Towards Automated Monitoring of Animal Movement using Camera Networks and AI

Sarah Bearman, Zhiang Chen, Harish Anand, Scott Sprague, Jeff Gagnon, Jnaneshwar Das

***Abstract*— We partnered with the Arizona Game and Fish Department to use computer vision techniques to enhance road ecology studies. Various structures, including overpasses, underpasses, escape ramps, and slope jumps have been constructed in order to facilitate animal movement across major highways and to mitigate animal-vehicle collisions. The successful functioning of these structures is monitored by placing between one and nine camera traps on each structure in order to capture how wildlife interacts with it. There are over 50 camera traps deployed across Arizona and a few neighboring states, resulting in tens of thousands of images being collected every 6-8 weeks. Our goal is to increase the efficiency of image processing and eliminate human error/bias by leveraging deep neural networks (Mask RCNN). Our results so far include an 88% detection accuracy and a 40% classification accuracy for 5 labeled species. Future work will focus on improving detection and classification, increasing the number of species we can identify, and identifying sex, age, and direction of travel. Longer term goals involve building a "smart" camera network to do real-time image processing on all of these structures.**

I. Introduction

Wildlife biology relies heavily on the use of camera traps for research. These camera traps typically use motion and infrared sensor technology to trigger a sequence of 3-5 images [1], and depending on the study area and species of interest, each camera trap can collect hundreds to thousands of images before an observer visits to download the captured data. Typically, multiple cameras are used in a study, and often, the data sets include significant amounts of false positives. Manually processing all of these images can be expensive in both time and money, and large data sets usually require more than one person to process the images. This can introduce observer-specific bias to the data, and fatigue can cause observers to label images as false negatives.

Our research focuses on large scale projects that encounter the aforementioned challenges. We have partnered with the Arizona Game and Fish Department (Contracts Branch), whose main focus is to monitor wildlife interactions with highways in order to make the most effective management decisions. These decisions consider factors such as habitat fragmentation, migration patterns, and human safety in terms of animal-vehicle collisions [2], [3]. The goal is to construct and monitor various structures, such as overpasses and underpasses, to allow wildlife to safely cross major highways (Fig. 1 , [4], [5], [6], [7]). Other structures include escape ramps and slope jumps, which allow animals that are already on the highway to safely exit, and prevents them from re-entering (Fig. 1).

Fig. 1: *Different types of animal passage structures with camera traps:* Overpasses (top left) and Underpasses (top right) are constructed to allow wildlife to cross major highways. Escape Ramps (bottom left) and Slope Jumps (bottom right) are constructed to allow animals on the highway to safely exit, as well as funnel wildlife to designated crossing structures.

These structures are located across the entire state of Arizona, and each one is equipped with one to nine cameras in order to monitor wildlife use/passage rates [5], [8], [7]. The Department is also monitoring structures in a few neighboring states, resulting in over 50 camera traps that collect thousands of images each over a span of 6-8 weeks. Given the nature of these structures, vegetation and vehicles often trigger the cameras, causing false positives to take up a significant amount of storage.

In this study, we are using computer vision techniques in order to improve the efficiency of image analysis and reduce user-specific bias while processing data sets. We are using the Mask RCNN network architecture [9] to count the number of animals in each image and identify five species of interest. Future aims are to increase the number of species we can identify, use morphological features to determine the sex and age of each animal, track individuals across a sequence of images, and to identify the direction of travel. This information is important as it gives insight into the wildlife demographics in the area, as well as indicate the effectiveness of the structure placements in facilitating wildlife movement and reducing animal-vehicle collisions.

II. Related Work

Data collection with mobile sensor networks has been explored with considerations to object localization and cov-

erage with bearing-only sensors such as cameras [10], and reconstruction of scalar fields such as environmental temperature [11]. In the arena of habitat monitoring, camera networks have enabled capturing of the spatio-temporal dynamics of terrestrial bird and mammal activity [12]. The popular use of camera traps has already inspired researchers to work on automated detection software, such as Animal Scanner [13]. This software uses deep neural networks to detect and classify humans, animals, and background images [13].

Beyond camera networks, other sensing modalities such as tagging, and telemetry from robotic vessels have been used for monitoring carp in Minnesota lakes [14]. Mola-Mola was tracked using underwater vehicles [15]. In precision agriculture applications, deep learning and computer vision algorithms have enabled counting of fruits [16], [17], [18], deweeding, and segmentation of flowers [19].

