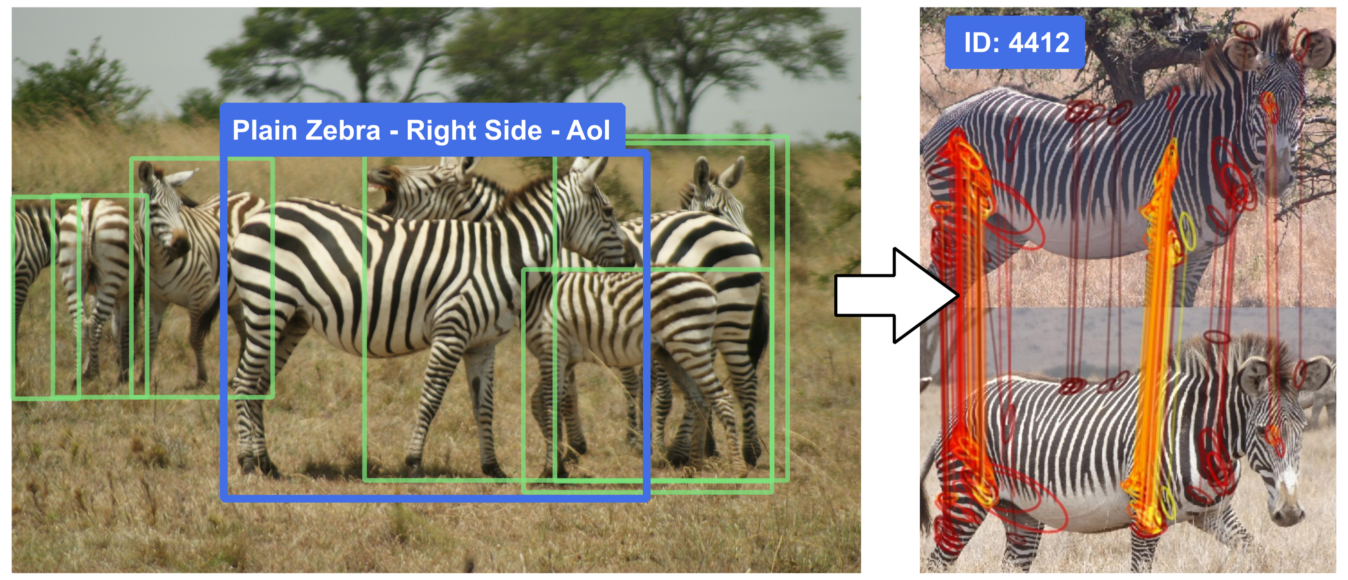
An Animal Detection Pipeline for Identification

Jason Parham*, Jonathan Crall, Daniel Rubenstein, Jason Holmberg, Tanya Berger-Wolf, and Charles Stewart

Paper https://lev.cs.rpi.edu/public/papers/parham_wacv_2018.pdf
Dataset http://lev.cs.rpi.edu/public/datasets/wild.tar.gz

INTRODUCTION



Detection Identification

Computer vision-based methods are being used increasingly as tools to assist wild animal object recognition. The ability to identify individual animals from images enables population surveys through sight-resight identification and forms the basis for demographic studies.

