Tailored object detection methods for trail cameras

First author, second author, third author, and Alastair Frank,

***Abstract*—Motion sensor camera traps are used extensively in the fields of zoology and ecology. Camera traps have been around for decades and have revolutionized wildlife research and conservation due to their ability to capture information with little expense, and minimal disturbance to the wildlife. Recently, the ability of computers to recognize certain aspects of images has led to the use of these techniques to save human time in examining the camera trap images. In this work object detection methods are applied to peregrine nest cameras. Our application of these techniques is able to achieve XX% accuracy. Additionally due to the nature of the sequence of images taken, any mistakes made can be rectified by using a time-sequence of the data.**

***Index Terms*—Deep learning, Animal identification, Camera-trap images, Constitutional Neural Networks**

I. Introduction

Surprisingly the concept of a camera trap has been around since the 1890s [1]. However camera traps would not find significant use for over a century later. Cameras traps have since been used increasingly by researchers to record and track different biological phenomena [2]. With the increase of the technolical ability of camera trap in the 1990s led to a flood of research. They have been used to study nest ecology [3], vertebrate activity [4], rare species identification [5], research of rare individuals [6], population estimation [7], and survivorship [8] among many others.

Camera traps have become more popular and have become a important piece of a lot of zoological research because it enables researchers to collect data inexpensively, inobstructively, and with high frequency [2]. This has led to the rapid expansion of the use of camera traps in research. As more research makes use of camera traps, difficulties dealing with the analysis of the data has arisen. The authors of [1] identified four problems restricting the use of camera trap in research: (1) cataloging imagery is slow and usually lags behind acquisition, (2) user entry is tedious and can be error prone, (3) inconsistent filing complicated data analysis and sharing, and (4) the struggle to keep pace with analysis and data management restricts the deployment of additional cameras. In [1] the solution presented is a systematic procedure of analysis, making use of different computer programs to help in the organization of the data.

However, systematic organization still requires the time of an expert to analyze the images. Not only does this tie up researchers from other tasks, but limits the number of cameras due to the shear number of images to be processed. With the advancement of computer vision techniques, several works have been done to make use of this to automatically identify images. Initially many techniques made use of hand-designed features to classify animals [9]. Most recently, deep learning has emerged as the most powerful computer vision technique. It has shown ability to solve many computer image problems and is able to surpass human experts in many ways. Deep learning has been used to distinguish birds from mammals [10] as well as in larger class problems such as [11]. Some of the best results were obtained on the SS data-set. Authors in [12] obtained a 57% correct in species identification of a 26 class subset of the data. Authors in [13] extended these results by obtaining 92% accuracy despite taking into account the full 48 class problem.

In comparison to this previous work, we apply DNN to relatively smaller problems, having at most three classes. The camera traps are nest-sites. Although less complicated than large class problems, we show higher quality in terms of class identification with significantly reduced training examples. Using the characteristics of the cameras which take three consecutive images during a motion event, allows the already high results to be supplemented and a large percent of the mistakes to be rectified creating a near expert level quality allowing completely automated image analysis and production of results. In addition, the nest cameras contain more movement and information than trail cameras, for example for the 28 cameras in 2015, the project captured over 2.1 million photos, making up 841 GB of information. Thus the different aspects of the situation requires a modified solution.

Most biological research groups will not have significant computational resources available. This would be even more notable in the nest traps, due to the shear number of images produced. Thus computational aspects should be taken into account. YOLOv3 [14] is a DNN which is designed to run fast than comparative techniques. This has a trade-off of quality and it has been shown to have slightly lower scores than other techniques such as XXX [14].

In Section II the arctic raptors project is introduced. Section III outlines the steps taken in the training and implementation of the computer vision techniques. Section IV-A describes the use of the time series of data to correct possible errors in object detection. Finally the paper is concluded in Section V.

II. Arctic Raptors

The arctic raptors research group is a longterm research project studying peregrine falcons in Arctic Canada. The primary study area consists of the region of Rankin Inlet, Nunavut. Positioned on the west coast of Hudsons Bay, and is home to the densest known population of Peregrine falcons in the arctic. The project is headed by the University of Alberta

R. S. C. Winter is with the Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: rswinter@ualberta.ca).
Kevin Hawkshaw, Erik Hedlin and Alastair Frank are with the Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: hawkshaw@ualberta.ca,hedlin@ualberta.ca,kfranke@ualberta.ca).
Manuscript received XXX, XXXX; revised XXX, XXXX.