III. Methods

Our current workflow includes image collection, annotation, training/validation, and inference.

A. System and Data Collection

Images are collected by mounting between one and nine Reconyx PC800 HyperFire Professional Semi-Covert Camera Traps onto the structures being monitored. The number of cameras mounted depends on the structure, and they are oriented in order to capture all of the animals that approach the structure, not just the animals that use the structure (Fig. 2).

The camera traps themselves are motion triggered and have a semi-visible infrared flash, producing color images during the day and black and white images at night [1]. The cameras are configured to take a sequence of either three or five images at 2fps depending on the structure. The images are collected roughly every 6-8 weeks by switching out SD cards.

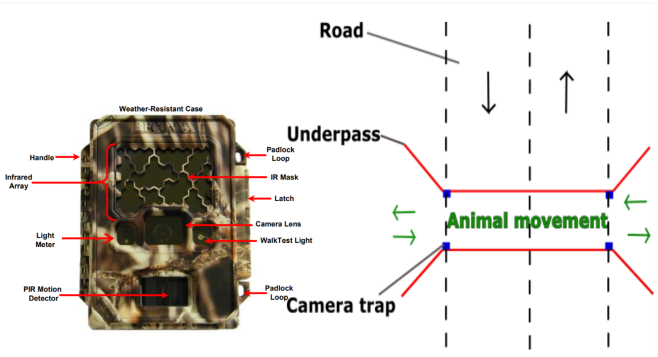

Fig. 2: *Monitoring animal passage structures with camera traps:* Reconyx PC800 HyperFire Professional Semi-Covert Camera Traps (shown on the left) are mounted on structures to monitor wildlife use. Shown on the right is a schematic of the camera layout of an Underpass in Oro Valley, Arizona. It has 4 cameras total (2 at each entrance/exit of the underpass) in order to capture animals that approach and cross the structure.

B. Data Analysis

For image annotation, we are using a self developed annotation tool (DeepGIS) that allows the user to draw a mask around the object of interest and attach a label (Fig. 3). We opted to use our own web-based labeling tool [20] instead of common labeling tools (such as LabelMe) due to data sharing restrictions with AZ Game and Fish. Additionally, the tool can be used to label datasets in any field of science; it is not exclusive to wildlife. For example, it was initially developed to label fruit for precision agriculture [20] and has since been used to label rocks along a fault scarp as well.

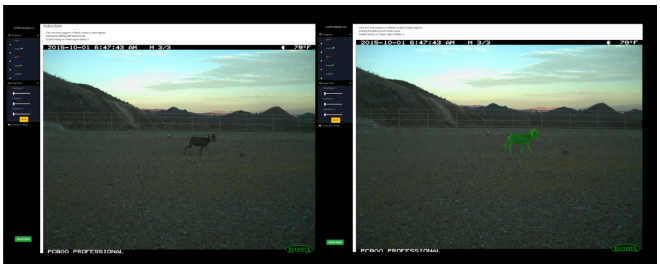

Fig. 3: *Annotating images using web-based labeling tool (deepgis) :* Images are uploaded to the database, and then a mask is drawn around the object of interest after selecting the appropriate label. The original image appears on the left, and the annotation appears on the right.

For the purposes of this study, we have started with 5 label choices for wildlife images: Deer, Elk, Sheep, Coyote, and Cattle. The label “Cattle” was initially “Wildlife” and has since been changed, although the results presented in this paper are from training on the “Wildlife” label. These five labels were chosen because wild ungulates (deer, elk, and bighorn sheep) are the main species of interest, and coyotes and cattle make up a significant portion of the images collected at certain structures.

We randomly selected 56 images to annotate, and these images happened to include 4 of the species of interest as well as ”blank” empty background photos (there were no images of coyotes). The 56 annotated images were then randomly divided into a Training dataset with 40 images and a Validation dataset with 16 images.

