An approach to rapid processing of camera trap images with minimal human input

Matthew Duggan[1], Melissa Groleau[1], Ethan Shealy[1], Lillian Self[1], Taylor Utter[1], Matthew Waller[1], Bryan Hall[2], Chris Stone[2], Layne Anderson[2], and Timothy Mousseau[1]

[1]University of South Carolina at Columbia
[2]South Carolina Army National Guard

March 23, 2021

Abstract

Point 1: Camera traps have become an extensively utilized tool in ecological research, but the processing of images created by a network of camera traps rapidly becomes an overwhelming task, even for small networks. Point 2: We used transfer training to create convolutional neural network (CNN) models for identification and classification. By utilizing a small dataset with less than 10,000 labeled images the model was able to distinguish between species and remove false triggers. Point 3: We trained the model to detect 17 object classes with individual species identification, reaching an accuracy of 92%. Previous studies have suggested the need for thousands of images of each object class to reach results comparable to those achieved by human observers; however, we show that such accuracy can be achieved with fewer images. Point 4: Additionally, we suggest several alternative metrics common to computer science studies to accurately evaluate the performance of such camera trap image processing models, as well as methods to adapt the model building process to two targeted purposes.

Introduction

Observational studies of wildlife occupancy and abundance are more important than ever as human disturbance has decreased wildlife population sizes by up to 60% globally in the last four decades (WWF 2018). These staggering declines have the of ecological monitoring through a variety of means including camera traps, mark-recapture methods, point counts, and line transects. Camera traps have become an especially useful for the rapid assessment of wildlife because they require fewer field hours than other common field methods, may be reviewed by other researchers, and minimize disturbance to the environment (Silveira et al. 2003, Steenweg et al. 2017, McCallum 2013). While camera traps are a useful tool for some ecological studies, processing massive quantities of images created by camera trap networks is a major limiting factor for researchers. Until methods are developed to efficiently process images, these limitations will persist in future studies and as camera trap networks become more complex.

Previous camera trap studies have noted factors which result in large accumulations of images. Wind, loose shrubbery, camera settings, and animal behavior specific to each camera site add noise to the dataset (Newey et al. 2015). The time involved in manually processing these false triggers, which often represent a majority of captured images, can delay analysis to the point where conclusions are no longer relevant. because a large expenditure of resources is often required to process images manually (Willi et al. 2019).

Increase in the use of camera traps for ecological studies has led to a push for standardized methods to improve the workflow of image analysis (Glover2019). One promising avenue for processing camera trap images is the utilization of artificial intelligence (AI) technology.

AI trained with convolutional neural networks (CNNs) has been employed and tested on several large datasets previously processed by citizen scientists. Swanson et al. (2015) trained and created a CNN for the Snapshot Serengeti dataset which consists of 3.2 million images collected over 99,241 camera trap days. The output of the neural network reached an accuracy of greater than 93.8% when compared to the records of citizen scientists. While several large-scale studies (e.g. Norouzzadeh et al. 2018) have achieved similar accurac on such large datasets, the training of these neural networks requires large numbers of images and substantial computer time to train the model. Such investments are often not feasible for smaller camera trap studies the current assumption that many thousands of images are needed to successfully train a model.

Only the largest camera trap studies have attempted to create their own neural networks, as it has been suggested that small clusters of images (˜1,000-5,000 images per species class) are not sufficient for deep learning (e.g. Norouzzadeh et al. 2018). Each model built by these large-scale studies must be tailored to particular set of species in order to properly function because neural networks are a complex series of algorithms that are used to detect specific features in supervised data. The neural network learns the features belonging to each species class, allowing it to differentiate between objects and the background of images while also classifying objects. Therefore, the model may not be similar enough to another study's range of objects and backgrounds to be useful, even in the same geographical location.

We suggest that the use of transfer training on neural networks has been overlooked for small scale camera trap studies. Adapting a neural network to a dataset by adjusting the final layers of the network through transfer learning and then reinforcement learning on a desired image set can be extremely useful, especially when data is scarce. e predict a premade neural network could achieve similar identification accuracy as neural networks trained with thousands of images while not requiring such a large memory footprint. Using a transfer-trained neural network allows camera trap surveys to be affordable, data efficient, and accessible to a broad range of projects.

Neural networks are used for all types of image processing and many are freely available through open-source software (e.g. Google, PyTorch, Keras). A premade neural network can be selected from an archive based on the types of images the network was built on; for instance, a neural network trained on animals/pets would be ideal for a camera trap project interested in identifying medium to large sized mammals. To mimic a small-scale camera trap study, we trained a premade, freely available neural network using less than 6,000 images from our larger dataset and achieved similar confidence in object identification as the previously mentioned large scale studies. Here we show that a small amount of diversified image can be as successful at eliminating false positives and identifying species as a model developed using many thousands of images.

Methods

Camera Trap Study

The subset of images used to train the model was pulled from a camera trap study consisting of 170 cameras, which were deployed for up to three years across two regions of South Carolina (see Supplementary material Appendix 1 for camera trap study details). We acquired images for the train and test datasets from 50 camera locations from each region within two separate one-month time frames. The complete consisted of 5,277 images of 17 classes, including images from both winter and summer months to account for seasonal background variation (Table 1). True negative images were not included because they would not assist in teaching the model about any of the species classes. A commonly used 90/10 split (e.g. Fink et al. 2019) was utilized to create the training and testing datasets from the selected images; 90% of images were used for training and 10% were used for testing.

