Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2

Michael A. Tabak, Mohammad S. Norouzzadeh, David W. Wolfson, Erica J. Newton, Raoul K. Boughton, Jacob S. Ivan, Eric A. Odell, Eric S. Newkirk, Reesa Y. Conrey, Jennifer L. Stenglein, Fabiola Iannarilli, John Erb, Ryan K. Brook, Amy J. Davis, Jesse S. Lewis, Daniel P. Walsh, James C. Beasley, Kurt C. VerCauteren, Jeff Clune, Ryan S. Miller

Running headline: Optimizing machine learning for camera traps

Abstract

1. Motion-activated wildlife cameras (or "camera traps") are frequently used to remotely and non-invasively observe animals. The vast number of images collected from camera trap projects have prompted some biologists to employ machine learning algorithms to automatically recognize species in these images, or at least filter-out images that do not contain animals. These approaches are often limited by model transferability, as a model trained to recognize species from one location might not work as well for the same species in different locations. Furthermore, these methods often require advanced computational skills, making them inaccessible to many biologists.

2. We used 3 million camera trap images from 18 studies in 10 states across the United States of America to train two deep neural networks, one that recognizes 58 species, the "species model," and one that determines if an image is empty or if it contains an animal, the "empty-animal model."

3. Our species model and empty-animal model had accuracies of 96.8% and 97.3%, respectively. Furthermore, the models performed well on some out-of-sample datasets, as the species model had 91% accuracy on species from Canada (accuracy range 36-91% across all out-of-sample datasets) and the empty-animal model achieved an accuracy of 91-94% on out-of-sample datasets from different continents.

4. Our software addresses some of the limitations of using machine learning to classify images from camera traps. By including many species from several locations, our species model is potentially applicable to many camera trap studies in North America. We also found that our empty-animal model can facilitate removal of images without animals globally. We provide the trained models in an R package (MLWIC2: Machine Learning for Wildlife Image Classification in R), which contains Shiny Applications that allow scientists with minimal programming experience to use trained models and train new models in six neural network architectures with varying depths.

Keywords

Computer vision, deep convolutional neural networks, image classification, machine learning, motion-activated camera, R package, remote sensing, species identification

Introduction

Motion-activated wildlife cameras (or "camera traps") are frequently used to remotely observe wild animals, but images from camera traps must be classified to extract their biological data (O'Connell, Nichols, & Karanth, 2011). Manually classifying camera trap images is an encumbrance that has prompted scientists to use machine learning to automatically classify images (Norouzzadeh et al., 2018; Willi et al., 2019), but this approach has limitations.

We address two major limitations of using machine learning to automatically classify animals in camera trap images. First, machine learning models trained to recognize species from one location and in one camera trap setup might perform poorly when applied to images from camera traps in different conditions. This "transferability problem" is thought to arise because different locations have different backgrounds (the part of the picture that is not the animal) and most models evaluate the entire image, including the background (Beery, Morris, & Yang, 2019; Miao et al., 2019; Norouzzadeh et al., 2019; Terry, Roy, & August, 2020; Wei, Luo, Ran, & Li, 2020). By including images from 18 different studies in North America, our objective was to train models with more variation in the backgrounds associated with each species. Furthermore, by training an additional model that distinguishes between images with and without animals, we provide an option that could be broadly applicable to camera trap studies worldwide. Second, the use of machine learning in camera trap analysis is often limited to computer scientists, yet the need for image processing exceeds the availability of computer scientists in wildlife research. To facilitate the use of these models by biologists with minimal programming experience, Machine Learning for Wildlife Image Classification (MLWIC2) includes an option to train and use models in user-friendly Shiny Applications (Chang, Cheng, Alaire, Xie, & McPherson, 2019), allowing users to point-and-click instead of using a command line. This facilitates easier site-specific model training when our models do not perform to expectations.

Materials and Methods

Camera trap images

Images were collected from 18 studies using camera traps in 10 states in the United States of America (California, Colorado, Florida, Idaho, Minnesota, Montana, South Carolina, Texas, Washington, and Wisconsin; Appendix S1). Images were either classified by a single wildlife expert or classified independently by two biologists, with discrepancies settled by a third. An image was classified as containing an animal if it contained any part of an animal. Our initial dataset included 6.3 million images but was unbalanced with most images from a few species (e.g., 51% of all images were *Bos taurus*). We rebalanced the number of images by species and site to ensure that no one species or site dominated the training process. Previous work suggested that training a model with 100,000 images per species produces good performance (Tabak et al., 2019); therefore, we limited the number of images for a single species from one location to 100,000. When > 100,000 images for a single species existed at one location, we randomly selected 100,000 of these images to include in the training/testing dataset. After rebalancing the data, we had a total of 2.98 million images; 90% were randomly selected for training, while 10% were used for testing. Images used in this study were either already a part of or were added to the North American Camera Trap Images dataset (lila.science/datasets/nacti; Tabak et al., 2019). Images from Canada were not used for training but were used to evaluate model transferability as an out-of-sample dataset.

