Spring 2021

Assessing Convolutional Neural Network Animal Classification Models for Practical Applications in Wildlife Conservation

Julia Larson
San Jose State University

ASSESSING CONVOLUTIONAL NEURAL NETWORK ANIMAL CLASSIFICATION MODELS FOR PRACTICAL APPLICATIONS IN WILDLIFE CONSERVATION

A Thesis

Presented to

The Faculty of the Department of Environmental Studies

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Julia Larson

May 2021

The Designated Thesis Committee Approves the Thesis Titled

ASSESSING CONVOLUTIONAL NEURAL NETWORK ANIMAL CLASSIFICATION MODELS FOR PRACTICAL APPLICATIONS IN WILDLIFE CONSERVATION

by

Julia Larson

APPROVED FOR THE DEPARTMENT OF ENVIRONMENTAL STUDIES

SAN JOSÉ STATE UNIVERSITY

May 2021

Lynne Trulio, Ph.D.	Department of Environmental Studies
Dustin Mulvaney, Ph.D.	Department of Environmental Studies
Philip Heller, Ph.D.	Department of Computer Science

ABSTRACT

ASSESSING CONVOLUTIONAL NEURAL NETWORK ANIMAL CLASSIFICATION MODELS FOR PRACTICAL APPLICATIONS IN WILDLIFE CONSERVATION

by Julia Larson

Convolution neural network models (CNNs) can successfully identify animal species in camera-trap images in simplified testing environments. CNN performance in more complex, realistic environments is understudied. Here the Wellington Camera Traps dataset was used to simulate a wildlife conservation project to detect invasive species at low population levels using camera-trap images and CNN models. Ten CNNs were developed and analyzed with seven testing datasets, simulating 13 possible project scenarios. Model performance was measured using standard computer science metrics, top-1, and top-5 accuracy, and two novel performance metrics developed for this research to directly reflect wildlife conservation goals, false alarm rate, and missed invasive rate. The highest performing models achieved 91.8% and 99.6% top-1 and top-5 accuracy; however, these models also had the highest missed invasive rates. This effect was related to the ratio of native to invasive species in the model's training images. As this ratio increased so did the model's top-1 and top-5 accuracy but also the missed invasive rate. Thus to achieve optimal performance when selecting or training a CNN for use in a wildlife camera-trap project the metric used to judge the performance of the model must be tailored to the specific goals of the project, and the distribution of species in the model's training images must match the distribution that will be seen in the project's camera-trap images.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my committee members Dr. Lynne Trulio, Dr. Dustin Mulvaney, and Dr. Philip Heller, for their help, guidance, and willingness to jump into an unfamiliar interdisciplinary project. Special thanks to Dr. Lynne Trulio for taking the deep dive; this research would have been infinitely harder without the extra effort you put forth to become familiar with the topics of this thesis. Additional thanks go to Greg Klein for his ceaseless belief in my abilities, to my family for their perpetual support, and Ping Ding for her friendship within and beyond SJSU. Finally, the students, staff, and faculty of San José State University are a brilliant example of the value of a diverse campus and I am honored to be part of such an institution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Evidence indicates that we are in a biodiversity crisis and are experiencing a human-induced global mass extinction event (Barnosky et al., 2011). Protecting rare wildlife species from threats is essential for biodiversity preservation. Primary drivers of extinction and biodiversity loss include the direct killing of species, the introduction of invasive populations of species, the introduction of pathogens, the fragmentation and destruction of habitat, the overuse of resources, and human-induced climate change (Barnosky et al., 2011). Camera-trap studies are a popular component of wildlife conservation programs in which motion-activated cameras are placed at a study site to capture images of passing animals (Burton et al., 2015). The resulting images are then used, among many applications, to study animal behavior, track individual animals, estimate species richness and abundance, and monitor for invasive species. All these applications are done in a cost-effective, unintrusive, and less time-intensive manner than traditional methods of observing animals using track censuses or direct counts (Silveira et al., 2003).

Technological advances in photography, including improvements in image quality, increases in the capacity to take and store images, and decreased overall costs, have contributed to the growing popularity of camera-traps projects (El Gamal, 2002). However, researchers and wildlife managers have quickly begun to amass more photographs than can be reasonably processed with available budgets and labor capacity (Tabek et al., 2018). The current process of reviewing and labeling thousands of images

is slow and tedious resulting in error-prone data entry and slow reaction times to conservation issues (Harris et al., 2010; Young et al., 2018).

Convolutional neural networks (CNNs), a type of machine learning computer vision model, promise to be an effective tool for automatically classifying (identifying) animal species in large numbers of images (Tabek et al., 2018). Given a batch of images the models could quickly produce summaries of species seen, alert managers to the presence of a target species or individual, and filter out empty images. This would not only remove the most error-prone stage of the process but also shrink the reaction time of project managers from months to minutes (Norouzzadeh et al., 2018). If animal classification models for camera-trap images become widespread, they would not only increase the effectiveness of wildlife conservation projects currently using camera-traps, but would allow for large-scale camera-trap projects, previously impossible to manage, and potentially free up time and resources for other biodiversity conservation projects.

In the future, efficiency could increase further if animal classification models are integrated into "smart" camera-trap systems. In a smart camera-trap system, immediately after images are taken by a camera in the field, the species in the images are classified by a software model and the project manager notified, in real-time, to the relevant information coming from the camera-trap (Glover‐Kapfer et al., 2019; Håvard, 2017). Such a system would fulfill the 2012 beliefs held by camera-trap researchers that in ten to twenty years camera-trapping would head in the direction of automation (identifying empty images, animal species, and individual animals), wireless capabilities via satellite or cellular networks, and real-time capabilities (when automation and wireless

capabilities are combined) (Glover‐Kapfer et al., 2019). As of 2021, advances have been made in all three areas. Automation has shown success with machine learning models for animal classification, while wireless capabilities are showing promise with several camera-traps commercially available that transmit images over cellular networks. Real-time capabilities that would be the mark of a true smart-camera-trap project are still in the works, with several companies working towards producing camera-traps with onboard computing for the automation of image processing. Researchers are beginning to design fully automated systems from camera-trap to automatic image processing in anticipation of fully viable wireless and real-time capabilities (Håvard, 2017).

Although animal classification models for camera-trap images are promising thus far their evaluation has been limited to performance metrics developed by the computer science field, not the conservation biology field. A shortcoming that means that how well these models work in applied wildlife conservation projects is unknown. To fully evaluate the effectiveness of these models new performance metrics are necessary that directly reflect wildlife conservation goals.

The research undertaken for this thesis advances the application of animal classification CNNs in the wildlife conservation field in two ways. First, this work builds and tests CNNs that reflect realistic field situations for a camera-trap project monitoring for invasive species. Second, custom performance metrics were developed to directly reflect the goals of efficiently detecting invasive species in camera-trap images, and compared to traditional computer science measurements of performance. Overall this

work helps wildlife managers determine the value of deploying a machine learning computer vision model for their wildlife conservation project.

Related Research

Machine Learning, Neural Networks, and Convolutional Neural Networks

Machine learning is the field of study in which computer models are designed to improve automatically through the process of training on example data (Géron, 2017; Mitchell, 1997). It emerged as a discipline from the more general subject of artificial intelligence after the second world war (Russell & Norvig, 2016). Neural networks are a popular category of machine learning models commonly used for classification tasks, in which the model is asked to identify the output class of new input (Murphy & Ebrary, 2012).

Neural networks are built from artificial neurons. The artificial neuron, inspired by the biological neuron, is a mathematical function proposed by Warren McCulloch and Walter Pitts in 1943 in which the weighted sum of the inputs plus a bias is run through an activation function to determine the output (Figure 1) (Russell & Norvig, 2016). The activation function controlling the output of a neuron varies depending on the model; popular activation functions include: logistic sigmoid, hyperbolic tangent (tanh), and rectified linear (ReLU) (Géron, 2017). Connected, artificial neurons become neural networks (Figure 2). The first neural network was a physical machine built by Marvin Minsky and Dean Edmonds in 1950 using vacuum tubes and parts from a surplus B-24 bomber (Russell & Norvig, 2016). Training a neural network is the process of adjusting the weights and biases of the neurons using training data to improve the accuracy of the model. Training is done in multiple rounds called epochs. After each epoch, the network's output error is calculated then a backpropagation algorithm determines how

much each neuron contributed to the error (Géron, 2017). The terms within a backpropagation algorithm can change during each training epoch to fine-tune the model, for example, the backpropagation algorithm Stochastic Gradient Descent has learning rate and weight decay terms (Géron, 2017).

