46.5 | 56.8 | 53.8 | 38.1 | 28.2 |
| **Ours** | **55.9** | **58.3** | **76.3** | **62.3** | **42.6** | **30.2** |



# Improves predominantly on challenging cases

(a) Object moving out of frame.

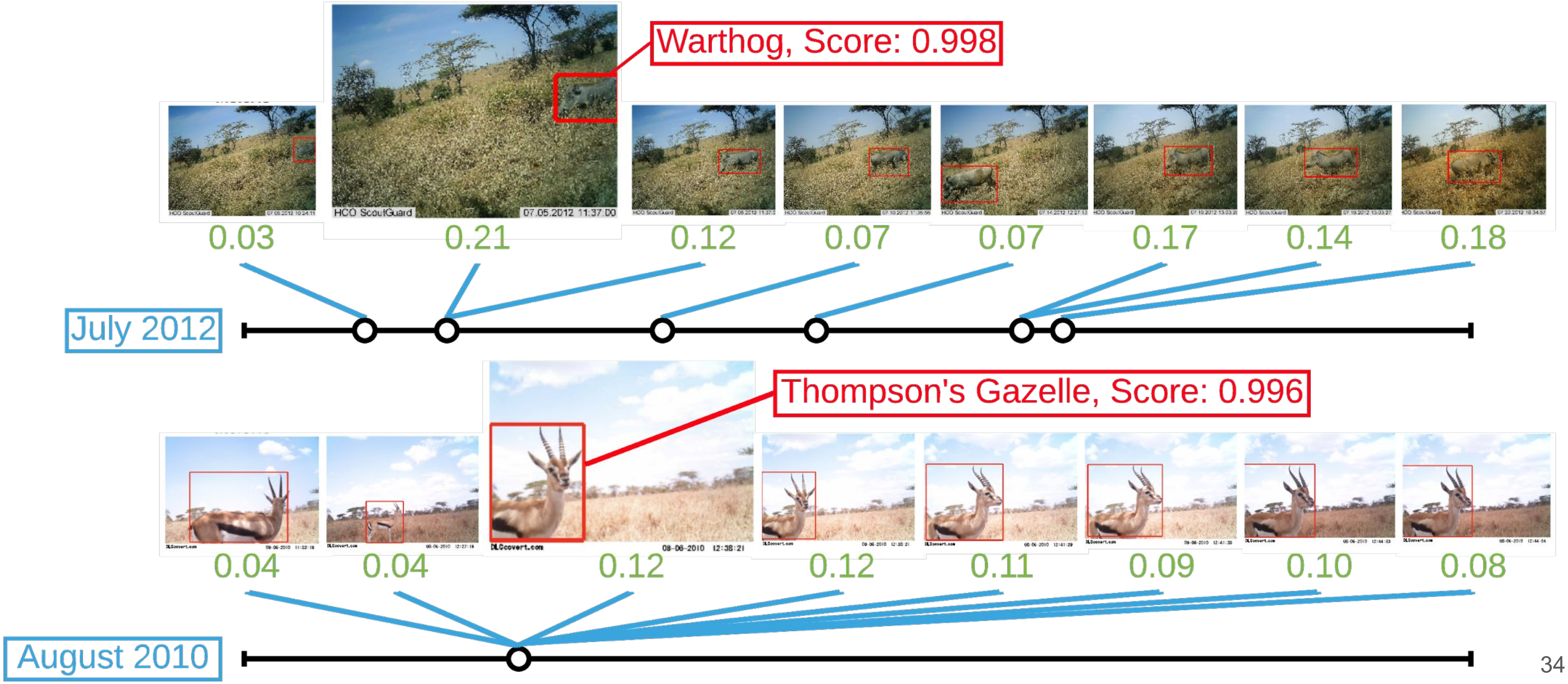
(b) Object highly occluded.

(c) Object far from camera.