Training models

We trained deep convolutional neural networks using the ResNet-18 architecture (He, Zhang, Ren, & Sun, 2016) in the Tensorflow framework (Adabi et al., 2016) on a high performance computing cluster, "Teton" (Advanced Research Computing Center, 2018). Models were trained for 55 epochs, with a ReLU activation function at every hidden layer and a softmax function in the output layer, mini-batch stochastic gradient descent with a momentum hyperparameter of 0.9 (Goodfellow, Bengio, &

Courville, 2016), a batch size of 256 images, and learning rates and weight decays that varied by epoch number (described in Appendix S2). We trained a species model, which contained classes for 58 species or groups of species and one class for empty images (Table 1). We also trained an empty-animal model that contained only two classes, one for images containing an animal, and the other for images without animals.

Model validation and transferability

We first evaluated our trained models by applying them to predicting species in the 10% of images that were withheld from training. Models were evaluated for each species using the recall, top-5 recall, and precision, which are values summarizing the number of true positives (TPs), false positives (FPs), and false negatives (FNs):

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

As recall is the proportion of images of each species that were correctly classified, top-5 recall is the proportion of images for each species in which one of the model's top five guesses is the correct species. We also calculated confidence intervals for recall and precision rates (Appendix S3). To evaluate transferability of the model, we conducted out-of-sample validation by applying our trained models to images from locations where the model was not trained. We evaluated the species model using four out-of-sample datasets from North America: the Caltech Camera Traps dataset (Beery, Van Horn, & Perona, 2018), the ENA24-detection dataset (Yousif, Kays, & He, 2019), the Saskatchewan, Canada dataset from this study, and the Missouri Camera Traps dataset (Zhang, He, Cao, & Cao, 2016). The empty-animal model was tested using the Wellington Camera Traps dataset from New Zealand (Anton, Hartley, Geldenhuis, & Wittmer, 2018), the Snapshot Serengeti dataset from Tanzania

(Swanson et al., 2015), and the Snapshot Karoo dataset from South Africa (http://lila.science/datasets/snapshot-karoo).

R package development

MLWIC2 was developed using the R packages Shiny (Chang et al., 2019) and ShinyFiles (Pedersen, Nijs, Schaffner, & Nantz, 2019) so the user can choose to either use a programming console or a graphical user interface. Users can navigate to locations on their computer using a browser window instead of specifying paths. The package can classify images at a rate of 2,000 images per minute on a laptop with 16 gigabytes of random-access memory. MLWIC2 will optionally write the top guess from each model and confidence associated with these guesses to the metadata of the original image file. The function "write_metadata" uses Exiftool (Harvey, 2016) to accomplish this. In addition, if scientists have labeled images, MLWIC2 has a Shiny app that allows users to train a new model to recognize species using one of six different convolutional neural network architectures (AlexNet, DenseNet, GoogLeNet, NiN, ResNet, and VGG) with different numbers of layers. Note that the time required to train a model depends on the number of images used for training and computing resources; operating MLWIC2 on a high-performance computing cluster requires programming experience.

Results

When we evaluated our models on the withheld images (within sample validation), we found an overall accuracy of 96.8% for the species model and 97.3% for the empty-animal model. Several species (six of 11) had recall of > 95% with fewer than 2,000 images used for training (Table 1; Fig.1). A confusion matrix (Appendix S4) depicts how all images of each species were classified by the species model. When evaluated on out-of-sample images, the species model accuracy ranged from 36.3% to 91.3% (Table 2), with top-5 accuracy ranging from 65.2% to 93.8% (Fig. 2), and the empty-animal model accuracy ranged from 90.6% to 94.1% (Table 2).