The basic process (Fig. 1) included selecting and labeling a subset of images from our camera trap image repository (See Supplementary material Appendix 1 for details) for transfer training, in order to adapt a

pre-made neural network to our image set. The subset of images used to train the model was pulled from a camera trap study consisting of 170 camera stations which had been deployed for up to three years in two regions of South Carolina (see Supplementary material Appendix 1 for camera trap study details). To begin, a subset of images was created by selecting 500 images of each species in a variety of positions within the field of view (Fig. 1, Step 1). In cases where classes (species being classified) reached 500 images, only images that contributed a unique perspective of the animal were added to the training dataset, in order to supply the model with a better generalization of the animal. The number of images in the training data set was limited to ensure the model did not favor one due to the number of images in the dataset. Despite adding more than 500 images to some classes, class weights were not influenced and remained comparable.

Feature Extraction

se of a supervised training process increases the accuracy of detection and classification by human-generated bounding boxes (Supplementary material Appendix 2). LabelImg (Tzutalin 2015), a graphical image annotation tool, was used to establish ground truths (locations of all objects in an image) and create the records needed for our supervised training process. This software allows a user to define a box containing the object and automatically generates a CSV file with the coordinates of the bounding box as well as the class defined by the user.

Classification Training

A transfer training process to adapt a premade neural network (Fig. 1, Step 3) was employed to create an identification and classification model. We transformed the CSV file generated by the feature extraction process into a compatible tensor dataset for the training process through the appropriate methodologies laid out in the Tensorflow (Abadi et al. 2015) package description. Tensorflow is an -source, experimental Python library from Google for identification and classification models. The Tensorflow transfer training process required a clone of the Tensorflow repository, in combination with a customized model configuration file defining parameters(Table 2).

Training Evaluation

The degree of learning that was completed after each step was analyzed using intersection over union (IOU) as training occurred (Krasin et al. 2017). A greater IOU equates to a higher overlap of generated predictions versus human labeled regions, thus indicating a better model (See Supplementary material Appendix 3). Observing an asymptote in IOU allowed for the determination of a minimum number of steps needed to train the model for each class and to assess which factors influenced the training process (e.g. feature qualities, amount of training images). Because the minimum step number was not associated with image quantity in determining step requirements, we relied on quality assessments, such as animal size and animal behavior.

Following training, final discrepancies between the model output and the labeled ground truths were summarized into confusion matrices (generated by scikit-learn, Table 3) including false positives, false negatives, true positives, true negatives, and misidentifications. Several metrics were calculated to evaluate aspects of model performance (Fig. 2). Relying on accuracy alone may result in an exaggerated confidence in the model's performance, so to avoid this bias, the model's precision, recall, and F-1 score were also calculated. Precision is a measure of FPs while recall is a measure of FNs, with F-1 being essentially an average of the two (Fig. 2). Due to the large proportion of TNs associated with camera trap studies, F-1 score does not include TNs in order to focus on measuring the detection of TPs.

In addition, the metrics were further separated into evaluations for identification and classification purposes. Identification (ID) models would focus only on finding objects and therefore deem misidentifications as correct because the object was found. Classification (CL) models would not deem misidentifications as correct. Finally, accuracy, precision, recall, and F-1 were calculated at a variety of confidence thresholds

(CT), a parameter constraining the lower limit of confidence necessary for a classification proposal, to determine the threshold that resulted in the highest value of the metric we wished to optimize.

Validation

To confirm results acquired from testing the model, it was essential to evaluate a validation set of images. This validation set was formed by randomly selecting five cameras from a 12-week period separate from the training dataset, but within the same larger dataset. The validation subset consisted of 10,983 images, including true negatives. The set ran using the optimal CT for F-1 score determined by the test data. These images were also labeled using labelImg to automate the calculation of evaluation metrics. The validation set scores and test scores should be compared to determine if the model is overfitted, meaning the test set is not representative of the validation set. Possible reasons for such a mismatch may be that the background environment has changed dramatically or species not included in the test set have appeared.

Results

Evaluation of Training

The performance of our model did not depend on the number of images used to train each species class (Fig. 4). , precision during the training process varied greatly species classes and was not a function of the number of images input into the model (Fig. 3). The class with the highest precision during training was armadillo (98%) with 186 images while grey squirrel had the lowest precision during training (30%), despite being trained on 318 images. The raccoon, turkey, and deer classes all resulted in comparably high precision values while being trained using 88, 430, and 1,109 images, respectively (Fig. 3). Five classes were trained using less than 60 images between the test and train dataset (Table 2, see Supplementary material Appendix 3 for all IOU graphs). Result metrics for these classes also varied as a function of species traits rather than number of images used to train the class.

Model Performance

To judge the performance of the model, we evaluated accuracy, precision, recall, and F-1 at several CT Metrics followed the same trends for both ID and CL purposes with CL values running slightly below ID values (Table 5). The test set produced recall values that were inversely related to the CTs, while the precision values were directly related; precision was highest at 0.95 CT (ID: 90%, CL: 88%) and recall was highest at 0.50 CT (ID: 96%, CL: 89%). F-1 score was highest at the 0.70 CT for ID (86%) and 0.90 CT for CL (83%). The difference between accuracy and F-1 values demonstrates the effect of TNs (Fig.). Accuracy and F-1 were highest at 0.90 CT for the test data; therefore, we decided to use 0.90 CT for the validation set. The validation test resulted in a 93% accuracy, 68% precision, 86% recall, and 76% F-1 score (Table).

Discussion

CNN Accessibility

This study demonstrates that AI-based identification and classification models are more accessible than previously thought. Until now, processing of camera trap images has been limited by human observers, expense, processing time, and ignorance of computer science techniques for in ecological studies. Employing labeling services (e.g. Google Cloud) can be unreliable for processing large datasets, and to have images labeled and processed currently costs approximately $0.05 per image (Google Cloud); which may not be practical when tens of thousands of images are involved.

