Received: 7 September 2018 | Accepted: 4 October 2018
DOI: 10.1111/2041-210X.13120

APPLICATION

Methods in Ecology and Evolution
BRITISH ECOLOGICAL SOCIETY

Machine learning to classify animal species in camera trap images: Applications in ecology

Michael A. Tabak[1,2] | Mohammad S. Norouzzadeh[3] | David W. Wolfson[1] | Steven J. Sweeney[1] | Kurt C. Vercauteren[4] | Nathan P. Snow[4] | Joseph M. Halseth[4] | Paul A. Di Salvo[1] | Jesse S. Lewis[5] | Michael D. White[6] | Ben Teton[6] | James C. Beasley[7] | Peter E. Schlichting[7] | Raoul K. Boughton[8] | Bethany Wight[8] | Eric S. Newkirk[9] | Jacob S. Ivan[9] | Eric A. Odell[9] | Ryan K. Brook[10] | Paul M. Lukacs[11] | Anna K. Moeller[11] | Elizabeth G. Mandeville[2,12] | Jeff Clune[3] | Ryan S. Miller[1]

[1]Center for Epidemiology and Animal Health, United States Department of Agriculture, Fort Collins, Colorado; [2]Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming; [3]Computer Science Department, University of Wyoming, Laramie, Wyoming; [4]National Wildlife Research Center, United States Department of Agriculture, Fort Collins, Colorado; [5]College of Integrative Sciences and Arts, Arizona State University, Mesa, Arizona; [6]Tejon Ranch Conservancy, Lebec, California; [7]Savannah River Ecology Laboratory, Warnell School of Forestry and Natural Resources, University of Georgia, Aiken, South Carolina; [8]Range Cattle Research and Education Center, Wildlife Ecology and Conservation, University of Florida, Ona, Florida; [9]Colorado Parks and Wildlife, Fort Collins, Colorado; [10]Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon, SK, Canada; [11]Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, Montana and [12]Department of Botany, University of Wyoming, Laramie, Wyoming

Correspondence
Michael A. Tabak
Email: tabakma@gmail.com
and
Ryan S. Miller
Email: ryan.s.miller@aphis.usda.gov

Funding information
U.S. Department of Energy, Grant/Award Number: DE-EM0004391; USDA Animal and Plant Health Inspection Service, National Wildlife Research Center and Center for Epidemiology and Animal Health; Colorado Parks and Wildlife; Canadian Natural Science and Engineering Research Council; University of Saskatchewan; Idaho Department of Game and Fish

Handling Editor: Theoni Photopoulou

Abstract

1. Motion-activated cameras ("camera traps") are increasingly used in ecological and management studies for remotely observing wildlife and are amongst the most powerful tools for wildlife research. However, studies involving camera traps result in millions of images that need to be analysed, typically by visually observing each image, in order to extract data that can be used in ecological analyses.
2. We trained machine learning models using convolutional neural networks with the ResNet-18 architecture and 3,367,383 images to automatically classify wildlife species from camera trap images obtained from five states across the United States. We tested our model on an independent subset of images not seen during training from the United States and on an out-of-sample (or "out-of-distribution" in the machine learning literature) dataset of ungulate images from Canada. We also tested the ability of our model to distinguish empty images from those with animals in another out-of-sample dataset from Tanzania, containing a faunal community that was novel to the model.
3. The trained model classified approximately 2,000 images per minute on a laptop computer with 16 gigabytes of RAM. The trained model achieved 98% accuracy at identifying species in the United States, the highest accuracy of such a model to date. Out-of-sample validation from Canada achieved 82% accuracy and correctly identified 94% of images containing an animal in the dataset from Tanzania. We provide an R package (Machine Learning for Wildlife Image Classification) that

allows the users to (a) use the trained model presented here and (b) train their own model using classified images of wildlife from their studies.

4. The use of machine learning to rapidly and accurately classify wildlife in camera trap images can facilitate non-invasive sampling designs in ecological studies by reducing the burden of manually analysing images. Our R package makes these methods accessible to ecologists.

KEYWORDS

artificial intelligence, camera trap, convolutional neural network, deep neural networks, image classification, machine learning, R package, remote sensing

1 | INTRODUCTION

Camera traps are increasingly used to remotely observe wildlife over large geographical areas with minimal human involvement and have made considerable contributions to ecology (Howe, Buckland, Després-Einspenner, & Kühl, 2017; O'Connell, Nichols, & Karanth, 2011; Rovero, Zimmermann, Bersi, & Meek, 2013). A common limitation is these methods lead to a large accumulation of images which must be first classified in order to be used in ecological studies (Niedballa, Sollmann, Courtiol, & Wilting, 2016; Swanson et al., 2015). The burden of manually viewing and classifying images often constrains studies by reducing the sampling intensity (e.g., number of cameras deployed), limiting the geographical extent and duration of studies. Recently, machine learning has emerged as a potential solution for automatically classifying images from camera traps (Chen, Han, He, Kays, & Forrester, 2014; Gomez Villa, Salazar, & Vargas, 2017; Norouzzadeh et al., 2018; Swinnen, Reijniers, Breno, & Leirs, 2014; Yu et al., 2013).

