

RESEARCH ARTICLE

Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies

Madeline C. Hayes,^{1,*} Patrick C. Gray,^{1,○} Guillermo Harris,² Wade C. Sedgwick,² Vivon D. Crawford,² Natalie Chazal,³ Sarah Crofts,⁴ and David W. Johnston^{1,○}

¹ Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University Marine Laboratory, Beaufort, North Carolina, USA

² Wildlife Conservation Society, Buenos Aires, Argentina

³ College of Sciences, Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA

⁴ Falklands Conservation, Stanley, Falkland (Malvinas) Islands

*Corresponding author: madeline.c.hayes@duke.edu

Submission Date: September 24, 2020; Editorial Acceptance Date: April 6, 2021; Published May 22, 2021

ABSTRACT

Population monitoring of colonial seabirds is often complicated by the large size of colonies, remote locations, and close inter- and intra-species aggregation. While drones have been successfully used to monitor large inaccessible colonies, the vast amount of imagery collected introduces a data analysis bottleneck. Convolutional neural networks (CNN) are evolving as a prominent means for object detection and can be applied to drone imagery for population monitoring. In this study, we explored the use of these technologies to increase capabilities for seabird monitoring by using CNNs to detect and enumerate Black-browed Albatrosses (*Thalassarche melanophris*) and Southern Rockhopper Penguins (*Eudyptes c. chrysochome*) at one of their largest breeding colonies, the Falkland (Malvinas) Islands. Our results showed that these techniques have great potential for seabird monitoring at significant and spatially complex colonies, producing accuracies of correctly detecting and counting birds at 97.66% (Black-browed Albatrosses) and 87.16% (Southern Rockhopper Penguins), with 90% of automated counts being within 5% of manual counts from imagery. The results of this study indicate CNN methods are a viable population assessment tool, providing opportunities to reduce manual labor, cost, and human error.

Keywords: Black-browed Albatross, convolutional neural network, deep learning, drone, population assessment, seabird monitoring, Southern Rockhopper Penguin

LAY SUMMARY

- We tested the viability of using deep learning coupled with drone imagery to monitor Black-browed Albatrosses and Southern Rockhopper Penguins.
- Many seabird colonies at the Falkland (Malvinas) Islands are large and remote, presenting challenges for long-term monitoring.
- We used convolutional neural networks to enumerate both species from drone imagery and compared automated counts to manual counts.
- Our results produced high accuracies and low percent difference with manual counts.
- Deep learning coupled with drone imagery shows great potential for the future of seabird monitoring, particularly in large and spatially complex colonies.

Los drones y el aprendizaje profundo producen un monitoreo preciso y eficiente de colonias de aves marinas de gran escala

RESUMEN

El monitoreo de las poblaciones de aves marinas coloniales es usualmente complicado por el gran tamaño de las colonias, los lugares remotos y la agregación densa inter- e intra-específica. Mientras que los drones han sido usados exitosamente para monitorear grandes colonias inaccesibles, la gran cantidad de imágenes colectadas introduce un cuello de botella en el análisis de los datos. Las redes neuronales convolucionales (RNC) están evolucionando como un medio prominente para la detección de objetos y pueden ser aplicadas a las imágenes de drones para el monitoreo poblacional. En este estudio, exploramos el uso de estas tecnologías para aumentar la capacidad de monitoreo de las

© The Author(s) 2021. Published by Oxford University Press for the American Ornithological Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

aves silvestres mediante el uso de RNC para detectar y contar a *Thalassarche melanophris* y *Eudyptes c. chrysocome* en una de sus colonias de cría más grandes, las Islas Malvinas. Nuestros resultados mostraron que estas técnicas tienen gran potencial para el monitoreo de aves marinas en colonias significativas y espacialmente complejas, produciendo precisiones para detectar y contar correctamente las aves en un 97.66% (*T. melanophris*) y 87.16% (*E. c. chrysocome*), con un 90% de los conteos automáticos estando dentro del 5% de los conteos manuales realizados a partir de las imágenes. Los resultados de este estudio indican que los métodos de RNC son una herramienta viable de evaluación poblacional, brindando oportunidades para reducir el trabajo manual, el costo y el error humano.

