Automated camera trap species recognition made easy: Using entry-level hardware and few training data

Paul Berry, Morgan Hauptfleisch, Nichola Knox

Computer vision methods used to analyse camera trap photos are usually computationally expensive, require large training datasets and typically focus on only one species per photograph or rely on static backgrounds between sequential images. In contrast, our proposed method requires only an entry-level computer and relatively few training data while handling multi-species photos with changing backgrounds. It is able to distinguish between four large mammal species common to the Iona–Skeleton Coast TFCA, namely giraffe, impala, oryx and zebra. Trained on readily available online images and applied to 4000 camera trap photos, the system yielded a recall of 59.1% in detecting the presence of animals in camera trap photos. Precision in detecting animals was 100% while precision in distinguishing between the four species of interest, namely giraffe, impala, oryx and zebra, was 96.8%. Based on the results, the method could be used to filter large raw datasets for photos containing animals, and to label or pre-label photos by species for further analysis. This may make it useful to aid in compiling species inventories, document animal migration, map species distributions and estimate densities.

Introduction

Camera traps are used throughout the world to monitor wildlife populations (O'Connell and Nichols, 2011; Trolliet et al., 2014) but tend to amass large numbers of photographs within relatively short periods of time (Swanson et al., 2015). The costly and time-consuming processing of large volumes of photographs limits this monitoring method (Harris et al., 2010; Newey et al., 2015). Recently, approaches to the visual analysis of camera trap photos using deep learning convolutional neural networks (CNNs) have been introduced (Norouzzadeh et al., 2018; Schneider, Taylor and Kremer, 2018; Nguyen et al., 2017; Yousif et al., 2019), but computation and hardware requirements may place them beyond the means of many camera traps users. Our objectives with this study thus were:

- to develop a computationally inexpensive computer vision method for the automatic recognition of multiple species in camera trap images and
- to test the efficacy of the method in monitoring the presence of large wildlife species.

Methods

Camera trapping

Sequences of 1000 photographs from two waterholes and two fence-crossing game trails (Figure 1) at the Etosha Heights private game reserve (19.24° S 15.08° E) were sampled, resulting in a sample size of 4000 images.

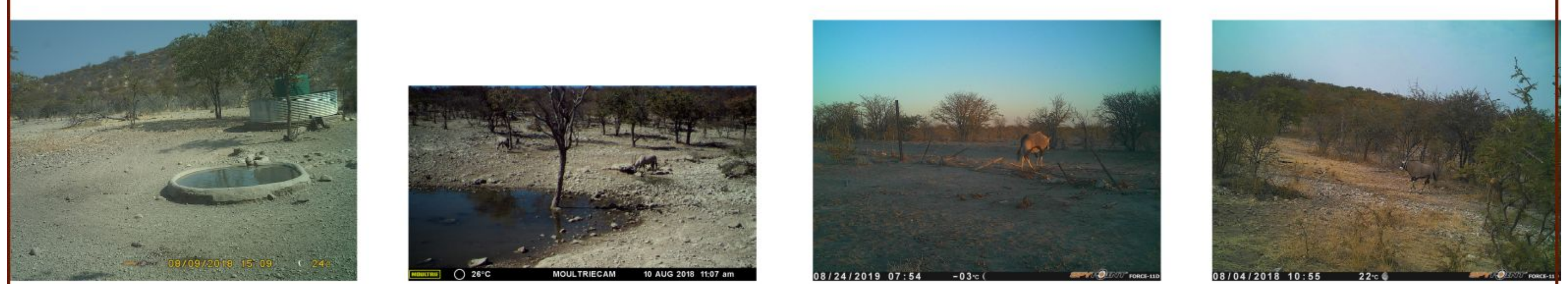

Fig. 1: The four camera trap sites used in this study.

Manual assessment

The presence of giraffe, impala, oryx and zebra was manually assessed for each photo. Other large mammal species which are either less common or do not occur at all in the Iona–Skeleton Coast TFCA were ignored.

Computer vision models

1. The YOLO v3 object detector (Redmon and Farhadi, 2018) was used to detect and localise animals in photos.
2. The Inception v3 image classification model (Szegedy et al., 2015) was used to identify by species the animals localised by the object detector.

Training

A total of 5291 training images (1462 images of giraffe, 1223 of impala, 1513 of oryx, and 1093 of zebra) were sourced online. These images were resized (resampled) to 299 × 299 pixels for input to the image classifier.

Fig. 2: Examples of training images sourced online (Wikimedia Commons).

To distinguish between the four large mammal species, an Inception v3 model trained on the ImageNet dataset (Deng et al., 2009) was used as a feature extractor. A softmax classifier was retrained to map the output of the feature extractor to the mammal species relevant to the study using Tensorflow (version 1.13.1). The image set used for training was randomly split into 80% for training, 10% for validation and 10% for testing. A learning rate of 0.01 and training and validation batch sizes of 100 images each were used and 4000 training iterations were run.

Inference

Step 1: Object detection

The OpenCV library (Bradski, 2000) inference module was used. Input images were resized (resampled) to 416 × 416 pixels. Output confidence threshold was set to 0.05 and the non-maximum suppression threshold was set to 0.4 (Nayak, 2018).

Step 2: Image classification

The sub-images of animals generated in Step 1 were input to the retrained image classifier at an input resolution of 299 × 299 pixels (Abadi et al., 2017) to distinguish between the four large mammal species of interest.