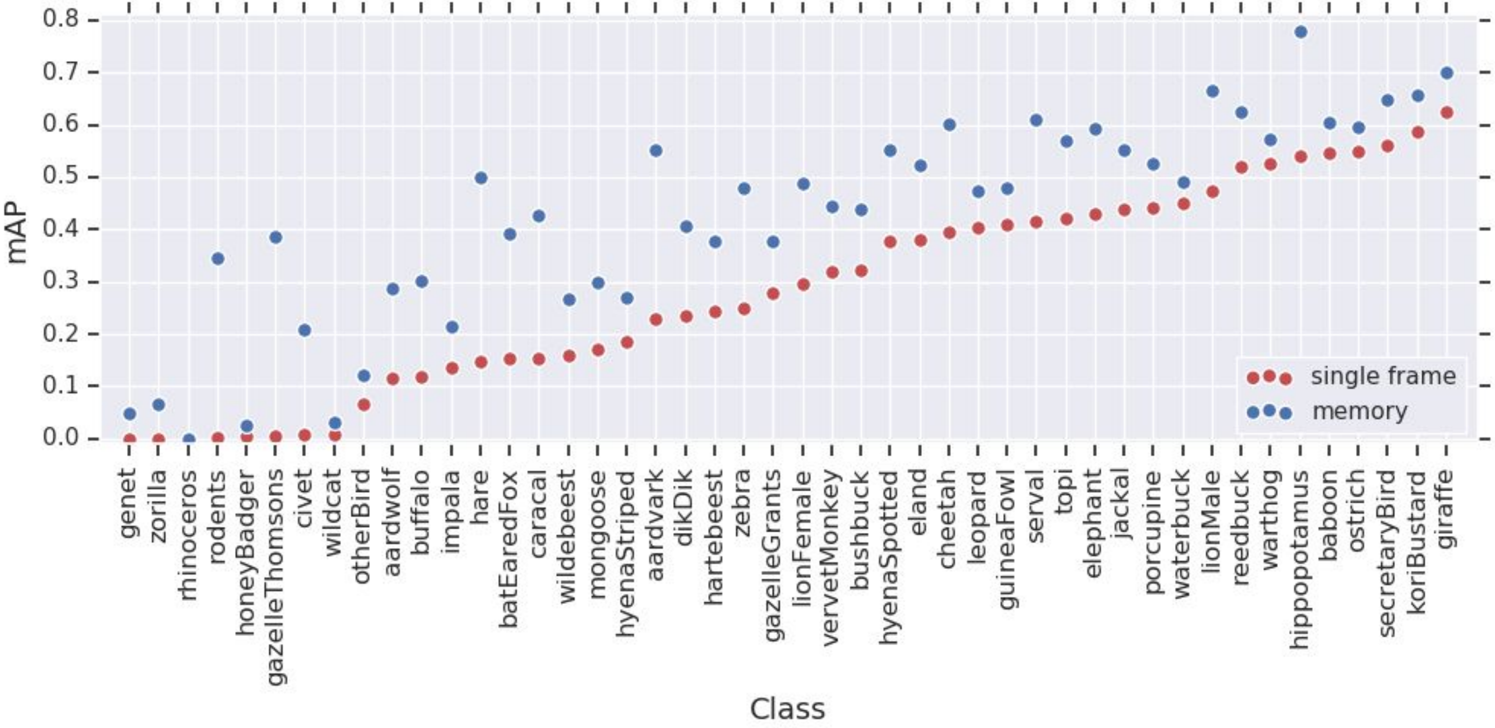
(d) Objects poorly lit.

(e) Background distractor.

