2021

Automatic Camera Trap Classification Using Wildlife-Specific Deep Learning in Nilgai Management

Matthew Kutugata

Jeremy Baumgardt

John A. Goolsby
United States Department of Agriculture

Alexis Racelis
The University of Texas Rio Grande Valley

Recommended Citation

Matthew Kutugata, Jeremy Baumgardt, John A. Goolsby, Alexis E. Racelis, M. Kutugata, A.E. Racelis, J. Baumgardt, A. Goolsby; Automatic Camera Trap Classification Using Wildlife-Specific Deep Learning in Nilgai Management. Journal of Fish and Wildlife Management 2021; doi: https://doi.org/10.3996/JFWM-20-076

Journal of Fish and Wildlife Management

Automatic Camera Trap Classification Using Wildlife-Specific Deep Learning in Nilgai Management

--Manuscript Draft--

Manuscript Number:	JFWM-20-076R1
Article Type:	Article
Keywords:	Boselaphus tragocamelus; camera trap; cattle fever ticks; deep learning; nilgai
Corresponding Author:	Matthew Kutugata Edinburg, Texas UNITED STATES
First Author:	Matthew Kutugata
Order of Authors:	Matthew Kutugata
	Jeremy Baumgardt
	John A. Goolsby
	Alexis E. Racelis
	M. Kutugata
	A.E. Racelis
	J. Baumgardt
	A. Goolsby
Abstract:	Camera traps provide a low-cost approach to collect data and monitor wildlife across large scales but hand-labeling images at a rate that outpaces accumulation is difficult. Deep learning, a subdiscipline of machine learning and computer science, has been shown to address the issue of automatically classifying camera trap images with a high degree of accuracy. This technique, however, may be less accessible to ecologists, to small scale conservation projects, and has serious limitations. In this study, a simple deep learning model was trained using a dataset of 120,000 images to identify the presence of nilgai Boselaphus tragocamelus, a regionally specific non-native game animal, in camera trap images with an overall accuracy of 97%. A second model was trained to identify 20 groups of animals and 1 group of images without any animals present, labeled as “none”, with an accuracy of 89%. Lastly, the multigroup model was tested on images collected of similar species but in the southwestern United States and resulted in significantly lower precision and recall for each group. This study highlights the potential of deep learning for automating camera trap image processing workflows, provides a brief overview of image-based deep learning, and discusses the often-understated limitations and methodological considerations in the context of wildlife conservation and species monitoring.

Automatic Camera Trap Classification Using Wildlife-Specific Deep Learning in Nilgai Management

Matthew Kutugata*, Jeremy Baumgardt, John A. Goolsby, and Alexis E. Racelis

M. Kutugata, A.E. Racelis
[1]University of Texas Rio Grande Valley, School for Earth, Environmental and Marine Sciences. 1201 W. University Dr., Edinburg, TX 78539, USA

J. Baumgardt
[2] Caesar Kleberg Wildlife Research Institute, Texas A&M University, Kingsville, TX 78363, USA

A. Goolsby
[3] United States Department of Agriculture, Agricultural Research Service, Knipling -Bushland U.S. Livestock Insects Research Laboratory, Cattle Fever Tick Research Laboratory, Edinburg, TX 78541, USA

Abstract

Camera traps provide a low-cost approach to collect data and monitor wildlife across large scales but hand-labeling images at a rate that outpaces accumulation is difficult. Deep learning, a subdiscipline of machine learning and computer science, has been shown to address the issue of automatically classifying camera trap images with a high degree of accuracy. This technique, however, may be less accessible to ecologists, to small scale conservation projects, and has serious limitations. In this study, a simple deep learning model was trained using a dataset of 120,000 images to identify the presence of nilgai *Boselaphus tragocamelus,* a regionally specific non-native game animal, in camera trap images with an overall accuracy of 97%. A second model was trained to identify 20 groups of animals and 1 group of images without any animals present, labeled as “none”, with an accuracy of 89%. Lastly, the multigroup model was tested on images collected of similar species but in the southwestern United States and resulted in significantly lower precision and recall for each group. This study highlights the potential of deep learning for automating camera trap image processing workflows, provides a brief overview of image-based deep learning, and discusses the often-understated limitations and methodological considerations in the context of wildlife conservation and species monitoring.

KEY WORDS *Boselaphus tragocamelus*, camera trap, cattle fever ticks, deep learning, nilgai, transfer learning

Received: October 2020; Accepted: July 2021; Published Online Early: July 2021; Published: xxx

Citation: Kutugata M, Baumgardt J, Goolsby JA, and Racelis AE. 2021. Automatic camera trap classification using wildlife-specific deep learning in nilgai management. *Journal of Fish and Wildlife Management* X(X):xx-xx; e1944-687X. https://doi.org/10.3996/JFWM-20-076

This Online Early paper will appear in its final typeset version in a future issue of the *Journal of Fish and Wildlife Management*. This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination, and proofreading process, which may lead to differences between this version and the Version of Record. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Introduction

Camera traps, wireless cameras placed on trees or posts activated via motion sensors, are an important tool for wildlife studies. They have been used to estimate population densities (Howe et al. 2017), create species lists and inventories in dense tropical environments (Srbek-Araujo and Chiarello 2005; Lading 2006), understand population size and distributions (O'Connell et al. 2010), and identify new species (Rovero and Rathbun 2006). Their relatively low-cost and ease make them scalable across large geographic regions. A common problem, however, is the rapid accumulation of images that outpace the ability of users to manually sort and label them (Swanson et al. 2016). To address this issue, deep learning, a subfield of machine learning, has been identified as a powerful technique to automate the process of classifying, or grouping, images by species (Gomez et al. 2016; Norouzzadeh et al. 2018; Willi et al. 2019). Applications of deep learning for camera trap classification have often relied on extremely large collections of images like Snapshot Serengeti (~7 million images) or the North American Camera Trap dataset (3.3 million images) for training (Swanson et al. 2015; Tabak et al. 2019; Schneider et al. 2020). Transfer learning, a deep learning technique that starts with pretrained models as a base for future learning, has been shown to overcome this problem. Both Schneider et al. (2020) and Shahinfa et al. (2020) found that only 1,000 images per class were needed to achieve an accuracy of 97% and 98%, respectively, for eight classes. Despite growing popularity, applications of transfer learning for rapid camera trap classification may still be considered beyond the expertise of many ecologists and conservation practitioners.

Our aim was to present an application of deep learning-based camera trap analysis using a small dataset of 120,000 images. We trained a model using transfer learning, evaluated its accuracy, and demonstrated its limitations when being applied on images outside the models training context. We leveraged a nature-specific model by Cui et al. (2018) as a base to further train a south Texas specific animal classifier. More specifically, we drew from a local database of camera trap images to train: 1) a binary classifier that discriminates between a single species, nilgai *Boselaphus tragocamelus*, an exotic bovid with expanding populations in south Texas and 2) a multigroup classifier for 20 animal groups and one “none” group. Lastly, we tested the model and its ability to generalize on images with similar classes but in different settings using the CalTech camera trap dataset collected in the southwestern United States (Beery et al. 2018).

Resources and further details about training and implementation can be found at the authors github repository (https://github.com/mkutu/Nilgai).