Discussion

In MLWIC2, we provide two trained machine learning models, one classifying species and another distinguishing between images with animals and those that are empty, with 97% accuracy, which can potentially be used to rapidly classify camera trap images from many locations. While the species model performed well on out-of-sample images from Saskatchewan, Canada (91% overall accuracy), the model performed poorly on some out-of-sample datasets (Table 2; Fig. 2). The discrepancy in model performance on images from different datasets indicates that transferability remains an issue and our species model will not be useful on all datasets; some users will need to train new models on images from their field sites, an option that is available in MLWIC2. Nevertheless, even in the Missouri dataset where our model performed worst, the top-5 accuracy, the rate at which the true species in an image was in the model's top-5 guesses, was 65% (Table 2). For some applications, e.g. detection of invasive or rare species, a good out-of-sample top-5 recall rate may be sufficient to address scientific questions or meet monitoring objectives. Additionally, our empty-animal model performed well at distinguishing empty images from those containing animals in datasets from three different countries (91-94% accuracy), indicating that this model may be broadly applicable for removing empty images from datasets globally. We propose a workflow for how users can apply these models to filter-out empty images and train new models as necessary (Fig. 3). By providing Shiny apps to train models and classify images, we make this technology accessible to more scientists with minimal programming experience. Our finding that high recall (> 95%) can be achieved with fewer than 2,000 images for some species (Table 1; Fig.1) suggests that smaller labeled image datasets can potentially be used to train models with this software.

Other researchers have developed models for recognizing animals in camera traps, with some success in out-of-sample identification. For example Zilong software accurately removed 85% of empty images (Wei et al., 2020), MegaDetector had a precision of 89-99% at detecting animals (Beery et al., 2019),

and MLWIC achieved an accuracy of 82% at out-of-sample species classification (Tabak et al., 2018, 2019). We hypothesize that our models performed well on some out-of-sample datasets (Snapshot Serengeti, Snapshot Karoo, Wellington, and Saskatchewan; Table 2) because they were trained using camera trap images from multiple locations with different camera placement protocols, allowing the model to develop a search image for each species in multiple backgrounds.

Transferability of machine learning models remains a complication for implementing these models more broadly to camera trap data and, in many cases, it is most productive for scientists to build models that are trained directly on their study sites (see Fig. 3). While such models will have less broad applicability (they are unlikely to be accurate globally), they can have high study-specific accuracies, thus reducing the burden of manual image classification.

Acknowledgements:

Contributions of JCB were partially supported by the DOE under Award Number DE-EM0004391 to the University of Georgia Research Foundation. Support for this research was provided by the USFWS Pittman-Robertson Wildlife Restoration Program and Wisconsin Department of Natural Resources. For supplying camera trap images, we thank USDA Forest Service: Rocky Mountain Research station; Montana Fish, Wildlife and Parks; Wyoming Game and Fish Department; Washington Department of Fish and Wildlife; Idaho Department of Fish and Game; Woodland Park Zoo.

Disclaimer: This manuscript was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information disclosed, or represents that its use not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof.

The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data Availability

The trained models described in this work are available in the MLWIC2 package (https://github.com/mikeyEcology/MLWIC2). Images used to train models are available in the North American Camera Trap Images dataset (lila.sciece/datasets/nacti).

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. (2018). *Teton Computing Environment, Intel x86_64 cluster*. Laramie, WY: University of Wyoming. Retrieved from https://doi.org/10.15786/M2FY47

Anton, V., Hartley, S., Geldenhuis, A., & Wittmer, H. U. (2018). Monitoring the mammalian fauna of urban areas using remote cameras and citizen science. *Journal of Urban Ecology*, *4*(1). doi:10.1093/jue/juy002

Beery, S., Morris, D., & Yang, S. (2019). Efficient Pipeline for Camera Trap Image Review. *ArXiv:1907.06772 [Cs]*. Retrieved from http://arxiv.org/abs/1907.06772

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in Terra Incognita (pp. 456–473). Presented at the Proceedings of the European Conference on Computer Vision (ECCV). Retrieved from http://openaccess.thecvf.com/content_ECCV_2018/html/Beery_Recognition_in_Terra_ECCV_2018_paper.html

Chang, W., Cheng, J., Alaire, J., Xie, Y., & McPherson, J. (2019). shiny: Web Application Framework for R (Version 1.4.0). Retrieved from https://CRAN.R-project.org/package=shiny

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

Harvey, P. (2016). *ExifTool*. Retrieved from https://exiftool.org/

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). IEEE. doi:10.1109/CVPR.2016.90

Miao, Z., Gaynor, K. M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M. S., … Getz, W. M. (2019). Insights and approaches using deep learning to classify wildlife. *Scientific Reports*, *9*(1), 1–9. doi:10.1038/s41598-019-44565-w

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2019). A deep active learning system for species identification and counting in camera trap images. *ArXiv:1910.09716 [Cs, Eess, Stat]*. Retrieved from http://arxiv.org/abs/1910.09716

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*, *115*(25), E5716–E5725. doi: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 ; New York: Springer.