Figure 1

Artificial Neuron

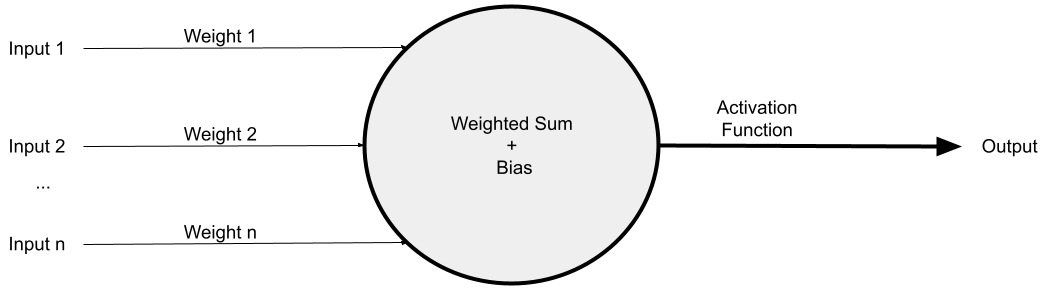
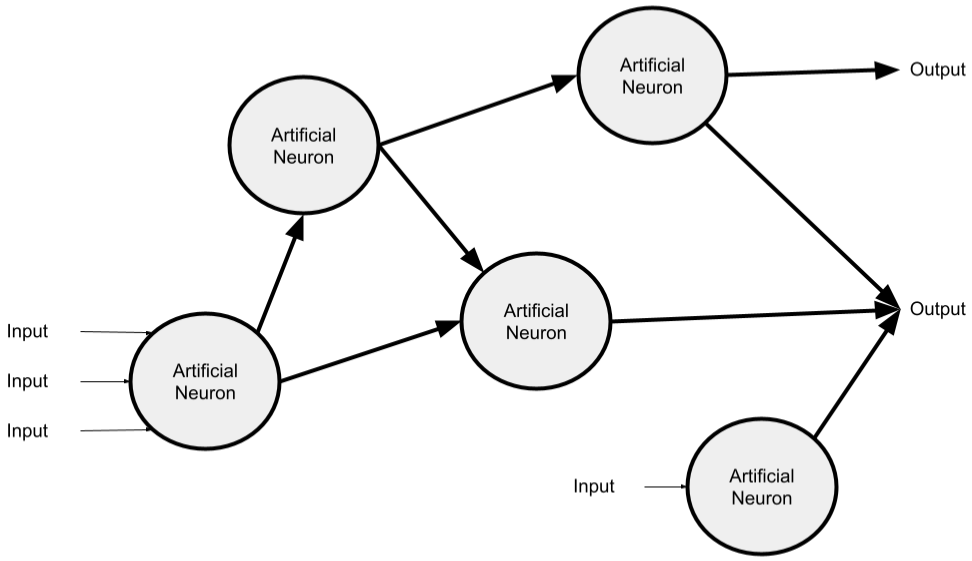


Figure 2

Neural Network



Neural networks are often categorized by their architecture which includes the number, arrangement, and connections between, artificial neurons in the network. Deep learning neural network architectures are characterized by two or more layers of neurons, called hidden layers, between the input and the output of the neural network. If every neuron in the layer is connected to every neuron in the next layer, the layer is fully connected (Figure 3). Fully connected layers increase the accuracy of a neural network but are computationally expensive and not always required (Géron, 2017). CNNs architectures are a subcategory of deep neural networks identified by the use of convolutional and pooling layers to learn complex patterns from simpler patterns in the input data using a limited number of neural connections (Figure 4) (Sewak et al., 2018). In convolutional layers, neurons are connected to a limited number of nearby neurons that summarize the input into feature maps (Géron, 2017). Pooling layers reduce the size of the feature maps to lower the computational requirements of the model (Géron, 2017). This architecture is inspired by the visual cortex of the mammalian brain and is commonly used in computer vision applications to enable a computer to interpret the contents of an image without human assistance (Lecun et al., 2015). Images are most frequently used as input for CNN models; however, it is possible to use any type of data that has an inherent underlying structure, including time series, text, weather maps, or audio files transformed into an image type structure (Sewak et al., 2018).

Figure 3

Deep Learning Neural Network

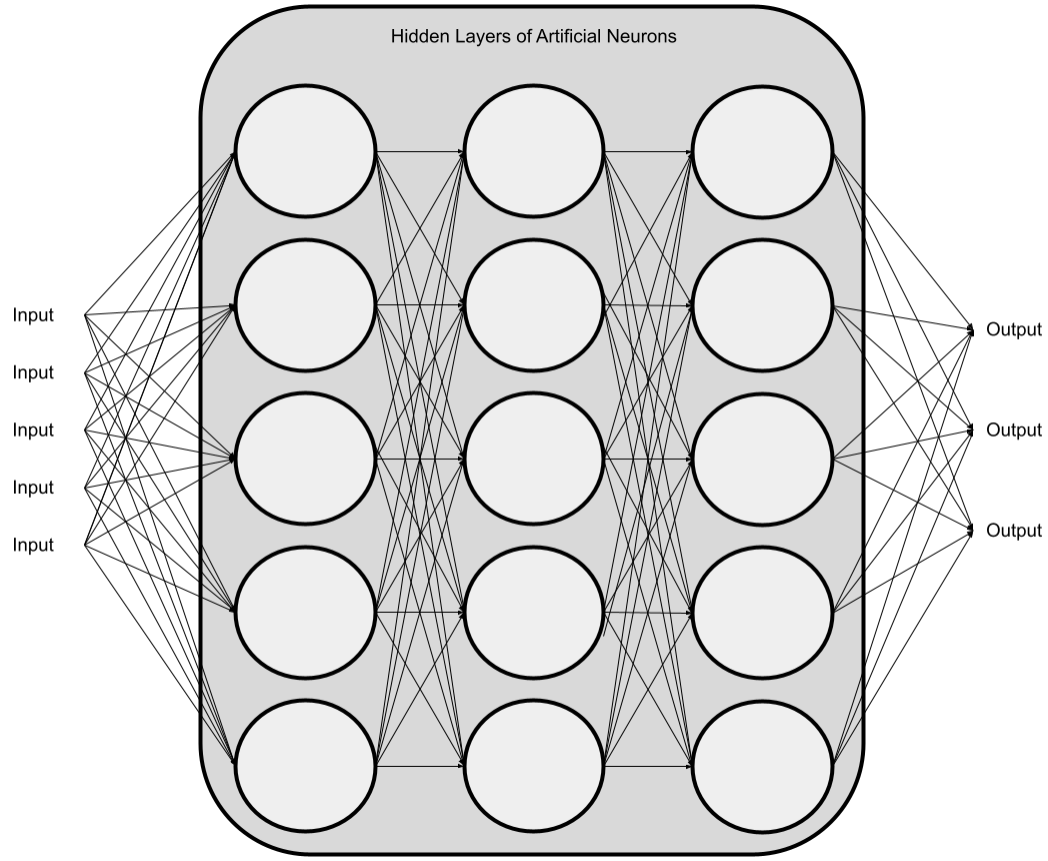
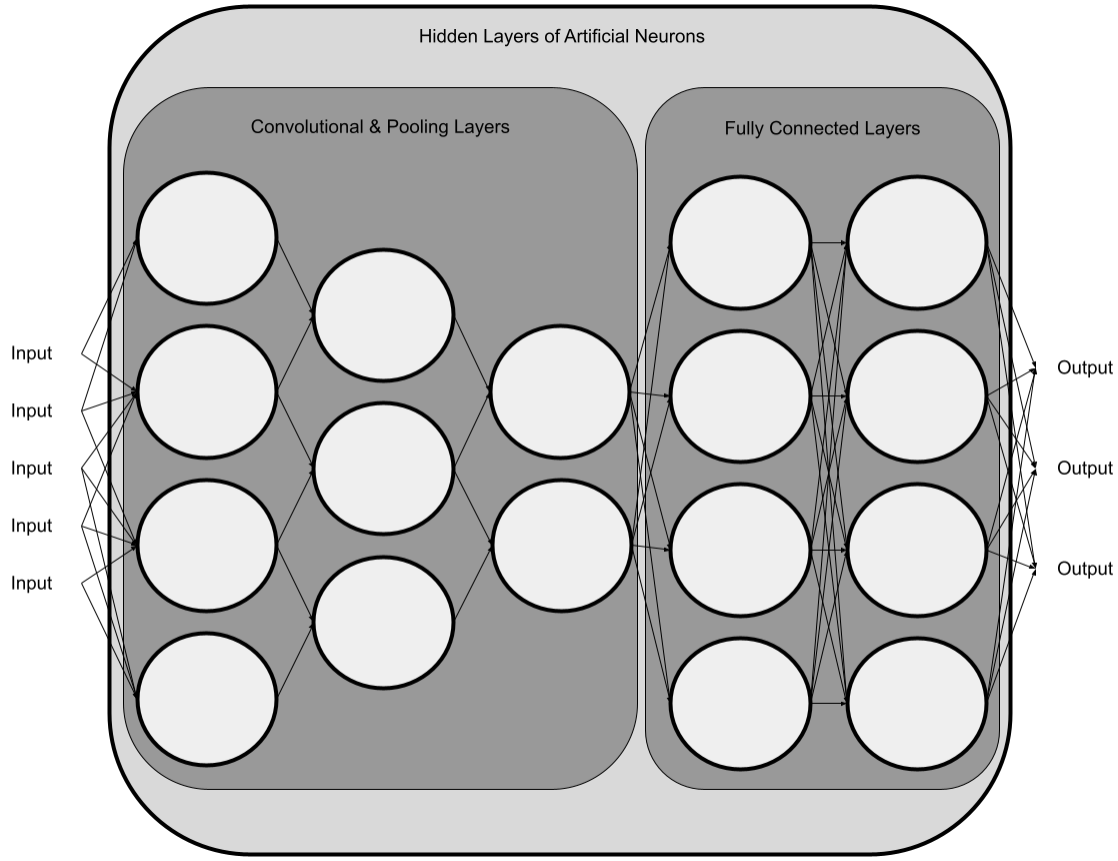


Figure 4

Convolutional Neural Network



Although the first CNNs were developed and successfully used for computer vision tasks in the 1980s, the computational barriers of the times prevented their widespread adoption (Fukushima, 1980). CNNs did not become practical until the 2000s when, alongside steady increases in computational power, Oh and Jung (2004) discovered that the implementation speed of CNNs could be increased twenty times by using Graphics Processing Units (GPUs). Later, pre-trained models for use in transfer learning became freely available on the internet greatly reducing the resources needed to build an accurate

CNN for a custom task. Transfer learning is when the weights and biases of a model pre-trained on a task are reused as part of a new model for a similar task; this process increases training speed and reduces the amount of training input needed to achieve an accurate model (Géron, 2017). Together these events precipitated the current abundance of research using CNNs; from 2000 to 2020 mentions of convolutional neural networks in abstracts increased from 5 to 26,862 (Digital Science, 2021).