Study Site

Image data were collected from motion-sensitive cameras placed in areas of known wildlife activity in Cameron County in the lower Rio Grande Valley of Texas from 2018 to 2019. This county is along the international border and characterized by a mosaic of shrubby plants, mesquite, and semi-arid vegetation. Free ranging nilgai native to the Indian subcontinent were introduced in the 1930s (Leslie 2008). Although there appears to be no competition with other native species, nilgai inhabit areas that support species of conservation concern such as northern populations of ocelot *Leopardus pardalis* and perhaps the Gulf Coast jaguarundi *Puma yagouaroundi cacomitli* (Schmidly 2004; Leslie 2016). Furthermore, recent studies reveal that nilgai are optimal hosts for the southern cattle-fever tick *Rhipicephalus microplus* and have exacerbated current efforts to eradicate this exotic pest of wildlife and livestock (Lohmeyer et al. 2018). As such, monitoring nilgai behavior, population, and distribution have important implications for both wildlife management and agriculture in the region (Foley et. al. 2017; Goolsby et al. 2019).

Methods

Image data and preprocessing

Images were randomly drawn for each group from a local database part of a multi-year field research project aimed at treating cattle fever tick-infested nilgai at fence crossings. Images were hand-labeled by research technicians with advanced experience in recognizing animals of interest. Images were labeled using the open-access Colorado Parks and Wildlife Photo Warehouse, a custom Microsoft Offices Access application designed specifically to store, manage, label, and analyze wildlife camera trap data (Ivan and Newkirk 2016). Three types of datasets were created necessary for training deep neural networks: 1) a large *training set* (~85% of total images) for model learning, 2) a smaller *validation set* (~5% of total images) for frequent testing and adjustment of model settings, and 3) a *test set* to evaluate the final trained model (~10% of total images). Separate train, validation, and test sets were created for each classifier.

Balancing training set

A balanced training set contains an even distribution of images across each group. The original raw image set of >2.5 million images was highly imbalanced with 84% (~2 million images) having no wildlife, which was labeled as "none". The top seven most common groups include feral pigs *Sus scrofa*, falsely triggered camera events, human activity, birds, nilgai, deer *Odocoileus virginianus*, and cattle. Camera trap datasets are often imbalanced because of wind, grass, or other nontarget objects that create false capture events. Training on the complete dataset would be problematic because models can favor groups with more examples while ignoring those with only a few (Norouzzadeh et al. 2018). Model adjustments could be made in such a way that "none" is chosen for most images resulting in a high overall accuracy. To correct the imbalance, we oversampled or sampled with replacement so each group had roughly the same number of images (He and Garcia 2009). For example, if the "dog" group only had 50 unique images, each was copied until the total number of images matched that of the most frequent occurring group. While this oversampling technique balances the dataset, it has drawbacks. Because images in rare groups are repeated, the model lacks robustness in these groups to generalize on new examples in the future. This might be an issue for conservation projects focusing on rare species that are important to monitor but rarely occur. For this study, however, the most important group, "nilgai", was one of the most frequently occurring. Still, to reduce the number of copies for oversampling, we lowered our total image set size from 2.5 million to 120,000 by taking slightly more than the next most frequent group ("human"). Additionally, a dataset of 120,000 images instead of 2.5 million lowered training time from weeks to days. Data was further altered by combining or eliminating groups. Four groups were combined – "feral cat", "ocelot", "bobcat" *Lynx rufus,* and "exotics, other" – to create the "cat" group while "unknown" and "squirrel" were eliminated. These groups either lacked sufficient examples or were mislabeled (e.g., an image of a "bobcat" was labeled as "ocelot"). Each capture event included three images taken in rapid successive order. Individual images, not capture events, were classified by research technicians, and contributed to the total dataset size and class count.

We applied four types of data augmentation, a technique commonly used to strengthen model predictions by slightly altering images. Images were rotated, shifted, sheared, and flipped both horizontally and vertically. Augmentation was performed for each training cycle and different augmentations were performed randomly for each image. Preprocessing also included rescaling pixel values between 0 and 1 and resizing the image from 2,048 x 1,152 to 299 x 299 pixels – standard procedures done to reduce the computational expense of training. The seven most common groups included feral hogs, a "none" group, human activity, birds, white-tailed deer, and cattle (Figure 1). Data preprocessing is an important step for reducing computational demands and increasing model robustness.

Deep learning

A subfield of machine learning, deep learning aims to extract information from big data by learning from successive layers of increasingly meaningful representations called features (Chollet 2018). A deep learning model, a neural network, is made up of many layers that are trained on labeled data and extract features hierarchically. Information from previous layers informs following layers and is stored in the form of weights to make predictions on new unlabeled data. The neural network uses predicted and actual values to calculate an error score that is propagated back through the network to adjust weight values. Learning occurs iteratively by updating weights in such a way that optimizes its ability to reduce its error score. Early layers are trained to react strongly to simple features like edges, lines, and sharp color gradients, while the final layer of a neural network infers probabilities of input features to a class like “nilgai” or “deer”. Features are distilled hierarchically from complex input images to a single prediction value (Figure 2; Toda and Okura 2019).

Training a neural network from scratch often requires large amounts of data. However, transfer learning, an approach useful for training on small datasets, applies the stored knowledge of a model pretrained on large generic data as a base for similar but more specific problems. Knowledge is transferred in the form of saved files that contain weights, complete or partial model architectures, and settings. Model parameters can be easily downloaded from open-source libraries and read into a new training instance. Feature extraction, the first step in using pretrained models, involves replacing and training only the final layer of a neural network on a new problem-specific dataset. The second process trains all layers including the newly added final layer. Network weights are adjusted making the model task-specific. Feature extraction must occur first since the final layer restricts overly large weight adjustments that could negatively affect inference or model prediction. Our model was pretrained by Cui et al. (2018), who used the iNaturalist 2017 dataset of 579,184 nature-specific objects including insects, mammals, and amphibians (Ueda 2017; Vanhorn et al. 2018). We then trained on a smaller but domain-specific dataset of south Texas wildlife (Figure 3).

Training and evaluation

We customized the InceptionV3 model, defined by its sequence and type of layers, to our unique number of groups (Szegedy et al. 2016). After each training cycle, we used the validation set to monitor performance and adjust model settings. In total, ~21-million weight parameters were updated until the model stopped improving on the validation set, roughly 24 hours for both the multi-label and binary classifier while using a single graphic processing unit. We evaluated each model after adjustments and training were completed by reporting prediction results on the test set – the number of true positives, true negatives, false positives, and false negatives - for each classifier. We calculated five common accuracy metrics: overall accuracy, precision, recall, harmonic mean using precision and recall known as F1 score, and Matthews correlation coefficient, an adjusted form of the phi coefficient (Table 1; Guilford 1954). A second test set, collected from the southwestern Unites States and known as the CalTech dataset, was used to further evaluate model robustness (Beery et al. 2018).