Pedersen, T. L., Nijs, V., Schaffner, T., & Nantz, E. (2019). shinyFiles: A Server-Side File System Viewer for Shiny (Version 0.7.5). Retrieved from https://CRAN.R-project.org/package=shinyFiles

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. doi:10.1038/sdata.2015.26

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., VerCauteren, K. C., Snow, N. P., … Miller, R. S. (2018). *MLWIC: machine learning for wildlife image classification in R*. Zenodo. doi:10.5281/zenodo.1445736

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., … Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, *10*(4), 585–590. doi:10.1111/2041-210X.13120

Terry, J. C. D., Roy, H. E., & August, T. A. (2020). Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods in Ecology and Evolution*, *11*(2), 303–315. doi:10.1111/2041-210X.13335

Wei, W., Luo, G., Ran, J., & Li, J. (2020). Zilong: A tool to identify empty images in camera-trap data. *Ecological Informatics*, *55*, 101021. doi:10.1016/j.ecoinf.2019.101021

Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., … Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, *10*(1), 80–91. doi:10.1111/2041-210X.13099

Yousif, H., Kays, R., & He, Z. (2019). Dynamic programming selection of object proposals for sequence-level animal species classification in the wild. *IEEE Transactions on Circuits and Systems for Video Technology*.

Zhang, Z., He, Z., Cao, G., & Cao, W. (2016). Animal Detection From Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Patch Verification. *IEEE Transactions on Multimedia*, *18*(10), 2079–2092. doi:10.1109/TMM.2016.2594138

Tables and Figures

Table 1: Mean recall and precision rates (along with 95% confidence intervals) for predicting species using the species model on the 10% of images that were withheld from training.

Class name (scientific name)	Number of training images	Recall	Precision
Accipitridae family (Accipitridae)	1,511	0.91(0.67,1)	0.94(0.89,0.97)
American crow (*Corvus brachyrhynchos*)	2,522	0.67(0.61,0.73)	0.7(0.64,0.75)
American marten (*Martes americana*)	51,081	0.96(0.95,0.97)	0.96(0.94,0.97)
Anatidae family (Anatidae)	1,071	0.97(0.92,0.99)	0.97(0.92,0.99)
armadillo (Cingulata)	8,947	0.94(0.59,0.99)	0.95(0.94,0.96)
bighorn sheep (*Ovis canadensis*)	1,189	1(0.97,1)	1(0.97,1)
black bear (*Ursus americanus*)	111,426	0.97(0.91,0.99)	0.99(0.91,0.99)
black-billed magpie (*Pica hudsonia*)	2,770	0.98(0.95,0.99)	0.96(0.91,0.99)
black-tailed jackrabbit (*Lepus californicus*)	5,617	0.95(0.93,0.96)	0.93(0.91,0.95)
black-tailed prairie dog (*Cynomys ludovicianus*)	43,999	0.93(0.93,0.94)	0.95(0.94,0.96)
bobcat (*Lynx rufus*)	31,634	0.96(0.95,0.99)	0.97(0.96,0.98)
California ground squirrel (*Otospermophilus beecheyi*)	30,301	1(1,1)	0.99(0.98,0.99)
California quail (*Callipepla californica*)	2,046	0.97(0.94,0.99)	0.99(0.97,1)
Canada lynx (*Lynx canadensis*)	15,119	1(0.99,1)	0.99(0.98,0.99)
cattle (*Bos taurus*)	269,963	0.97(0.93,0.98)	0.98(0.77,0.99)
Clark's nutcracker (*Nucifraga columbiana*)	2,785	0.94(0.91,0.96)	0.92(0.87,0.95)