Animal Classification Models for Camera-trap Images

Machine learning models are increasingly used for animal classification in conservation biology to identify animals in images (Nguyen et al., 2017; Norouzzadeh et al., 2018; Tabak et al., 2018). Animal classifiers receive images as input and output a list of likelihoods indicating how confident the model is that each possible animal class is present in each image. Top-1 and top-5 accuracy are traditionally used to evaluate animal classification models (Table 1) (Thoma, 2017). These metrics are useful by providing standard metrics for comparing animal classification models to each other and if the goal of the project is to estimate species richness or abundance. However, to better measure model effectiveness for monitoring invasive species, this work developed and measured two custom performance metrics, false alarm rate and missed invasive rate. False alarm rate reflects wasted time and resources. Missed invasive rate, accounts for the high ecological cost of undetected invasive animals (Table 1) (Thoma, 2017).

Table 1

Model Performance Metrics

Performance metric	Equation	Field	Status
top-1 accuracy	$\frac{\#\ of\ times\ the\ top\ prediction\ is\ correct}{\#\ of\ images\ tested}$	computer science	standard
top-5 accuracy	$\frac{\#\ of\ times\ the\ correcsmit\ class\ is\ in\ the\ top\ 5\ predictions}{\#\ of\ images\ tested}$	computer science	standard
false alarm rate	$\frac{\#\ of\ empty\ or\ native\ images\ labeled\ as\ invasive}{\#\ of\ images\ tested}$	conservation biology	custom
missed invasive rate	$\frac{\#\ of\ invasive\ images\ labeled\ as\ native\ or\ empty}{\#\ of\ invasive\ images\ tested}$	conservation biology	custom

Animal classification in machine learning computer vision models is categorized into fine and coarse-grained tasks, each of which requires distinct methodologies. Fine-grained animal classification tasks attempt to distinguish between similar species (mouse versus rat) or individual animals within a species and are considered to be more difficult, while coarse-grained animal classification tasks attempt to classify animals into general groups (rodents, cats, birds, etc.). The work presented here focuses on coarse-grained animal classification tasks. Images are often but not always cropped to the area of interest before the classification step of the model. Cropping removes irrelevant background information to help with animal identification across different contexts but increases the overall complexity of making and using a model (Norouzzadeh et al., 2021).

In 2013, Yu and colleagues made one of the earliest attempts to classify animals in camera-trap images using external feature extraction methods, manually cropped images,

and a support vector machine (SVM) and achieved a classification top-1 accuracy of 82% on 18 animal classes. Soon after, Chen et al. (2014) attempted to fully automate the process by using a CNN and a novel external animal cropping algorithm developed by Ren et al. (2013) specifically for the highly dynamic scenes typically found in camera-trap images. Although Chen et al. (2014) only achieved a top-1 accuracy of 38% on twenty animal classes, their methodology contained the key traits that would be vital for upcoming success in the field.

Although Sharath Kumar et al. (2015) made the animal classification task simpler by using hand taken images instead of camera-trap images, their work was still important, as it demonstrated that animal classification tasks were within the realm of possibility. Their model reached a top-1 accuracy of 82% on 25 animal classes, using a K-Nearest Neighbor classifier model, an external feature extraction algorithm, and a fast but still human-reliant cropping algorithm. Villa et al. (2017) relied on manual cropping and achieved a top-1 accuracy of 88.9% on 26 animal classes. Despite the time-consuming disadvantage of manual cropping, their research demonstrated that newly developed CNN models could classify animals in camera-trap images with a high degree of accuracy.

In 2017, researchers developed fully automated external cropping algorithms for animals in camera-trap images using segmentation algorithms (Zhu et al., 2018; Giraldo-Zuluaga et al., 2017a). Giraldo-Zuluaga et al. (2017b) used an animal detection segmentation algorithm they had developed (Giraldo-Zuluaga et al., 2017a) to create a species classification model with a top-1 accuracy of 92.65% on ten mammal classes.

Yousif et al. (2017) built a block-based algorithm to automatically draw bounding boxes around animals with an accuracy rate of 83.78% that assisted a CNN to reach a top-1 accuracy of 95.6% --but only on three coarse classes (animal, background, and human).

Most recently researchers have explored the ability of CNNs to detect the location of the animal in the image, in addition to classifying the animal, with either a single CNN or two separate CNNs. The single CNN route uses a CNN architecture designed with integrated object (animal) detection to classify the camera-trap image. Schneider et al. (2018a) used this approach to develop a faster region-CNN that achieved a top-1 accuracy of 93.0% and 76.7% on two models trained with different, relatively small datasets (946 and 4,432 images respectively). The CNN animal classifier developed by Tabak et al. (2018) achieved a top-1 accuracy of 94% on 27 animal classes when tested on an in-sample testing dataset (a dataset containing images from the same camera-traps as the images in the model's training dataset); on an out-of-sample testing dataset (a dataset containing images from different camera-traps as the images in the model's training dataset) the model achieved a top-1 accuracy of 82%. This model also used the single CNN approach to animal detection and itself can be used as a binary animal detection classifier (animal or no animal) on images from a completely different ecosystem from which it was trained, with an accuracy of 94% (Tabak et al., 2018).

Two papers have explored the use of two separate CNNs for detecting and classifying animals in camera-trap images. The first CNN is a binary animal detection classifier that filters out empty images and in some models crops the image to the animal. The second CNN takes the non-empty images and classifies them by species or group. A CNN

binary animal detection classifier developed by Nguyen et al. (2017), had an accuracy of 96.6% that assisted a species classifier to reach a top-1 accuracy of 90.4% on three animal classes. Norouzzadeh et al. (2018) did a similar project, creating a binary CNN animal detection classifier with an accuracy of 96.8%, that assisted their species classifier to reach a top-1 accuracy of 93.8% on 48 animal classes. The two-step approach is the method used in this paper.

Challenges of Camera-trap Images for Convolutional Neural Networks

CNN animal classifiers for camera-trap images must address generalization issues like all machine learning models. Beery et al. (2018) demonstrated that state-of-the-art neural network classifiers experience accuracy drop-offs when classifying images from camera sites on which they were not trained, with errors increasing up to 140%. Tabak et al. (2018) confirmed this trend, seeing a top-1 accuracy in their model at 94% for in-sample images versus 82% for out-of-sample images.

Van Horn and Perona (2017) demonstrated that, when it comes to top-1 accuracy, it is better to train image classifiers on image datasets that reflect the long-tailed distributions found in the field instead of artificially balancing the datasets, except when the lack of balance is extreme. This idea was supported by Nguyen et al. (2017), who experienced a drop in accuracy when balancing their dataset, as well as by Villa et al. (2017), whose model was trained on an extremely unbalanced dataset and had top-1 and top-5 accuracies of only 35.4% and 60.4%.

Invasive Predators as a Global Problem

The unchecked activities of humans have resulted in the earth's sixth global mass extinction event (Barnosky et al., 2011). The conservative calculated rate of global extinctions is desperately fast and will only increase as the currently high rate of population losses transform into extinctions (Ceballos et al., 2015; Ceballos et al., 2017). In comparison to other species, terrestrial vertebrates on islands are at higher risk of extinction because of multiple factors, invasive species are one major factor (Spatz et al., 2017).

Invasive species occur when humans introduce into a landscape novel species that harm endemic species (Beck et al., 2008). They can also occur when the population of an endemic species is increased through anthropogenic means to a level that harms other endemic species (Colautti & MacIsaac, 2004). Western colonization around the world is responsible for most invasive species seen today; colonists enabled the spread of invasive species by altering landscapes and/or by introducing invasive populations of species, either unintentionally as stowaways or intentionally as food sources, pets, or weapons for disrupting local systems (McNeely, 2001).

Invasive populations damage ecosystems in many ways including direct consumption, the spread of disease, competition for resources, and the altering of ecosystem processes, species abundance, or habitat (Mack et al., 2000). Invasive populations of predators introduced into previously predator-free islands can be especially destructive to endemic species, which often live in only a few locations and lack predator defenses (Russell et al., 2017). The introduction of rats has been particularly damaging with approximately

40-60% of all globally recorded bird and reptile extinctions partially attributed to invasive populations of rats (U.S. Fish and Wildlife Service, 2007). Other destructive invasive species commonly targeted for removal include hedgehogs, rabbits, possums, mice, stoats, ferrets, weasels, foxes, deer, domestic cats, goats, and horses (DIISE, 2018).

While the spread of invasive species globally is an ongoing issue, many of the original vectors through which invasive populations of animals established on small, uninhabited islands (shipwrecks and the establishment of food sources for months-long ship journeys) have mostly disappeared. Several organizations have started to remove invasive predators from islands to restore natural systems and save threatened species. So far, approximately 1,233 mammalian eradications have been conducted on 806 islands with a success rate of 88% (DIISE, 2018). Governments are increasingly aware of the threat invasive species pose to functional ecosystems and are taking action. Aotearoa - New Zealand, a prime example, has an ambitious Predator Free 2050 goal to remove the most destructive invasive predators from the entire country by 2050 (New Zealand Government, 2017). However, with costs into the millions and the labor-intensive requirements of current eradication methods, the number of eradication campaigns completed is limited, despite the great need (Brooke et al., 2007). If more or larger islands are to be freed of invasive animals, new or enhanced methodologies are required.