Results

The trained binary classifier achieved an overall accuracy of 0.97, F1 score of 0.97 and Matthews correlation coefficient of 0.94 indicating the classifier was able to generalize on new images from the same area and accurately predict the presence of a nilgai. During training, we found a ~15% increase in validation accuracy from the first to second stage. Recall (0.98) was slightly larger than precision (0.96), which is favorable for this unique task. The occasional instance of deer or cattle classified as a nilgai is preferred because these images will likely be reviewed and ‘caught’ by research technicians. A lost and uncounted nilgai image, however, is more detrimental to overall project goals. For multigroup problems, the average of Matthews correlation coefficient is a more appropriate evaluation metric because it pools the performance over all samples and groups. Our multigroup classifiers achieved an average Matthews correlation coefficient of 0.89. Group-wise test results and evaluation metrics show that two of the most highly correlated classes – "skunk" *Mephitis mephitis* and "tortoise" *Gopherus berlandieri* - were the most imbalanced with each having <22 images (Table 2). The three most common groups in our dataset – "nilgai", "deer", and "none" – were strongly correlated. The multigroup classifier was successful in classifying 21 groups (Figure 4). For the second evaluation using the CalTech dataset, classes were adjusted to complement those of the south Texas dataset. Dissimilar classes were removed ("bat", "lizard", "badger"), similar classes were combined ("car" and "human"), and classes were renamed when appropriate ("bobcat" to "cat"). The average Matthews correlation coefficient for the CalTech dataset was 0.22, further inspection of the other four metrics by class also indicated very poor performance (Table 3).

Discussion

Our aim was to test if a small number of hand-labeled camera trap images could be used to train a deep learning model to automatically detect wildlife including a specific species. We also explored the limits of our model by testing on a dataset that was not used in training, had similar species, but different context. Class imbalance played a major role in skewing the performance of the model on rare classes where test images were found to be similar to training images. For example, a tortoise's slow movement was enough to trigger the camera sensor multiple times, which resulted in many nearly identical images. Because rare groups contained even fewer number of images in the test set, it was difficult to evaluate their accuracy. Addressing the class imbalance issue is an important factor for improving results. Applying a technique like emphasis sampling attempts to increase prediction accuracy by duplicating, or emphasizing, only images that have been misclassified instead of oversampling all rare groups (Norouzzadeh et al. 2018). This approach is more dynamic because it balances data as needed by responding to prediction results. Alternatively, multiple data sources can be combined to add images to rare classes from other camera trap datasets (Swanson et al. 2015; LILA BC 2019). However, this approach risks introducing too dissimilar environmental setting, images, and class types. Secondly, evaluating a second dataset allowed us to illustrate the model's lack of location invariance or inability to generalize on new images with conditions not represented in the training set (Beery et al. 2018). The strength of the model to make accurate predictions under a diverse set of conditions depends on how well those conditions are represented in the training data. Lastly, adopting a trained model into an automatic camera trap classification workflow should be closely monitored by inspecting important and rare groups for anomalies or regularly testing on a subset of new images. As our study shows, new camera angles, species, or locations pose challenges to accurate classifications. Transfer learning has the potential to save time and resources typically required to hand-label camera trap images. A simple trained classifier making predictions on 3,000 raw images saves roughly 12 personnel hours. Applications of deep learning, while traditionally left to experts in computer vision, have become less complicated with the emergence of publicly available datasets and open-source software. Likewise, we include our code, trained model, instructions, and a set of sample images that we hope improve the transfer of knowledge from academia to the field.

Supplemental Material

Please note: The Journal of Fish and Wildlife Management is not responsible for the content of functionality of any supplemental material. Queries should be directed to the corresponding author.

Data S1. A set of two IPython notebooks to automatically classify and evaluate sample images using a deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. The model is designed to classify images as wildlife or as being empty (a false camera trigger event). Notebooks use additional supplemental data such as input weight files, a sample repository of images, and true image labels to evaluate predictions. The notebooks generate a new set of folders for each class, copies input images, and places them in folders based on predicted group. The notebooks generate figures of the distribution of predictions across animal groups. A csv file containing true image labels is applied to generate an evaluation report. Available at: https://github.com/mkutu/Nilgai/tree/master/notebooks (15.57 MB IPYNB).

Data S2. A sample of 222 new images from the camera trap dataset collected in the lower Rio Grande Valley in Texas in 2018 and 2019. This sample, along with true label information (also provided in the supplemental material) can be used to test the deep learning model to automatically classify images as wildlife or as being empty (a false camera trigger event). Available at: https://github.com/mkutu/Nilgai/tree/master/images/images (28.5 MB JPG).

Data S3. The csv file contains true image label information for evaluating the accuracy of a deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. Available at: https://github.com/mkutu/Nilgai/blob/master/notebooks/image_labels.csv (7.19 KB CSV).

Data S4. A set of two .h5 files that contain the stored weights and model settings created by training a deep learning model on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. Available at: https://github.com/mkutu/Nilgai/tree/master/model (250.6 MB H5).

Text S1. A “README.md” text file with instructions for creating a virtual environment needed for running a deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. A virtual environment allows users to install dependencies, small pieces of software in the form of source code, that are required to run Python programs without making major changes to the users’ systems. Instructions outline the procedures for setting up environments for both Windows and Mac OSX operating systems. Notes on trouble shooting are also included. Available at: https://github.com/mkutu/Nilgai/blob/master/README.md (3.03 KB MD).

Text S2. A “requirements.txt” text file used to install the required Python dependencies, small pieces of software in the form of source code, inside the virtual environment. Dependencies are required to run the deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. Available at: https://github.com/mkutu/Nilgai/blob/master/requirements.txt (113 BYTES TXT).

Acknowledgements

Game camera images and initial processing was supported through appropriated research project 3094-32000-042-00-D, Integrated Pest Management of Cattle Fever Ticks. This article reports results of research only and mention of a proprietary product does not constitute an endorsement or recommendation by the U.S. Department of Agriculture for its use. U.S. Department of Agriculture is an equal opportunity provider and employer. Special thanks to Amelia Berle for data management, and research technicians who spent countless hours labeling images. Additional thanks to Dr. Rupesh Kariyat and Dr. Christofferson for providing access to computing equipment. We would also like to thank the Journal reviewers and Associate Editor for their commitment to open-access that ensures applied conservation science remains accessible to all. Matthew Kutugata was supported by USDA National Institute of Food and Agriculture Grant #2016-38422-25543.

Any use of trade, product, website, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Ahmed M, Najmul Islam AKM. 2020. Deep learning: hope or hype. Annals of Data Science 7:427–432.

Beery S, Horn GV, Perona P. 2018. Recognition in terra incognita. Proceedings of the Computer Vision - ECCV 2018 - 15th European Conference.

Chollet F. 2018. Deep learning with Python. Shelter Island, New York: Manning Publications Co.

Cui Y, Song Y, Sun C, Howard A, Belongie S. 2018. Large scale fine-grained categorization and domain-specific transfer learning. In Conference on Computer Vision and Pattern Recognition.

Foley AM, Goolsby JA, Ortega-S. A, Ortega-S. JA, Pérez de León A, Singh NK, Schwartz A, Ellis D, Hewitt DG, Campbell TA. 2017. Movement patterns of nilgai antelope in South Texas: implications for cattle fever tick management. Preventive Veterinary Medicine 146:166–172.