The original images and annotations were used as inputs for training the deep neural network Mask RCNN, which is a combination of Faster R-CNN and FCN [9]. It begins by detecting regions of interest (called a Regional Proposed Network), and then generates a bounding box and predicts a classification label. Then, FCN does instance segmentation within the bounding box to generate a pixel-by-pixel segmentation mask over the object of interest [9]

To begin the training process, Mask RCNN was initialized with weights pretrained on COCO 2017 [21]. We then trained Mask RCNN with our Training dataset using a process composed of two phases. In the first phase, Faster RCNN was trained for object detection with a learning rate of 0.001. Then, Faster RCNN and FCN were jointly trained with a smaller learning rate of 0.0001. We used stochastic gradient optimization in both training phases with learning

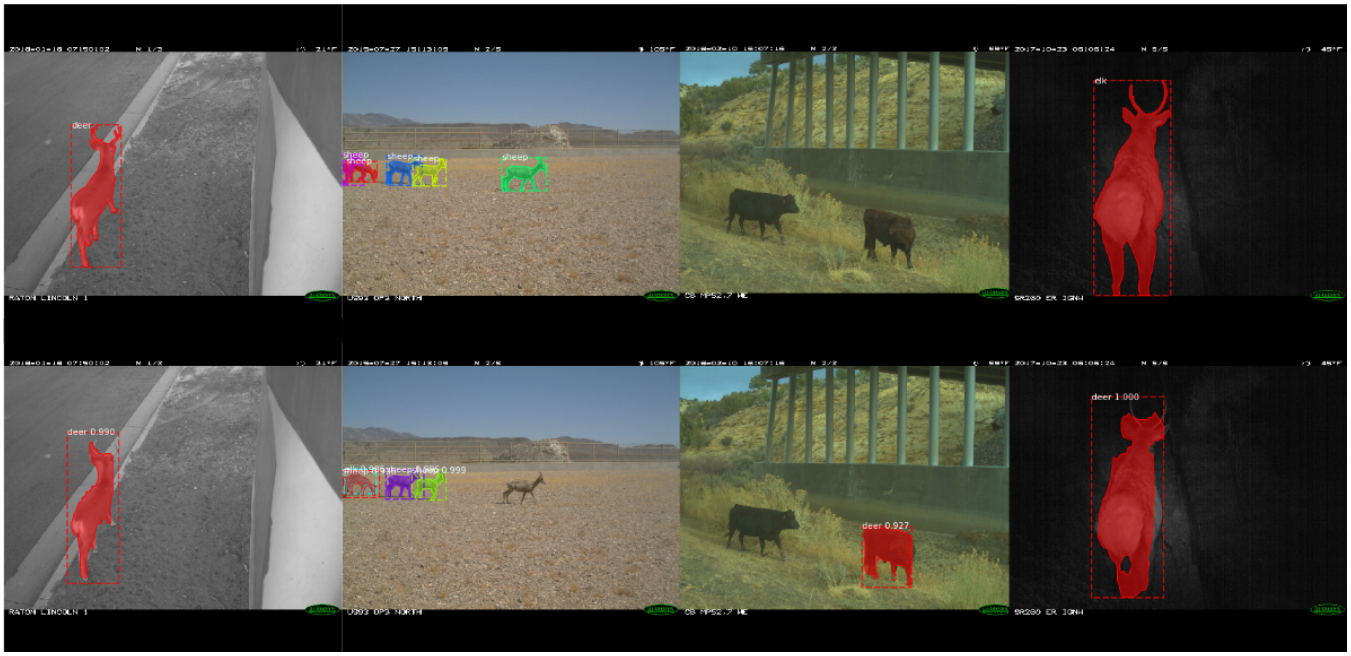

Fig. 4: *Mask RCNN validation predictions*. Bounding box, classification, and segmentation predictions were made on the 16 validation images. The top row shows the original annotations drawn, and the bottom row shows the associated predictions. The example on the left demonstrates correct classification and detection. The second image shows detection, missed detection, correct classification, and misclassification all in one image. Detection, missed detection, and misclassification occur in the third image, and detection and misclassification are shown in the last image.

momentum 0.9 and weight decay 0.0001. We then verified the generalization of the neural network with the Validation dataset, which had not been exposed to the network during training.

It is important to note that although we are currently only interested in the detection information, the segmentation mask generated by the network could be useful for identifying morphological features that can help distinguish the sex and relative age (juvenile vs adult) of an animal, as well as the orientation of an animal in an image.

IV. Results and Analysis

There were 17 animals total in the 16 validation images, and the prediction results are shown in Table I. The network could detect 15 out of the 17 animals, and the classification accuracy among the 15 detections was 40% (Fig. 4). Five of the validation images were blank background and the network did not predict any false positives.

