Recognition in Terra Incognita

Sara Beery, Grant Van Horn, and Pietro Perona

Presenters: Amit Parikh & Taixi Lu

Motivation

State of the art recognition models perform great in **common** contexts…

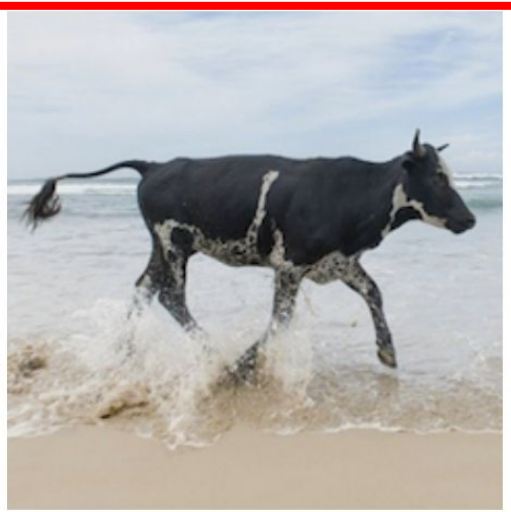
(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

…but, in **uncommon** contexts they fail to detect (B) or perform poorly (C)

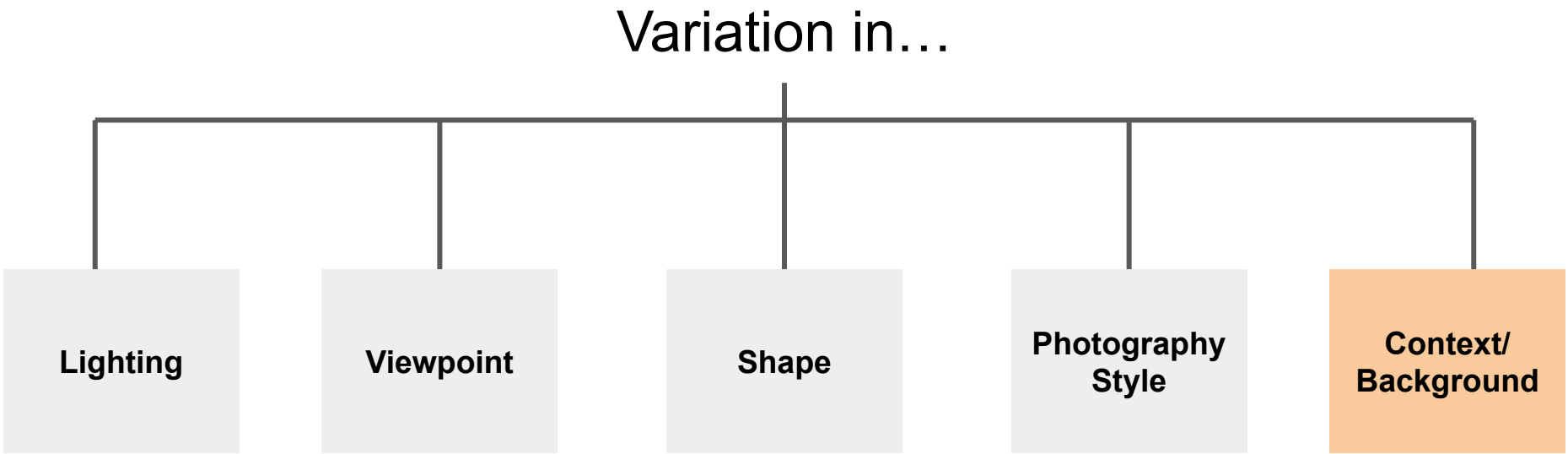
(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98



(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

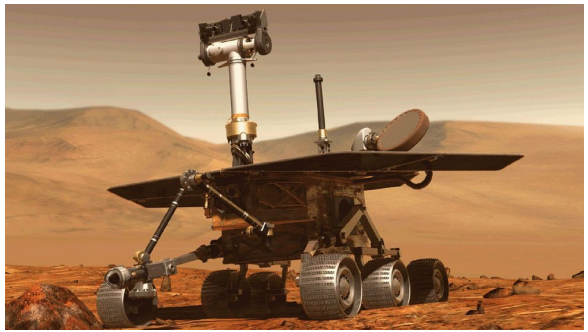
(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

What makes generalization difficult?



Scenarios with varying backgrounds

Surveillance



Automated Exploration

Home Automation

Camera Traps

Deployed in wide variety of backgrounds

Controlled for lighting

Eliminates photographer bias

The Dataset

Existing Datasets

Common Benchmarks

- Taken in different locations by different people
- Biased towards well-lit images with subjects centered

Natural World Datasets

- Large amount of diversity in location & perspective
- Mostly taken by humans under good lighting

Snapshot Serengeti Dataset

- Largest-scale camera trap dataset ever collected
- Not yet suitable for controlled experiments

