

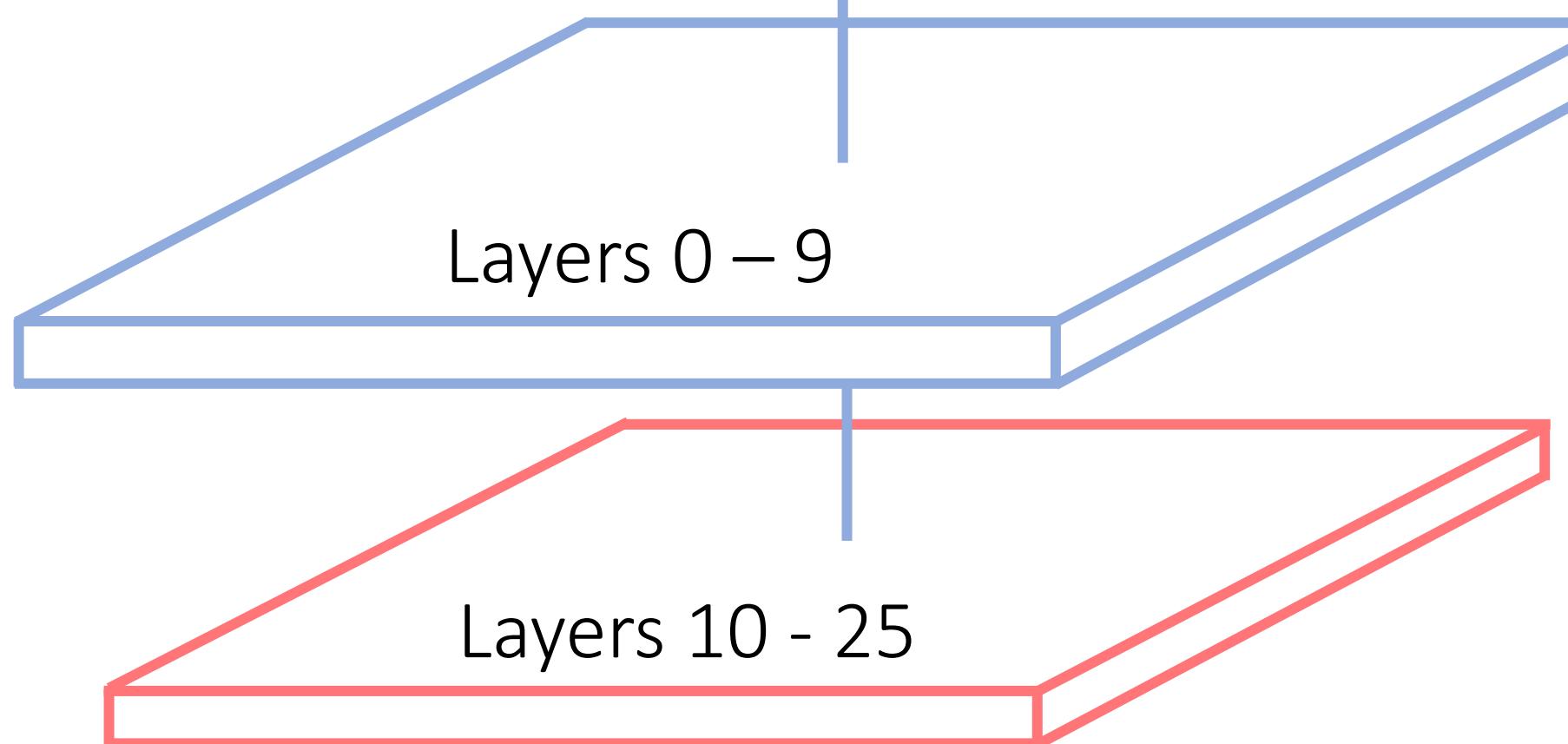
# Automated Animal Detection in Camera Trap Imagery

Joshua Liu  
Washington University in St. Louis

## Motivation

Currently, humans are responsible for labeling camera trap imagery, a process that consumes significant amounts of time and restricts ecologists' ability to make timely decisions.