We sought to develop a machine learning approach that can be applied across study sites and provide software that ecologists can use for identification of wildlife in their own camera trap images. Using over three million identified images of wildlife from camera traps from five locations across the United States, we trained and tested deep learning models that automatically classify wildlife. We provide an R package (Machine Learning for Wildlife Image Classification [MLWIC]) that allows researchers to classify camera trap images from North America or train their own machine learning models to classify images.

2 | MATERIALS AND METHODS

2.1 | Camera trap images

Species in camera trap images from five locations across the United States (California, Colorado, Florida, South Carolina and Texas) and one location from Canada (Saskatchewan) were identified manually by researchers (see Appendix S1 for a description of each field location). Images were either classified by a single wildlife expert or evaluated independently by two researchers; any conflicts were decided by a third observer (Appendix S1). If any part of an animal (e.g., leg or ear) was identified as being present in an image, this was included as an image of the species. If an image did not contain any animals, it was classified as empty. The images from Canada were not used for training, but were used as an out-of-sample dataset for validation. This resulted in a total of 3,741,656 classified images that included 27 species or groups (see Table 1) across the study locations. We present these images and their classifications for other scientists to use for model development as the North American Camera Trap Images (NACTI) dataset. To increase processing speed, images were resized to 256 × 256 pixels following the methods and using the Python script of Norouzzadeh et al. (2018). To have a more robust model, we randomly applied different label-preserving transformations (cropping, horizontal flipping, and brightness and contrast modifications), called data augmentation (Krizhevsky, Sutskever, & Hinton, 2012).

We randomly selected 90% of the classified images for each species or group to train the model and 10% of the images to test it. However, we wanted to evaluate the model's performance for each species present at each study site, so we used conditional sampling in which we altered training–testing allocation for the rare situations (four total instances) where there were few classified images of a species at a site. Specifically, with 1–9 classified images for a species at a site (two instances), we used all of these images for testing and none for training (the model was trained using only images of these species from other sites); for site-species pairs with 10–30 images (two instances), 50% were used for training and testing; and for >30 images per site for each species, 90% were allocated to training and 10% to testing (Appendices S3–S7 show the number of training and test images for each species at each site). This resulted in 3,367,383 images used to train the model and 374,273 images used for testing.

2.2 | Machine learning process

As machine learning methods are new to many ecologists, we provide a brief introduction in a supplement (Appendix S2). Following Norouzzadeh et al., we trained a deep convolutional neural network (ResNet-18) architecture (He, Zhang, Ren, & Sun, 2016) using the TensorFlow framework (Adabi et al., 2016) using Mount