An increasingly accurate and efficient method of image processing is transfer training (e.g. Deepak et al. 2019, Swati et al. 2019, Shi et al. 2019), which is an especially desirable technique for studies with limited

data (Shin et al. 2016). Despite improvements in this training architecture, the use of these methods in ecology has been limited. Transfer training saves time and reduces data requirements, allowing for smaller studies to spend less time processing while still calibrating the architecture with specific images and training the model on a percentage of their complete dataset. Additionally, transfer training prevents overfitting of the model, which can be an issue when using a smaller number of images (Deepak and Ameer 2019, Han et al. 2018).

A smaller image set allows the model to be more flexible, making it more applicable for ecologists than other advanced machine learning techniques (Xie et al. 2016). Feature extraction with transfer training provides camera trap projects an alternative option to starting a CNN architecture from scratch, instead opting to use a pre-trained CNN product (e.g. Microsoft MegaDetector) or unsupervised learning techniques (e.g. cluster analysis).

By using open-source programs and premade neural nets, models can be built to simply remove images without animals or to fully automate the classification of species. This study, along with similar studies (e.g. Tabek et al. 2019), provides evidence that a reliable identification and classification model can be created with open-source tools (e.g. Tensorflow) by using transfer learning and premade neural networks. Further, we completed this process using a very limited set of images and achieved encouraging results. This technology could be especially desirable for researchers wishing to eliminate false positives as well as to quickly sort and label species classes.

Calibration Analysis

Currently, accuracy is the standard metric to evaluate classification models for camera trap studies (Gomez et al. 2016, Norouzzadeh et al. 2018, Swanson et al. 2015). We suggest the optimization of customized models also be based on F-1-score rather than relying on accuracy alone, because accuracy can be heavily biased by TNs (Wolf et al. 2006). This the greater than 20% difference between our test accuracy (TNs excluded) and validation accuracy (TNs included).

The metrics used to optimize a model will depend on the purpose of the project and the resources available to the researcher. The F-1-score can be broken down into precision and recall, both of which can be optimized for different purposes. In a study focusing on rare species (e.g. Alexander et al. 2016, Karanth et al. 1995), precision should be optimized to ensure the detection of all possible occurrences of animals. Alternatively, recall should be optimized if processing time is limited and every image of an animal is not essential for the global analysis. Optimizing recall is ideal for a general survey of common, easily identified animals (e.g. Chitwood et al. 2017).

Optimizing Model Performance

Analyzing model performance during training is especially useful to determine which classes the model is not identifying and is easily visualized using IOU graphs. Precision during training did not seem to depend on the number of images used to train each class; rather, the type of object the class refers to was most important in determining the model. Objects with unique shapes, color patterns, and textures (e.g. turkey and armadillo) were detected by the model more easily (Fig.). The model was not as successful with objects that were small and difficult to distinguish from the background (e.g. grey squirrel), similar to another class (e.g. coyote and dog), or when train examples were highly variable in the subjects within the same class (e.g. humans and vehicles).

Depending on the aim of the study, the choice of metric allows the researcher to facilitate either an ID or CL model. Certain camera trap studies benefit greatly from automating the removal of TNs, especially when focusing on topics such as camera trap effectiveness (e.g. Ferreira-Rodríguez et al. 2019, Edwards et al. 2016) or instances where human-supervised processing will be required to extract details such as behavior. To focus a model on detection of objects rather than classification, researchers should focus on metrics associated with ID. The use of this type of identification model would allow researchers to decrease processing time

and ensure detection of objects while not being overly concerned with the accuracy of species classification by the model. Alternatively, studies focusing on general ecosystem monitoring (e.g. Steenweg et al. 2017, Jiménez et al. 2010) or density of common species (e.g. Parsons et al. 2017) would benefit from a CL model, and should use CL metrics to build a model fully capable of both identifying and classifying species.

Several methods may be employed to adjust the model's parameters. CTs are a simple way to a model to reach the desired metric's optimal value. If optimization cannot be reached by of CTs the model can be further improved by adding images to classes which the model consistently predicts incorrectly. This will help the model learn from the dataset and objects

As biodiversity declines worldwide (Kolbert 2014), employing commonly used computer science techniques in future camera trap studies will greatly enhance our ability to monitor wild populations.

Conclusions

1. Transfer training with bounding boxes is successful and requires far fewer training images than traditional model building.
2. Identification and classification models built using transfer training and small image sets can be very successful with species that are easily distinguished. Species that are more difficult to distinguish can also be identified but require more training images.
3. The traditional metric of accuracy can give a false sense of confidence in a model because of inflation by true negatives. F-1 should be used for general purposes because it is not biased by true negatives.
4. Studies focusing on simply removing true negatives do not require high model performance studies attempting to classify species .

Acknowledgments

We thank the South Carolina Army National Guard for funding this project and their assistance with field work throughout this project. This project would not have been possible without the support of the Biology Department at the University of South Carolina (UofSC) and undergraduate funding through the UofSC Honors College and UofSC's Office of Undergraduate Research. The Samuel Freeman Charitable Trust and American Council of Learned Societies provided essential support for this project. We also thank Gabriella Spatola (UofSC), Sarah Doyle (UofSC), and Luke Wilde (UofSC) for their comments and feedback throughout the writing process.

Authors' Contributions

MD conceived the presented ideas in discussion with TM. MD designed the model and computational framework with assistance from LS and TU. MD and MG wrote the manuscript with support from MW, ES, and TM. The camera trap project in the present study was made possible by BH, CS, LA, MG, and TM; with field assistance and data collection provided by all authors. TM secured funding provided by BH, CS, and LA. TM and MG supervised the project. All authors contributed to the drafts of the publication.