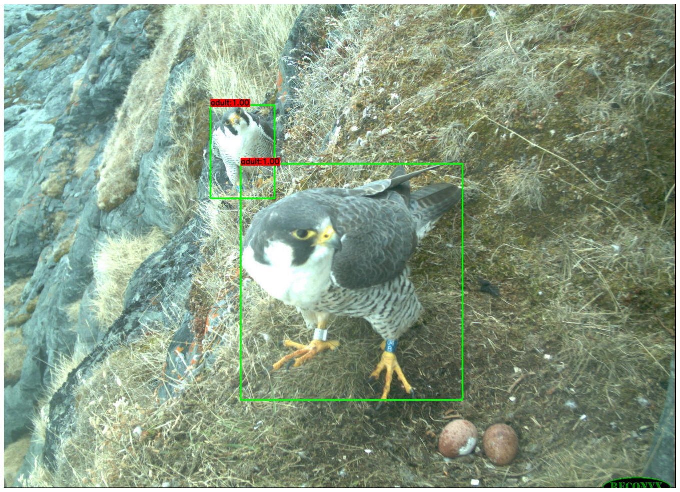

Fig. 1. Adults detection.

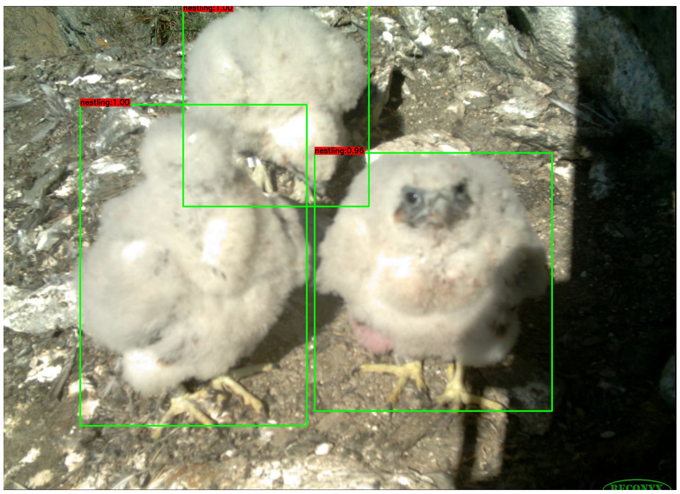

Fig. 2. Detection of nestlings

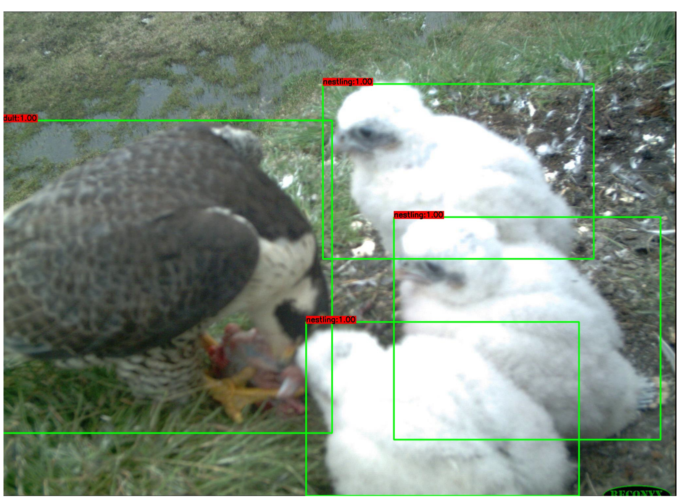

Fig. 3. Detection showing nestlings and adults.

Fig. 4. Egg detection.

Fig. 5. Band detection.

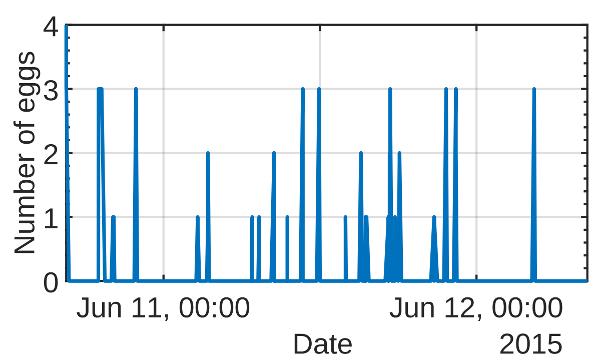

Fig. 6. Number of eggs detected over time. This value can be used to determine the incubating behavior.

and the University of Quebec at Rimouski, and is a project member of ArcticNet network. One of the research areas is the nesting program. This involves the study of the survival rate and nesting behavior of the raptors. The study involves 36 camera traps, each consisting of a camera aimed at a potential or suspected nesting site. The cameras consist of Reconyx HP2X cameras. The cameras are set to acquire one photo every 15 minutes and three photos one second apart in the case the camera is triggered by motion.