common raven (*Corvus corax*)	21,134	0.99(0.91,0.99)	0.99(0.98,1)
coyote (*Canis latrans*)	41,512	0.96(0.94,0.98)	0.97(0.96,0.99)
Cricetidae and Muridae families	1,254	0.93(0.87,0.96)	0.83(0.7,0.94)
dog (*Canis familiaris*)	1,136	0.82(0.7,0.98)	0.78(0.6,0.99)
domestic sheep (*Ovis aries*)	16,340	0.99(0.99,1)	0.99(0.99,1)
donkey (*Equus asinus*)	2,403	0.99(0.97,1)	0.94(0.9,0.96)
elk (*Cervus canadensis*)	112,389	0.97(0.95,0.98)	0.99(0.86,0.99)
empty (no animal)	907,096	0.97(0.93,0.98)	0.95(0.92,0.97)
fisher (*Pekania pennanti*)	7,697	0.98(0.97,0.99)	0.99(0.96,1)
golden-mantled ground squirrel (*Callospermophilus lateralis*)	1,587	0.89(0.83,0.92)	0.86(0.81,0.91)
grey fox (*Urocyon cinereoargenteus*)	16,094	0.98(0.96,0.99)	0.97(0.95,0.99)
grey jay (*Perisoreus canadensis*)	3,776	0.97(0.87,0.98)	0.94(0.8,0.98)
grey squirrel (*Sciurus carolinensis*)	24,677	0.98(0.64,0.99)	0.98(0.64,0.99)
grizzly bear (*Ursus arctos horribilis*)	843	0.99(0.94,1)	0.99(0.94,1)
Gunnison's prairie dog (*Cynomys gunnisoni*)	17,393	0.83(0.82,0.85)	0.93(0.91,0.94)
horse (*Equus ferus*)	3,644	0.94(0.53,0.97)	0.95(0.45,0.98)
human (*Homo sapiens*)	139,983	0.98(0.97,0.98)	0.98(0.97,0.99)
Marmota genus (*Marmota* spp.)	1,497	0.98(0.95,0.99)	0.95(0.91,0.98)
moose (*Alces alces*)	11,741	0.99(0.97,1)	0.99(0.97,1)
mountain lion (*Puma concolor*)	13,900	0.96(0.95,0.97)	0.97(0.96,0.98)
mule deer (*Odocoileus hemionus*)	91,068	0.98(0.95,0.99)	0.98(0.93,0.99)
opossum (Didelphimorphia)	5,782	0.94(0.76,0.98)	0.97(0.87,0.99)

other grouse (Tetraoninae)	4,237	0.97(0.91,0.99)	0.98(0.96,0.99)
other mustelids (Mustelidae)	2,467	0.89(0.85,0.92)	0.91(0.85,0.96)
other passerine birds (Passeriformes)	3,363	0.86(0.81,0.9)	0.88(0.75,0.94)
porcupine (Erethizontidae and Hystricidae)	6,608	0.97(0.82,0.99)	0.98(0.96,0.98)
prairie chicken (*Tympanuchus cupido*)	815	1(0.96,1)	0.98(0.93,1)
pronghorn (*Antilocapra americana*)	57,953	0.98(0.97,0.98)	0.99(0.98,0.99)
raccoon (*Procyon lotor*)	51,439	0.9(0.83,0.99)	0.93(0.91,0.99)
red fox (*Vulpes vulpes*)	43,433	0.98(0.96,0.99)	0.98(0.97,0.99)
red squirrel (*Tamiasciurus hudsonicus*)	21,586	0.85(0.84,0.96)	0.86(0.88,0.97)
river otter (*Lontra canadensis*)	1,821	0.96(0.92,0.98)	0.97(0.93,0.98)
snowshoe hare (*Lepus americanus*)	37,467	0.97(0.94,0.99)	0.97(0.95,0.98)
Steller's jay (*Cyanocitta stelleri*)	1,844	0.91(0.8,0.98)	0.96(0.87,1)
striped skunk (*Mephitis mephitis*)	12,416	0.98(0.9,0.99)	0.97(0.96,0.98)
swift fox (*Vulpes velox*)	3,266	0.85(0.81,0.88)	0.95(0.92,0.97)
Sylvilagus family	6,385	0.93(0.82,0.99)	0.94(0.86,0.97)
Totals	2,682,380	0.97	0.97
vehicle (truck, ATV, car)	32,912	0.97(0.96,0.98)	0.97(0.97,0.98)
white-tailed deer (*Odocoileus virginianus*)	88,531	0.93(0.83,1)	0.97(0.84,0.99)
wild pig (*Sus scrofa*)	243,344	0.98(0.98,0.99)	0.99(0.98,1)
wild turkey (*Meleagris gallopavo*)	15,686	0.94(0.88,0.99)	0.98(0.95,1)
wolf (*Canis lupus*)	3,070	0.96(0.88,1)	0.95(0.8,1)
wolverine (*Gulo gulo*)	18,810	0.98(0.96,1)	0.98(0.97,0.99)

Table 2: Out-of-sample validation results. All out-of-sample images are available from lila.science/datasets.

