

# Efficient Pipeline for Camera Trap Image Review

Sara Beery

sbeery@caltech.edu

California Institute of Technology

Pasadena, California

Dan Morris

dan@microsoft.com

Microsoft AI for Earth

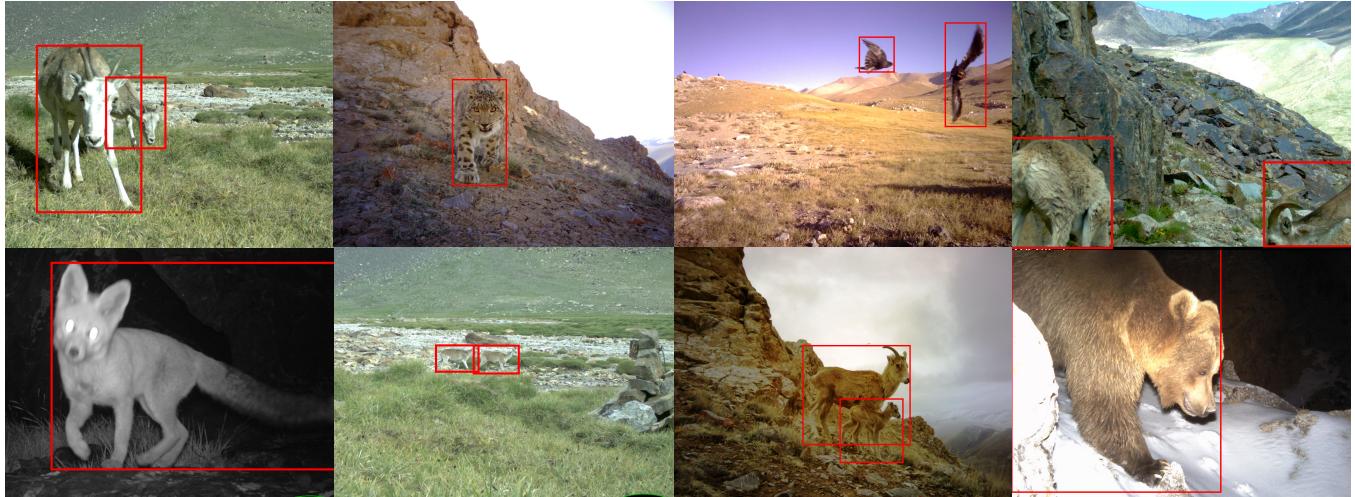
Redmond, Washington

Siyu Yang

yasiyu@microsoft.com

Microsoft AI for Earth

Redmond, Washington



**Figure 1:** Example results from our generic detector, on images from regions and/or species not seen during training.

## ABSTRACT

Biologists all over the world use camera traps to monitor biodiversity and wildlife population density. The computer vision community has been making strides towards automating the species classification challenge in camera traps [1, 2, 4–16], but it has proven difficult to apply models trained in one region to images collected in different geographic areas. In some cases, accuracy falls off catastrophically in new region, due to both changes in background and the presence of previously-unseen species. We propose a pipeline that takes advantage of a pre-trained general animal detector and a smaller set of labeled images to train a classification model that can efficiently achieve accurate results in a new region.

## CCS CONCEPTS

- Computing methodologies → Machine learning.

## KEYWORDS

detection, classification, camera traps, biodiversity monitoring

## 1 INTRODUCTION

Camera traps are heat- or motion-activated cameras placed in the wild to monitor and investigate animal populations and behavior. They are used to locate threatened species, identify important habitats, monitor sites of interest, and analyze wildlife activity patterns. At present, the time required to manually review images severely limits productivity. Additionally, ~70% of camera trap images are empty, due to a high rate of false triggers.

Previous work has shown good results on automated species classification in camera trap data [8], but further analysis has shown that these results do not generalize to new cameras or new geographical regions [3]. Additionally, these models will fail to recognize any species they were not trained on. In theory, it is possible to re-train an existing model in order to add missing species, but in practice, this is quite difficult and requires just as much machine learning expertise as training models from scratch. Consequently, very few organizations have successfully deployed machine learning tools for accelerating camera trap image annotation.

We propose a different approach to applying machine learning to camera trap projects, combining a *generalizable detector* with *project-specific classifiers*.

We have trained an animal detector that is able to find and localize (but not identify) animals, even species not seen during training, in diverse ecosystems worldwide. See Fig. 1 for examples of the detector on camera trap images from regions and/or species not seen during training. By first finding and localizing animals, we are able to:

- (1) drastically reduce the time spent filtering empty images, and
- (2) dramatically simplify the process of training species classifiers, because we can crop images to individual animals (and thus classifiers need only worry about animal pixels, not background pixels).

With this detector model as a powerful new tool, we have established a modular pipeline for on-boarding new organizations and building project-specific image processing systems.

## 2 PIPELINE

We break our pipeline into four stages: data ingestion, animal detection, classifier training, and application to new data.

### 2.1 Data ingestion

First we transfer images to the cloud, either by uploading to a drop point or by mailing an external hard drive. Data comes in a variety of formats; we convert each dataset to the COCO-Camera Traps format, i.e., we create a JSON file that encodes the annotations and the image locations within the organization’s file structure.

### 2.2 Animal detection

We next run our (generic) animal detector on all the images to locate animals. We have developed an infrastructure for efficiently running this detector on millions of images, dividing the load over multiple nodes.