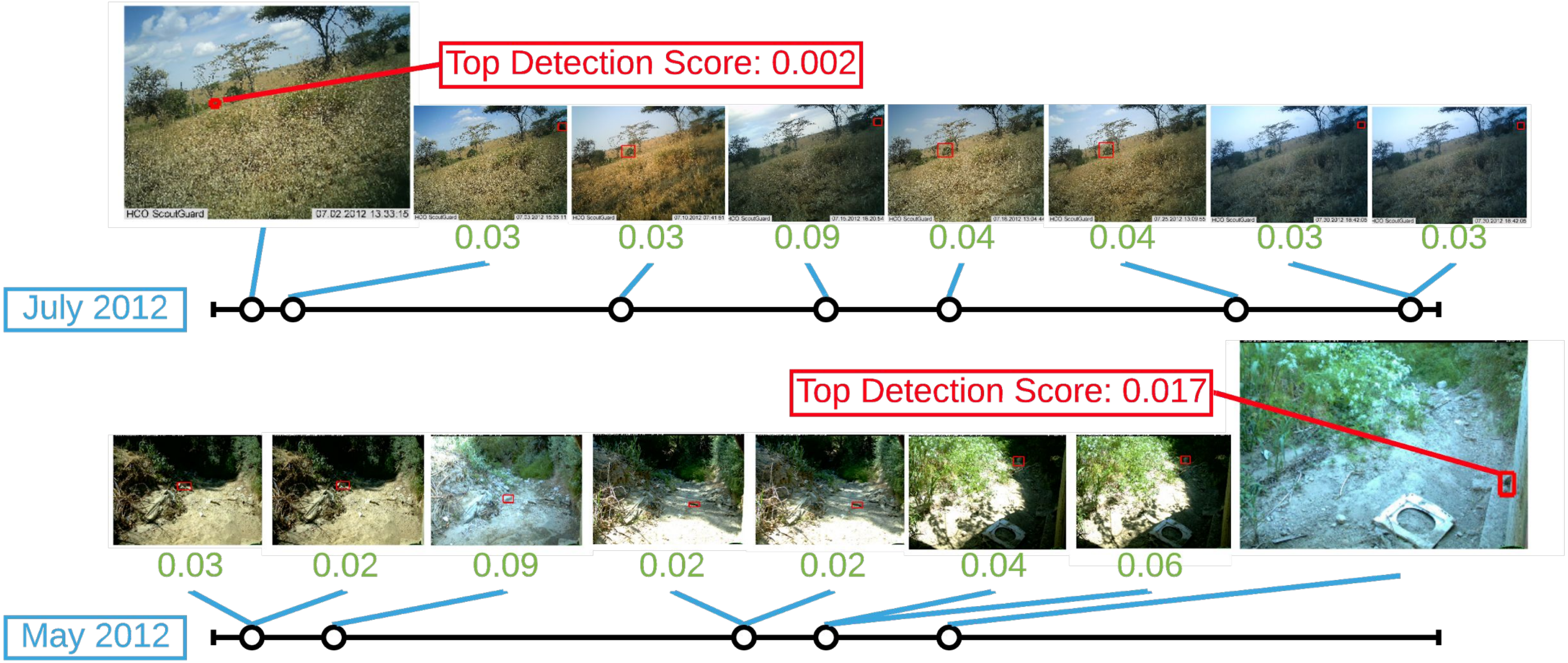
# Attention is temporally adaptive to relevance