Palabras clave: aprendizaje profundo, drone, *Eudyptes c. chrysocome*, evaluación poblacional, monitoreo de aves marinas, red neuronal convolucional, *Thalassarche melanophris*

INTRODUCTION

Accurate wildlife population assessments are crucial for effective conservation and ecosystem management, particularly of focal species whose abundance can indicate the condition of the more complex community (Zacharias and Roff 2001). Seabird population dynamics have proven to be successful indicators of ecological change due to tightly coupled dependence on oceanographic conditions and key roles as marine predators (Diamond and Devlin 2003, Hazen et al. 2019). Seabird populations are sensitive to environmental change across spatial and temporal scales, and specifically vulnerable to anthropogenic stressors like climate change and fisheries competition and bycatch (Bost and LeMaho 1993, Croxall et al. 2012). Colonial seabirds can be more easily monitored than many other marine megafauna species, and many long-term studies have linked changes in seabird demographic parameters to several threats including the effects of climate change on marine ecosystems (Barbraud and Weimerskirch 2001, Weimerskirch et al. 2003). Yet, many seabird species often breed in large numbers at inaccessible locations closely aggregated together, sometimes with other species, presenting challenges for traditional ground-based surveys (Rush et al. 2018).

Unoccupied aircraft systems (UAS) or drones are a rapidly evolving technology that has been successfully used for surveying a variety of marine species, including cetaceans (Aniceto et al. 2018), dugongs (Hodgson et al. 2013), seals (Seymour et al. 2017), and seabirds (Chabot et al. 2015, Rush et al. 2018). These surveys can often be completed with consumer, off-the-shelf (COTS) drones that are relatively inexpensive yet collect high spatial and temporal resolution imagery (Linchant et al. 2015). The use of drones to monitor seabird colonies, when compared to ground counting methods, can significantly increase the total colony areas surveyed, increase accuracy of counts, reduce direct disturbance and are a viable option for long-term monitoring, but the amount of data collected can introduce a data analysis bottleneck as manual counting of wildlife is labor-intensive (Lyons et al. 2019).

Automated counting of avifauna from aerial imagery has been effectively applied to many different species and locations (Abd-Elrahman et al. 2005, Descamps

et al. 2011, Groom et al. 2011, Ratcliffe et al. 2015). Many computer-automated approaches have focused on spectral thresholding, object-based image analysis, and traditional supervised machine learning, although these methods do not work well if there are multiple species present in imagery and have been most successfully applied to smaller colonies (Chabot and Francis 2016, Hong et al. 2019).

Unlike traditional supervised machine learning approaches, supervised deep learning algorithms, also called neural networks, can effectively learn representative features directly from data with a practitioner only providing the training labels (Akçay et al. 2020). Once trained, a network can then extract similar features from new unseen data for classification and regression problems. Most recent advances are from convolutional neural networks (CNN), which ingest imagery and iteratively scan with a series of learned filters (the convolutional layers) that transform the input data into higher-level features representing aspects that are important for the task the CNN is being trained to do (LeCun et al. 2015). Once learned, these filters help the CNN detect relevant features in previously unseen imagery despite changes in lighting, layout, or exact geometry. CNNs are increasingly important tools in remote sensing and ecology because the convolutional layers incorporate spatial context, which helps identify relevant ecological patterns and processes (Brodrick et al. 2019). CNNs are particularly useful for object detection and, from there, automated enumeration of wildlife.

CNN-based object detection is a rapidly growing research area and numerous methods have proven successful in studies to count wildlife in drone imagery. For example, Gray et al. (2019) used a CNN to automate species identification and measurement of cetaceans in drone imagery, finding that the CNN correctly predicted whale species in 98% of images. Borowicz et al. (2018) used a neural network to detect Adélie Penguins (*Pygoscelis adeliae*) in drone imagery within 10% accuracy of ground counts, resulting in the discovery of a new Adélie Penguin hotspot. Furthermore, Gray et al. (2018) used a CNN to detect sea turtles in drone imagery, finding that the model reduced manual analysis burden to 1.5% of the initial amount of time required to manually count sea turtles. While deep learning and drones show great promise for the future of wildlife monitoring, there are still ecological and technical

limitations. Some species may be too small or elusive, a drone may introduce disturbance, and standard drone sensor visibility can be limited in both marine and terrestrial environments (Johnston 2019).