Dataset	Number of images tested	Model tested	Accuracy	Top-5 accuracy*
Snapshot Karoo (South Africa)	38,101	empty-animal	0.906	
Snapshot Serengeti (Tanzania)	104,651	empty-animal	0.941	
Wellington (New Zealand)	266,966	empty-animal	0.939	
Caltech Camera Traps (USA)	218,147	species	0.562	0.744
ENA24-Detection (USA)	5,285	species	0.507	0.649
Missouri Camera Traps (USA)	5,008	species	0.362	0.652
Saskatchewan (Canada)	5,200	species	0.913	0.938

*Top-5 accuracy is not relevant for the empty-animal model because there are only two classes

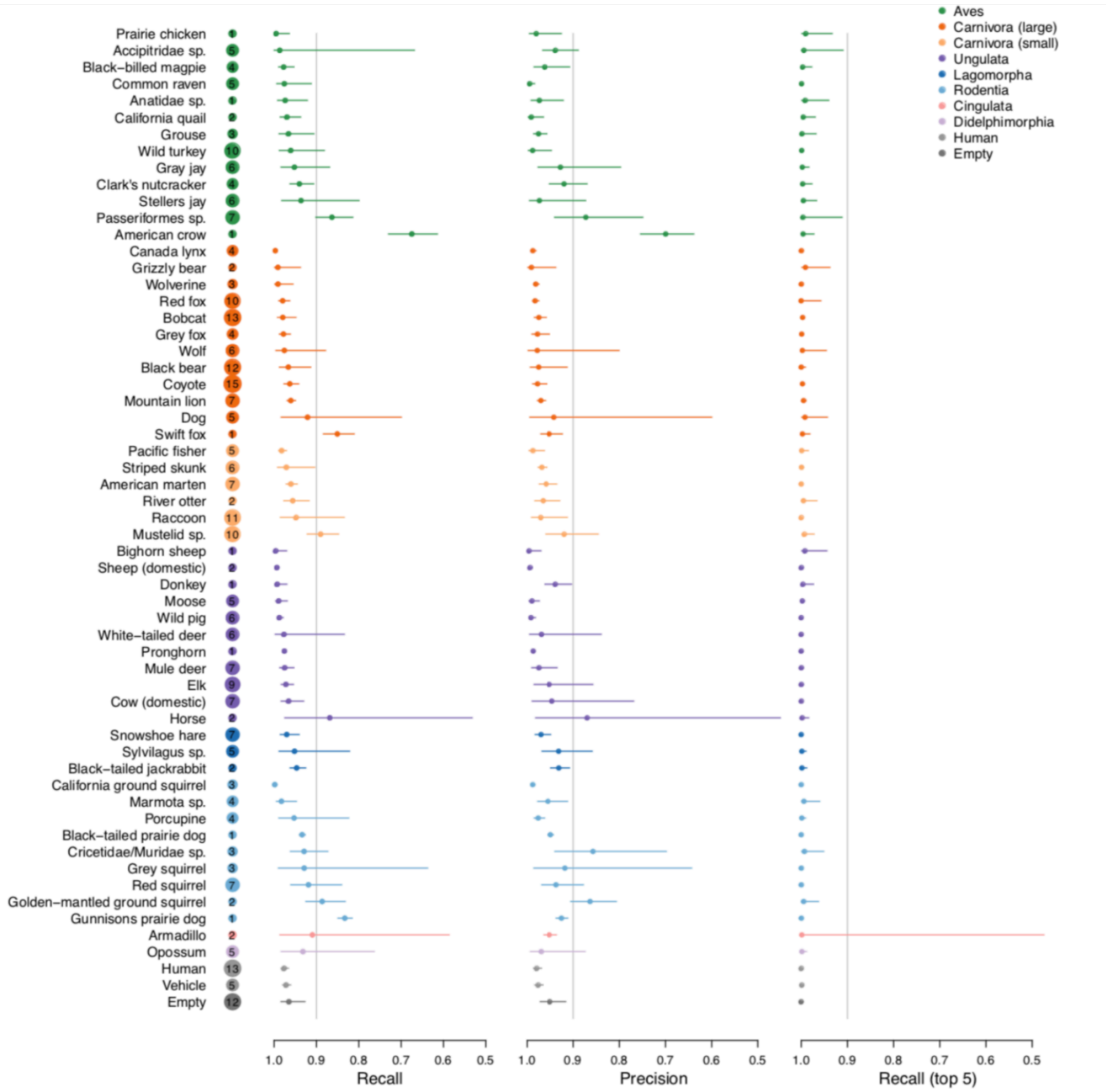

Fig. 1: Within sample validation of the species model revealed high recall and precision for most species. Median values across datasets are presented along with 95% confidence intervals. The number of datasets for each species is included in the circle next to the species name (circle sizes are proportional to the number of datasets containing each species).