Data Availability

Data are available upon request from the corresponding author (Timothy Mousseau, *mousseau@sc.edu*) and will be deposited online to the EDI repository upon acceptance of the paper.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Israd, M., Jia, Y., Jozefowics, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems* . Software available from tensorflow.org. arxiv:1603.04467

Alexander, J. S., Zhang, C., Shi, K., & Riordan, P. (2016). A granular view of a snow leopard population using camera traps in Central China.*Biol. Conserv. 197* , pp. 27-31. https://doi.org/10.1016/j.biocon.2016.02.023

Chitwood, M. C., Lashley, M. A., Higdon, S. D., DePerno, C. S., & Moorman, C. E. (2020). Raccoon Vigilance and Activity Patterns When Sympatric with Coyotes. *Diversity* , 12(9), pp. 341. https://doi.org/10.3390/d12090341

Edwards, S., Gange, A. C., & Wiesel, I. (2016). An oasis in the desert: the potential of water sources as camera trap sites in arid environments for surveying a carnivore guild. *J. Arid Environ.* , *124* , pp. 304-309. https://doi.org/10.1016/j.jaridenv.2015.09.009

Ferreira-Rodríguez, N., & Pombal, M. A. (2019). Bait effectiveness in camera trap studies in the Iberian Peninsula. *Mammal Res.* ,*64(2),* pp. 155-164. https://doi.org/10.1007/s13364-018-00414-1

Fink, G. A., Frintrop, S., & Jiang, X. (2019). Pattern recognition: 41st DAGM German Conference. Dortmund, Germany. *Spring Nature* . pp. 394. ISBN: 978-3-030-33676-9

Glover-Kapfer, P., Soto-Navarro, C.A., & Wearn, O.R. (2019). Camera-trapping version 3.0: current constraints and future priorities for development. *Remote Sens. Ecol. Conserv.* , 5, pp. 209-223. https://doi.org/10.1002/rse2.106

Gomez, A., Diez, G., Salazar, A., & Diaz, A. (2016). Animal identification in low quality camera-trap images using very deep convolutional neural networks and confidence thresholds.*International Symposium on Visual Computing.* pp. 747-756. https://doi.org/10.1016/j.ecoinf.2017.07.004

Han, D., Liu, Q., & Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Syst. Appl. 95* , pp. 43-56. https://doi.org/10.1016/j.eswa.2017.11.028

Jimenez, C. F., Quintana, H., Pacheco, V., Melton, D., Torrealva, J., & Tello, G. (2010). Camera trap survey of medium and large mammals in a montane rainforest of northern Peru. *Rev Peru. Biol.* ,*17* (2), pp. 191-196. https://doi.org/10.15381/RPB.V17I2.27

Karanth, K. U. (1995). Estimating tiger populations from camera-trap data using Michler capture models. *Biol. Conserv.* , *71* (3), pp. 333-338.

https://doi.org/10.1016/0006-3207(94)00057-W

Kolbert, E. (2014). *The sixth extinction: An unnatural history* . New York: Henry Holt and Company.

Krasin I., Duerig T., Alldrin N., Ferrari V., Abu-El-Haija S., Kuznetsova A., Rom H., Uijlings J., Popov S., Veit A., Belongie S., Gomes V., Gupta A., Sun C., Chechik G., Cai D., Feng Z., Narayanan D., Murphy K. (2017). *OpenImages: A public dataset for large-scale multi-label and multi-class image classification* . Available from https://github.com/openimages

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks.*ACM* , *60* (6), pp. 84-90. https://doi.org/10.1145/3065386

McCallum, J. (2013). Camera trap use and development in field ecology.*Mammal Rev* ., 43, pp. 196-206. https://doi.org/10.1111/j.1365-2907.2012.00216.x

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.*Proc. Natl. Acad. Sci* , 115(25). E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, I., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imag.* , 35(5), pp. 1285-1298. https://doi.org/10.1109/tmi.2016.2528162

Silveira, L., Jacomo, A. T., & Diniz-Filho, J. A. F. (2003). Camera trap, line transect census and track surveys: a comparative evaluation.*Biol. Conserv* ., 114(3), pp. 351-355. https://doi.org/10.1016/s0006-3207(03)00063-6

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., Townsend, S. E., Carbone, C., Rowcliffe J. M., Whittington, J., Brodie, J., Royle J. A., Switalski, A., Clevenger, A. P., Helm, N. & Rich, L.N. (2017). Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. *Front. Ecol. Environ* ., 15(1), pp. 26-34. https://doi.org/10.1002/fee.1448

Tzutalin. LabelImg. Git code (2015). Available from https://github.com/tzutalin/labelImg

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., & Fortson, L. (2019) Identifying animal species in camera trap images using deep learning and citizen science.*Methods Ecol. Evol.* , 10, pp. 80-91. https://doi.org/10.1111/2041-210X.13099

Wolf, C., & Jolion, J. M. (2006). Object count/area graphs for the evaluation of object detection and segmentation algorithms. *Int. J. Doc. Anal. Recognit.* , *8* (4), pp. 280-296. arXiv:1807.01544v2

WWF (2018) *Living Planet Report 2018: Aiming higher* (eds. Grooten N & Almond REA). WWF International, Gland, Switzerland.

Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2015). Transfer learning from deep features for remote sensing and poverty mapping. In:*Proceedings 30th AAAI Conference on Artificial Intelligence* , 30(1). arXiv:1510.00098.