At the beginning of the arctic raptors project, researchers in the field would visit the nests on a daily basis and track the nestling survivability and behavior []. When camera traps were applied they allowed a much better understanding of the mortality, if a nest was taken by a fox sometime between researchers visit, it could now be known.

Some problems of interest that could be automated could be the amount of time parents spend on the nest, the hatching dates, the nestling mortality.

III. Methods

For the problems of interest which where outlined in Section II can be solved using object detection of a three class problem. Where eggs, adult birds, and nestlings, bands are the four classes. Different numbers of images for the different classes were obtained from the dataset for training purposes. This set consisted of 91 images with bands, 106 images with eggs, and 450 images containing either adults and nestlings. These images were augmented with bounding boxes of the objects of interest and then the training for the darknet training was undertaken. Thanks to the relative size of the class number, the training converged rapidly towards a high quality result.

IV. Results

A. Time analysis

The cameras used in this study take three pictures when detecting motion as well as a image every 15 minutes if no motion is detected. This time series data can be used to increase the correct decisions made when examining a single image. Firstly, the adult birds are larger and move through more of the frame at any point. Therefore, depending on the nestlings age, a motion event can be more likely attributed to the presence of an adult. Additionally, in a motion event three images can be used to see if an adult is present. The motion event usually takes the form of an adult coming to the nest, a feeding event, or an adult leaving the nest. Thus if the three images contain an adult in the first two images and no detection in the third image, that might indicate that the adult has left the nest site. If the next image taken 15 minutes later, also contains a low probability of an adult, together these results can be added together to make a conclusion on what event took place.

V. Conclusion

This paper demonstrates the high accuracy of quality of results from trail cameras used in conjunction with convolution deep learning methods. Due to limited class size and high correlated training sets, the results demonstrate a XX% overall accuracy rate. Due to the nature of the cameras, which take three photos in rapid succession during a motion trigger event, the time series data can be used to increase the quality of the results further, to XX%. For many research groups who have more homogeneous environments and a limited number of classes these techniques could be easily modified for their use. Although training of a DNN generally requires significant computational resources, making use of YOLOv3 minimizes the computational burden during image analysis, making this a possibility for a larger number of biological based research groups.

References

[1] G. Harris, R. Thompson, J. L. Childs, and J. G. Sanderson, “Automatic storage and analysis of camera trap data,” *The Bulletin of the Ecological Society of America*, vol. 91, no. 3, pp. 352–360, 2010.

[2] A. F. O’Connell, J. D. Nichols, and K. U. Karanth, *Camera traps in animal ecology: methods and analyses*. Springer Science & Business Media, 2010.

[3] R. E. Major and G. Gowing, “An inexpensive photographic technique for identifying nest predators at active nests of birds.,” *Wildlife Research*, vol. 21, no. 6, pp. 657–665, 1994.

[4] S. M. Carthew and E. Slater, “Monitoring animal activity with automated photography,” *The Journal of Wildlife Management*, pp. 689–692, 1991.

[5] A. K. Surridge, R. J. Timmins, G. M. Hewitt, and D. J. Bell, “Striped rabbits in southeast asia,” *Nature*, vol. 400, no. 6746, p. 726, 1999.

[6] D. Martyr, “Important findings by ffi team in kerinci seblat, sumatra, indonesia,” 1997.

[7] A. F. O’Connell Jr, N. W. Talancy, L. L. Bailey, J. R. Sauer, R. Cook, and A. T. Gilbert, “Estimating site occupancy and detection probability parameters for meso-and large mammals in a coastal ecosystem,” *The Journal of Wildlife Management*, vol. 70, no. 6, pp. 1625–1633, 2006.

[8] K. U. Karanth, J. D. Nichols, N. S. Kumar, and J. E. Hines, “Assessing tiger population dynamics using photographic capture–recapture sampling,” *Ecology*, vol. 87, no. 11, pp. 2925–2937, 2006.

[9] K. R. Swinnen, J. Reijniers, M. Breno, and H. Leirs, “A novel method to reduce time investment when processing videos from camera trap studies,” *PloS one*, vol. 9, no. 6, p. e98881, 2014.

[10] A. Gomez, G. Diez, A. Salazar, and A. Diaz, “Animal identification in low quality camera-trap images using very deep convolutional neural networks and confidence thresholds,” in *International Symposium on Visual Computing*, pp. 747–756, Springer, 2016.

[11] G. Chen, T. X. Han, Z. He, R. Kays, and T. Forrester, “Deep convolutional neural network based species recognition for wild animal monitoring,” in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 858–862, IEEE, 2014.

[12] A. G. Villa, A. Salazar, and F. Vargas, “Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks,” *Ecological Informatics*, vol. 41, pp. 24–32, 2017.

[13] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, “Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. E5716–E5725, 2018.

[14] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.