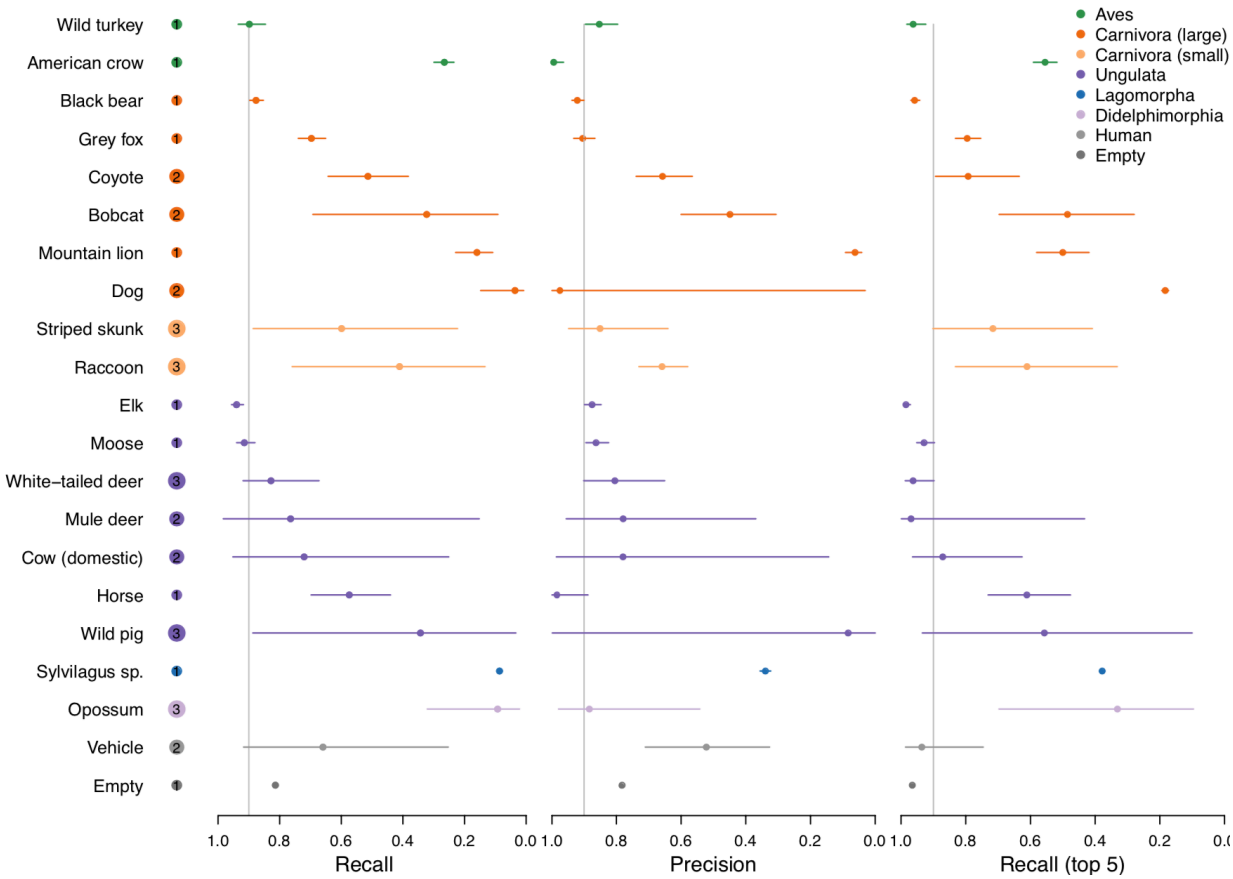

Fig. 2: Species model out-of-sample validation revealed variable recall and precision rates across species. Median values across datasets are presented along with 95% confidence intervals. The number of datasets for each species is included in the circle next to the species name.