# Snapshot Serengeti mAP improves for all classes



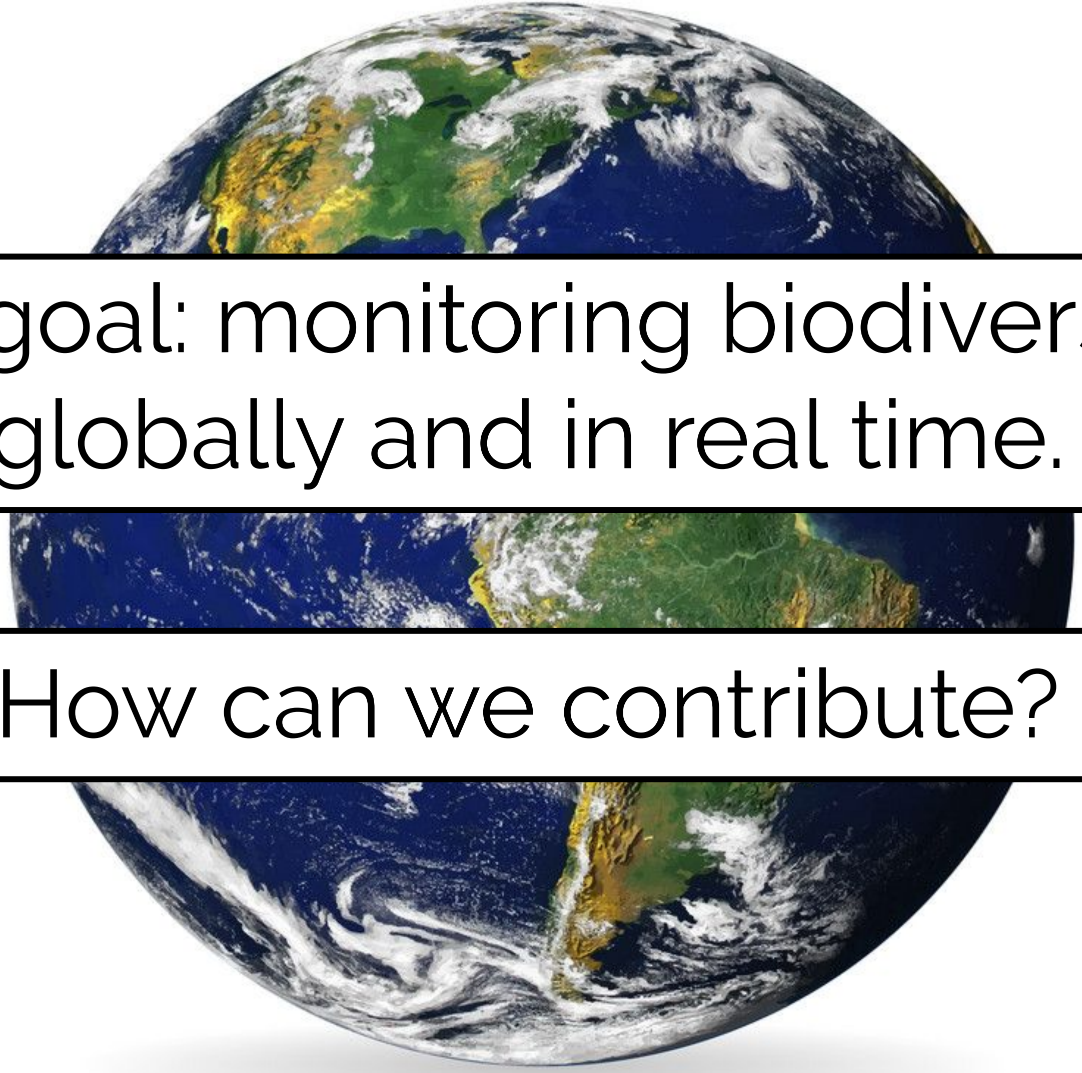
# Background classes are learned without supervision



# Static passive monitoring sensors



- Sparse, irregular frame rate
- Power, computational, and memory constraints.
- Much of the data is “empty”



Big goal: monitoring biodiversity, globally and in real time.

How can we contribute?

# Current Biodiversity AI Competitions


Global camera traps (WCS) + RS

**https://www.kaggle.com/c/iwildcam-2020-fgvc7**

## GeoLifeCLEF 2020

Location-Based
Species
Recommendation

2M Species Observations + RS + LC + Covariates

**https://www.imageclef.org/GeoLifeCLEF2020**

# Acknowledgements

Microsoft

AI for Earth

Caltech
Vision Lab

CALIFORNIA INSTITUTE OF TECHNOLOGY
1891

WILD ME

NSF

NATIONAL
PARK
SERVICE

IDAHO
FISH & GAME

USGS
science for a changing world

MPALA
THE MPALA RESEARCH CENTRE &
THE MPALA WILDLIFE FOUNDATION

LILA BC

Labeled Information Library of Alexandria: Biology and Conservation