Image #	Original Label	Prediction
0	1 deer	1 elk
1	5 sheep	3 sheep, 1 elk, 1 undetected
2	1 deer	1 deer
3	Blank	Blank
4	2 unlabeled cows	1 deer, 1 elk
5	Blank	Blank
6	1 elk	1 deer
7	2 unlabeled cows	1 deer, 1 undetected
8	1 deer	1 deer
9	1 sheep	1 elk
10	Blank	Blank
11	1 elk	1 elk
12	1 sheep	1 deer
13	Blank	Blank
14	Blank	Blank
15	1 elk	1 deer

TABLE I: *Detection and classification predictions for Validation dataset*. This table compares the original label and the Mask RCNN prediction for each image in the Validation dataset.

In order to improve detection and classification accuracy, we are going to start by annotating and training on more images. Of the 40 images we trained on, 6 contained elk, 14 contained deer, 8 contained sheep, 3 contained cows, and 9 were blank. Several images had more than one animal, so the network was trained on 11 elk labels, 21 deer labels, 14 sheep labels, and 13 blank labels.

As previously mentioned, the fifth label we trained on was “Wildlife” instead of “Cattle”, therefore the images that contained cows were treated as blank since the label “Wildlife” is inaccurate. The label has since been fixed, but this error may account for some of the missed detections and misclassification we see in the results.

Deer and Elk are very similar morphologically, and it can be challenging for even human observers to distinguish between the two species, especially under poor lighting conditions. Increasing the number of elk and deer annotations will hopefully increase the classification accuracy, however we may need to annotate specific features that are unique to each species. For example, Mule Deer are much smaller and have a black tip on the end of their tails, while elk are much larger and have white rear ends with very small white tails.

Annotating morphological features may also help with more specific species and sex identification. For example, the current label “Deer” is a general term, and in the future we would like to be able to distinguish between Mule Deer and White-tailed Deer. As one can tell from the name, White-tailed deer have fluffy white tails, so annotating these features may help with specific species classification. Additionally, annotating antlers may help with identifying male from female ungulates.

V. Future Work

There are several threads we would like to focus on for future work. As previously stated, we will begin by annotating and training on more images in order to improve the detection and classification accuracy for the five labels we currently have. We will also begin working on our annotation tool to create more labels, as well as improve the user interface to allow anyone using the site to create their own labels. Annotating and training on additional species will follow as new labels are created.

We will also focus on tracking individuals across a sequence of images. Work similar to this has been done in precision agriculture, but instead of a moving camera tracking static fruits, a static camera will track moving animals [22]. This information is important so we don't inflate the number of animals interacting with each structure, and to determine if the animal successfully crossed the structure or if it turned away.

This also introduces the importance of identifying the direction of travel. For structures like Escape Ramps and Slope Jumps, an animal crossing the structure does not always mean the structure successfully served its function. If an animal crossed from the right-of-way(ROW) to the non-right-of-way (NROW) side, then the structure successfully served its function. However, if an animal crosses from the NROW to the ROW then the structure failed and adjustments need to be made.

We would also like to use the segmentation masks to try to identify the relative age of an animal (juvenile versus adult). Younger animals should generate smaller masks, but the size of the mask is dependent on the animal's proximity to the camera. To address this, we will try to use the camera field of view specifications and known distances between the camera and a marker in order to determine how close an animal is to the camera.

Longer term goals rely on the success of the features described above. Ultimately we would like to develop a "smart" camera that can be deployed by itself, or as part of a network. It would be able to classify all of the information above in real time, which would cut down on processing time as well as automatically delete false positives in order to save storage. Additionally, structures that have more than one camera sometimes capture an animal entering on one side but not exiting on the other side, and vice versa. In order to increase the likelihood of capturing an animal on both sides of the structure, the cameras could signal each other to turn on when one side has been triggered.

References

[1] R. Inc., *RECOYNYX Hyperfire High Performance Cameras Instruction Manual*, RECONYX Inc.

[2] J. W. Gagnon, C. D. Loberger, S. C. Sprague, K. S. Ogren, S. L. Boe, and R. E. Schweinsburg, "Cost-effective approach to reducing collisions with elk by fencing between existing highway structures," *Human-Wildlife Interactions*, vol. 9, no. 14, 2015.

[3] N. L. Dodd, J. W. Gagnon, S. Boe, K. Ogren, and R. E. Schweinsburg, *Wildlife-vehicle collision mitigation for safer wildlife movement across highways : State Route 260*, 2012, no. 603.