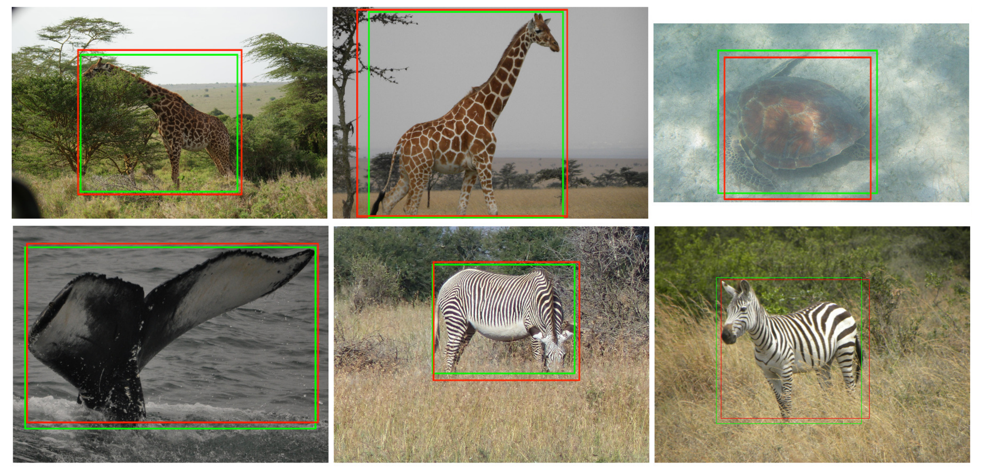
Detection includes the obvious steps of finding animals in images, determining their species, and placing bounding boxes around them, creating what we refer to as an *annotation*. But, the problem is more complex than this, especially when large data volumes gathered by non-specialists are considered: there may be multiple (or no) animals from several different species in an image; some annotations may have poor quality while others may show only parts of an animal due to self occlusion, or occlusion by other animals or vegetation; and may be seen from a range of viewpoints and poses, only some of which show identifiable information. We propose a five stage detection pipeline (middle, top) to address these challenges and a new dataset to evaluate its performance.

METHODS - IMAGE CLASSIFICATION

The purpose of the image classifier is to predict the existence of species of interest within an image. We structure the classifier to predict a multi-label, multi-class vector where the species flag is set to 1 if *at least one* animal of that species exists in the image and 0 otherwise. Our implementations are built using Theano and Lasagne **[1, 2].**

The image classifier can be thought of as a high-pass filter to prevent irrelevant images from being processed further.

METHODS - ANNOTATION LOCALIZATION



The annotation localization network design is based on the You Only Look Once (YOLO, version 1) network by **[3]**. The network's goal is to perform bounding box regression (left) and species classification around all objects of interest, the result being a collection of sub-regions that can be cropped into a candidate list of object annotations.

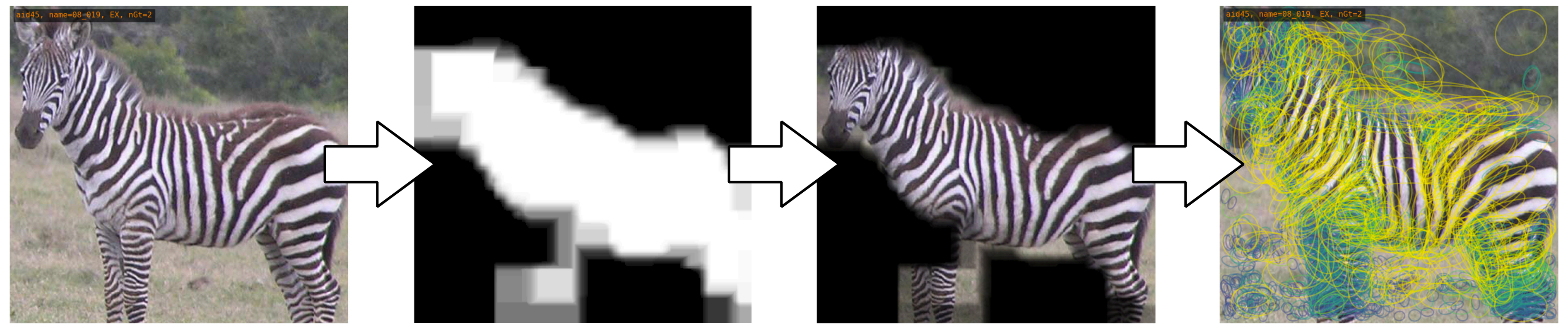
The predicted bounding boxes by the localization network have associated species label classifications. Since we are performing annotation classification anyway, we can treat these localizations simply as salient object detections.

METHODS - ANNOTATION CLASSIFICATION

The annotation classification network architecture is very similar to the image classification component except that it performs a standard single-label, multi-class classification. We intentionally train a separate set of weights for the convolutional feature extractors in each component. The primary task of annotation classification is to correctly label the annotation's species and the correct viewpoint together. Poor scoring boxes do not continue in the pipeline.