One particular challenge in eradication campaigns is confirming that all invasive animals have been successfully removed. Eradication campaigns commonly consist of two steps. First, there is a removal step, during which invasive animals are removed frequently through trapping or poisoning, followed by a monitoring step during which

camera-traps are set out to monitor the island for individual animals that were missed. Typically, someone must physically return to each camera between every one to three months to retrieve the images. Processing the images then takes several more months. During this time, missed animals can reproduce and more effort and money must be spent to initiate another removal campaign.

To stimulate novel improvements to the monitoring step, Conservation X Labs, an organization that runs challenges to encourage the development of new conservation technologies, and Island Conservation, a non-profit organization that removes invasive animals from islands, have published an open challenge called "Confirming Zero" to help find solutions. Implementing a CNN to automatically identify invasive animals in camera-trap images quickly and immediately after retrieval would be one such solution.

Objectives

Despite the accuracy of CNN models in automatically identifying animals in camera-trap images, wildlife managers are not widely adopting CNN models for their camera-trap projects. Hindering their adoption is a lack of information on how the variations in building, training, using, and measuring model performance affect their utility in realistic wildlife conservation projects. More generally, wildlife managers may not know that these models can be effective and, if they do, there is no clear guidance on how to select the right model and train it for their specific project.

To address these immediate challenges, I trained several CNN animal classification models to test the effectiveness of the models in detecting invasive species in a simulated invasive species monitoring camera-trap project (Table 2). The models addressed the following questions using two standard computer science performance metrics, top-1 and top-5 accuracy, and two custom conservation biology metrics, false alarm rate, and missed invasive rate:

Q1 - How do model performance metrics change if a model's output classes are grouped into invasive, native, or empty classes or left as individual species classes?

Q2 - How do model performance metrics change as the maximum number of training images of each output class from each camera site scales logarithmically both in models with grouped and individual output classes?

Q3 - How do model performance metrics change when novel invasive animal species classes not present in the training dataset are included in the testing dataset?

Q4 - How do model performance metrics compare between custom and off-the-shelf biome models and between large and small project-specific models?

Table 2

Research Questions and Models

Research question	Model		Training dataset[a]	Testing dataset[b]
Q1	grouped		A	A
	individual		B	B
Q2	10 image grouped		C	A
	100 image grouped		D	A
	1,000 image grouped		E	A
	10 image individual		F	B
	100 image individual		G	B
	1,000 image individual		H	B
Q3	large project	without novel species	I	D
		with novel species	I	E
	small project	without novel species	J	F
		with novel species	J	G
Q4	biome	custom	B	B
		off-the-shelf	B	C
	large project-specific		I	D
	small project-specific		J	F

Note. Some training/testing dataset combinations were used multiple times to answer different research questions.

[a] Table 3. [b] Table 4.

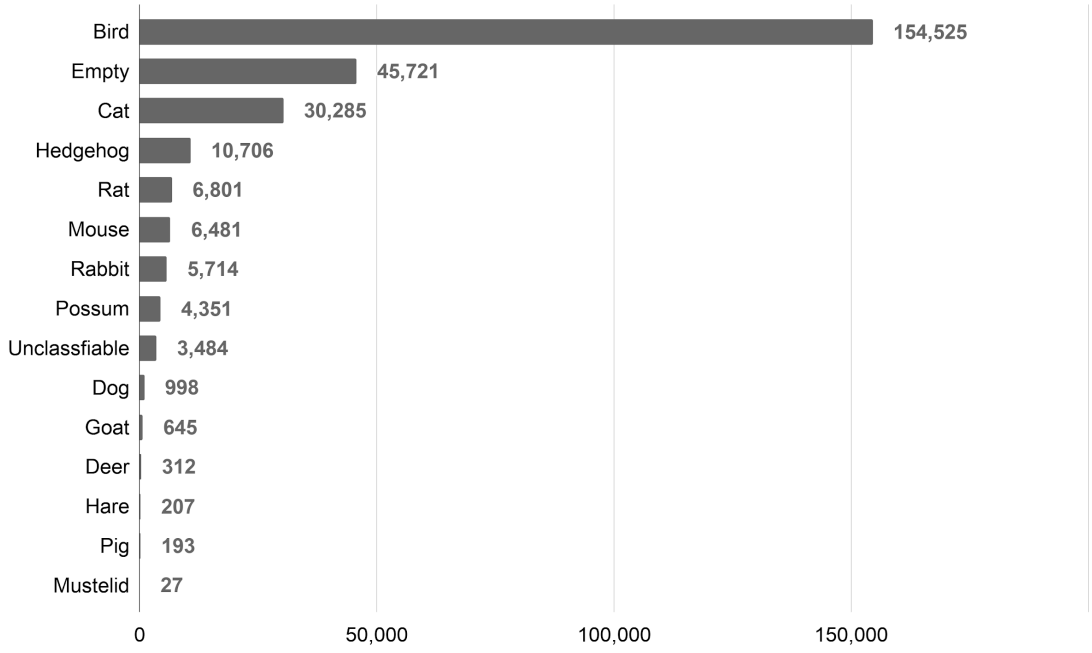
Methods

Camera-trap Image Dataset

Training and testing images were selected from the Wellington Camera Traps dataset, the publicly available labeled camera-trap image dataset with the most relevant species for the invasive species monitoring scenario at the time this project started (Anton et al., 2018). This dataset is available via the Labeled Information Library of Alexandria: Biology and Conservation (http://lila.science/), and consists of 270,450 images taken of 17 animal classes at 182 camera sites in Aotearoa - New Zealand with 17% of images labeled empty. Images were taken in sequences of three and labeled by citizen scientists and/or professional ecologists from Victoria University of Wellington. Images in the same sequence were given the same label even if the animal did not appear in all images. Importantly, the dataset accurately reflects the unbalanced distribution of animal species caught in camera-trap images in general (Figure 5), the different frequencies at which each animal type appears at each camera-trap site, the high frequency of empty images, and the different number of images captured at each camera-trap site (Van Horn & Perona, 2017).

Figure 5

Species Distribution of the Wellington Camera Traps Dataset



To increase the size of the training datasets and to increase CNN accuracy, standard data augmentation techniques were applied to all training images (Shorten et al., 2019). Augmentation included random horizontal flipping, cropping, and color jitter. Images labeled as rat, Norway rat, and ship rat were combined in a single "rat" label, hare and rabbit were combined into a single "hare" label. Each image was given two labels for different models—an individual label specifying the species or group of species and a grouped label where birds were labeled as "native," empty frames labeled as "empty," and all others (unclassifiable, cat, deer, dog, hare, hedgehog, mouse, mustelid, pig, possum, and rat) labeled as "invasive."

Training and Testing Datasets

All images from the Wellington Camera Traps dataset were first put through Microsoft's MegaDetector, an animal detection classifier for camera-trap images, to crop the animal from the image (Beery et al., 2019). MegaDetector uses a Faster-RCNN architecture with an InceptionResNetv2 base network (He et al., 2016; Szegedy et al., 2016). Images labeled as empty were cropped using the boundary boxes (crop points) from the last detected animal. The Megadetector detected an animal in 91.2% of the Wellington Camera Traps images containing animals.

Next, the cropped images were split into training datasets (Table 3) and testing datasets (Table 4) in such a way as to both mimic a real invasive species monitoring project and meet standard CNN dataset practices using the reasoning and process detailed in Figure 6. For each camera site in the dataset, images were sorted by date, from oldest to most recent, to reflect the reality that camera-trap images classified using the model will be taken later than the images used in training. In training datasets A (grouped) and B (individual) the included images were the first 80% of the images by date independent of output class. Training datasets C through E (grouped) and F through H (individual) included the *first* ten, one hundred, and one thousand images of each output class. All images except those labeled empty, bird, or hedgehog were removed from training dataset F to make training dataset I. Hedgehogs were selected because of the high number of available images in the database and their importance as an invasive species (*Predator Free 2050*, 2017). Training dataset J was a reduced dataset of training dataset I using only images from the ten camera sites with the most hedgehog images.

Table 3

Model Training Datasets

ID	Maximum training images per class per camera	Camera site count	Image count	Invasive to non-invasive ratio	Output Classes		
					Type	Count	Labels
A[a]	all available	171	196,957	0.35	grouped	3	native[b], invasive[c], empty
B[a]	all available	171	196,957	0.35	individual	13	empty, bird, unclassifiable, cat, deer, dog, hare, hedgehog, mouse, mustelid, pig, possum, rat
C	10	171	6,435	0.74	grouped	3	native[b], invasive[c], empty
D	100	171	51,717	0.47	"	"	"
E	1,000	171	214,403	0.37	"	"	"
F	10	171	10,205	1.76	individual	13	empty, bird, unclassifiable, cat, deer, dog, hare, hedgehog, mouse, mustelid, pig, possum, rat
G	100	171	64,617	0.84	"	"	"
H	1,000	171	216,417	0.39	"	"	"
I	10	171	4,089	0.11	individual	3	empty, bird, hedgehog
J	10	9	316	0.44	"	"	"

[a] Training datasets A and B consist of the same images but with different output class labels. [b] Animals included in the native output class are birds. [c] Animals included in the invasive output class are unclassifiable, cat, deer, dog, hare, hedgehog, mouse, mustelid, pig, possum, and rat.

