Fine-Grained Counting with Crowd-Sourced Supervision

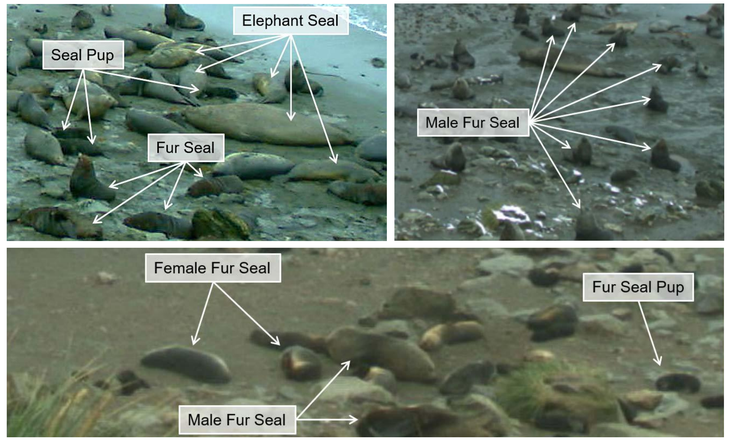
Justin Kay
Ai.Fish

Catherine M. Foley
Stony Brook University

Tom Hart
University of Oxford

CV4Animals Workshop

Motivation: Seal Counting and Classification

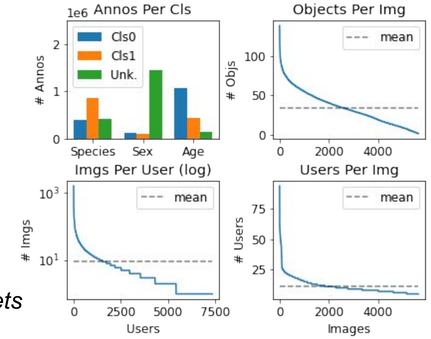


- "Sentinel species": populations reflect conditions of the environment.
- **Important attributes (species, sex, and age)** are **fine-grained and difficult to classify.**
- Crowdsourced data from past ecological studies: **1.7M annotations from 7,364 users!**

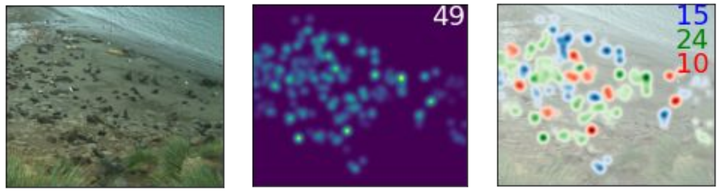
Dataset

# Classes	8
# Images	5,633
# Objects	~192k
# Annotations	1.7M

~50% larger than existing fine-grained counting datasets



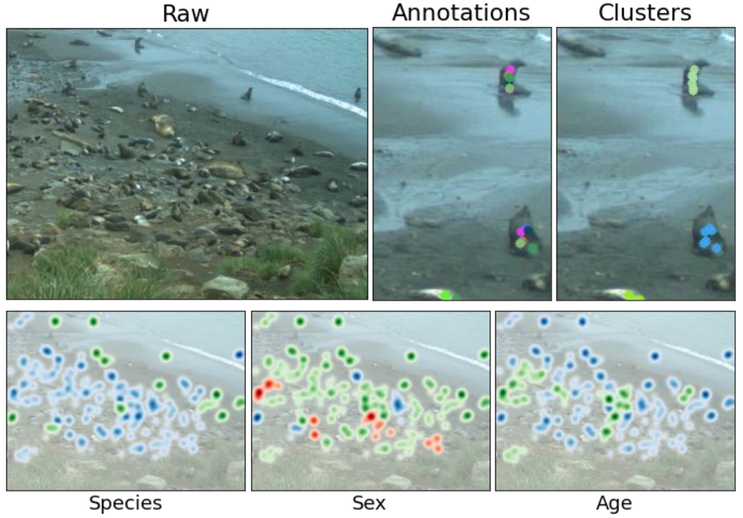
Computer Vision Approach: Fine-Grained Counting



Raw Image | Overall Density / Count | Class Densities / Counts

How to generate ground truths?

- **Our approach:** cluster annotations (capture target shape/size); majority vote for classification; some classifications are still "unknown"



Blue/Green: Binary classifications for each attribute; Red: "Unknown"

Metrics

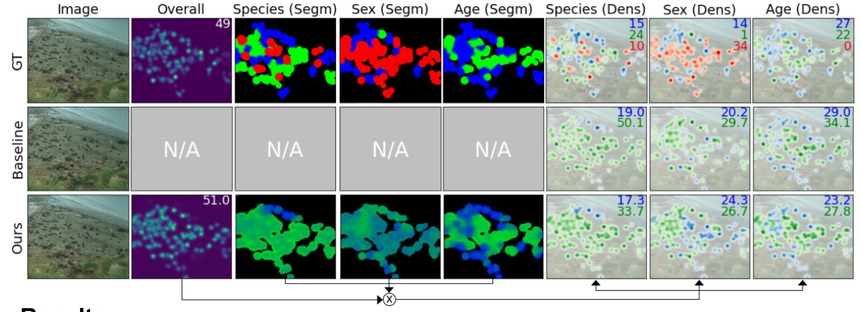
- *MAE* for overall count error
- Fine-grained count error: **C**ategory-averaged **M**asked **MAE** :

$$CMMAE = \frac{1}{A}\sum_{a=1}^{A}\frac{1}{C_a}\sum_{c=1}^{C_a} MMAE_{a,c}$$

C classes for A attributes; macro-average of MAE with "unknown" regions masked

Preliminary Methods and Results

- Train with density maps based on clustered annotations.
- Add branch to fully convolutional counting network† to predict in parallel: (a) overall density, (b) soft multiclass segmentations. Element-wise multiply outputs. Ensures consistent counts across attributes.
- Mask segmentation and fine-grained counting loss in "unknown" regions.



Results:

- 8% relative improvement in *MAE* & *CMMAE* over baseline.
- 17% / 20% relative improvement on male / female, the most "unknown" classes.

Method	MAE	CMMAE	Species	Sex	Age
Baseline	8.88	5.98	5.55 9.99	4.47 4.68	**4.66** 6.54
Ours	**8.15**	**5.52**	**5.38** **9.23**	**3.70** **3.76**	4.68 **6.37**

Next Steps

- Collect expert ground truth annotations (in progress!).
- Investigate methods that incorporate spatiotemporal context.
- Methods with localization would be useful for downstream tasks.

†Li et al. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. CVPR 2018.