The Falkland (Malvinas) Islands are home to the world's largest colonies of Black-browed Albatrosses (*Thalassarche melanophris*) and the second largest colonies of Southern Rockhopper Penguins (*Eudyptes c. chrysocome*) (Baylis 2012). Accordingly, the conservation status of both species is dependent on local population trends. Since 1990, the Falkland Islands Seabird Monitoring Programme has been monitoring Black-browed Albatrosses and Southern Rockhopper Penguins annually at select colonies, with an archipelago-wide census conducted every 5 years until 2010, when it was shifted to species-specific censuses (Crofts and Stanworth 2019). The results of the 2015 Black-browed Albatross census are still being analyzed.

The Black-browed Albatross, classified as Least Concern (LC) by the International Union for Conservation of Nature's (IUCN) Red List of Threatened Species, breeds at 12 distinct sites in the Falkland Islands, with the largest colony on Steeple Jason Island. The census in 2010 estimated between 474,000 and 535,000 breeding pairs across the entire Falkland Islands (Wolfaardt 2012). Approximately 0.5% of the breeding population is monitored annually (Crofts and Stanworth 2019). Aerial surveys from occupied aircraft and ground-based surveys have been used independently to complete and compare archipelago-wide censuses. Traditional ground-based field counts and counts from images, including drones, were most recently used for the annual surveys at Steeple Jason, with 10% of the annually monitored population counted with photographs (Crofts and Stanworth 2019). Census results suggest that the population numbers were increasing between 2000 and 2010, which may have reflected the ongoing efforts to reduce incidental seabird bycatch in the Falkland Islands and other regional fisheries (Moreno et al. 2008, Wolfaardt 2012).

Similar to Black-browed Albatrosses, the largest colony in the Falkland Islands of Southern Rockhopper Penguins, classified as Vulnerable (VU) by IUCN, is on Steeple Jason Island (Baylis 2012). The most recent complete survey utilized traditional ground-based field counts and photographic counts to estimate 319,000 breeding pairs at the Falkland Islands. Approximately 2.6% of this population is monitored annually and these surveys are conducted with traditional ground-based field counts and aerial imagery including, more recently, drone counts (Crofts and Stanworth 2019). The 5 sub-colony areas surveyed in 2019 using drone counts correspond to about half of the total estimated colony area. Archipelago-wide censuses in 2000 and 2005 estimated a steep decline in the breeding population, but the most recent 2010 census indicated a 50.6% increase in breeding pairs between 2005 and 2010 (Baylis

et al. 2013). Southern Rockhopper Penguin populations are threatened by the effects of changing sea surface temperatures that influence their prey abundance and availability, as well as the impact of oil pollution at sea (Pütz et al. 2002, Dehnhard et al. 2013). A starvation event was identified as the possible cause (Morgenthaler et al. 2018) of a drop of 31% in the annually monitored breeding population in 2016 (Crofts and Stanworth 2017).

While annually monitored sites for both Black-browed Albatrosses and Southern Rockhopper Penguins account for a small portion of total Falkland Islands populations of both species, they are designed to detect finer resolution population fluctuations at selected sites to represent changes for the Falkland Islands as a whole (Baylis 2012). Archipelago-wide censuses, carried out less frequently, are used to assess overall population abundances and corroborate any significant population trends detected in the annual surveys (Huin and Reid 2006). Combined, both selected site surveys and larger archipelago censuses are critical components to understand the species population dynamics on a long-term scale, however, logistical limitations and associated costs can be a tradeoff between the level of survey effort and survey frequency, particularly for large, remote, and numerous widespread colonies, such as those on the Falkland Islands (Wolfaardt 2012). Incorporating newer technologies such as drones and automated counting can significantly advance effective seabird monitoring at the Falkland Islands.