Table 4

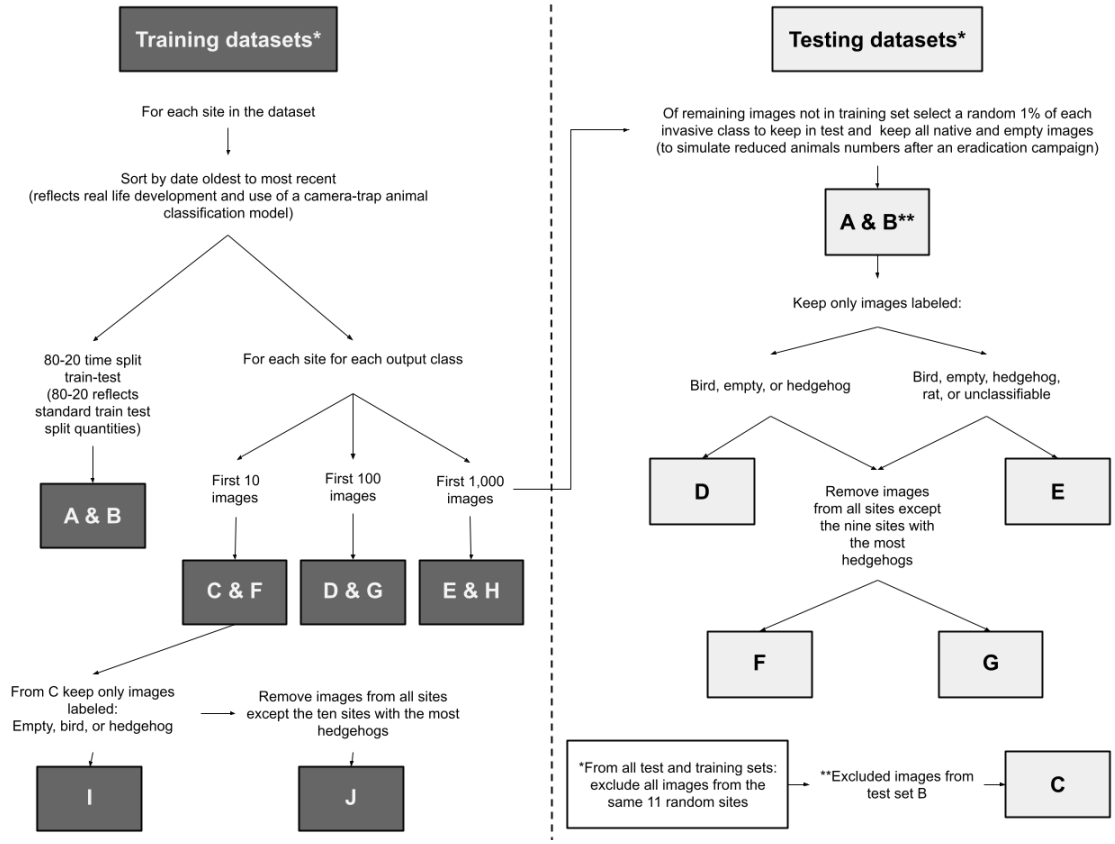
Testing Datasets

ID	Camera site count	Image count	Invasive to non-invasive ratio	Sample type	Classes		
					Type	Count	Labels
A[a]	171	27,170	0.03	in-sample	grouped	3	native[b], invasive[c], empty
B[a]	171	27,170	0.03	in-sample	individual	13	empty, bird, unclassifiable, cat, deer, dog, hare, hedgehog, mouse, mustelid, pig, possum, rat
C	10	428	0.03	out-of-sample	individual	13	"
D	171	26,436	0.01	in-sample	individual	3	empty, bird, hedgehog
E	171	26,681	0.01	in-sample	individual	5	empty, bird, hedgehog, rat, unclassifiable
F	9	3,585	0.00	in-sample	individual	3	empty, bird, hedgehog
G	9	3,596	0.01	in-sample	individual	5	empty, bird, hedgehog, rat, unclassifiable

[a] Testing datasets A and B consist of the same images but with different output class labels. [b] Animals included in the native output class are birds. [c] Animals included in the invasive output class are unclassifiable, cat, deer, dog, hare, hedgehog, mouse, mustelid, pig, possum, and rat.

Figure 6

Process for Splitting Camera Trap Images into Training and Testing Datasets



Note. Process executed using a Python script.

All testing datasets were made from images not included in any of the training datasets. For testing datasets A (grouped) and B (individual) images labeled empty or bird, and a random 1% of each invasive class (which simulates the reduced animal numbers after an eradication campaign) were included. Testing datasets D and E contained only images labeled empty, bird, or hedgehog, and empty, bird, hedgehog, rat, or unclassifiable, respectively. Testing datasets F and G were the same as D and E respectively but only contained images from the ten camera sites with the most hedgehog

images. The final step for all training and testing datasets was to exclude all images from the same 11 (1 for training dataset J and testing datasets F and G) random camera sites. The excluded images from testing dataset B were used to create testing dataset C, an out-of-sample test for the off-the-shelf biome model.

Model Development and Training

All CNN models were trained using Pytorch, a python package for building, training, and testing neural networks with GPU acceleration (Paszke et al., 2019), on a custom computer with an Intel Core i7-8700K, Nvidia GeForce GTX 1080 Ti GPU, and 32 GB of RAM. The models use the ResNet-18 architecture (He et al., 2016) pretrained on the imagenet database (Deng et al. 2009) with a rectified linear activation function (ReLU), Stochastic Gradient Descent with momentum for backpropagation (Goodfellow et al., 2016), and learning rates and weight decays matching those of Norouzzadeh et al. (2018) and Tabak et al. (2018).

Model Analysis

Models were evaluated on the following performance metrics: top-1 and top-5 accuracy, false alarm rate, and missed invasive rate (Table 1). Top-5 accuracy is only reported for models with more than five output classes (models with five or fewer output classes will always have top-5 accuracies of 100%). To answer the research questions some of the models were analyzed multiple times using different testing datasets (Table 2).

Description of Models

Ten CNN models using ten training datasets (Table 3) and seven testing datasets (Table 4) in 13 combinations were developed and tested to assess each of the objectives of this research. Some of the model/testing dataset combinations were used in multiple research questions but were assigned different model names for each research question for clarity (Table 2).

To assess research question 1, two models were trained—grouped and individual—to determine what type of output classes worked best for the invasive species monitoring scenario (Table 2). The training (A and B) and testing (A and B) datasets for both models contained the same images but were sorted into different output classes (Table 3, Table 4). For the grouped model the species were grouped into empty, native, and invasive classes; for the individual model, the species were left as individual classes.

Six models were trained to address research question 2, to help managers decide between the desired accuracy of the model and the resources needed to collect and label images for training (Table 2). Three models (10/100/1,000 image grouped) used grouped output classes with training datasets C through E, another three (10/100/1,000 image individual) used individual output classes with training datasets F through H (Table 3). All six models included in their training datasets a maximum of ten, one hundred, or one thousand images of each output class from each camera site appropriately matched to their model name (Table 3). Due to the naturally uneven distribution of animal species in the camera-trap images and the rarity of many species, most class/site combinations did not reach one thousand or even one hundred training images. The testing datasets (A and

B) for all six models contained the same images (also the same images as the question 1 testing datasets) but were sorted into different output classes as appropriate (Table 4).

To assess research question 3, two models (large project and small project) were trained using datasets I and J and tested on two testing datasets each (D through G) to see how the performance metrics reacted when novel invasive animal species classes not present in the training dataset were included in the testing dataset (Table 2). These models could not classify the novel species correctly because they had no matching output class, instead, they selected an output class from those present in their training dataset. The training datasets consisted only of bird, hedgehog, and empty images, from 171 camera sites for the large project and nine camera sites for the small project (Table 3). The four testing datasets consisted of images taken from matching numbers of camera sites; one testing dataset of each pair contained images with novel invasive animal species and species in the training dataset (birds, hedgehogs, empty, rats, and unclassifiable), the other only species in the training dataset (birds, hedgehogs, and empty) (Table 4).

Finally, four models were trained and tested mapping to four realistic field situations for invasive species monitoring camera-trap projects (Table 2). Two scenarios represented biome models with 13 individual species output classes covering any likely species in the region, both scenarios used training dataset B (Table 3). One biome model used testing dataset B to represent a project where a custom model was trained using images from the project's own camera sites (in-sample), the other used testing dataset C to represent a project where an off-the-shelf model was used which was not trained using

images from the project's own camera sites (out-of-sample) (Table 4). The two other scenarios represented custom models with only the three individual species output classes (birds, hedgehogs, and empty) specifically required for the project (Table 3). The large project-specific model included 171 camera sites using training dataset I; the small project-specific model included only nine camera sites using training dataset J (Table 3). The project-specific models used testing datasets D and F which matched their training datasets in both output classes and the number of camera sites (Table 4).

Results

Among the 13 model/testing dataset combinations, top-1 accuracy ranged from 59.1 to 91.8%, top-5 accuracy from 94.7 to 99.6%, false alarm rate from 1.5 to 20.8%, and missed invasive rate from 0.0 to 44.2% (Table 5). A model's performance metrics were related; models with higher top-1 accuracies had lower false alarm rates but higher missed invasive rates (Figure 7). Top-1 and top-5 accuracy rates for individual output classes within models showed that output classes with more training images had higher accuracies, with some output classes reaching 100% in both top-1 and top-5 accuracy (Figure 8).

Table 5

Model Results

Model	Top-1		Top-5[a]	False alarm		Missed invasive	
	Accuracy	Count	Accuracy	Rate	Count	Rate	Count
grouped	91.8%	24,938	-	1.5%	417	23.5%	205
individual & custom biome	91.3%	24,817	99.6%	1.9%	510	24.0%	210
off-the-shelf biome	81.5%	349	98.1%	1.6%	7	34.1%	14
10 image grouped	71.2%	19,251	-	16.1%	4,371	10.8%	94
100 image grouped	78.5%	21,336	-	9.5%	2,568	13.8%	121
1,000 image grouped	85.0%	23,084	-	2.6%	713	23.9%	209
10 image individual	69.0%	18,752	94.7%	19.1%	5,185	8.5%	74
100 image individual	74.6%	20,272	98.0%	8.7%	2,367	15.8%	138
1,000 image individual	83.1%	22,589	99.2%	4.1%	1,127	22.3%	195
large project without novel species	77.6%	20,512	-	2.7%	708	14.3%	20
large project with novel species & large project-specific	77.1%	20,567	-	2.5%	662	44.2%	170
small project without novel species	60.4%	2,165	-	20.5%	736	0.0%	0
small project with novel species & small project-specific	59.6%	2,145	-	20.8%	749	33.3%	7

[a] Top-5 accuracy only reported for models with more than five output classes.