## Approach

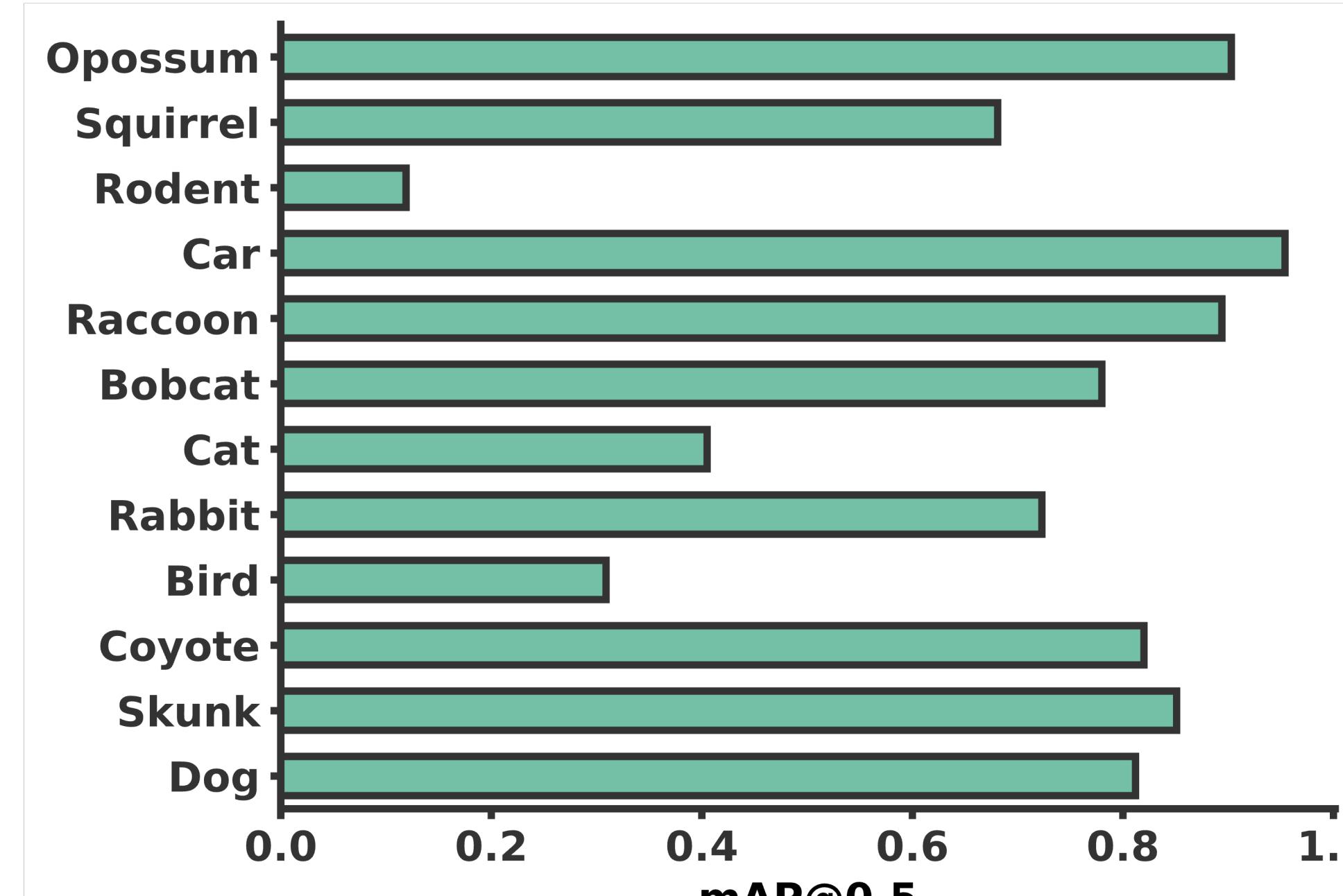


- We used the yolov5 [1] architecture to fine-tune Mega Detector [3].
- The first 10 layers (backbone) were frozen.