The purposes of the present study were to (1) collect population assessment quality drone imagery of all the known colonies of Black-browed Albatrosses and Southern Rockhopper Penguins at Steeple and Grand Jason Islands, (2) build and train deep learning models to detect and enumerate individuals of both species in drone imagery, and (3) deploy these models and evaluate accuracy to develop a cost-effective method for seabird colony monitoring of both species in a conservation stronghold habitat.

METHODS

Study Area

Drone surveys at both albatross and penguin colonies were focused on Steeple Jason Island and Grand Jason Island in the Falkland (Malvinas) Islands (Figure 1). The Jason Islands are located north-west of West Falkland (51.077°S, 60.969°W) and their coastlines consist of rocky shores and steep cliffs rising to tussock grass-covered slopes. Black-browed Albatrosses and Southern Rockhopper Penguins arrive at these islands to lay eggs in early October and early November, respectively (Baylis 2012). Black-browed Albatross nests are built with mud and guano, whereas Southern Rockhopper Penguin nests are ill-defined scrapes. Nesting sites are reused every year for both species. Based



FIGURE 1. Map of the study sites at the Jason Islands. Colony names are outlined in Table 1 based on the corresponding numbers.

on the last island-wide census of Black-browed Albatrosses in 2010, there were an estimated 201,986 to 214,203 breeding pairs on Steeple Jason and 89,580 breeding pairs on Grand Jason (Wolfaardt 2012). Censuses of Southern Rockhopper Penguins at Steeple Jason estimated 121,396 breeding pairs in 2010 and 10,496 breeding pairs at Grand Jason in 2005 (Baylis 2012).

UAS Imagery Collection

Aerial imagery was collected with a DJI Phantom (Shenzhen DJI Sciences and Technologies, Nanshan, Shenzhen, China), a consumer drone with a 9-mm fixed focal length lens, a resolution of $4,864 \times 3,648$ pixels, and a flight time of 20–25 min. This drone was used due to its vertical takeoff and landing capabilities in rough terrain. Imagery was collected between November 11 and 21, 2018 for all but one section of the colonies on Steeple Jason. Imagery was collected again between November 3 and 13,

2019 for the colony area not surveyed in 2018 on Steeple Jason plus 4 smaller subsets of those areas flown in 2018. All colony areas on Grand Jason were surveyed in 2019 during this time. These surveys were conducted both years during the incubation period for both species, consistent with the previous timing of censuses (Wolfaardt 2012), where a single member of the Black-browed Albatross pair remains sitting on the nest and Southern Rockhopper Penguins are present at the nest as pairs. The 2018 imagery was collected at an average resolution of 5 cm pixel^{-1} , whereas the 2019 imagery was collected at an average of 1 cm pixel^{-1} . The average flight altitude is directly related to the resolution and was around 90 m in 2018 and 60 m in 2019 but varied greatly by colony area because of the uneven terrain. Flight paths over all sites were conducted in overlapping parallel lines following predefined patterns set with DroneDeploy (<https://www.dronedeploy.com/product/mobile/>). Flight characteristics from each colony

TABLE 1. Overview of flight characteristics for each colony area, numbered by site in Figure 1. Ground sampling distance (GSD) is related to flight height and is the distance between 2 consecutive pixels on the ground. A GSD value of 5 means that 1 pixel in the imagery represents 5 cm on the ground.

Colony area name	Date	Colony area size (km ²)	GSD (cm)
1. Steeple Jason North	November 13, 2018	0.17	5
2. Steeple Jason South	November 13, 2018	0.12	5
3. Steeple Jason Bubble	November 8, 2019	0.009	0.64
4. Steeple Jason Hump	November 8, 2019	0.011	0.52
5. Steeple Jason Blob	November 8, 2019	0.009	0.74
6. Steeple Jason N Tip Gully	November 9, 2019	0.14	1.20
7. Steeple Jason South Neck	November 8, 2019	0.19	1.50
8. Grand Jason SW Right Third	November 12, 2019	0.086	1.80
9. Grand Jason SW Middle Third	November 12, 2019	0.061	0.95
10. Grand Jason SW Left Third	November 12, 2019	0.077	1.14
11. Grand Jason SE Blob	November 10, 2019	0.014	0.48
12. Grand Jason SE Colony	November 12, 2019	0.095	1.17

area (labeled in Figure 1) are presented in Table 1. Flights were conducted at different altitudes to determine the lowest resolution threshold at which a deep learning model could accurately detect seabirds; training on images from different resolutions creates a more robust model.