Figure 7

Top-1 Accuracy as a Function of False Alarm and Missed Invasive Rate

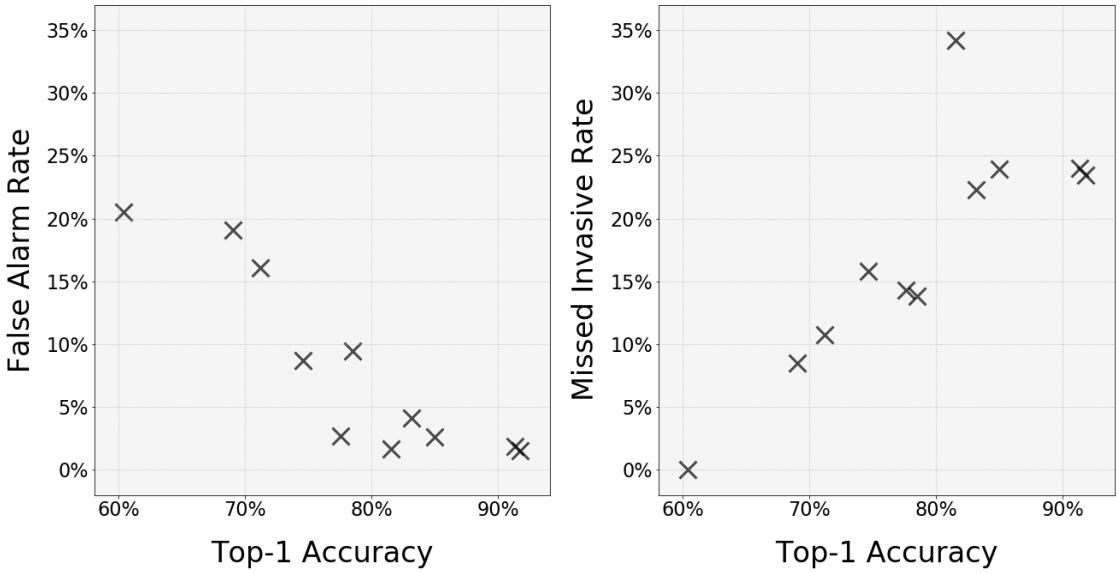


Figure 8

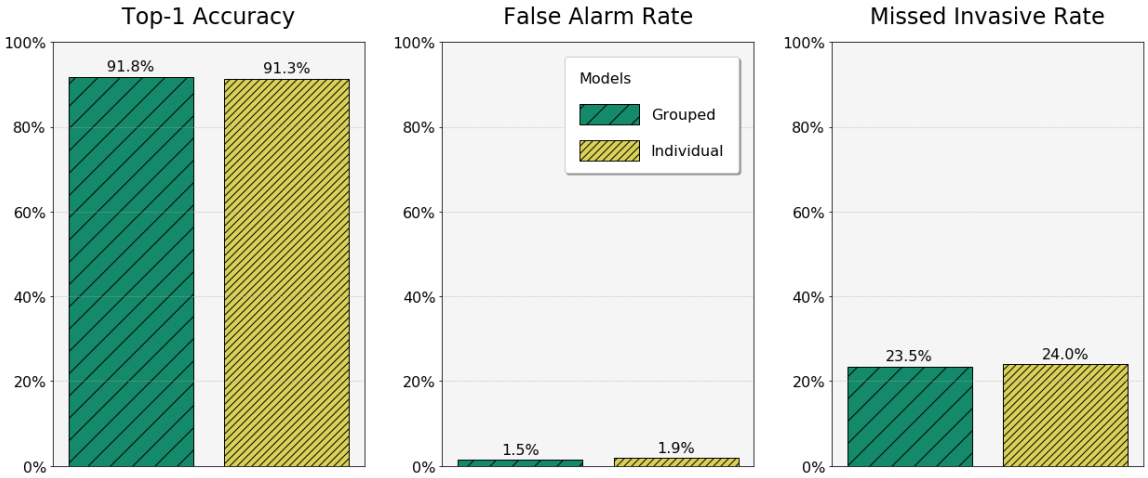
Top-1 and Top-5 Accuracy by Output Class

Note. Each marker represents a model output class of either an individual species or a group of species.

Both models trained to address question 1 (grouped and individual), performed similarly with the grouped model demonstrating slightly higher top-1 accuracy and slightly lower false alarm and missed invasive rates (Figure 9).

Figure 9

Grouped Versus Individual Models

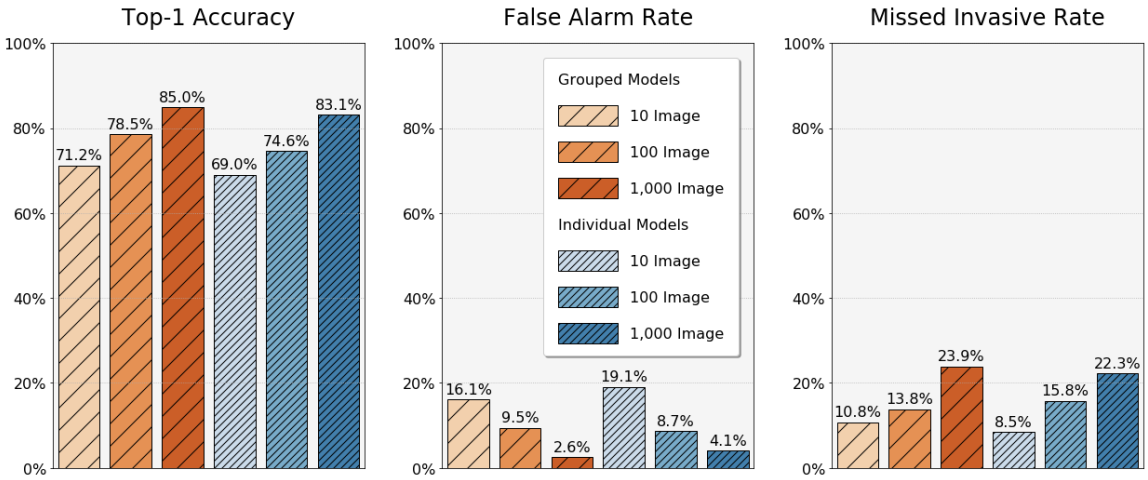


Note. Both models used the same training datasets but with different output classes.

Performance metrics for the six models used to assess question 2 (10/100/1,000 image grouped and individual) demonstrated that increasing the maximum training images per class per camera increased top-1 accuracy and decreased false alarm rates, but increased missed invasive rates (Figure 10). The three grouped models performed similarly compared to their matched individual models with the grouped models demonstrating slightly higher top-1 accuracy and slightly lower false alarm and missed invasive rates (Figure 10).

Figure 10

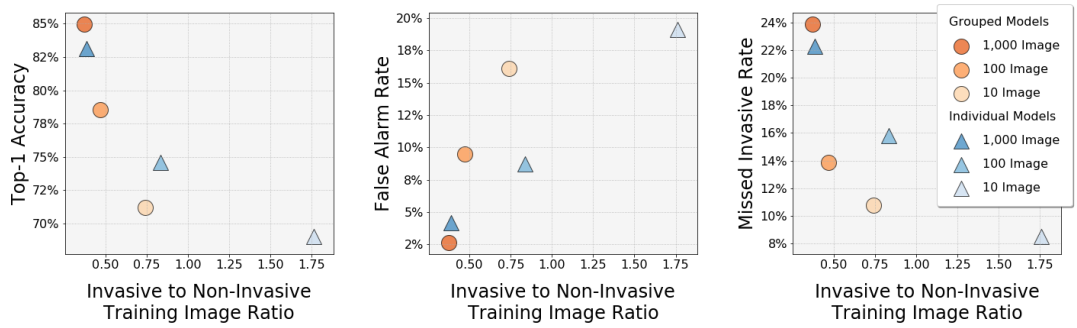
Effects of Maximum Number of Training Images per Class per Camera Site



Because there were different numbers of each species at each camera site in the training datasets, as the maximum number of training images per class per camera increased, the ratio of invasive to non-invasive (empty and bird) images in the training dataset decreased. An increase in the ratio of invasive to non-invasive images was associated with a decrease in top-1 accuracy and a decrease in missed invasive species rates, but an increase in false alarm rates (Figure 11). Note, a ratio of 1.0 does not mean an equal number of training images in each output class but an equal number of invasive images of any of the 11 invasive classes and non-invasive images of either of the two classes empty or bird.