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. Pages 747–756 in Bebis G, Boyle R, Parvin B, Koracin D, Porikli F, Skaff S, Entezari A, Min J, Iwai D, Sadagic A, Scheidegger C, Isenberg T, editors. International symposium on visual computing. Cham, Switzerland: Springer International Publishing.

Goolsby J, Cantu D, Vasquez A, Racelis A. 2019. Development of a remotely activated field sprayer and evaluation of temperature and aeration on the longevity of Steinernema riobrave entomopathogenic nematodes for treatment of cattle fever tick-infested nilgai. Subtropical Agriculture and Environments 70: 1-5.

Guilford JP. 1954. Psychometric methods. New York City: McGraw-Hill

He H, Garcia EA. 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21:1263–1284.

Howe EJ, Buckland ST, Després- Einspenner ML, Kühl HS. 2017. Distance sampling with camera traps. Methods in Ecology and Evolution 8:1558–1565.

Ivan JS, Newkirk ES. 2016. Cpw photo warehouse: a custom database to facilitate archiving, identifying, summarizing and managing photo data collected from camera traps. Methods in Ecology and Evolution 7:499–504.

Lading E. 2006. Camera trapping and conservation in lambir hills national park, Sarawak. The Raffles Bulletin of Zoology 54:469–475

[LILA BC] 2019. Labeled information library of Alexandria: biology and conservation. Available: http://lila.science/datasets (July 2021).

Leslie DM Jr. 2008. Boselaphus tragocamelus (Artiodactyla: bovidae). Mammalian Species 813:1–16.

Leslie DM Jr. 2016. An international borderland of concern: conservation of biodiversity in the lower Rio Grande Valley. Reston, VA: U.S. Geological Survey. Scientific Investigations Report 2016-5078

Lohmeyer KH, May MA, Thomas DB, Pérez de León AA. 2018. Implication of nilgai antelope (Artiodactyla: bovidae) in reinfestations of Rhipicephalus (boophilus) microplus (Acari: ixodidae) in South Texas: a review and update. Journal of Medical Entomology 55:515–522.

Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, 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:E5716-E5725.

O’Connell AF, Nichols JD, Karanth KU. 2010. Camera traps in animal ecology: methods and analyses. New York: Springer Science + Business Media.

Rovero F, Rathbun GB. 2006. A potentially new giant sengi (elephant-shrew) from the Udzungwa Mountains, Tanzania. Journal of East African Natural History 95(2): 111-115.

Schmidly DJ, Bradley RD. 2004. The mammals of Texas. Austin, Texas: University of Texas Press.

Srbek-Araujo AC, Chiarello AG. 2005. Is camera-trapping an efficient method for surveying mammals in neotropical forests? A case study in South-Eastern Brazil. Journal of Tropical Ecology 21:121–125.

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 2:1-14.

Swanson A, Kosmala M, Lintott C, Packer C. 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conservation Biology 30:520–531.

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. 2016. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Tabak MA, Norouzzadeh MS, Wolfson DW, Sweeney SJ, Vercauteren KC, Snow NP, Halseth JM, Salvo PAD, Lewis JS, White MD, et al. 2019. Machine learning to classify animal species in camera trap images: applications in ecology. Methods in Ecology and Evolution 10:585–590.

Toda Y, Okura F. 2019. How convolutional neural networks diagnose plant disease. Plant Phenomics 2019.

Ueda K. 2017. iNaturalist research-grade observations. Online database: occurrence dataset. Available: https://doi.org/10.15468/ab3s5x (July 2021).

Van Horn G, Mac Aodha O, Song Y, Cui Y, Sun C, Shepard A, Adam H, Perona P, Belongie S. 2018. The INaturalist species classification and detection dataset. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Willi M, Pitman RT, Cardoso AW, 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 in Ecology and Evolution 10:80–91.

Table 1

Evaluation metric	Description	Equation
Accuracy	Calculates the ratio of all correct predictions out of all instances	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	Calculates the ratio of true positives to total positives	$\frac{TP}{TP + FP}$

Recall	Calculates the ratio of true positives to all conditional positives	$\frac{TP}{TP + FN}$
F1 score	Uses precision and recall to apply a harder penalty when one measure improves at the expense of another	$2 \cdot \frac{Precision\ \cdot\ Recall}{Precision + Recall}$
Matthews correlation coefficient	Correlation between true and predicted results using values between -1 and +1	$\frac{TP \cdot TN - \ FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$

Table 2

Class	Images	TN	TN	FP	FN	Precision	Recall	Accuracy	F1	MCC
Armadillo	262	241	9717	21	21	0.92	0.92	1.00	0.92	0.92
Birds	856	781	9070	74	75	0.91	0.91	0.99	0.91	0.91
Cat	449	366	9508	43	83	0.90	0.82	0.99	0.85	0.85
Cattle	1325	1142	8589	86	183	0.93	0.86	0.97	0.90	0.88

Coyote	489	421	9444	67	68	0.86	0.86	0.99	0.86	0.86
Deer	867	743	8998	135	124	0.85	0.86	0.97	0.85	0.84
Dog	99	88	9900	1	11	0.99	0.89	1.00	0.94	0.94
Horse	12	12	9983	5	0	0.71	1.00	1.00	0.83	0.84
Humans	869	784	9101	30	85	0.96	0.90	0.99	0.93	0.93
Mouse	683	582	9277	40	101	0.94	0.85	0.99	0.89	0.89
Nilgai	805	700	9057	138	105	0.84	0.87	0.98	0.85	0.84
None	857	770	8984	159	87	0.83	0.90	0.98	0.86	0.85
Opossum	201	182	9759	40	19	0.82	0.91	0.99	0.86	0.86
Pig	788	713	9144	68	75	0.91	0.91	0.99	0.91	0.90
Rabbit	561	524	9367	72	37	0.88	0.93	0.99	0.91	0.90
Raccoon	584	535	9374	42	49	0.93	0.92	0.99	0.92	0.92
Rat	537	501	9384	79	36	0.86	0.93	0.99	0.90	0.89
Skunk	21	20	9979	0	1	1.00	0.95	1.00	0.98	0.98
Spider	19	17	9977	4	2	0.81	0.90	1.00	0.85	0.85
Tortoise	12	12	9987	1	0	0.92	1.00	1.00	0.96	0.96
Turkey	153	150	9827	20	3	0.88	0.98	1.00	0.93	0.93

Table 3

Class	Images	TP	TN	FP	FN	Precision	Recall	Accuracy	F1	MCC
Armadillo	0	0	9872	128	0	0.00		0.99	0.00	
Birds	414	82	9381	205	332	0.29	0.20	0.95	0.23	0.21
Cat	541	129	9164	295	412	0.30	0.24	0.93	0.27	0.23