Image Processing and Manual Counts

All aerial imagery was processed into orthomosaics with ± 1 m horizontal accuracy using the structure from motion software Pix4D (<https://www.pix4d.com/product/pix4dmapper-photogrammetry-software>; version 4.5.6). Some of the survey sites presented challenges for automated stitching of images, including ghosting artifacts and edge effects. These orthomosaics were manually edited to create a clear picture of each colony area.

The CNNs must first be trained on a subset of images with the objects of interest manually labeled. Orthomosaics were split into smaller tiles and both bird species were manually marked using the VGG Image Annotator software (<https://www.robots.ox.ac.uk/~vgg/software/via/>; version 3.0.8). All individual birds present of both species were marked regardless of body shape or position. The same observer analyzed all images, although it is expected that counts from different observers would be similar as the birds are relatively easy to identify. All tiles were created with a 60-pixel overlap for Black-browed Albatross detection and a 30-pixel overlap for Southern Rockhopper Penguin detection, which were slightly above the average pixel size of birds in the images. It took 10 hr to manually label the Black-browed Albatrosses and 20 hr to manually label the Southern Rockhopper Penguins. All labeled tiles were split into 80-10-10 training data, validation data, and testing data (Figure 2). Training data are used to train the deep learning model, while the validation data are used to determine the stopping point of training, and the testing dataset of previously unseen data is used to determine the final accuracy of the model. Manually marking the training samples also provides manual counts for total site

population, which is a useful metric to compare the performance of the CNN to the “true” values.

Total Black-browed Albatross counts in the Steeple Jason imagery from 2018 were conducted using a density-based estimation in ArcGIS Pro (<https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>; version 2.3.0). First, total colony areas were delineated manually. A grid overlay was then used to split the total area into continuous 15×15 m quadrats. Utilizing similar techniques to estimate nest densities as in the 2005 census (Huin and Reid 2006), 16 quadrats were selected that were distributed relatively evenly on each island, none of which are within 5 m of the edge of the colony, resulting in $\sim 35\%$ of the total colony by area surveyed. Our use of quadrat rows, similar to strip transects, was preferred because it accounts for lower densities near colony borders (Croxall and Prince 1979). All individual albatrosses were digitized in these quadrat rows and the total numbers were divided by the colony area surveyed to generate an albatross per square meter metric. The total colony area was then multiplied by this metric to estimate the total count of individual albatross numbers on Steeple Jason North and South. Variance in the estimated number of nests for each colony can be calculated by the methods outlined in the 1987 census (Thompson and Rothery 1991), but this requires multiple measurements of colony area and multiple counts by 2 observers. We could not determine the uncertainty in our method as the colony area was measured once and nests were counted once by one observer.

CNN Architecture and Training

Object detection architectures can be categorized as either one-stage or two-stage. Two-stage architectures split potential objects into either foreground or background classes, and then classify all foreground objects into the specific classes of interest, while one-stage detectors do not have this first step. Two-stage detectors are known to be more accurate but computationally intensive.