Figure 11

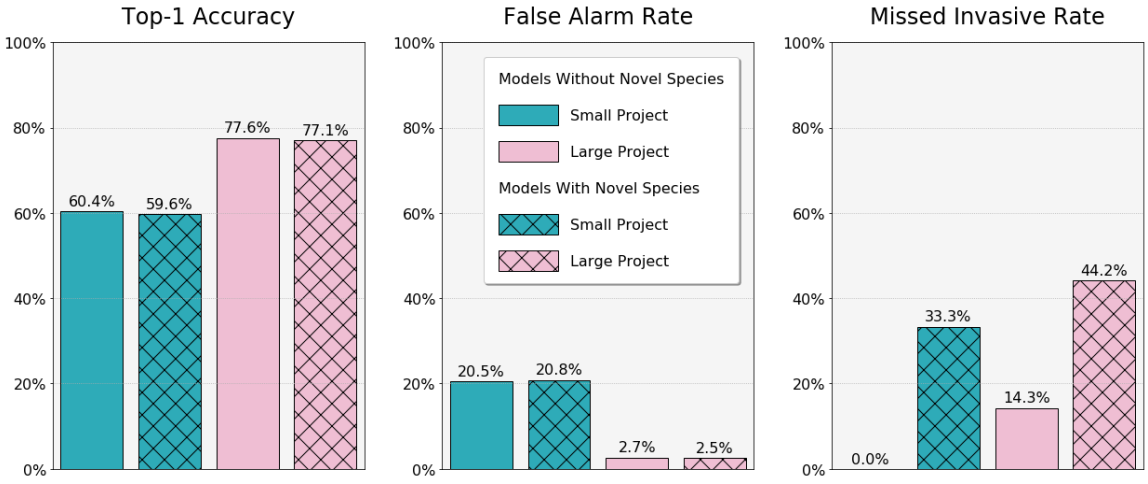
Effects of Invasive to Non-invasive Animal Image Ratio in the Training Dataset



When the large and small project models were tested with and without novel invasive species classes in the testing dataset top-1 accuracy and false alarm rates were very similar. However, the missed invasive rate was markedly higher with the novel species included; in the large project model, the missed invasive rate was three times higher with the novel species than without (Figure 12).

Figure 12

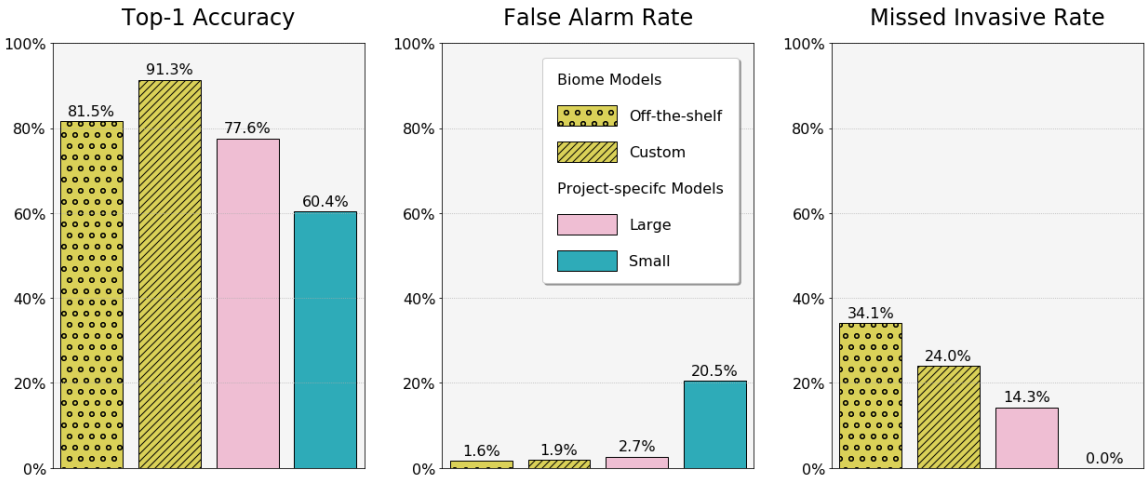
Effects of Novel Invasive Species in the Testing Dataset



Assessing the four models representing the realistic field situations for invasive species monitoring camera-trap projects showed that the custom biome model had the highest top-1 accuracy (Figure 13). The model with the lowest false alarm rates was the off-the-shelf biome model, and the model with the lowest missed invasive rate was the small project-specific model (Figure 13). No one model maximized beneficial results for all three metrics; rather, each model had different strengths and weaknesses. Comparing the two biome models showed that top-1 accuracy and missed invasive rate performance dropped on the out-of-sample testing dataset compared to the in-sample testing dataset, with increased error rates of 112.6% and 42.1% respectively, while false alarm rate performance improved on the out-of-sample testing dataset with decreased error rates of 15.8% (Figure 13).

Figure 13

Biome and Project-specific Models



Discussion

The overarching purpose of this research was to provide direction for managers on the most accurate models for analyzing camera-trap images, given their project goals. Specifically, this work focused on how to best use computer vision animal classification models to detect invasive species at low population levels. What emerged clearly from the results of this research is that wildlife managers selecting a model for use for any camera-trap project must assess the tradeoffs both between different performance metrics—such as top-1 accuracy, false alarms, and missed invasive species—as well as the costs of model procurement, difficulty of use, and time needed to produce a model.

Overall, the models in this study performed on par with animal classification models in other recent camera-trap research when assessed via top-1 and top-5 accuracy, the only performance metrics available for comparison (Norouzzadeh et al., 2018; Schneider et al., 2018a; Tabak et al., 2018). As generally expected in machine learning, output classes with more training images demonstrated higher accuracy, and this research showed that pattern (Goodfellow et al., 2016). Results of this study also revealed an increase in top-1 error rates when moving from an in-sample testing dataset to an out-of-sample testing dataset as was reported by Beery et al. (2018) and Tabak et al. (2018). However, top-1 and top-5 accuracy are not the only metrics of importance, when false alarm and missed invasive rates are taken into account a more complex picture emerges, where the best model is not simply the one with the most training images.

Of the eight models tested on the datasets containing the same images (grouped, individual, 10/100/1,000 image grouped and individual), the best models concerning

top-1 accuracy (91.8%) and top-5 accuracy (99.6%) were the grouped and individual models, respectively. The grouped model also had the lowest false alarm rate (1.5%) of the eight models. In contrast, the model that performed best in regards to the lowest missed invasive rate (8.5%) was the 10 image individual model. In general, models that performed well on the top-1 and top-5 accuracy metrics also performed well on the false alarm metric but performed poorly in the missed invasive metric. This trade-off is likely caused by the ratio of invasive to non-invasive images in the training and testing datasets, and the sensitivity of the performance metrics to this ratio.

Although this research did not look specifically at the effects of balancing the number of images of each output class included in the training dataset in isolation from the total number of training images, the models with a higher ratio of invasive to non-invasive training images, but fewer total training images—due to the undersampling of common classes—had lower top-1 and top-5 accuracies compared to models without undersampling where the ratio was lower but the number of training images higher. This same pattern was noted by Nguyen et al. (2017) and Van Horn and Perona (2017). However, the models with higher ratios of invasive to non-invasive images had lower missed invasive rates. This suggests that artificially balancing a dataset by undersampling common classes is only a drawback if the most important metric is overall accuracy. However, if not missing a rare species, such as an invasive species at low population levels, is vital, an artificially balanced dataset decreases the chance of missing an instance of that species, albeit at the cost of more false alarms. When taken in context with the machine learning adage that the more relevant training data provided to the

model the more accurate the model, the definition of relevant relates to the metric of most importance. If a wildlife manager's top priority is lowering missed invasive rates, there is a point at which the addition of empty and native species images to the training dataset is not relevant, marked by the deterioration of the missed invasive rate even if other metrics continue to improve.

Selecting an appropriate performance metric in regards to the distribution of output classes in the expected testing dataset is also important. When images of novel invasive species not seen in the training datasets were added to the testing datasets of large and small project models, top-1 and top-5 accuracy dropped only a little but missed invasive rates increased a lot. The missed invasive rate metric is more sensitive to the errors caused by the novel species because its denominator is the number of invasive images in the testing dataset (a small number), not the total number of images in the testing dataset (a large number). Wildlife managers most concerned about missing an infrequent species should judge models based on a metric that is more, rather than less, sensitive to the appropriate errors.

Grouping the output classes did improve the performance of the models in all three metrics, but only by a small amount. If this effect holds for other datasets, the value of this improvement depends on the specifics of the project, specifically whether the small performance improvement is worth the loss of species-specific information, the cost of manually retrieving the species-specific information if desired, and the cost of obtaining a grouped output class model if one is not already available. Small improvements in performance metrics are not always worth the cost of procuring a new model.

The results of this study also indicated that grouping output classes together is not a substitute for including training images of each species the model will encounter. As shown by the increase in missed invasive rates in the models tested on novel invasive species, the novel invasive images (rat and unclassifiable) were not conveniently classified as the only available invasive species output class, hedgehog. Although a rat may appear visually closer to a hedgehog than a bird to the human eye, the computer's "eye" does not come to the same conclusion. This also suggests that the models are representing the grouped output class "invasive" as a positive label with multiple visual forms (hedgehog or rat or cat) as opposed to a negative form (not a bird or empty), which matches the generalized understanding that CNNs are analyzing visual patterns and not using other forms of logic.

Finally, if expertise in computer science is available, a custom-trained model is preferable to an off-the-shelf one for multiple reasons. First, the same biome model performed better on top-1 accuracy and missed invasive rate when used as a custom model (tested on in-sample images) than when used as an off-the-shelf model (tested on out-of-sample images). Second, the images used for training the model can be customized to reflect the goals of the project and the performance metric of choice. Even models with only a few hundred training images can outperform an off-the-shelf model trained on thousands of images, as long as the training images are those that are most relevant to the performance metric of choice.

However, given the significant amount of time and expertise in computer science required to train a custom model, the performance benefits might not outweigh the costs.

Especially since off-the-shelf models still perform well and might be accurate enough for some wildlife use cases. As supported by this research, the likelihood of an off-the-shelf model working well enough for a camera-trap project is increased if the distribution of species in the project's camera-trap images matches the species distribution seen in both the training and testing datasets of the chosen off-the-shelf model, and the metrics used for assessing the performance of the off-the-shelf model coincides with the goals of the project.