TABLE 1 Model performance for each species or group

Species or group name	Scientific name	Number of training images	Number of test images	Recall	Top-5 recall	Precision	False-positive rate	False-negative rate
Moose	*Alces alces*	8,967	997	0.98	1.00	0.98	0.02	0.02
Cattle	*Bos taurus*	1,817,109	201,903	0.99	1.00	0.99	0.01	0.01
Quail	*Callipepla californica*	2,039	236	0.91	0.96	0.93	0.07	0.09
Canidae	Canidae	20,851	2,321	0.89	0.99	0.93	0.07	0.11
Elk	*Cervus canadensis*	185,390	20,606	0.99	1.00	0.99	0.01	0.01
Mustelidae	Mustelidae	1,991	223	0.77	0.99	0.87	0.13	0.23
Corvid	Corvidae	4,037	452	0.84	1.00	0.80	0.20	0.16
Armadillo	*Dasypus novemcinctus*	8,926	993	0.89	0.99	0.93	0.07	0.11
Turkey	*Meleagris gallopavo*	3,919	447	0.90	1.00	0.90	0.10	0.10
Opossum	*Didelphis virginiana*	1,804	210	0.79	0.96	0.88	0.12	0.21
Horse	*Equus* spp.	2,517	281	0.94	0.99	0.94	0.06	0.06
Human	*Homo sapiens*	88,667	9,854	0.96	1.00	0.97	0.03	0.04
Rabbits	Leporidae	17,768	1,977	0.95	1.00	0.96	0.04	0.05
Bobcat	*Lynx rufus*	22,889	2,554	0.91	0.99	0.94	0.06	0.09
Striped skunk	*Mephitis mephitis*	10,331	1,154	0.95	0.98	0.96	0.04	0.05
Rodent	Rodentia	3,279	366	0.79	0.98	0.88	0.12	0.21
Mule deer	*Odocoileus hemionus*	87,700	8,543	0.98	1.00	0.98	0.02	0.02
White-tailed deer	*Odocoileus virginianus*	87,900	1,360	0.94	1.00	0.95	0.05	0.06
Raccoon	*Procyon lotor*	42,948	4,781	0.90	1.00	0.89	0.11	0.10
Mountain lion	*Puma concolor*	13,272	1,484	0.92	0.98	0.97	0.03	0.08
Squirrel	*Sciurus* spp.	59,072	6,566	0.97	1.00	0.95	0.05	0.03
Wild pig	*Sus scrofa*	287,017	31,893	0.98	1.00	0.98	0.02	0.02
Fox	*Vulpes vulpes* and *Urocyon Cinereoargenteus*	10,749	1,204	0.91	0.99	0.94	0.06	0.09
Black bear	*Ursus americanus*	79,628	8,850	0.95	1.00	0.98	0.02	0.05
Vehicle		23,413	2,602	0.93	1.00	0.95	0.05	0.07
Bird	Aves	61,063	6,787	0.94	1.00	0.95	0.05	0.06
Empty		414,119	46,016	0.96	1.00	0.94	0.06	0.04
Total		3,367,365	364,660	0.98	1.00	0.98		

Moran, a high performance computing cluster (Advanced Research Computing Center, 2012). We used the ReLU activation function, 55 epochs, a backpropagation algorithm of Stochastic Gradient Descent with Momentum (Goodfellow, Bengio, & Courville, 2016), and the learning rate (η) and weight decay varied by epoch number as described in Appendix S8.

In Appendix S2, we describe the calculation of metrics including accuracy, recall, precision and false-positive and false-negative error rates. Briefly, recall and precision are measures of the model's performance at correctly identifying each species. We fit generalized additive models (GAMs) to the relationship between recall and the logarithm (base 10) of the number of images used to train the model; see Appendix S9 for a description of this model. We also calculated the recall and rates of error specific to each of the five datasets from which images were acquired.

2.3 | Model validation

To evaluate how the model would perform for a completely new study site in North America, we used a dataset of 5,900 classified images of ungulates (moose, cattle, elk and wild pigs) from Saskatchewan, Canada, by running the trained model on these images. We also evaluated the ability of the model to operate on images with a completely different species community (from Tanzania) to determine the model's ability to correctly classify images as having an animal or being empty when encountering new species that it has not been trained to recognize. This was done using 3.2 million classified images from the Snapshot Serengeti dataset (Swanson et al., 2015).

3 | RESULTS

Our model performed well, achieving 97.6% accuracy of identifying the correct species with the top guess. The top-5 accuracy was >99.9%. Figure 1 provides examples of image classification by the model. The model confidence in the correct answer varied, but was mostly >95%; see Figure 2 for confidences for each image for three example species. In Appendix S10, we present a confusion matrix comparing the classifications by the model with those from manual classification. Supporting a similar finding for camera trap images in Norouzzadeh et al. (2018), and a general trend in deep learning (Goodfellow et al., 2016), species and groups that had more images available for training were classified more accurately (Figure 3, Table 1). GAMs relating the number of training images with recall predicted 95% recall could be achieved when

(a) Correct classification by model

(b) Incorrect classification by model

(a)

Model Guess	Confidence (%)
Wild pig	96.11
Cattle	2.38
Empty	1.49
White-tailed deer	<0.1
Moose	<0.1

Answer from human classifiers: Wild pig

(b)

Model Guess	Confidence (%)
Wild pig	48.82
Cattle	31.27
Moose	16.93
Black bear	2.51
Bobcat	0.51

Answer from human classifiers: Cattle

FIGURE 1 Examples of images that could be difficult to classify. The model correctly identifies a wild pig (a) by seeing only its hindquarters and tail (right side of image). The model incorrectly classifies a cattle as a wild pig (b), as only an ear is visible in the image; note that the model has relatively low confidence in the top guess for this image. Nevertheless, cattle are within the top-5 guesses for this image, so while it is incorrect, it counts towards the top-5 recall for cattle