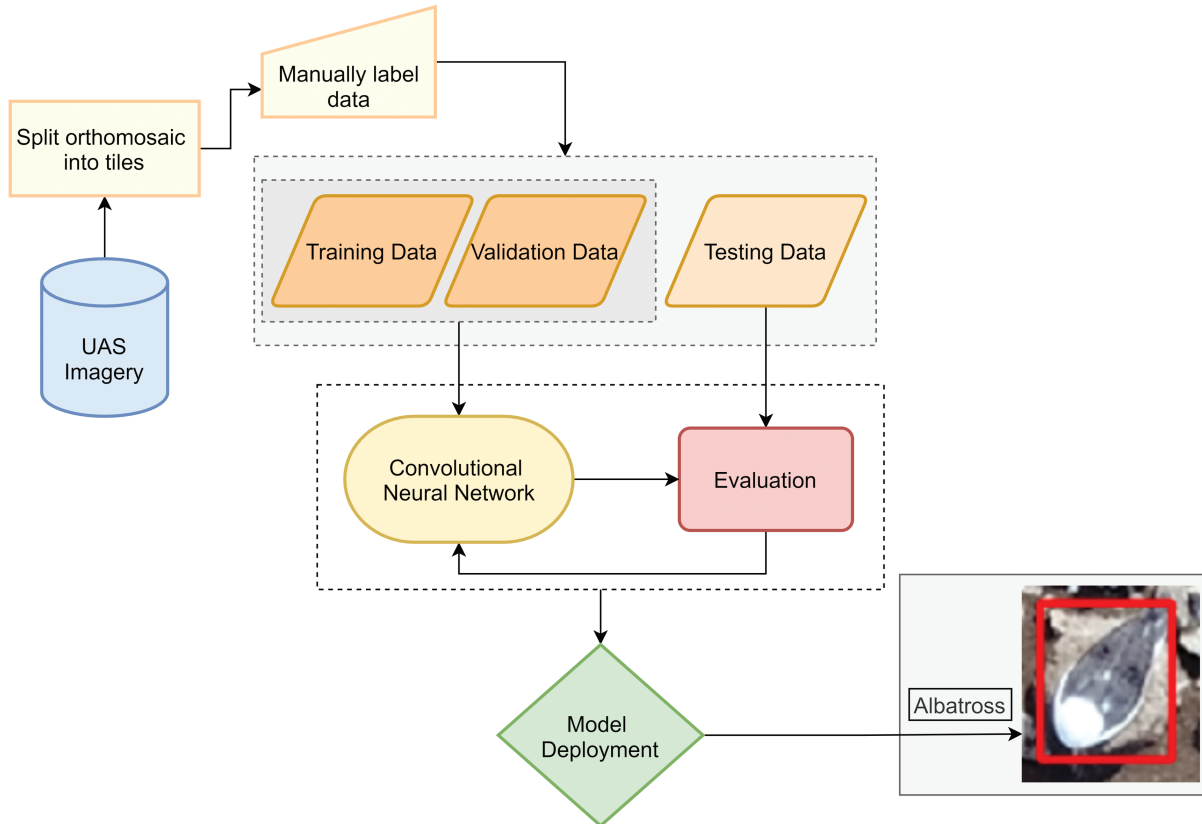


FIGURE 2. Overview of CNN development and deployment for Black-browed Albatross detection. The development cycle is the same for Southern Rockhopper Penguin detection.

Typically, a deep learning model is chosen by considering the tradeoff between speed and accuracy. Faster R-CNN (Ren et al. 2017) has remained a top choice and exceeded all other models on overall accuracy and ability to detect small objects when it was initially published but remains a slower model (Huang et al. 2017). Although one-stage detectors like YOLO (Redmon et al. 2016) and SSD (Liu et al. 2016) yield faster inference times, their accuracies are often 10–40% below that of two-stage detectors (Lin et al. 2017a). The present study employed the Keras implementation (<https://github.com/fizyr/keras-retinanet>) of the one-stage RetinaNet object detection architecture with the ResNet-50-FPN backbone (Lin et al. 2017a) to achieve high accuracy without decreasing computational efficiency. RetinaNet was the first one-stage detector to match the accuracy of more complex two-stage detectors like Faster R-CNN and outperformed all previous one-stage and two-stage detectors in the speed vs. accuracy tradeoff on the Common Objects in Context (COCO) benchmark (Lin et al. 2017a). The ResNet-50-FPN backbone was chosen over the larger ResNet-101-FPN because it has the fastest runtime while achieving similar accuracy on a 500-pixel image scale (Lin et al. 2017a), which is the largest size of the tiles used in this study. It is worth noting that for many

applied ecology problems the exact model is often of less importance than the quality of data, labels, and the time spent collecting and refining these data (Sun et al. 2017, Christin et al. 2019).