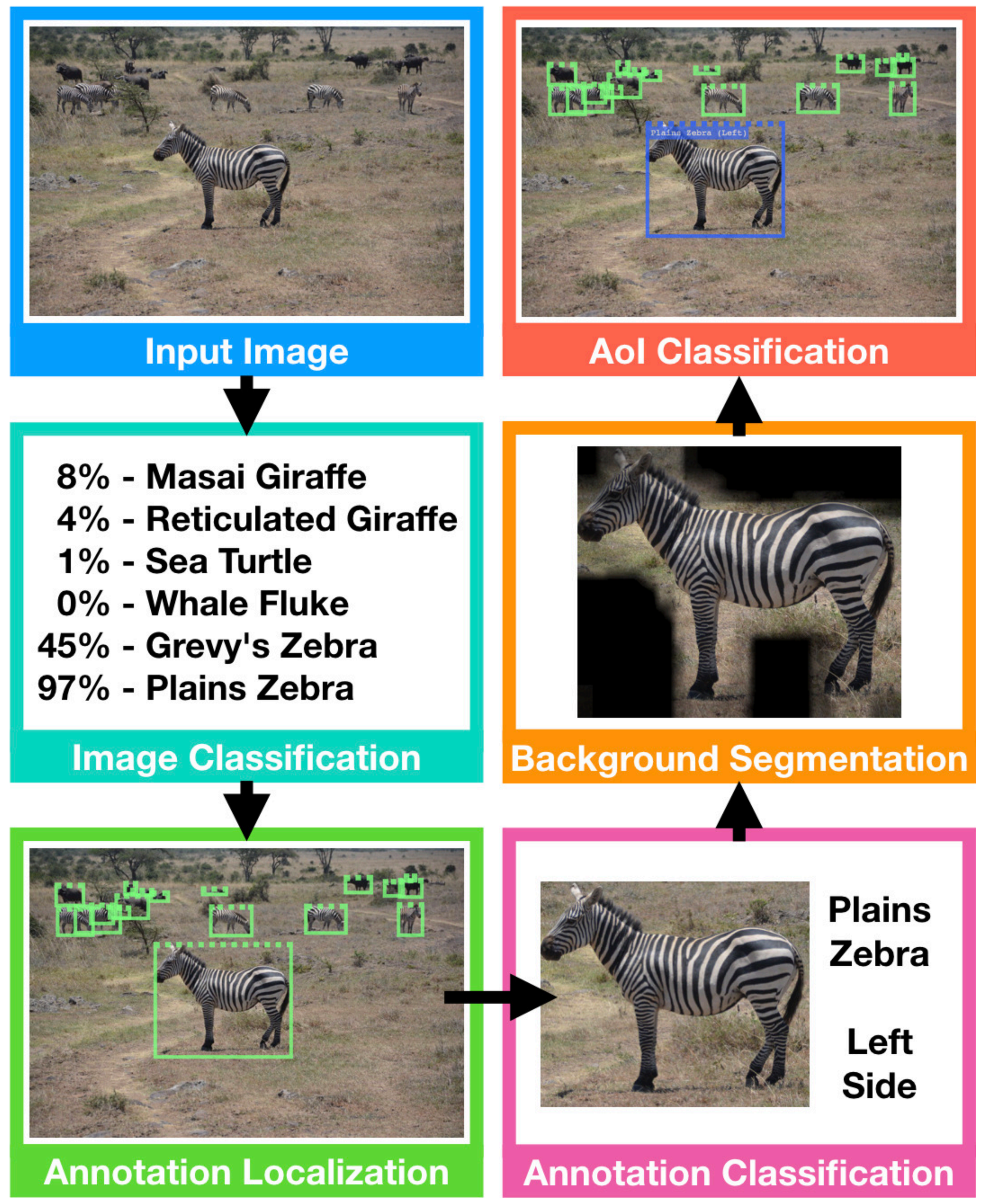
METHODS - BACKGROUND SEGMENTATION

The annotation background segmentation neural networks are a distinct type of architecture called a Fully Convolutional Neural Network **[4]**. Our method does not require fully-segmented data, but instead performs coarse background segmentation as a binary classification. The goal of this component is to produce a species-specific background mask, which can be used to eliminate or down-weight distracting non-animal pixel information (below).


Input Background Mask Blended Mask Weighted SIFT key-points

We do not have fully-segmented ground-truth for a quantitative evaluation of our network's background segmentation performance, as shown in the paper. We do provide examples, however, from the detection pipeline (middle, bottom).