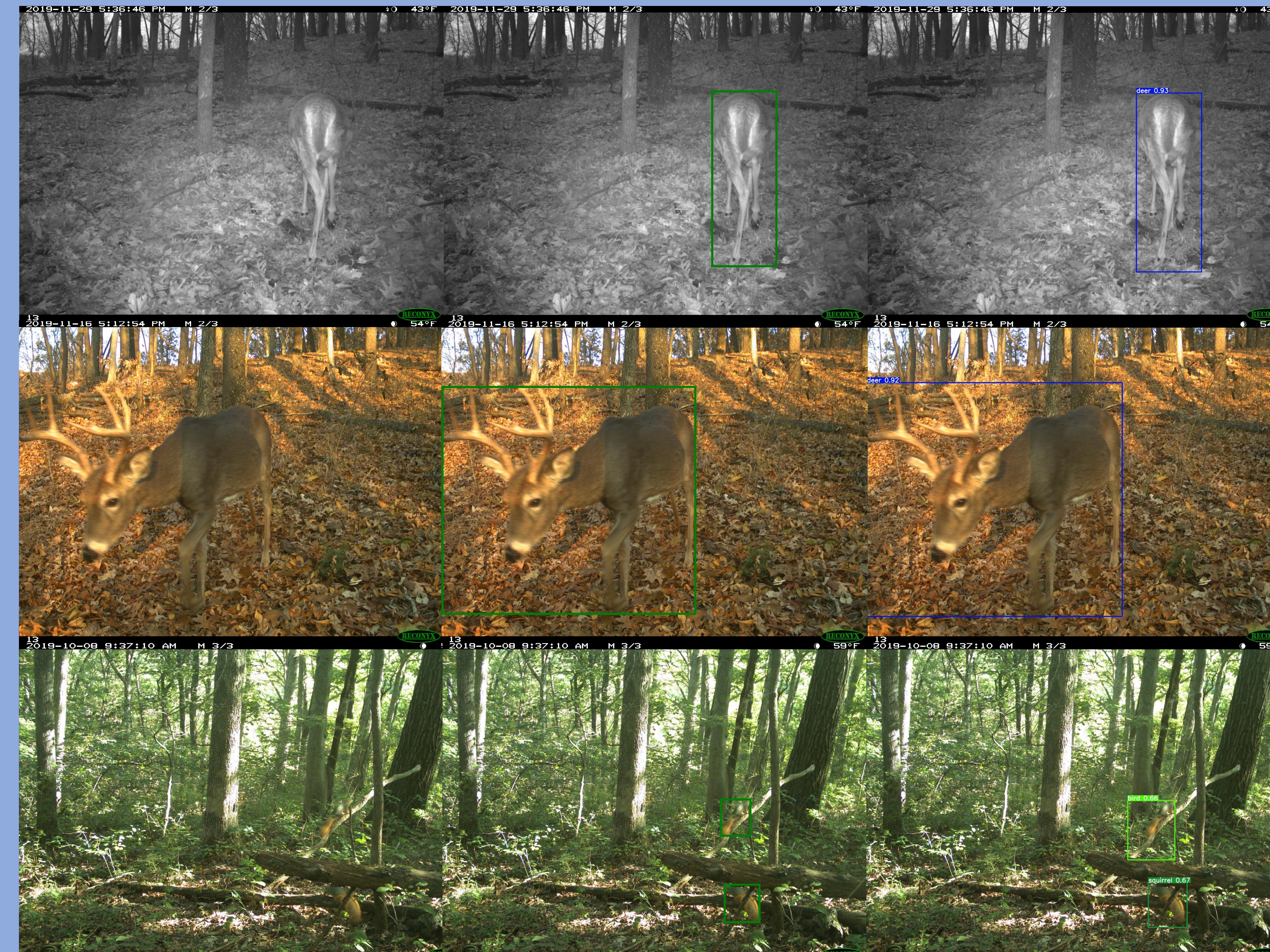
## Evaluation

- Caltech Camera Traps Dataset [2]
  - 63,025 images with bounding boxes
  - Split by camera location:
    - Train: 43,528 images
    - Val: 19,497 images
- Optimized with SGD for 10 epochs in batch sizes of 4 images on a NVIDIA A100