Object detection problems involve both a regression of the corners of the bounding box around the object and a classification problem to decide what is in that box. Neural networks are trained for this task by minimizing a loss function. A loss function is used to evaluate how well the network output compares to the expected output (i.e. the training label). In typical classification problems, this is done by comparing the predictions (which are probabilities between 1 and 0 that the input data belong to that class) to the true label (a vector where the correct class is 1 and all others are 0). Loss is minimized when the model has high confidence in the correct class. For regression problems, the mean squared error between the target value and the predicted value is defined as the loss. Typically, the loss function for CNNs conducting object detection is defined by simply adding the classification and regression loss. When minimizing loss, the entire network can be thought of as a function whose input is each learnable parameter in the neural network (often in the millions) and whose output is a loss value. This is often called the

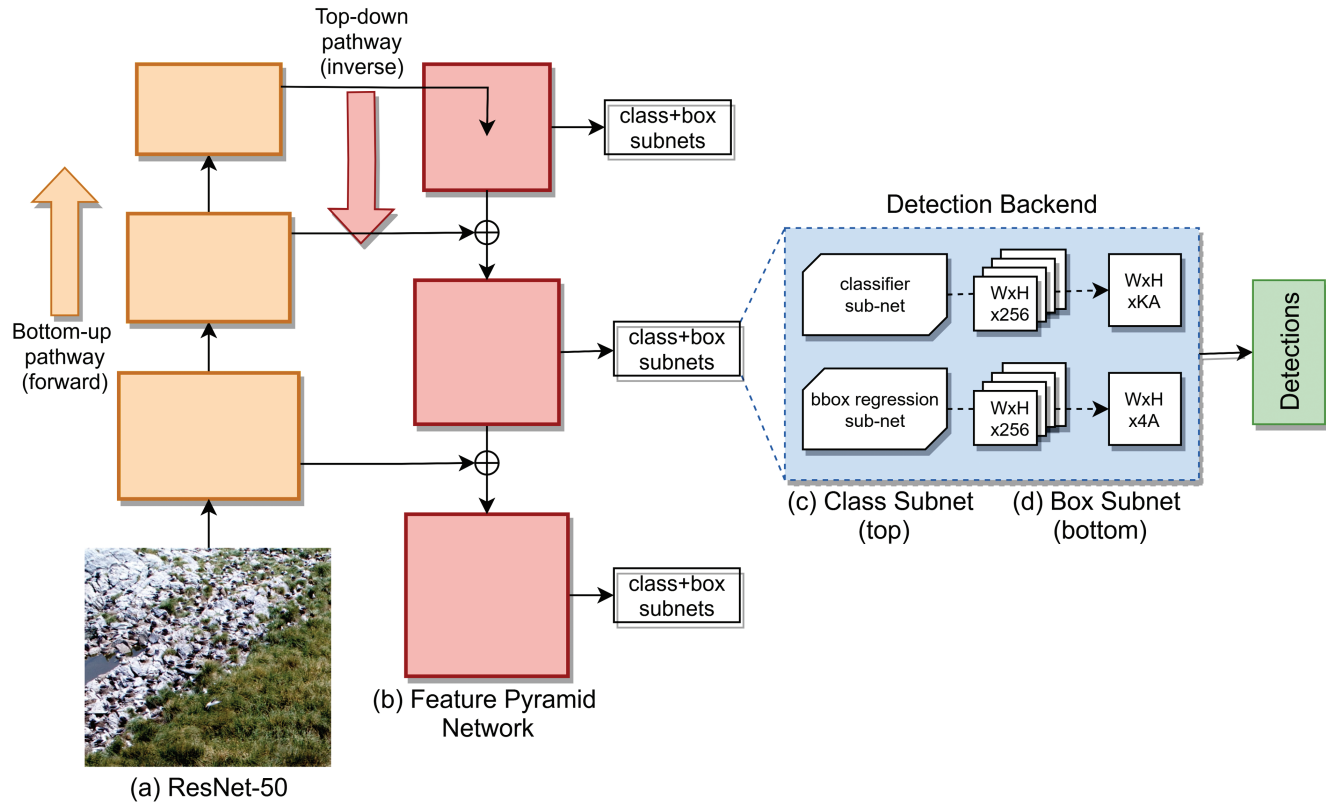


FIGURE 3. Overview of the RetinaNet architecture, consisting of the backbone network and 2 subnetworks for classification and regression. (A) ResNet-50, the convolutional network, generates feature maps at different scales. The spatial resolution decreases and the semantic value increases as it moves up the pathway. (B) The FPN feature extractor combines low-resolution, semantically strong features with high-resolution, semantically weak features to generate multi-scale feature maps. The top-down pathway up samples the spatially coarser feature maps and the lateral connections merge top-down and bottom-up layers. The default anchors, predefined bounding boxes used to capture the scale and aspect ratio of specific object classes, are moved around the multi-scale feature maps and each level of the FPN encodes information at a different scale, informing overall object detection. (C) The classification subnet is a fully convolutional network attached to each FPN level. The output feature map is shape $(W \times H \times KA)$, where $W \times H$ is related to the input feature map and KA is the number of object classes and anchor boxes, respectively. Each anchor box detects the existence of objects from K classes. (D) The box regression subnet (class-agnostic bounding box regressor) is a fully convolutional network to each pyramid level that is identical to the classification subnet but terminates in $4A$ linear outputs per spatial location. The 4 outputs for each A anchor predict the relative offset between the anchor and ground truth box.