Evaluation

The computer inference for each photograph and species (giraffe, impala, oryx and zebra) was established as true positive (TP), true negative (TN), false positive (FP) or false negative (FN). The following performance metrics were calculated:

$$Recall = \frac{TP}{TP+FN} \qquad Precision = \frac{TP}{TP+FP} \qquad Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Species other than these four were excluded in the evaluation of species recognition performance.

Results

The object detector yielded a recall of 59.1% and a precision of 100% in detecting animals. Recall in classifying photos according to species ranged from 46.4 to 68.0% while precision ranged from 90.5% to 100% (Table 1).

Table 1: The number of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), recall, precision and accuracy for four large mammal species.

Species	TP	TN	FP	FN	Precision	Recall	Accuracy
giraffe	133	3714	14	139	0.905	0.489	0.962
impala	69	3859	6	66	0.920	0.511	0.982
oryx	595	3108	17	280	0.972	0.680	0.926
zebra	315	3321	0	364	1.000	0.464	0.909
overall	1112	14002	37	849	0.968	0.567	0.945

Typically, not all individuals of a species were detected, but those detected were usually correctly classified by species. In some cases, the method detected poorly exposed animals while in other cases it failed to detect animals obvious to the eye (Figure 3).

Fig. 3: Examples of object detection.

Discussion

We demonstrated that camera trap photos can be automatically classified by species using an entry-level computer and around 1000 training images per species. The YOLO v3 object detector (Redmon and Farhadi, 2018) was available pre-trained and thus ready to use. The Inception v3 image classifier (Szegedy et al., 2015) was retrained, by transfer learning, with relatively little computational effort, on the four large mammal species of interest.

Although animals were detected in only 59.1% of photographs in which they occurred, the visit of a species to a camera trap site in many cases can result in several photographs being taken, arguably increasing the probability that the species is detected per visit. Notably, precision in detecting animals was 100%, meaning that if the system claimed there was an animal present, this was correct.

Our proposed computer vision method has the potential to assist to with a variety of tasks, by filtering out photographs that contain animals, classifying the predominant large mammals by species, or a combination of these. It could serve at least as a preliminary classification of photographs, which then could be verified and corrected manually.

The compilation of standardised, publicly accessible training sets for a range of species could be considered for future work. Researchers could then select ready-made training sets on which to train the image classifier for identifying species relevant to their projects.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irv- ing, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. and Zheng, X., 2017. Tensor- Flow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). Savannah, GA: USENIX As- sociation, pp.265–283.

Bradski, G., 2000. The OpenCV Library. http://www.drdobbs.com/open-source/the-opencv- library/184404319 [Accessed 11 April 2019].

Deng, J., Dong, W., Socher, R., Li, L.J., Kai Li and Li Fei-Fei, 2009. ImageNet: A large- scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, pp.248–255. Available from: https://doi.org/10.1109/CV PR.2009.5206848.

Harris, G., Thompson, R., Childs, J.L. and Sanderson, J.G., 2010. Automatic storage and analysis of camera trap data. The Bulletin of the Ecological Society of America, 91(3), pp.352– 360. Available from: https://doi.org/10.1890/0012-9623-91.3.352.

Nayak, S., 2018. Deep learning based object detection using YOLOv3 with OpenCV (Python 14 / C++). https://www.learnopencv.com/deep-learning-based-object-detection-using-yolov3- with-opencv-python-c/ [Accessed 30 January 2019].

Newey, S., Davidson, P., Nazir, S., Fairhurst, G., Verdicchio, F., Irvine, R.J. and Wal, R. van der, 2015. Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective. Ambio, 44(4), pp.624–635. Available from: https: //doi.org/10.1007/s13280-015-0713-1.

Nguyen, H., Maclagan, S.J., Nguyen, T.D., Nguyen, T., Flemons, P., Andrews, K., Ritchie, E.G. and Phung, D., 2017. Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Tokyo, Japan: IEEE, pp.40–49. Available from: https://doi.org/10.1109/DSAA.2017.31.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C. and Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera- trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), pp.E5716–E5725. Available from: https://doi.org/10.1073/pnas.1719367115.

O'Connell, A.F. and Nichols, J.D., eds., 2011. Camera traps in animal ecology: Methods and analyses. Tokyo: Springer. Redmon, J. and Farhadi, A., 2018. YOLOv3: An incremental improvement. arXiv:1804.02767. 1804.02767.

Schneider, S., Taylor, G.W. and Kremer, S.C., 2018. Deep learning object detection methods for ecological camera trap data. arXiv:1803.10842 [cs]. 1803.10842.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. and Packer, C., 2015. Snap- shot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific Data, 2(150026), pp.1–14. Available from: https: //doi.org/10.1038/sdata.2015.26.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2015. Rethinking the Inception architecture for computer vision. arXiv:1512.00567 [cs]. 1512.00567.

Trolliet, F., Huynen, M.C., Vermeulen, C. and Hambuckers, A., 2014. Use of camera traps for wildlife studies. A review. Biotechnology, Agronomy, Society and Environment, 18(3), pp.446–454.

Yousif, H., Yuan, J., Kays, R. and He, Z., 2019. Animal Scanner: Software for classifying humans, animals, and empty frames in camera trap images. Ecology and Evolution, 9(4), pp.1578–1589. Available from: https://doi.org/10.1002/ece3.4747.

SCIONA
BRIDGING THE KUNENE

Funded by
the European Union

NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

INSTITUTO SUPERIOR DE CIÊNCIAS DE EDUCAÇÃO
ISCED
HUÍLA