Figures

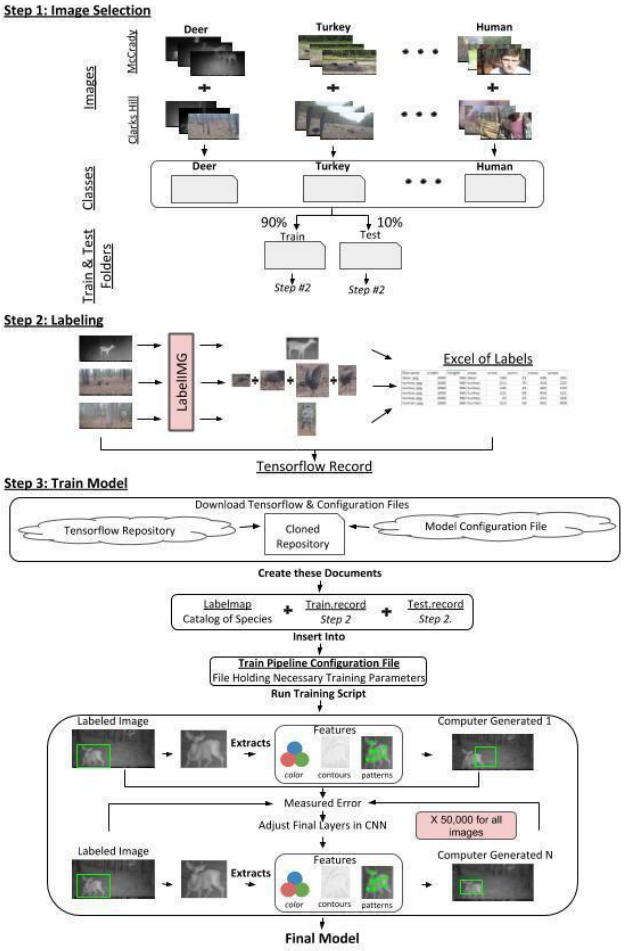


Figure 1. Diagram of image collection and training process. The visual representation demonstrates the main ideas of selecting and organizing up to 500 images for each class, employing transfer training, and producing the final identification model that is set to classify animals within the camera trap study.

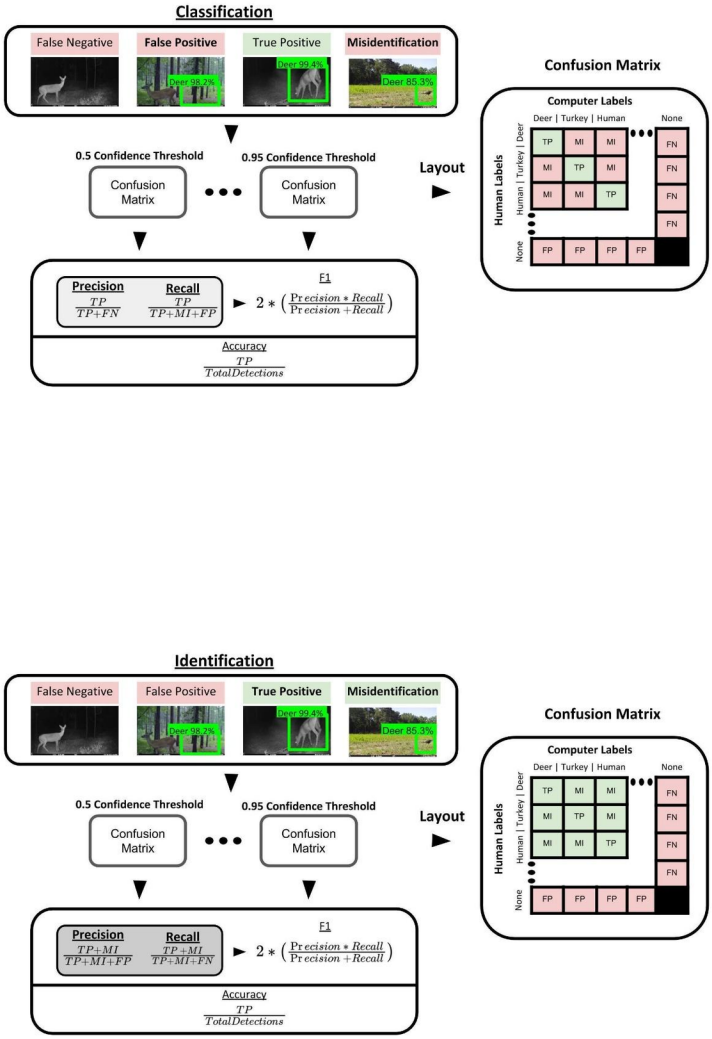
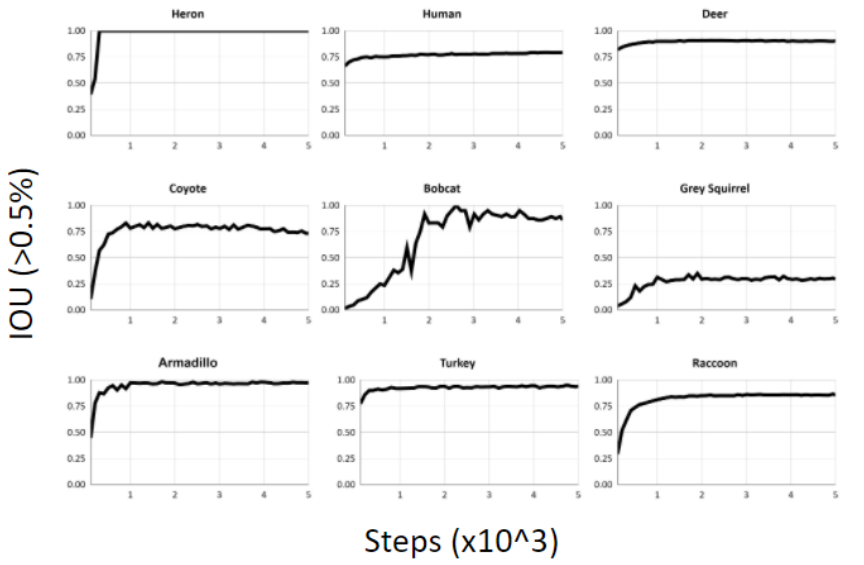