Cattle	689	223	8760	551	466	0.29	0.32	0.90	0.31	0.25
Coyote	765	427	8762	473	338	0.47	0.56	0.92	0.51	0.47
Deer	484	278	8663	853	206	0.25	0.57	0.89	0.34	0.33
Dog	196	12	9803	1	184	0.92	0.06	0.98	0.12	0.24
Horse	0	0	9999	1	0	0.00		1.00	0.00	
Human	203	192	9374	423	11	0.31	0.95	0.96	0.47	0.53
Nilgai	0	0	9318	682	0	0.00		0.93	0.00	
None	5175	2501	4437	388	2674	0.87	0.48	0.69	0.62	0.44
Opossum	667	125	9274	59	542	0.68	0.19	0.94	0.29	0.34
Pig	0	0	8710	1290	0	0.00		0.87	0.00	
Rabbit	535	193	9034	431	342	0.31	0.36	0.92	0.33	0.29
Raccoon	449	140	9428	123	309	0.53	0.31	0.96	0.39	0.39
Rodent	176	12	9667	157	164	0.07	0.07	0.97	0.07	0.05
Skunk	76	9	9914	10	67	0.47	0.12	0.99	0.19	0.23
Squirrel	171	0	9829	0	171		0.00	0.98	0.00	
Tortoise	0	0	9969	31	0	0.00		1.00	0.00	

Table 1. Five metrics used to evaluate the accuracy of a deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. The five metrics include overall accuracy, precision, recall, harmonic mean using precision and recall known as the F1 score, and Matthews correlation coefficient. Descriptions and equations have

also been provided. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), were gathered from prediction results.

Table 2. Evaluation results of a deep learning model trained and tested on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. Results of predictions made on new images not included in training were compared with their true labels to calculate overall accuracy, precision, recall, harmonic mean using precision and recall known as the F1 score (F1), and Matthews correlation coefficient (MCC). The precision, recall, accuracy, and F1 score (F1) are ratios from 0 to 1 while MCC is between -1 and 1.

Table 3. Evaluation results for a deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019 but tested on the CalTech camera trap dataset (Beery et al. 2018). The CalTech dataset was collected in the southwestern United States in 2018, contains similar animal groups, but includes conditions and backgrounds which are absent in the original Texas training set. Results of predictions made on images not included in training were compared with their true labels to calculate overall accuracy, precision, recall, harmonic mean using precision and recall known as the F1 score (F1), and Matthews correlation coefficient (MCC). The precision, recall, accuracy, and F1 score (F1) are ratios from 0 to 1 while MCC is between -1 and 1.

Figure 1. Examples of cropped and resized camera trap images collected in the lower Rio Grande Valley of Texas in 2018 and 2019 and used for training a deep learning model that can automatically classify new images of wildlife. The top seven most common animal groups in the image dataset include; A) feral pigs labeled "pigs", B) falsely triggered capture events without

animals as "none", C) signs of human activity as “human”, D) “bird”, E) “nilgai”, *Boselaphus tragocamelus*, F) white-tailed deer *Odocoileus virginianus* as “deer”, and G) “cattle”.

Figure 2. Inside a deep learning model trained on camera trap images collected in the lower Rio Grande Valley in Texas in 2018 and 2019. The trained model identifies important image patterns, or features, associated with each class to make predictions. Image data are distilled to

its representative features; filtering layers extract meaningful characteristics (highlighted in yellow), a flattening layer transforms a 3-dimensional array of feature values into 2-dimensions, and the final connected layer produces predicted model probabilities by class ending with an output label, “Nilgai” *Boselaphus tragocamelus*. Parentheses indicate the dimensions of image data (width, length, channel).

Figure 3. Transfer learning was performed by updating a model pretrained on a larger iNaturalist dataset using a small but regionally specific camera trap dataset collected in the lower Rio

Grande Valley in Texas in 2018 and 2019 to automatically classify new, unlabeled images (Ueda 2017). Transfer learning applies the learned features of large datasets to a more specific task.

Figure 4. A random sample of 16 model predictions illustrates the performance of a deep learning model trained on camera trap image collected in the lower Rio Grande Valley in 2018

and 2019. The trained model was designed to classify images into 20 animal groups and 1 empty "none" group. Sample test images were drawn from the original dataset but were not included for training. Classifier predictions are titled for each image. In this sample, a single incorrectly labeled image, middle-right, predicted as "pig" was in fact an image of "nilgai" *Boselaphus tragocamelus* as shown by the white arrow.