To further judge the acceptability of using any animal classification model, it is valuable to translate the performance metric rates into expected counts given the number of images in a project. For example, if the testing images from the small project-specific model were classified using the off-the-shelf model, there would be approximately 52 false alarms and 11 missed invasive images, compared to the original 736 false alarms and no missed invasive species images. The particular resources and goals of a project dictate which model is preferable.

Recommendations

Although the pursuit of a perfect computer vision model for detecting and classifying camera-trap images is worthy, current models already work well enough for many applications. The challenge is to find and train a model that works for a specific project. As the research here highlights, evaluating tradeoffs to select an appropriate model can be difficult. Custom models offer higher performance, depending on the performance metric, even with limited training images; and they offer the chance to customize the performance metrics. But custom models require more time and expertise. Off-the-shelf models have less flexible performance metrics and offer somewhat lower performance, but require less time, money, and expertise to implement.

If the costs and expertise required to create a custom model are beyond reach, managers should look for an off-the-shelf model that:

- Reports the performance metric most important to the project or provides the information to calculate that metric.
- Uses a training dataset with a similar distribution of images per species as the expected distribution in the project's own camera-trap images.

When it is possible to build a custom model:

- Use training images from camera-traps at the project site depicting each species the model will classify.
- Include in the training dataset all available images of the most important species to the performance metric of choice, including images from other projects.
- Limit the number of less relevant classes to maintain a balanced class ratio.

The work of improving machine learning computer vision animal classification models for camera-trap images can be advanced by exploring the use of different metrics, such as the missed invasive rate reported here, to inform the training process of the model. Other techniques worth exploring include adjusting the data augmentation techniques used on individual output classes that are more or less relevant to the performance metrics, and further methods for handling high-class imbalance in training datasets.

It is important to incorporate new machine learning classification and detection models as they continue to evolve and improve into wildlife conservation applications. A worthy area of further research is the study of how building and using these models can be simplified to the point that they are within reach of all wildlife managers and their projects. To do so will require bringing wildlife managers together with user experience designers and engineers to design an end-to-end smart camera-trap system that is accurate and usable.

Finally, wildlife managers without enough expertise in the computer science field to build a model should reach out to local college and university computer science departments. There are many graduate students, and even undergraduate students, capable and willing to take on applied projects that will benefit the environment. Such challenges allow students to use their skills in service of our environment and to teach them the power they have to help. There is a growing movement of people in all disciplines who want to use their skills for good. Fostering these interdisciplinary

connections and integrating environmental awareness into all human activities, will be vital as we tackle the unprecedented environmental threats we face.

Literature Cited

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

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*(7336), 51–57. https://doi.org/10.1038/nature09678

Beck, K. G., Zimmerman, K., Schardt, J. D., Stone, J., Lukens, R. R., Reichard, S., Randall, J., Cangelosi, A. A., Cooper, D., & Thompson, J. P. (2008). Invasive species defined in a policy context: Recommendations from the federal invasive species advisory committee. *Invasive Plant Science and Management*, *1*(4), 414–421. https://doi.org/10.1614/IPSM-08-089.1

Beery, S., Morris, D., & Yang, S. (2019). *Efficient pipeline for camera trap image review*. arXiv. https://arxiv.org/abs/1907.06772

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in terra incognita. *Computer Vision – ECCV* 2018, 11220, 472–489. https://doi.org/10.1007/978-3-030-01270-0_28

Brooke, M. de L, Hilton, G. M, & Martins, T. L. F. (2007). The complexities of costing eradications: A reply to Donlan Wilcox. *Animal Conservation*, *10*(2), 157–158. https://doi.org/10.1111/j.1469-1795.2007.00107.x

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., Boutin, S., & Stephens, P. (2015). REVIEW: Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *The Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Ceballos, G., Ehrlich, P. R, Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), Article E1400253–E1400253. https://doi.org/10.1126/sciadv.1400253

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences - PNAS*, *114*(30), Article E6089–E6096. https://doi.org/10.1073/pnas.1704949114

Chen G., Han, T. X., He Z., Kays, R., & Forrester, T. (2014). Deep convolutional neural

network based species recognition for wild animal monitoring. *2014 IEEE International Conference on Image Processing (ICIP)*, 858–862. https://doi.org/10.1109/ICIP.2014.7025172

Colautti, R. I., & MacIsaac, H. J. (2004). A neutral terminology to define 'invasive' species. *Diversity and Distributions*, *10*(2), 135–141. https://doi.org/10.1111/j.1366-9516.2004.00061.x

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Digital Science, Dimensions - free version (2021). Retrieved February 24, 2021, from https://app.dimensions.ai

The Database of Island Invasive Species Eradications, developed by Island Conservation, Coastal Conservation Action Laboratory UCSC, IUCN SSC Invasive Species Specialist Group, University of Auckland and Landcare Research New Zealand (2018). Retrieved February 24, 2021, from http://diise.islandconservation.org

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

El Gamal, A. (2002). Trends in CMOS image sensor technology and design. *Technical Digest - International Electron Devices Meeting*, 805–808. https://doi.org/10.1109/IEDM.2002.1175960

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Glover‐Kapfer, P., Soto‐Navarro, C. A., & Wearn, O. R. (2019). Camera‐trapping version 3.0: Current constraints and future priorities for development. *Remote Sensing in Ecology and Conservation*, *5*(3), 209–223. https://doi.org/10.1002/rse2.106

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

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

Håvard, T. (2017). *Unified detection system for automatic, real-time, accurate animal detection in camera trap images from the Arctic tundra*. [Master's thesis, The Arctic

University of Norway, Tromsø, Norway]. https://hdl.handle.net/10037/11218

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications*, *10*(3), 689–710. https://doi.org/10.1890/1051-0761(2000)010[0689:BICEGC]2.0.CO;2

Mcneely, J. (2001). *The great reshuffling: Human dimensions of invasive alien species*. IUCN.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

New Zealand Government (2017, June). *Predator free 2050*. https://www.doc.govt.nz/Documents/our-work/predator-free-2050.pdf

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

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

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021). A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, *12*(1), 150–161. https://doi.org/10.1111/2041-210X.13504

Oh, K., & Jung, K. (2004). GPU implementation of neural networks. *Pattern Recognition*, *37*(6), 1311–1314. https://doi.org/10.1016/j.patcog.2004.01.013

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). *PyTorch: An imperative style, high-performance deep learning library*. arXiv. https://arxiv.org/abs/1912.01703

Ren, X., Han, T. X., & He, Z. (2013). Ensemble video object cut in highly dynamic scenes. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1947–1954. https://doi.org/10.1109/CVPR.2013.254

Russell, J. C., Meyer, J., Holmes, N. D., & Pagad, S. (2017). Invasive alien species on islands: Impacts, distribution, interactions and management. *Environmental Conservation*, *44*(4), 359–370. https://doi.org/10.1017/S0376892917000297

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Schneider, S., Taylor, G. W., & Kremer, S. (2018). Deep learning object detection methods for ecological camera trap data. *Proceedings - 2018 15th Conference on Computer and Robot Vision CRV 2018*, 321–328. https://doi.org/10.1109/CRV.2018.00052

Sewak, M., Karim, Md. R., & Pujari, P. (2018). *Practical convolutional neural networks: Implement advanced deep learning models using Python*. Packt Publishing.

Sharath Kumar, Y. H., Manohar, N., & Chethan, H. K. (2015). Animal classification system: A block based approach. *Procedia Computer Science*, *45*, 336–343. https://doi.org/10.1016/j.procs.2015.03.156

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Silveira, L., Jácomo, A. T. A., & Diniz-Filho, J. A. F. (2003). Camera trap, line transect census and track surveys: A comparative evaluation. *Biological Conservation*, *114*(3), 351–355. https://doi.org/10.1016/S0006-3207(03)00063-6

Spatz, D. R., Zilliacus, K. M., Holmes, N. D., Butchart, S. H. M., Genovesi, P., Ceballos, G., Tershy, B. R., & Croll, D. A. (2017). Globally threatened vertebrates on islands with invasive species. *Science Advances*, *3*(10), e1603080–e1603080. https://doi.org/10.1126/sciadv.1603080

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). *Inception-v4, Inception-ResNet and the impact of residual connections on learning*. arXiv.

https://arxiv.org/abs/1602.07261

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., … Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, *10*(4), 585–590. https://doi.org/10.1111/2041-210X.13120

Thoma, M. (2017). *Analysis and optimization of convolutional neural network architectures*. arXiv. https://arxiv.org/abs/1707.09725

U.S. Fish and Wildlife Service (2007). *Restoring wildlife habitat on Rat Island: Environmental assessment*. https://digitalcommons.unl.edu/usfwspubs/77

Van Horn, G., Perona, P. (2017). *The devil is in the tails: Fine-grained classification in the wild*. arXiv. https://arxiv.org/abs/1709.01450

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

Young, S., Rode‐Margono, J., & Amin, Rajan. (2018). Software to facilitate and streamline camera trap data management: A review. *Ecology and Evolution*, *8*(19), 9947–9957. https://doi.org/10.1002/ece3.4464

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

Zhu, C., Li, T. H., & Li, G. (2018). Towards automatic wild animal detection in low quality camera-trap images using two-channeled perceiving residual pyramid networks. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, *2018*, 2860–2864. https://doi.org/10.1109/ICCVW.2017.337

Appendix A

Epoch number	Weight decay	Learning rate
1-18	0.0005	0.01
19-29	0.0005	0.005
30-43	0	0.001
44-52	0	0.0005
53-55	0	0.0001

Note. Model training variables learning rate and weight decay by training epoch are the same as those used by Norouzzadeh et al. (2018) and Tabak et al. (2018).