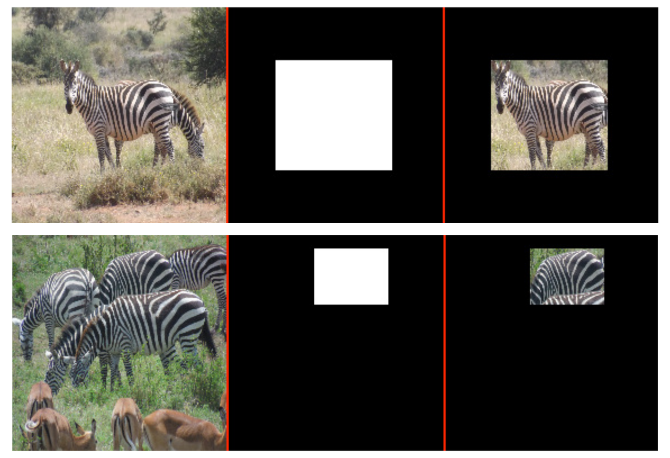
PIPELINE OVERVIEW



DETECTION PIPELINE EXAMPLES


blue bounding boxes indicate Annotations of Interest

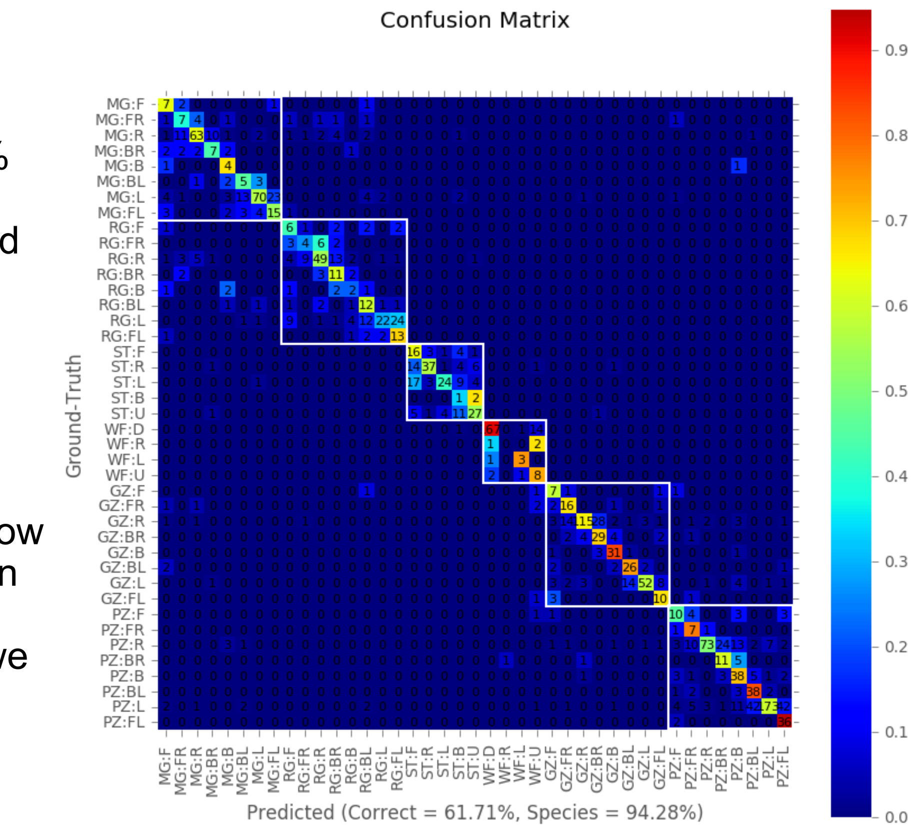
METHODS - AOI CLASSIFICATION



The novel task for the AoI classification network to solve is to predict an a posteriori decision concerning the composition of an image: "why did the photographer take this picture?" The figure to the left shows the training data for a true AoI (top) and an annotation that is not an AoI (bottom).

The overarching motivation for AoI classification is to prioritize further processing on only the most identifiable annotations. While the concept of identifiability is algorithm-dependent, we structure AoI as a generalized, easy-to-determine proxy. The end result of the AoI classifier is eliminating the need to perform processing on background and partially-visible animals, which cause confusion and increases the need for a *human-in-the-loop* reviewer.