Figure 2. Diagram illustrating calculation of each metric used in training (train and test) data: Precision, Recall, Accuracy and F1 (range of 0 to 1). For identification purposes, misidentifications are counted as correct (green in confusion matrix) because the animal was detected; whereas, for classification purposes, misidentifications are counted as incorrect (red in confusion matrix) because the object was not classified correctly. True Positives (TP), False Positives (FP), and False Negatives (FN) are represented in the confusion matrix with True Negatives (TN) not present in training data. Adjusting confidence thresholds (range of 0.5 to 0.95) optimizes the model for specific applications.



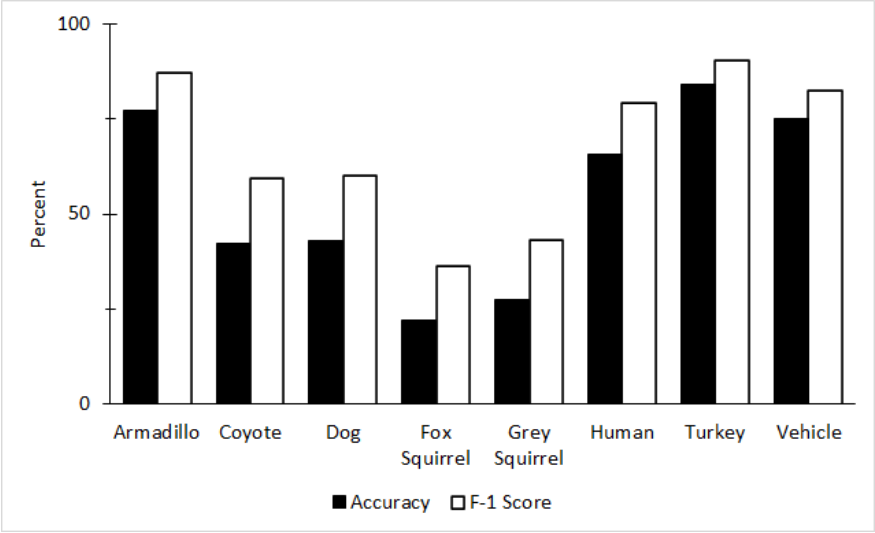


Figure . Comparison of select classes at 0.95 confidence threshold (CT) from test output. F-1 values (white) are consistently higher than the accuracy (black).

Tables

Table 1. Distribution of image subset for train and test datasets by class.

Class	Train Images	Train Objects	Test Images	Test Objects
Armadillo	186	186	21	21
Bobcat	18	18	4	4
Coyote	162	171	18	18

Crow	39	59	11	13
Deer	1109	1379	136	159
Dog	86	114	18	21
Fox Squirrel	79	79	17	18
Grey Fox	88	88	11	11
Grey Squirrel	318	327	32	34
Heron	52	52	3	3
Human	822	1948	89	194
Opossum	18	18	3	3
Rabbit	269	278	17	17
Raccoon	200	208	26	26
Skunk	17	17	2	2
Turkey	430	879	43	80
Vehicle	780	2962	84	271
Total	4673	8783	535	895

Table 2. Details about model training and hardware used.

CPU	Windows 10 Intel i9-9
RAM	64GB
GPU	Nvidia 2070 super 8GB
Batch Size (images per training round)	4
Epoch Steps (complete cycle through training data)	50,000
Train Configuration	Faster R-CNN Inception v2
Training Evaluation	Every 1,000 steps
Evaluation Configuration	Open Images V2 Detection Metric

Summary of averages at each confidence threshold (CT).

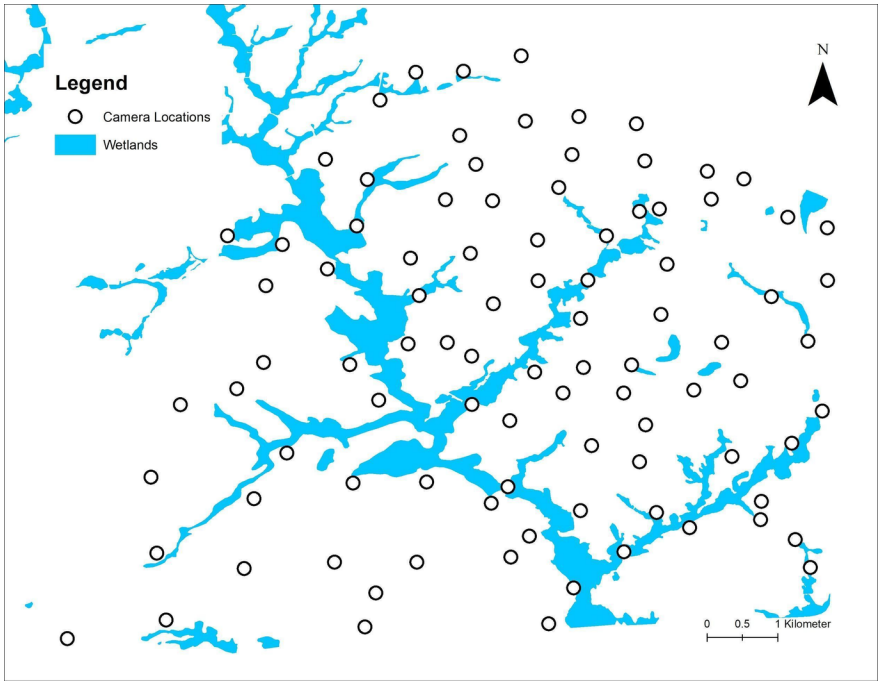
Confidence Threshold	Test - Identification **Accuracy**	Test - Identification **Precision**	Test - Identification **Recall**	Test - **F-1**
0.50	75	75	96	84
0.60	71	79	94	85
0.70	72	81	91	86
0.80	73	84	88	86
0.90	73	88	83	85
0.95	71	90	78	84
Validation - Classification 0.90 Confidence Threshold	Validation - Classification 0.90 Confidence Threshold	Validation - Classification 0.90 Confidence Threshold	Validation - Classification 0.90 Confidence Threshold	Validat 0.90 C