We find that a single detector works for a broad range of regions and species. If the detection results (as validated by the organization) are not sufficiently accurate, it is possible to collect annotations for a small set of their images and fine-tune the detector. Typically these annotations would be fed back into a new version of the general detector, improving results for subsequent projects.

### 2.3 Classifier training

Using species labels provided by the organization, we train a (project-specific) classifier on the cropped-out animals.

### 2.4 Application to new data

We use the general detector and the project-specific classifier to power tools facilitating accelerated verification and image review, e.g., visualizing the detections, selecting images for review based on model confidence, etc.

## 3 CASE STUDY: THE IDAHO DEPARTMENT OF FISH AND GAME

We applied our pipeline to 4.8 million images collected by the Idaho Department of Fish and Game (IDFG) from six regions in the state, of which 0.76 million have image-level species labels. Spreading the load over 16 nodes, each with one GPU, it took under three days to perform detection on this batch of images. By filtering out images without confident detections, we have eliminated some 80% of images (estimated by the project owner at IDFG) from manual review as this study contained a large percentage of empty frames. The average precision for animal detections ranges from 0.885 to 0.988 for different regions, evaluated against species labels as an indication of animal presence.

Notably, the version of the detector used was not trained with any camera trap images with snow but performed very well on such images in the IDFG data. The detector was also able to find animals in night images that reviewers would have missed without adjusting the exposure of the image. False positives were a problem where branches and rocks were misidentified as animals. A post-processing step to remove detections that appear in the same position for many frames in a row alleviated this issue. We are in

the process of training the project-specific classifier for IDFG, and preliminary results for species classification are promising.

## 4 CONCLUSIONS

We propose a pipeline that allows us to train classifiers for new camera trap projects in an efficient way, first leveraging a generic animal detection model to localize animals and remove empties, then training a project-specific classifier using the localized images of animals and their image-level labels. We present this as a new approach to structuring camera trap projects, and aim to formalize discussion around the steps that are required to successfully apply machine learning to camera trap images.

Our code and models are available at [github.com/microsoft/cameratraps](https://github.com/microsoft/cameratraps), and public datasets used for training are available at [lila.science](https://lila.science).

## REFERENCES

- [1] Sara Beery, Yang Liu, Dan Morris, Jim Piavis, Ashish Kapoor, Markus Meister, and Pietro Perona. 2019. Synthetic Examples Improve Generalization for Rare Classes. *arXiv preprint arXiv:1904.05916* (2019).
- [2] Sara Beery, Grant Van Horn, and Pietro Perona. 2018. Recognition in terra incognita. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 456–473.
- [3] Sara Beery, Grant Van Horn, and Pietro Perona. 2018. Recognition in Terra Incognita. In *The European Conference on Computer Vision (ECCV)*.
- [4] Guobin Chen, Tony X Han, Zhihai He, Roland Kays, and Tavis Forrester. 2014. Deep convolutional neural network based species recognition for wild animal monitoring. In *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 858–862.
- [5] Jhony-Heriberto Giraldo-Zuluaga, Augusto Salazar, Alexander Gomez, and Ángelica Diaz-Pulido. 2017. Camera-trap images segmentation using multi-layer robust principal component analysis. *The Visual Computer* (2017), 1–13.
- [6] Kai-Hsiang Lin, Pooya Khorrami, Jiangping Wang, Mark Hasegawa-Johnson, and Thomas S Huang. 2014. Foreground object detection in highly dynamic scenes using saliency. In *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 1125–1129.
- [7] Agnieszka Miguel, Sara Beery, Erica Flores, Loren Klemesrud, and Rana Bayrakcsmith. 2016. Finding areas of motion in camera trap images. In *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 1334–1338.
- [8] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences* 115, 25 (2018), E5716–E5725.
- [9] Xiaobo Ren, Tony X Han, and Zhihai He. 2013. Ensemble video object cut in highly dynamic scenes. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 1947–1954.
- [10] Alexandra Swanson, Margaret Kosmala, Chris Lintott, Robert Simpson, Arfon Smith, and Craig Packer. 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific data* 2 (2015), 150026.
- [11] Alexander Gomez Villa, Augusto Salazar, and Francisco Vargas. 2017. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics* 41 (2017), 24–32.
- [12] Michael J Wilber, Walter J Scheirer, Phil Leitner, Brian Heflin, James Zott, Daniel Reinke, David K Delaney, and Terrance E Boult. 2013. Animal recognition in the mojave desert: Vision tools for field biologists. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. IEEE, 206–213.
- [13] Hayder Yousif, Jianhe Yuan, Roland Kays, and Zhihai He. 2017. Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. In *Circuits and Systems (ISCAS), 2017 IEEE International Symposium on*. IEEE, 1–4.
- [14] Xiaoyuan Yu, Jiangping Wang, Roland Kays, Patrick A Jansen, Tianjiang Wang, and Thomas Huang. 2013. Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing* 2013, 1 (2013), 52.
- [15] Zhi Zhang, Tony X Han, and Zhihai He. 2015. Coupled ensemble graph cuts and object verification for animal segmentation from highly cluttered videos. In *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2830–2834.
- [16] Zhi Zhang, Zhihai He, Guitao Cao, and Wenming Cao. 2016. Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification. *IEEE Transactions on Multimedia* 18, 10 (2016), 2079–2092.