RESULTS



Our method is able to achieve a localization mAP of 81.67%, a species and viewpoint annotation classification accuracy of 94.28% and 87.11%, respectively, and an AoI accuracy of 72.75% across 6 animal species of interest. We also introduce the Wildlife Image and Localization Dataset (WILD), which contains 5,784 images and 12,007 labeled annotations across 28 classification species and a variety of challenging, real-world detection scenarios.

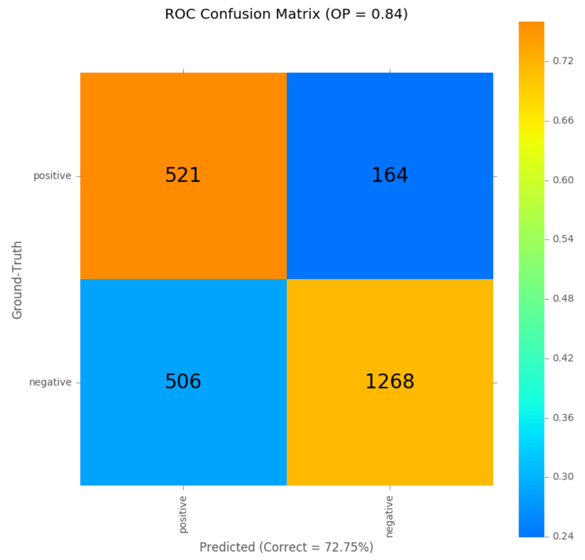
The overall accuracy of species and viewpoint combination classifications is 61.71% over 42 distinct categories (right). The accuracy improves from this baseline when we take into account how viewpoint variance impacts identification (i.e. a ±45% degree shift in yaw is acceptable for giraffes and plains), which achieves a "fuzzy" accuracy of 87.11%. Further, AoI accuracy goes up to 93.33% if we treat false-negatives as the true errors of filtering.

WILD DATASET

Species	Images	Annots.	AoIs
Masai Giraffe	1,000	1,468	611
Reticulated Giraffe	1,000	1,301	595
Sea Turtle	1,000	1,002	567
Whale Fluke	1,000	1,006	595
Grevy's Zebra	1,000	2,173	669
Plains Zebra	1,000	2,921	561
TOTAL	**5,784**	**9,871**	**3,598**

We created a new ground-truthed dataset called WILD; WILD is comprised of photographs taken by biologists, wildlife rangers, citizen scientists **[5]**, and conservationists, and captures detection scenarios that are uncommon in publicly-available computer vision datasets like PASCAL, ILSVRC, and COCO. In WILD all of the images in the dataset were taken *in situ* by on-the-ground photographers (a breakdown of images and annotations to left). We distribute the dataset in the PASCAL VOC format, with additional metadata attributes to mark viewpoints and AoI flags.

CONCLUSION



We evaluated five detection components against WILD, a new dataset of real-world animal sightings that focuses on challenging detection scenarios. The end result of our proposed pipeline is a collection of novel annotations of interest (AoI) with species and viewpoint labels. The output AoIs, for example, could be fed as input data into an appearance-based identification system (confusion matrix to right).

The goal of our method is to increase the reliability and automation of animal censusing studies and to provide better ecological information to conservationists. Future work can be focused on improvements to WILD, improving the accuracy of AoI classification, and performing a comprehensive identification performance study.

REFERENCES

[1] J.Bergstra,O.Breuleux,F.Bastien,P.Lamblin,R.Pascanu, G. Desjardins, J.Turian, D.Warde-Farley, and Y. Bengio.Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference*, volume 4 of *SciPy '10*, pages 3–10, 2010. 00548 Oral Presentation.
[2] S. Dieleman, J. Schlter, C. Raffel, E. Olson, S. K. Snderby, D. Nouri, and others. Lasagne: First release., 2015. DOI: 10.5281/zenodo.27878.
[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi.You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640:1–10, 2015.
[4] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.
[5] A. Irwin. Citizen Science: A Study of People, Expertise and Sustainable Development. Environment and Society. Rout- ledge, 1995.

UIC
WILDBOOK
KENYA WILDLIFE SERVICE
NVIDIA