Supplementary Information

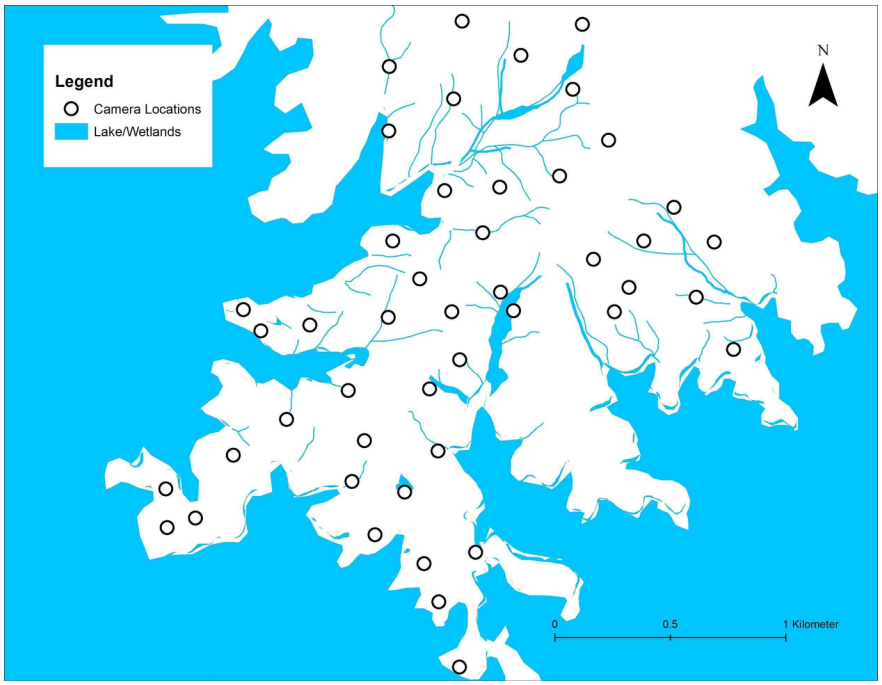
Appendix 1: Camera Trap Study

The images used in this study were all pulled from an ongoing camera trap study on South Carolina Army National Guard (SCARNG) property. This camera trap study covers two locations in the midlands of South Carolina (SI Figs 1 and 2) with the same habitat types and species composition (SI Table 1). Images were

collected and hand-processed by students at the University of South Carolina from McCrady SCARNG training center since October 2017 and from Clark's Hill SCARNG training center since January 2019. Details for each site can be found in SI Table 2.



SI Figure 1. Map of camera locations at McCrady SCARNG training center in South Carolina.



SI Figure 2. Map of camera locations at Clark's Hill SCARNG training center in South Carolina.

SI Table 1. Reference sheet for the name of fauna species and their binomial nomenclature in South Carolina.

Common Name Key	**Common Name Key**
Common Name	Binomial Nomenclature
Armadillo	*Dasypus novemcinctus*
Bobcat	*Lynx rufus*
Coyote	*Canis latrans*
Crow	*Corvus brachyrhynchos*
Deer	*Odocoileus virginianus*
Dog	*Canis familiaris*
Fox Squirrel	*Sciurus niger*
Grey Fox	*Urocyon cinereoargenteus*
Grey Squirrel	*Sciurus carolinensis*
Heron	*Ardea herodias*
Human	*Homo sapiens*
Opossum	*Didelphis virginiana*
Rabbit	*Sylvilagus floridanus*
Raccoon	*Procyon lotor*
Skunk	*Mephitis mephitis*
Turkey	*Meleagris gallopavo*

SI Table 2. Study details for McCrady and Clark's Hill SCARNG training centers in South Carolina.

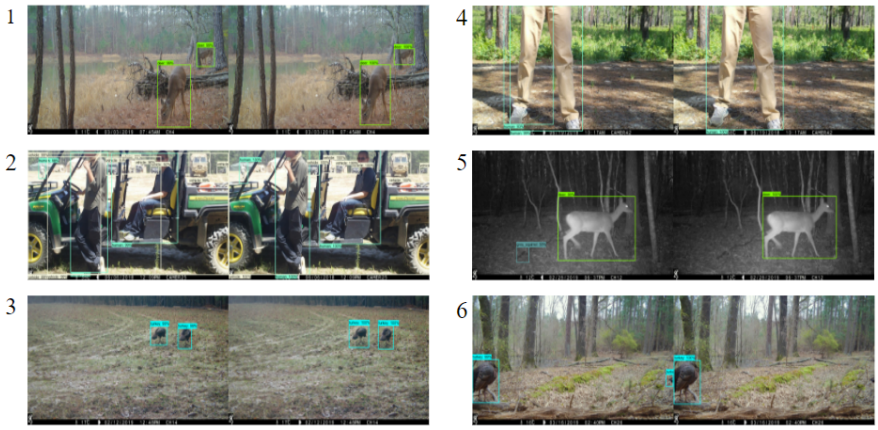
		McCrady	**Clark's Hill**
Study Began	Study Began	October 2017	January 2019
Total Species	Total Species	24	24
Total Individuals	Total Individuals	78,359	4,002
Total Mammal Species	Total Mammal Species	18	15
Individuals of Common Species	*Armadillo*	205	113
	Boar	1	0
	Bobcat	65	10
	Coyote	1,075	123
	Deer	66,128	2,340
	Fox Squirrel	1,776	70
	Grey Fox	153	121
	Grey Squirrel	1,127	510
	Opossum	187	12
	Rabbit	247	171
	Raccoon	965	123
	Turkey	3,352	123