CCT-20: CalTech Camera Traps-20

CalTech Camera Traps

- 243,187 Images
- 140 Camera Locations

Subset to American Southwest

CCT-20

- 57,868 Images
- 20 Camera Locations
- 16 Species Classes
- Bounding Boxes

Challenges

(1) Illumination

(2) Blur

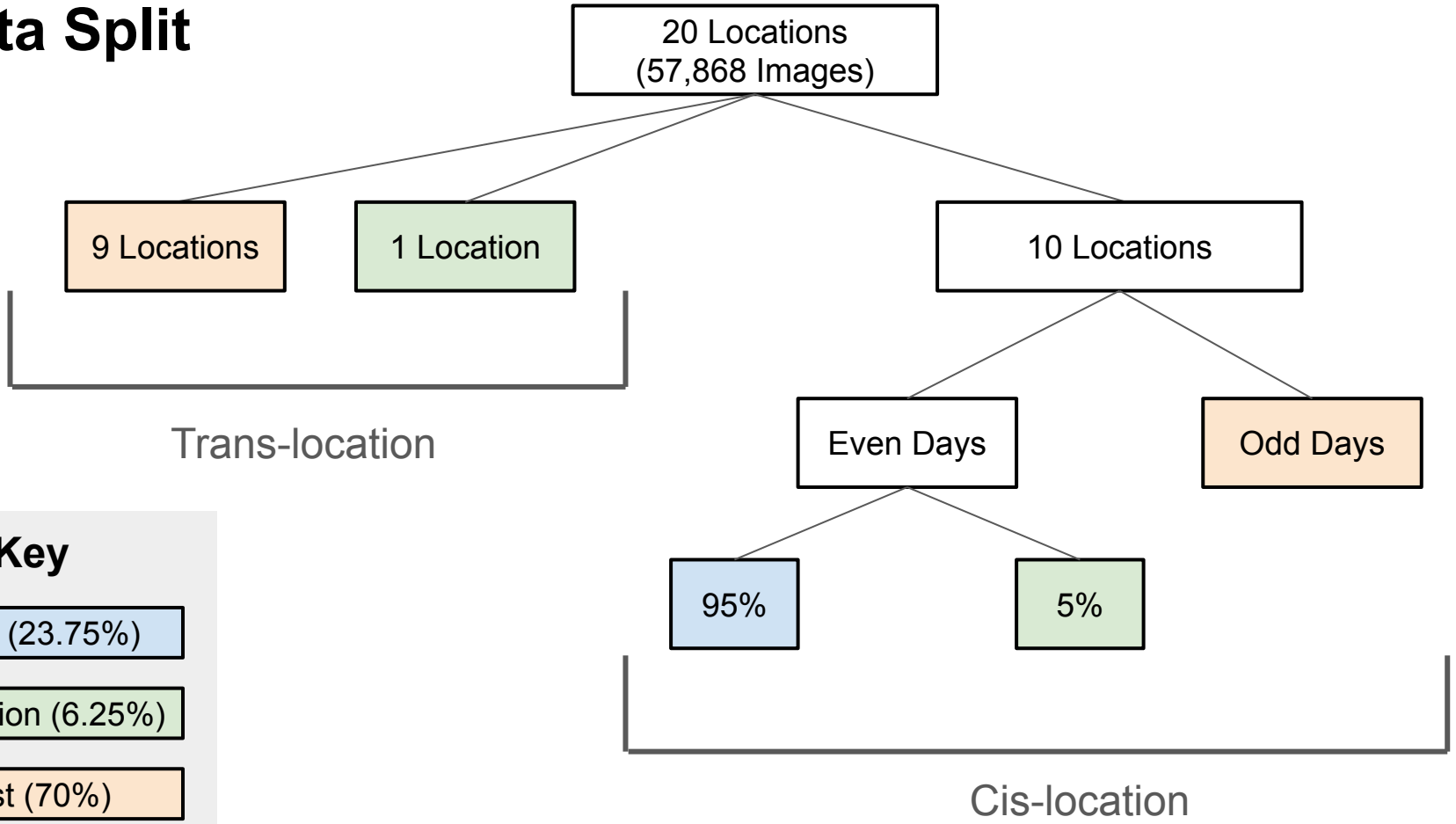
(3) ROI Size

(4) Occlusion

(5) Camouflage

(6) Perspective

Data Split

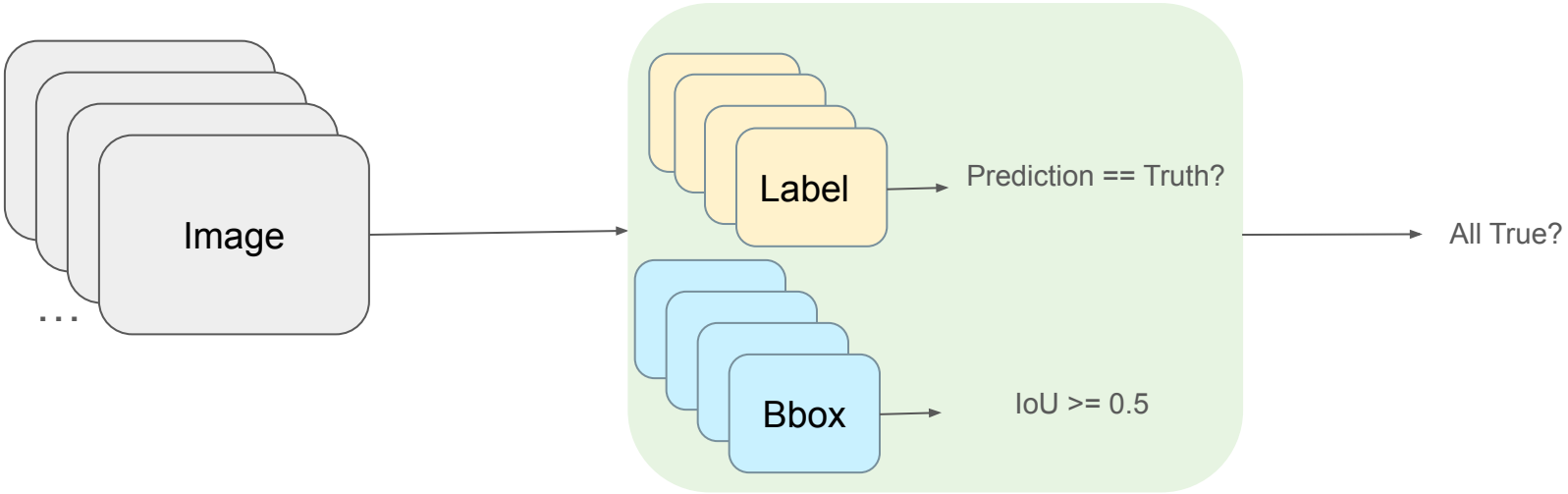


Key

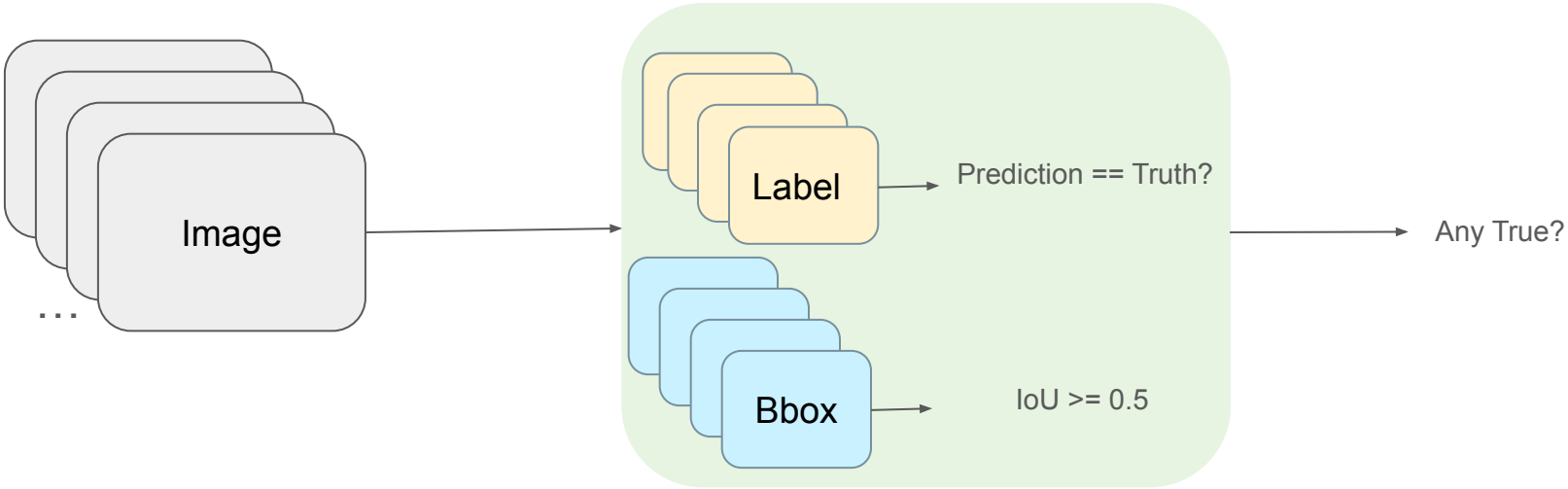
- Train (23.75%)
- Validation (6.25%)
- Test (70%)

Evaluation Protocol

Most Confident



Oracle



Experiments

Classification

	Cis-Locations		Trans-Locations		Error Increase	
Sequence Information	Images	Bboxes	Images	Bboxes	Images	Bboxes
None	19.06	8.14	41.04	19.56	115%	140%
Most Confident	17.7	7.06	34.53	15.77	95%	123%
Oracle	14.92	5.52	28.69	12.06	92%	118%

Detection

	Cis-Locations		Trans-Locations		Error Increase	
Sequence Information	ResNet	Inception	ResNet	Inception	ResNet	Inception
None	77.10	77.57	70.17	71.37	30%	27.6%
Most Confident	84.78	86.22	84.09	85.44	4.5%	5.6%
Oracle	94.95	95.04	92.13	93.09	55.8%	39.3%

Detection mAP at IoU=0.5 across experiments.

Number of Training Examples