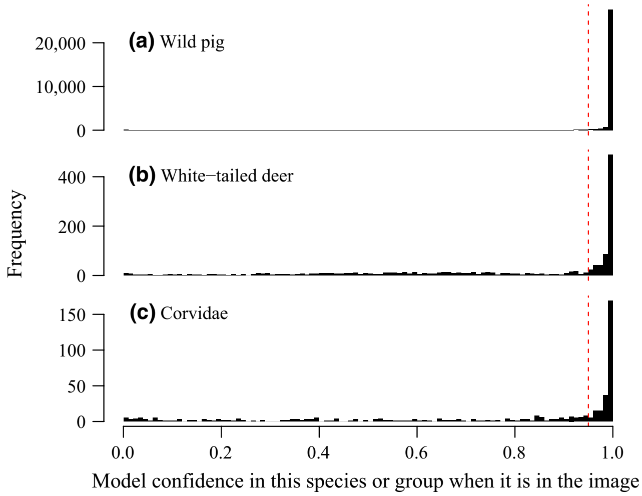

FIGURE 2 Histograms represent the confidence assigned by all of the top-5 guesses by the model for each of these three example species when it was present in an image. The dashed line represents 95% confidence; the majority of model-assigned confidences were greater than this value

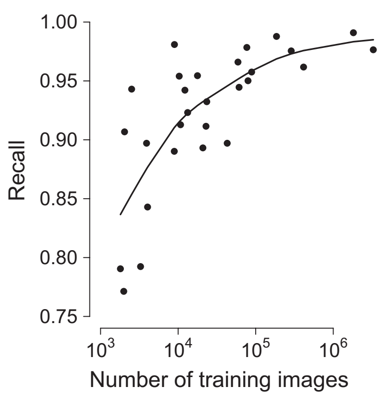

FIGURE 3 Model recall (the ability of the model to recognize species) increased with the size of the training dataset for that species. Points represent each species or group of species. The line represents the result of generalized additive models relating the two variables (see Appendix S9 for details)

approximately 54,000 training images were available for a species or group. However, for several species and groups, 95% recall was achieved with fewer than 50,000 images (Figure 3). We found there was not a large effect of daytime versus night-time on accuracy in the model as daytime accuracy was 98.2% and night-time accuracy was 96.6%. The top-5 accuracies for both times of day were ≥99.9%. When we subsetted the testing dataset by study site, we found that site-specific accuracies ranged from 90% to 99% (Appendices S3–S7).

When we conducted out-of-sample validation by using our model to evaluate images of ungulates from Canada, we achieved an overall accuracy of 81.8% with a top-5 accuracy of 90.9%. When we tested the ability of our model to accurately predict the presence or absence of an animal in the image using the Serengeti Snapshot dataset, we found that 85.1% were classified correctly as empty, while 94.3% of images containing an animal were classified as containing an animal. Our trained model was capable of classifying approximately 2,000 images per minute on a Macintosh laptop with 16 gigabytes of RAM.

4 | DISCUSSION

To our knowledge, our model achieved the highest accuracy (97.6%) to date in using machine learning to classify wildlife in camera trap images (a recent paper achieved 95% accuracy; Norouzzadeh et al., 2018). This model performed almost as well during the night as during the day (accuracy = 97% and 98%, respectively). We provide this model as an R package (MLWIC), which is especially useful for researchers studying the species and groups available in this package (Table 1) in North America, as it performed well (82% accuracy) in classifying ungulates in an out-of-sample test of images from Canada. The model can also be valuable for researchers studying other species by removing images without any animals from the dataset before beginning manual classification, as we achieved high accuracy in separating empty images from those containing animals in a dataset from Tanzania. This R package can also be a valuable tool for any researchers that have classified images, as they can use the package to train their own model that can then classify any subsequent images collected.

The ability to rapidly identify millions of images from camera traps can fundamentally change the way ecologists design and implement wildlife studies. The burden of classifying images from camera traps has led ecologists to limit the duration and size of camera trap studies (Kelly et al., 2008; Scott et al., 2018). By removing this burden, camera traps can be applied in more studies including monitoring invasive or sensitive species, long-term ecological research and small-scale occupancy studies.

ACKNOWLEDGEMENTS

We thank the hundreds of volunteers and employees who manually classified images and deployed camera traps. We thank Dan Walsh for facilitating cooperation amongst groups. Camera trap projects were funded by the U.S. Department of Energy under award # DE-EM0004391 to the University of Georgia Research Foundation; USDA Animal and Plant Health Inspection Service, National Wildlife Research Center and Center for Epidemiology and Animal Health; Colorado Parks and Wildlife; Canadian Natural Science and Engineering Research Council; University of Saskatchewan; and Idaho Department of Game and Fish.