Appendix 2: Bounding Boxes

We used bounding boxes to establish ground truths in our study to increase the value of images, allowing us to use far fewer images to train our model. Bounding boxes provide the model with the location of each object dictating the bounds of the object and background noise (SI Fig. 3, human labeled). Providing the model with images without bounding boxes makes it more difficult for the model to distinguish commonality in patterns of similar objects and would further complicate identification when repeated, uncorrelated,

confounding objects or background noise are present.

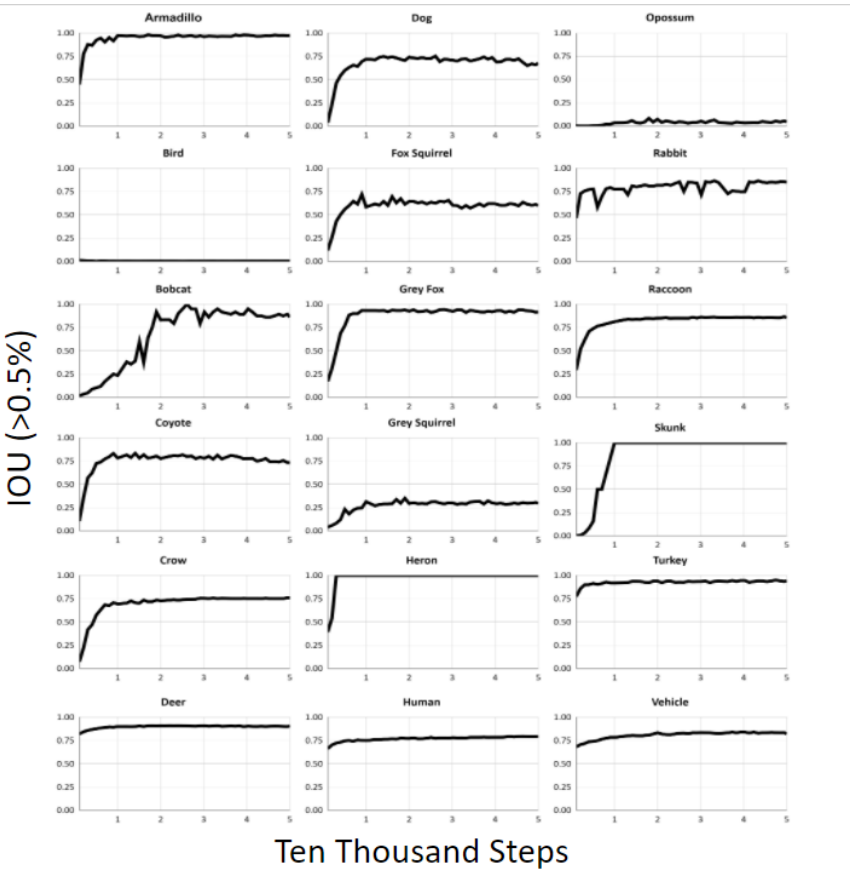
Once trained, the model will identify and classify all objects by placing bounding boxes: a box, the corresponding label of that object, and a feature score. A feature score is the percent likelihood that the object detected reflects the respective label. Our model correctly identified the objects in images 1-3. The model can be more precise than human labelers in finding objects, for example, image 6 displays correctly labeled tail feathers of a turkey that were not labeled correctly by human labelers. Additionally, the model may pick up objects incorrectly (image 5) with low confidence. The confidence threshold (CT) was set at 50%, so any objects detected with over 50% confidence were displayed. This CT can be adjusted to negate low confidence objects, but during training can give insights into errors that may impact validation accuracies and F-1 score. For example, in image 5, images with the same background or images of grey squirrel can be added to further distinguish the misidentified object. Image 4 shows an example of object splitting, when one object is identified by two bounding boxes. Object splitting creates problems with counting the correct number of individuals in an image. Again, adding additional similar images of an event where object splitting can occur can increase the chances of correct bounding boxes. These types of discrepancies suggest the need for a combination of human labelers and AI prescreening for a completely thorough analysis of camera trap imagery.



SI Figure 3. Six images were randomly selected from the test set during training that evaluate the performance of training on the final step (50,000). On each image, the left side is the computer-generated image and the right side is the human labeled image.

Appendix 3: Intersection over union

The model was evaluated throughout the training process using intersection over union (IOU), the degree of overlap between human labeled and computer-generated identifications. Higher IOU represents a greater overlap of the two. For our model, IOU did not depend on the number of images input for training; rather, the uniqueness of objects due to shape and texture was the determining factor. IOU graphs for all object classes are displayed in Supplementary material Fig. 4.



SI Figure 4. All 17 classes' intersection over union (IOU) graphs of greater than 50% overlap over 50,000 steps conducted in training. A class for 'birds' was also created but was not included in our analyses.