Figures

Figure 1

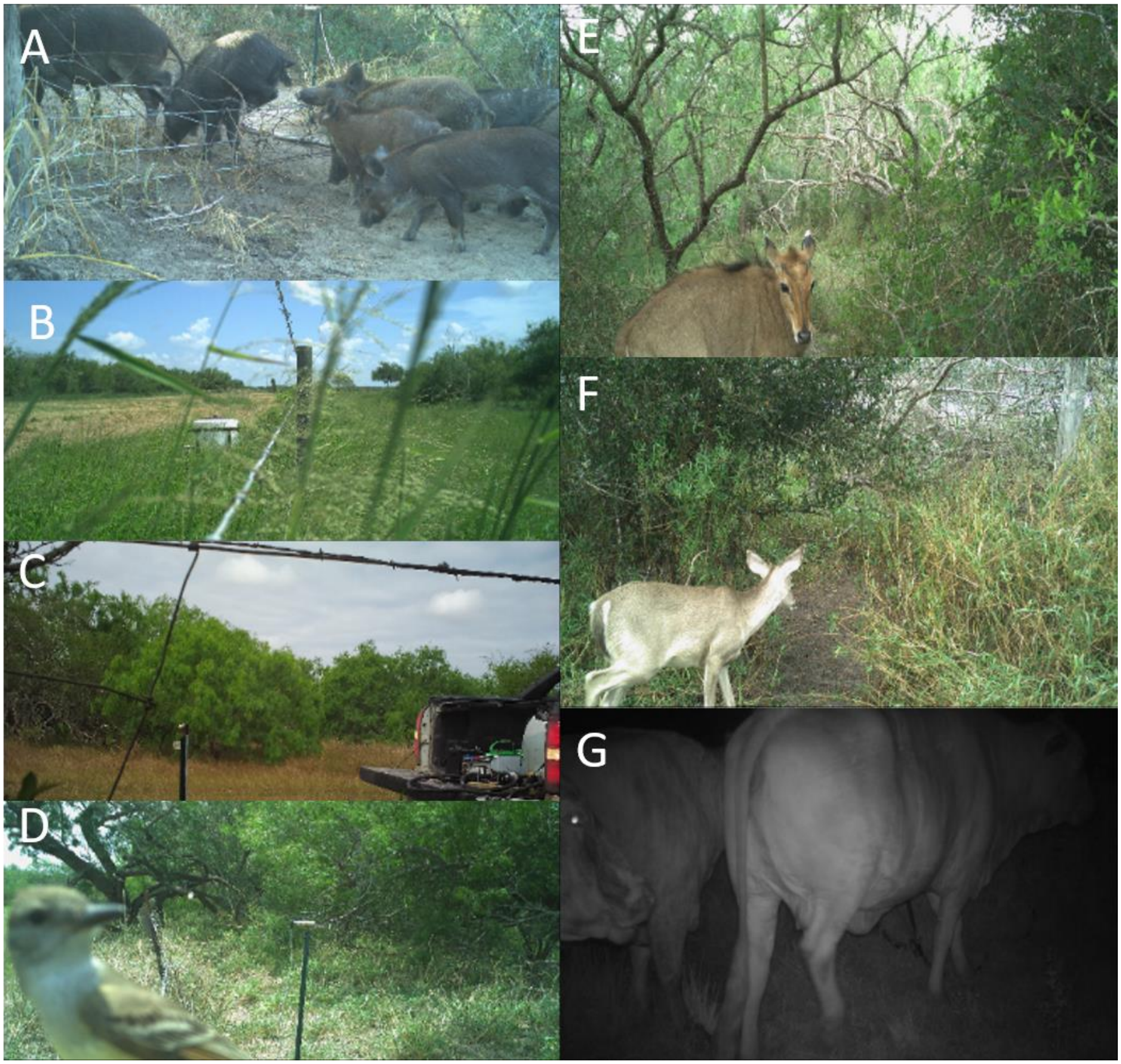

Figure 2

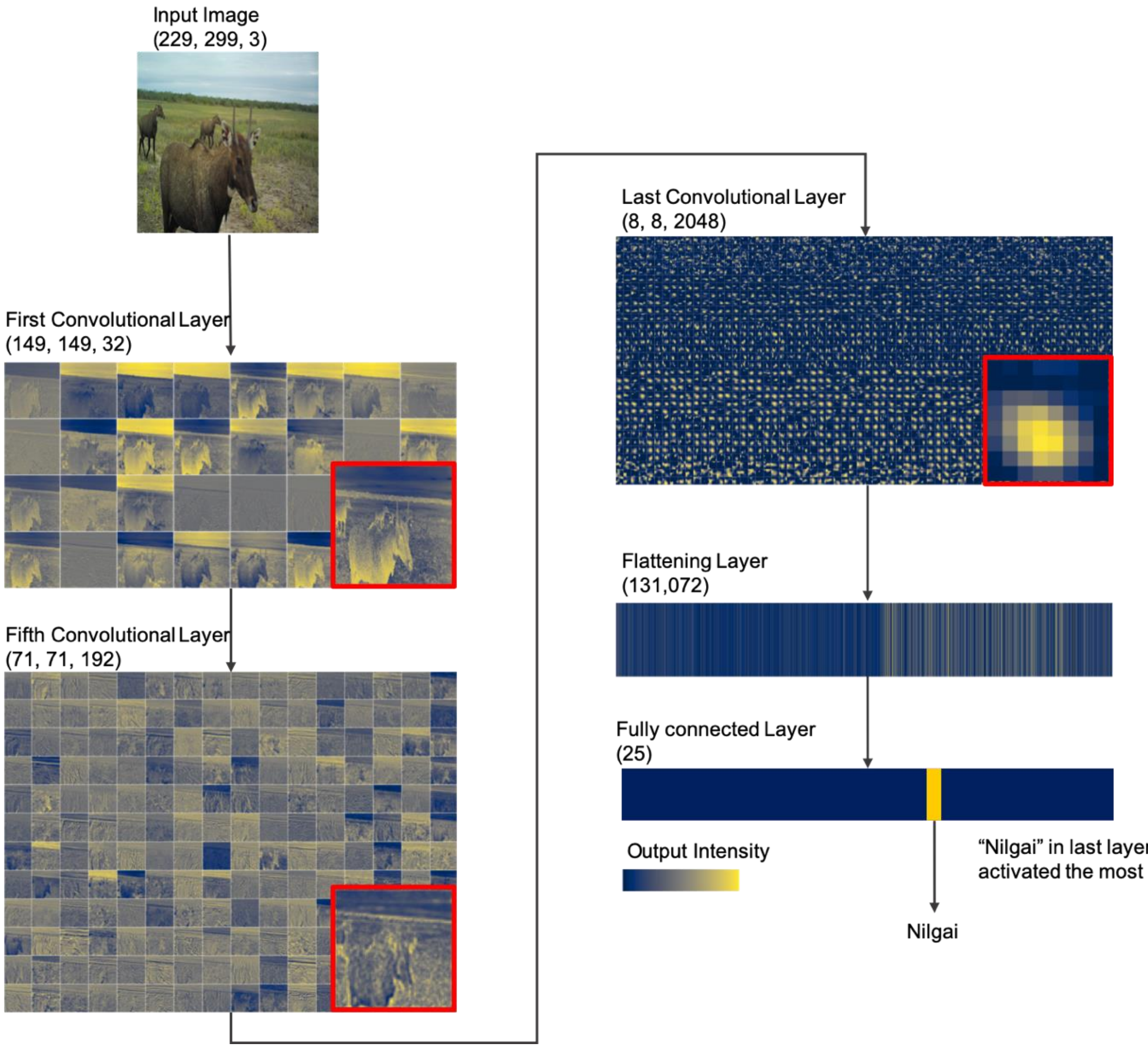
Input Image
(229, 299, 3)
First Convolutional Layer
(149, 149, 32)
Fifth Convolutional Layer
(71, 71, 192)
Last Convolutional Layer
(8, 8, 2048)
Flattening Layer
(131,072)
Fully connected Layer
(25)
Output Intensity
"Nilgai" in last layer
activated the most
Nilgai

Figure 3

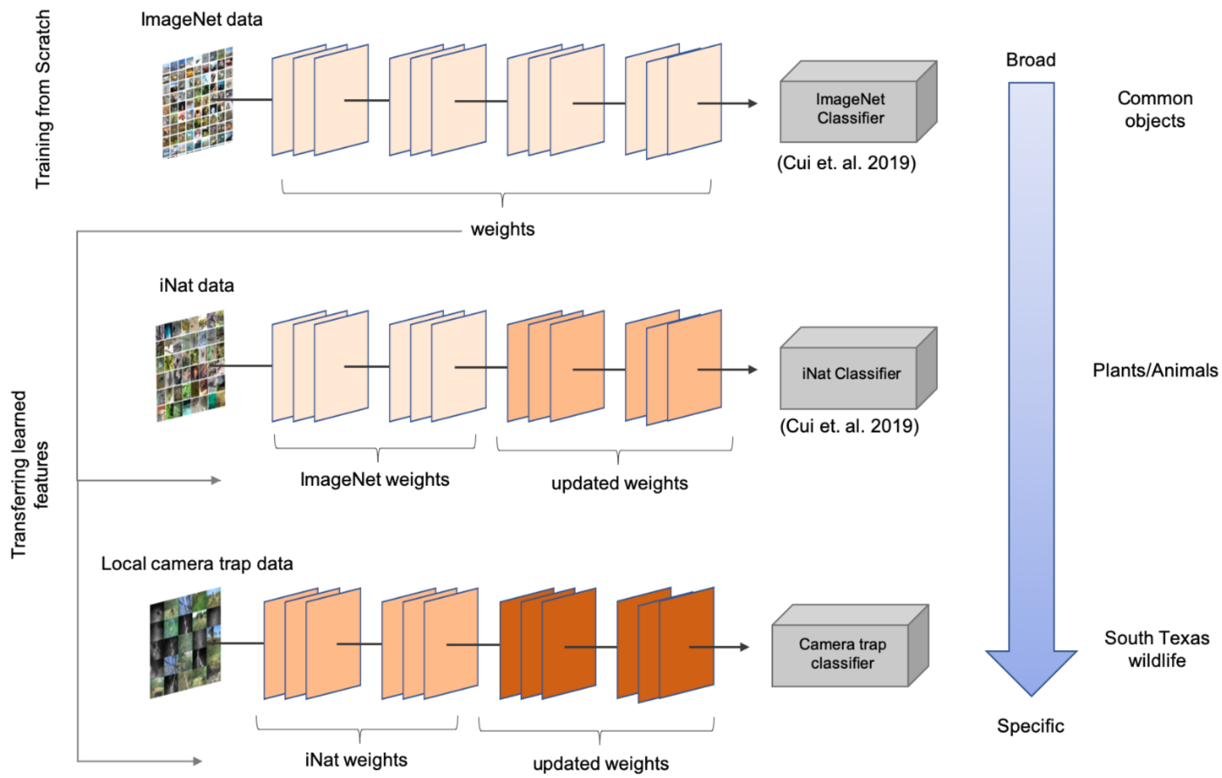
Training from Scratch
ImageNet data
ImageNet Classifier
(Cui et. al. 2019)
weights
Transferring learned features
iNat data
iNat Classifier
(Cui et. al. 2019)
ImageNet weights
updated weights
Local camera trap data
Camera trap classifier
iNat weights
updated weights
Broad
Specific
Common objects
Plants/Animals
South Texas wildlife