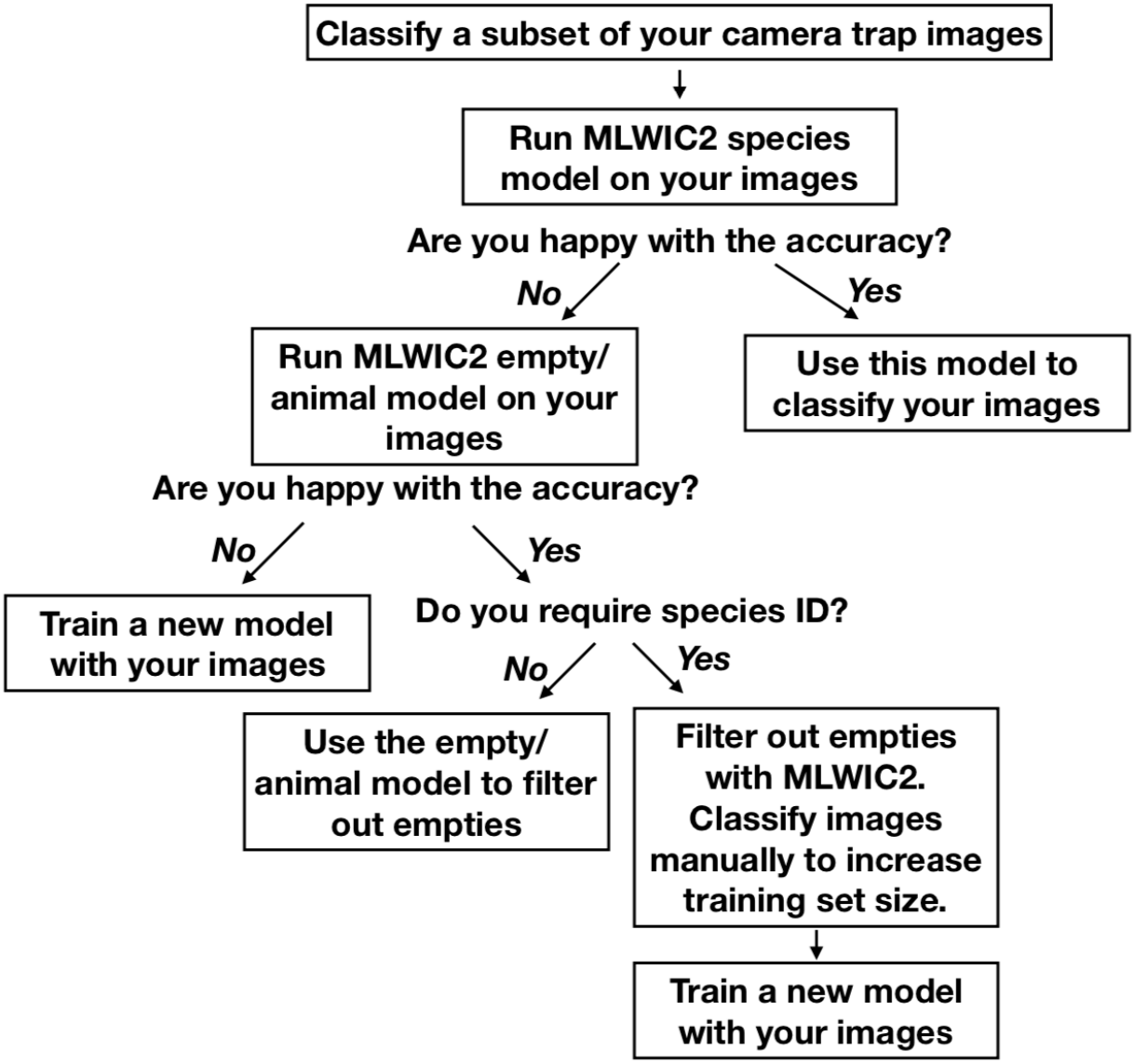

Fig. 3: Proposed workflow for using MLWIC2 models when classifying camera trap images.

Supporting Information

Appendix S1: Information for each of the 18 studies that produced camera trap images used in this paper. The final 59 columns are the number of images of each species (or group of species).

Appendix S2: Learning rate and weight decay for each epoch in the model training process.

Appendix S3: Calculation of pooled recall and precision rate and corresponding confidence intervals.

Appendix S4: Confusion matrix depicting the number of images of each species (or group of species) that were classified by the species model as each species (or group of species). Columns are the ground truth labels from human observers; rows are predictions from the model.