AUTHORS' CONTRIBUTIONS

M.A.T., R.S.M., K.C.V., N.P.S., S.J.S. and D.W.W. conceived of the project; D.W.W., J.S.L., M.A.T., R.K.B., B.W., P.A.D., J.C.B., M.D.W., B.T., P.E.S., N.P.S., K.C.V., J.M.H., E.S.N., J.S.I., E.A.O., R.K.B., P.M.L.

and A.K.M. oversaw collection and manual classification of wildlife in camera trap images from the study sites; M.S.N. and J.C. developed and programmed the machine learning models; M.A.T. led the analyses and writing of the R package; E.G.M. assisted with R package development and computing; M.A.T. and R.S.M. led the writing. All authors contributed critically to drafts and gave final approval for submission.

DATA ACCESSIBILITY

The trained model is available in the R package MLWIC from GitHub (https://github.com/mikeyEcology/MLWIC; https://doi.org/10.5281/zenodo.1445736). We provide the >3.7 million classified images as the North American Camera Trap Images (NACTI) dataset in the Labeled Information Library of Alexandria: Biology & Conservation (LILA:BC) digital repository (available online at http://lila.science/datasets/nacti).

ORCID

Michael A. Tabak http://orcid.org/0000-0002-2986-7885

Nathan P. Snow http://orcid.org/0000-0002-5171-6493

REFERENCES

Adabi, M., Barhab, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Zheng, X. (2016). *TensorFlow: a system for large-scale machine learning* (Vol. 16, pp. 265–283). Presented at the 12th USENIX Symposium on Operating Systems Design and Implementation, USENIX Association.

Advanced Research Computing Center. (2012). *Mount Moran: IBM system X cluster*. Laramie, WY: University of Wyoming. Retrieved from https://arcc.uwyo.edu/guides/mount-moran

Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014). *Deep convolutional neural network based species recognition for wild animal monitoring* (pp. 858–862). IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/icip.2014.7025172

Gomez Villa, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, *41*, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (1st ed.). Cambridge, MA: MIT Press.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778). IEEE. https://doi.org/10.1109/cvpr.2016.90

Howe, E. J., Buckland, S. T., Després-Einspenner, M.-L., & Kühl, H. S. (2017). Distance sampling with camera traps. *Methods in Ecology and Evolution*, *8*(11), 1558–1565. https://doi.org/10.1111/2041-210X.12790

Kelly, M. J., Noss, A. J., Di Bitetti, M. S., Maffei, L., Arispe, R. L., Paviolo, A., … Di Blanco, Y. E. (2008). Estimating puma densities from camera trapping across three study sites: Bolivia, Argentina, and Belize. *Journal of Mammalogy*, *89*(2), 408–418. https://doi.org/10.1644/06-MAMM-A-424R.1

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).Retrieved from https://papers.nips.cc/book/advances-in-neural-information-processing-systems-25-2012.

Niedballa, J., Sollmann, R., Courtiol, A., & Wilting, A. (2016). camtrapR: An R package for efficient camera trap data management. *Methods in Ecology and Evolution*, *7*(12), 1457–1462. https://doi.org/10.1111/2041-210X.12600

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

O'Connell, A. F., Nichols, J. D., & Karanth, K. U. (Eds.) (2011). *Camera traps in animal ecology: Methods and analyses*. Tokyo, Japan; New York, NY: Springer.

Rovero, F., Zimmermann, F., Bersi, D., & Meek, P. (2013). "Which camera trap type and how many do I need?" A review of camera features and study designs for a range of wildlife research applications. *Hystrix, the Italian Journal of Mammalogy*, *24*(2), 1–9.

Scott, A. B., Phalen, D., Hernandez-Jover, M., Singh, M., Groves, P., & Toribio, J.-A. L. M. L. (2018). Wildlife presence and interactions with chickens on Australian commercial chicken farms assessed by camera traps. *Avian Diseases*, *62*(1), 65–72. https://doi.org/10.1637/11761-101917-Reg. 1

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, *2*, 150026. https://doi.org/10.1038/sdata.2015.26

Swinnen, K. R. R., Reijniers, J., Breno, M., & Leirs, H. (2014). A novel method to reduce time investment when processing videos from camera trap studies. *PLoS ONE*, *9*(6), e98881. https://doi.org/10.1371/journal.pone.0098881

Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., & Huang, T. (2013). Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing*, *2013*(1), 52. https://doi.org/10.1186/1687-5281-2013-52

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Tabak MA, Norouzzadeh MS, Wolfson DW, et al. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods Ecol Evol*. 2018;00:1–6. https://doi.org/10.1111/2041-210X.13120