Figure 4


cattle
cattle
none
nilgai
bird
pig
rabbit
~~pig~~
nilgai
bird
deer
rabbit
cat
human
none
nilgai
coyote

Click here to access/download
Supplemental Material
requirements.txt

Click here to access/download
Supplemental Material
bi_nil.h5

Click here to access/download
Supplemental Material
multi_nil_100K_89acc.h5

Click here to access/download
Supplemental Material
image_labels.csv

Click here to access/download
Supplemental Material
make_binary_prediction.ipynb

Click here to access/download
Supplemental Material
make_multi_predictions.ipynb

Click here to access/download
Supplemental Material
0214.jpg

Click here to access/download
Supplemental Material
0122.jpg

Click here to access/download
Supplemental Material
010.jpg

Click here to access/download
Supplemental Material
074.jpg

Click here to access/download
Supplemental Material
012.jpg

Click here to access/download
Supplemental Material
066.jpg

Click here to access/download
Supplemental Material
024.jpg

Click here to access/download
Supplemental Material
0148.jpg

Click here to access/download
Supplemental Material
0120.jpg

Click here to access/download
Supplemental Material
0160.jpg

Click here to access/download
Supplemental Material
0161.jpg

Click here to access/download
Supplemental Material
0183.jpg

Click here to access/download
Supplemental Material
052.jpg

Click here to access/download
Supplemental Material
095.jpg

Click here to access/download
Supplemental Material
061.jpg

Click here to access/download
Supplemental Material
05.jpg

Click here to access/download
Supplemental Material
0119.jpg

Click here to access/download
Supplemental Material
0126.jpg

Click here to access/download
Supplemental Material
088.jpg

Click here to access/download
Supplemental Material
0114.jpg

Click here to access/download
Supplemental Material
0142.jpg

Click here to access/download
Supplemental Material
047.jpg

Click here to access/download
Supplemental Material
084.jpg

Click here to access/download
Supplemental Material
079.jpg

Click here to access/download
Supplemental Material
018.jpg

Click here to access/download
Supplemental Material
0221.jpg

Click here to access/download
Supplemental Material
011.jpg

Click here to access/download
Supplemental Material
077.jpg

Click here to access/download
Supplemental Material
0136.jpg

Click here to access/download
Supplemental Material
0166.jpg

Click here to access/download
Supplemental Material
00.jpg

Click here to access/download
Supplemental Material
0203.jpg

Click here to access/download
Supplemental Material
053.jpg

Click here to access/download
Supplemental Material
0162.jpg

Click here to access/download
Supplemental Material
0113.jpg

Click here to access/download
Supplemental Material
0182.jpg

Click here to access/download
Supplemental Material
068.jpg

Click here to access/download
Supplemental Material
048.jpg

Click here to access/download
Supplemental Material
0207.jpg

Click here to access/download
Supplemental Material
0132.jpg

Click here to access/download
Supplemental Material
0128.jpg

Click here to access/download
Supplemental Material
03.jpg

Click here to access/download
Supplemental Material
0130.jpg

Click here to access/download
Supplemental Material
0194.jpg

Click here to access/download
Supplemental Material
0217.jpg

Click here to access/download
Supplemental Material
0218.jpg

Click here to access/download

Supplemental Material

0145.jpg

Click here to access/download
Supplemental Material
0117.jpg

Click here to access/download
Supplemental Material
0147.jpg

Click here to access/download
Supplemental Material
049.jpg

Click here to access/download
Supplemental Material
082.jpg

Click here to access/download
Supplemental Material
0170.jpg

Click here to access/download
Supplemental Material
027.jpg

Click here to access/download
Supplemental Material
091.jpg

Click here to access/download
Supplemental Material
013.jpg

Click here to access/download
Supplemental Material
030.jpg

Click here to access/download
Supplemental Material
096.jpg

Click here to access/download
Supplemental Material
043.jpg

Click here to access/download
Supplemental Material
044.jpg

Click here to access/download
Supplemental Material
0198.jpg

Click here to access/download
Supplemental Material
036.jpg

Click here to access/download
Supplemental Material
0133.jpg

Click here to access/download
Supplemental Material
0135.jpg

Click here to access/download
Supplemental Material
073.jpg

Click here to access/download
Supplemental Material
09.jpg

Click here to access/download
Supplemental Material
0216.jpg

Click here to access/download
Supplemental Material
0163.jpg

Click here to access/download
Supplemental Material
028.jpg

Click here to access/download
Supplemental Material
085.jpg

Click here to access/download
Supplemental Material
0192.jpg

Click here to access/download
Supplemental Material
0191.jpg

Click here to access/download
Supplemental Material
0111.jpg

Click here to access/download
Supplemental Material
0152.jpg

Click here to access/download
Supplemental Material
0201.jpg

Click here to access/download
Supplemental Material
071.jpg

Click here to access/download
Supplemental Material
0107.jpg

Click here to access/download
Supplemental Material
076.jpg

Click here to access/download
Supplemental Material
090.jpg

Click here to access/download
Supplemental Material
0138.jpg

Click here to access/download
Supplemental Material
0220.jpg

Click here to access/download
Supplemental Material
050.jpg

Click here to access/download
Supplemental Material
0202.jpg

Click here to access/download
Supplemental Material
0196.jpg

Click here to access/download
Supplemental Material
0169.jpg

Click here to access/download
Supplemental Material
0101.jpg

Click here to access/download
Supplemental Material
0197.jpg

Click here to access/download
Supplemental Material
087.jpg

Click here to access/download
Supplemental Material
01.jpg

Click here to access/download
Supplemental Material
0209.jpg

Click here to access/download
Supplemental Material
0168.jpg

Click here to access/download
Supplemental Material
0156.jpg

Click here to access/download
Supplemental Material
094.jpg

Click here to access/download
Supplemental Material
083.jpg

Click here to access/download
Supplemental Material
0187.jpg

Click here to access/download
Supplemental Material
014.jpg

Click here to access/download
Supplemental Material
0219.jpg

Click here to access/download
Supplemental Material
0103.jpg

Click here to access/download
Supplemental Material
0199.jpg

Click here to access/download
Supplemental Material
031.jpg

Click here to access/download
Supplemental Material
07.jpg

Click here to access/download
Supplemental Material
035.jpg

Click here to access/download
Supplemental Material
057.jpg

Click here to access/download
Supplemental Material
0151.jpg

Click here to access/download
Supplemental Material
0177.jpg

Click here to access/download
Supplemental Material
070.jpg

Click here to access/download
Supplemental Material
0193.jpg

Click here to access/download
Supplemental Material