[4] N. L. Dodd and J. W. Gagnon, "Influence of underpasses and traffic on white-tailed deer highway permeability," *Wildlife Society Bulletin*, vol. 35, no. 3, pp. 270–281, 2011.

[5] N. L. Dodd, J. W. Gagnon, A. Manzo, and R. Schweinsburg, "Video surveillance to assess highway underpass use by elk in arizona," *Journal of Wildlife Management*, vol. 21, no. 2, pp. 637–645, 2007.

[6] J. Gagnon, T. Theimer, N. Dodd, and R. Schweinsburg, "Traffic volume alters elk distribution and highway crossings in arizona," *Journal of Wildlife Management*, vol. 71, no. 7, pp. 2318–2323, 2007.

[7] J. Gagnon, T. Theimer, N. Dodd, A. Manzo, and R. Schweinsburg, "Effects of traffic on elk use of wildlife underpasses in arizona," *Journal of Wildlife Management*, vol. 71, no. 7, pp. 2324–2328, 2007.

[8] J. Gagnon, N. Dodd, K. Ogren, and R. Schweinsburg, "Factors associated with use of wildlife underpasses and importance of long-term monitoring," *Journal of Wildlife Management*, vol. 75, no. 6, pp. 1477–1487, 2011.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2018.

[10] P. Tokekar and V. Isler, "Sensor placement and selection for bearing sensors with bounded uncertainty," *2013 IEEE International Conference on Robotics and Automation*, pp. 2515–2520, 2013.

[11] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, June 2008. [Online]. Available: http://dl.acm.org/citation.cfm?id=1390681.1390689

[12] R. Kays, B. Kranstauber, P. Jansen, C. Carbone, M. Rowcliffe, T. Fountain, and S. Tilak, "Camera traps as sensor networks for monitoring animal communities," in *2009 IEEE 34th Conference on Local Computer Networks*, Oct 2009, pp. 811–818.

[13] H. Yousif, J. Yuan, R. Kays, and Z. He, "Animal scanner: Software for classifying humans, animals, and empty frames in camera trap images," *Ecology and Evolution*, vol. 9, no. 4, pp. 1578–1589, 2019.

[14] P. Tokekar, D. Bhadauria, A. Studenski, and V. Isler, "A robotic system for monitoring carp in minnesota lakes," *Journal of Field Robotics*, vol. 27, no. 6, pp. 779–789, 2010.

[15] L. L. Sousa, F. López-Castejón, J. Gilabert, P. Relvas, A. Couto, N. Queiroz, R. Caldas, P. S. Dias, H. Dias, M. Faria, F. Ferreira, A. S. Ferreira, J. Fortuna, R. J. Gomes, B. Loureiro, R. Martins, L. Madureira, J. Neiva, M. Oliveira, J. Pereira, J. Pinto, F. Py, H. Queirós, D. Silva, P. B. Sujit, A. Zolich, T. A. Johansen, J. B. de Sousa, and K. Rajan, "Integrated monitoring of mola mola behaviour in space and time," *PLOS ONE*, vol. 11, no. 8, pp. 1–24, 08 2016.

[16] P. Roy, N. Stefas, C. Peng, H. Bayram, P. Tokekar, and V. Isler, "Robotic surveying of apple orchards," *University of Minnesota, Department of Computer Science, Tech. Rep.*, 2015.

[17] R. Barth, J. IJsselmuiden, J. Hemming, and E. Van Henten, "Data synthesis methods for semantic segmentation in agriculture: A capsicum annuum dataset," *Computers and Electronics in Agriculture*, vol. 144, pp. 284–296, 2018.

[18] Q. Wang, S. Nuske, M. Bergerman, and S. Singh, "Automated crop yield estimation for apple orchards," in *Experimental robotics*. Springer, 2013, pp. 745–758.

[19] P. A. Dias, A. Tabb, and H. Medeiros, "Apple flower detection using deep convolutional networks," *Computers in Industry*, vol. 99, pp. 17 – 28, 2018.

[20] S. W. Chen, S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, C. J. Taylor, and V. Kumar, "Counting apples and oranges with deep learning: A data-driven approach," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 781–788, April 2017.

[21] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014.

[22] J. Das, G. Cross, C. Qu, A. Makineni, P. Tokekar, Y. Mulgaonkar, and V. Kumar, "Devices, systems, and methods for automated monitoring enabling precision agriculture," in *Automation Science and Engineering (CASE), 2015 IEEE International Conference on*. IEEE, 2015, pp. 462–469.