## Results



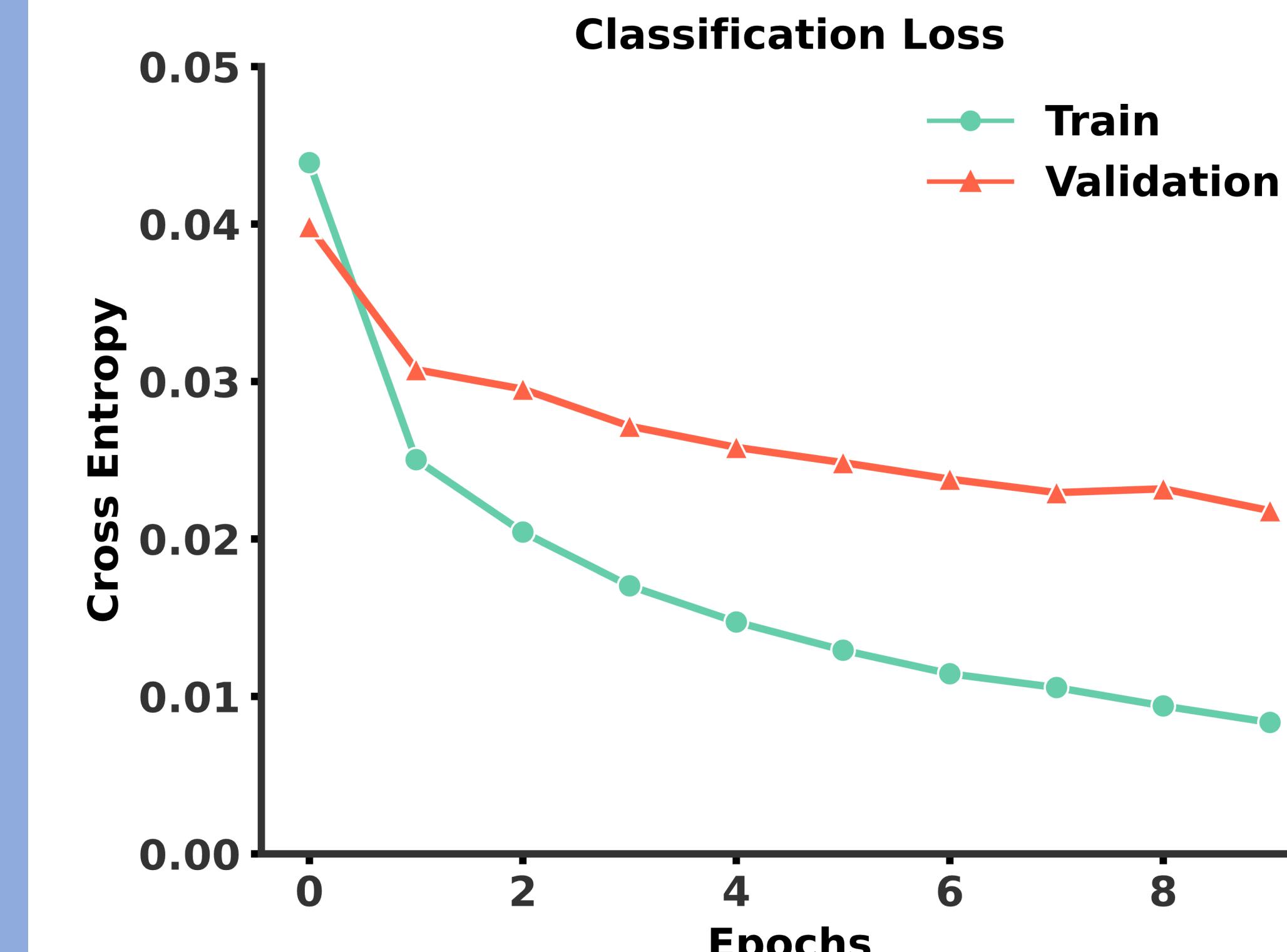
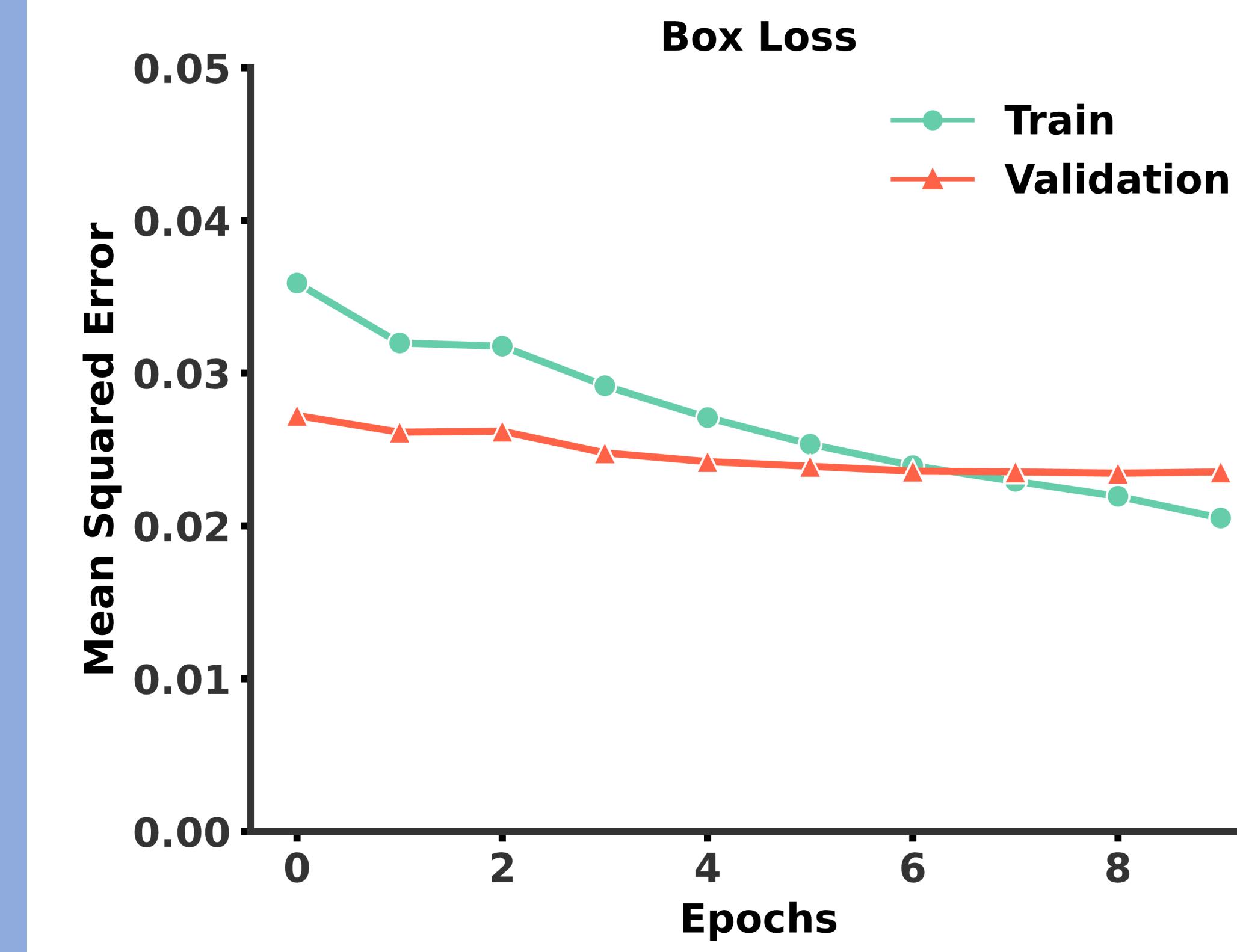
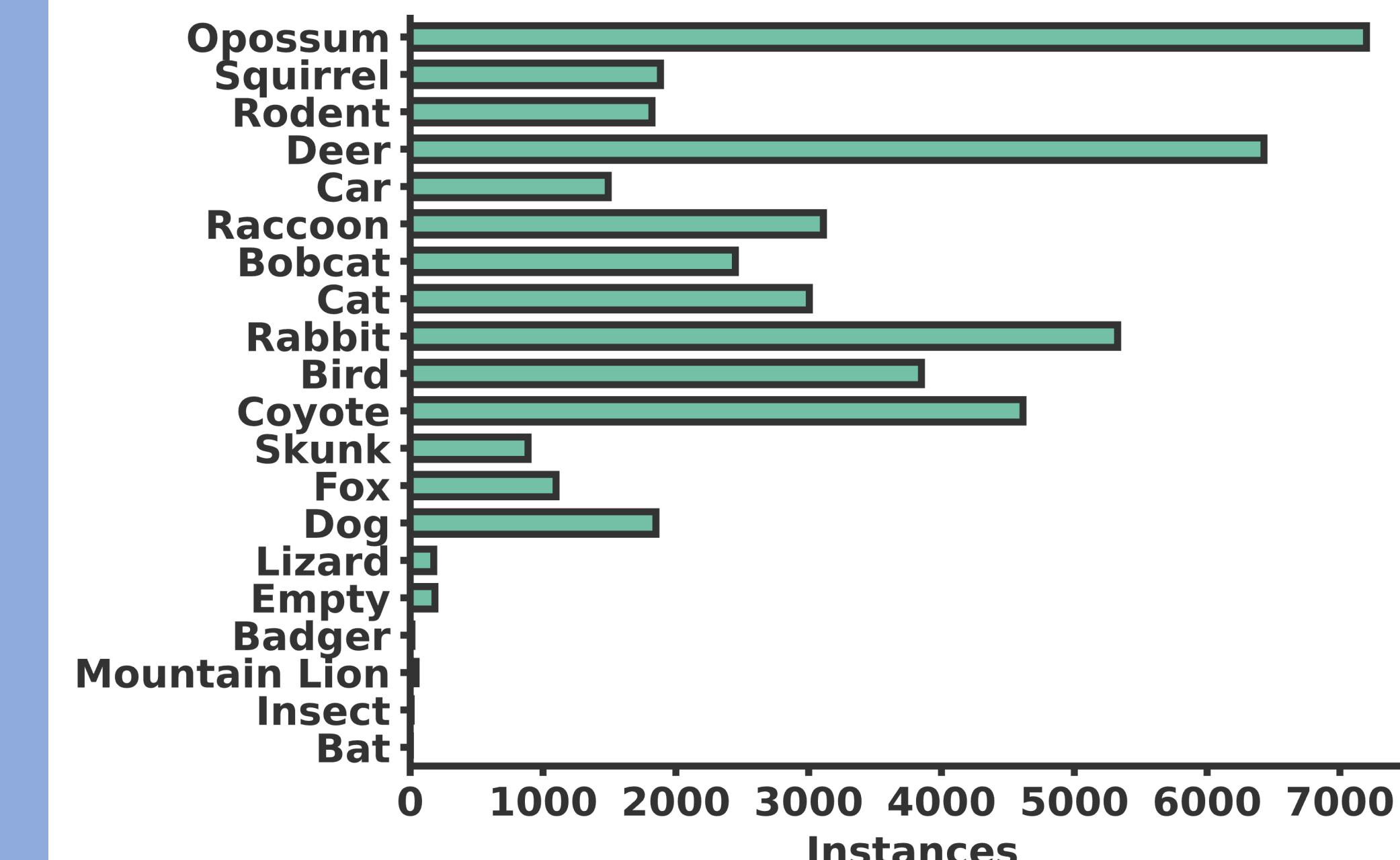
# We use object detection models to drastically improve throughput of camera trap annotations.



Raw

Mega Detector

Fine Tuned



## Next Steps

- Use a weighted loss function
- Include spatio-temporal information in training and inference
- Implement Active Learning
- Create a more balanced dataset

• Sri Kumar Sastry, Nathan Jacobs

## References

1. Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934.
2. Sara Beery, Grant van Horn, and Pietro Perona. 2018. Recognition in Terra Incognita. arXiv:1807.04975.
3. Sara Beery, Dan Morris, and Siyu Yang. 2019. Efficient Pipeline for Camera Trap Image Review. arXiv:1907.06772.