081.jpg

Click here to access/download
Supplemental Material
016.jpg

Click here to access/download
Supplemental Material
056.jpg

Click here to access/download
Supplemental Material
0179.jpg

Click here to access/download
Supplemental Material
0149.jpg

Click here to access/download
Supplemental Material
099.jpg

Click here to access/download
Supplemental Material
040.jpg

Click here to access/download
Supplemental Material
0184.jpg

Click here to access/download
Supplemental Material
058.jpg

Click here to access/download
Supplemental Material
0109.jpg

Click here to access/download
Supplemental Material
033.jpg

Click here to access/download
Supplemental Material
022.jpg

Click here to access/download
Supplemental Material
0139.jpg

Click here to access/download
Supplemental Material
075.jpg

Click here to access/download
Supplemental Material
0125.jpg

Click here to access/download

Supplemental Material

0102.jpg

Click here to access/download
Supplemental Material
0131.jpg

Click here to access/download
Supplemental Material
020.jpg

Click here to access/download
Supplemental Material
0215.jpg

Click here to access/download
Supplemental Material
065.jpg

Click here to access/download
Supplemental Material
029.jpg

Click here to access/download
Supplemental Material
0195.jpg

Click here to access/download
Supplemental Material
032.jpg

Click here to access/download
Supplemental Material
0173.jpg

Click here to access/download
Supplemental Material
092.jpg

Click here to access/download
Supplemental Material
045.jpg

Click here to access/download
Supplemental Material
067.jpg

Click here to access/download
Supplemental Material
0127.jpg

Click here to access/download
Supplemental Material
0189.jpg

Click here to access/download
Supplemental Material
019.jpg

Click here to access/download
Supplemental Material
0159.jpg

Click here to access/download
Supplemental Material
0121.jpg

Click here to access/download
Supplemental Material
042.jpg

Click here to access/download
Supplemental Material
093.jpg

Click here to access/download
Supplemental Material
080.jpg

Click here to access/download
Supplemental Material
06.jpg

Click here to access/download
Supplemental Material
0185.jpg

Click here to access/download
Supplemental Material
086.jpg

Click here to access/download
Supplemental Material
038.jpg

Click here to access/download
Supplemental Material
017.jpg

Click here to access/download
Supplemental Material
08.jpg

Click here to access/download
Supplemental Material
051.jpg

Click here to access/download
Supplemental Material
054.jpg

Click here to access/download
Supplemental Material
0106.jpg

Click here to access/download
Supplemental Material
0181.jpg

Click here to access/download
Supplemental Material
0174.jpg

Click here to access/download
Supplemental Material
0146.jpg

Click here to access/download
Supplemental Material
0100.jpg

Click here to access/download
Supplemental Material
02.jpg

Click here to access/download
Supplemental Material
0105.jpg

Click here to access/download
Supplemental Material
0115.jpg

Click here to access/download
Supplemental Material
0150.jpg

Click here to access/download
Supplemental Material
0213.jpg

Click here to access/download
Supplemental Material
060.jpg

Click here to access/download
Supplemental Material
0171.jpg

Click here to access/download

Supplemental Material

046.jpg

Click here to access/download
Supplemental Material
0108.jpg

Click here to access/download
Supplemental Material
0175.jpg

Click here to access/download
Supplemental Material
078.jpg

Click here to access/download
Supplemental Material
0172.jpg

Click here to access/download
Supplemental Material
0123.jpg

Click here to access/download
Supplemental Material
0165.jpg

Click here to access/download
Supplemental Material
0118.jpg

Click here to access/download
Supplemental Material
0212.jpg

Click here to access/download
Supplemental Material
0129.jpg

Click here to access/download
Supplemental Material
0140.jpg

Click here to access/download
Supplemental Material
023.jpg

Click here to access/download
Supplemental Material
0164.jpg

Click here to access/download
Supplemental Material
0134.jpg

Click here to access/download
Supplemental Material
063.jpg

Click here to access/download
Supplemental Material
0176.jpg

Click here to access/download
Supplemental Material
0137.jpg

Click here to access/download
Supplemental Material
072.jpg

Click here to access/download
Supplemental Material
0155.jpg

Click here to access/download
Supplemental Material
025.jpg

Click here to access/download
Supplemental Material
0143.jpg

Click here to access/download
Supplemental Material
0167.jpg

Click here to access/download
Supplemental Material
0153.jpg

Click here to access/download
Supplemental Material
0104.jpg

Click here to access/download
Supplemental Material
0144.jpg

Click here to access/download
Supplemental Material
04.jpg

Click here to access/download
Supplemental Material
0154.jpg

Click here to access/download
Supplemental Material
0158.jpg

Click here to access/download
Supplemental Material
0116.jpg

Click here to access/download
Supplemental Material
0186.jpg

Click here to access/download
Supplemental Material
089.jpg

Click here to access/download
Supplemental Material
055.jpg

Click here to access/download
Supplemental Material
037.jpg

Click here to access/download
Supplemental Material
0112.jpg

Click here to access/download
Supplemental Material
0141.jpg

Click here to access/download
Supplemental Material
0208.jpg

Click here to access/download
Supplemental Material
062.jpg

Click here to access/download
Supplemental Material
015.jpg

Click here to access/download
Supplemental Material
064.jpg

Click here to access/download
Supplemental Material
0188.jpg

Click here to access/download
Supplemental Material
059.jpg

Click here to access/download
Supplemental Material
041.jpg

Click here to access/download
Supplemental Material
026.jpg

Click here to access/download
Supplemental Material
0157.jpg

Click here to access/download
Supplemental Material
098.jpg

Click here to access/download
Supplemental Material
0180.jpg

Click here to access/download
Supplemental Material
0205.jpg

Click here to access/download
Supplemental Material
0206.jpg

Click here to access/download
Supplemental Material
039.jpg

Click here to access/download
Supplemental Material
0211.jpg

Click here to access/download
Supplemental Material
097.jpg

Click here to access/download
Supplemental Material
0124.jpg

Click here to access/download
Supplemental Material
0110.jpg

Click here to access/download
Supplemental Material
0204.jpg

Click here to access/download
Supplemental Material
034.jpg

Click here to access/download
Supplemental Material
0190.jpg

Click here to access/download
Supplemental Material
0178.jpg

Click here to access/download
Supplemental Material
069.jpg

Click here to access/download
Supplemental Material
0200.jpg

Click here to access/download
Supplemental Material
README.txt

Click here to access/download
Response to Reviewers.
JFWM Response to Reviewers.docx