cost function. The loss function and cost function must be differentiable so that gradient descent can be used to minimize the whole cost function. Gradient descent optimizes the cost function by updating the training parameters, or weights, to step down the gradient until the lowest point of the function is reached. These steps down the gradient are calculated using the training samples. Weights are the learnable parameters of a deep learning model that transform the input image into output classifications and bounding boxes. In practice, stochastic gradient descent is often the preferred optimization algorithm as it randomly selects one training sample at each iteration instead of calculating loss on all the training samples at each step, significantly reducing computations particularly in large datasets (Bottou 2010).

RetinaNet is able to achieve state-of-the-art performance by utilizing a concept called focal loss to rescale the

loss function. Focal loss improves prediction accuracy by reshaping the classification loss to pay less attention to background examples (decreasing their influence in the loss function) and focusing on challenging foreground examples (increasing their influence in the loss function) (Lin et al. 2017a). RetinaNet is composed of a feature pyramid network (FPN) on top of a feedforward CNN architecture, plus 2 task-specific network branches for classification and bounding box regression (Figure 3). The CNN takes an input image and processes it through several convolutional filters, each outputting a feature map (Figure 3A). The feature maps of the first few layers capture high-level features, such as color gradients and edges, and the later layers create smaller but deeper feature maps that capture more abstract features representative of the final classes (e.g., wings, nest structure). The FPN combines the smaller and more precise feature maps with the larger more context-aware

feature maps through several up-sampling levels (Lin et al. 2017b), resulting in multi-scale feature maps (Figure 3B). At each of these levels, several “anchors” are moved around the feature maps. The anchors are rectangles of preset sizes and aspect ratios, defined by the user based on the expected size of the objects in the imagery, which act as the initial bounding boxes of predicted objects. Our anchors were optimized for small object detection based on the Zlocha et al. (2019) framework. These multi-scale features and anchors are then fed through 2 final branches made up of additional convolutions and pooling. The first branch is for classification which predicts the probability of object presence at each spatial location for each of the A anchors and K object classes (Figure 3C). The probability, or the confidence score, is simply the highest activation from the network’s output neurons. The output activation is passed through a sigmoid function which “squishes” this final output value for each neuron into a range of 0–1 and the highest value is the assigned class. The classification subnet implements focal loss as the loss function by down-weighting the importance of well-classified background samples, preventing a large number of background samples from overwhelming the detector during training, and reducing the effect on model weights. The second subnet is the regression branch that predicts x_1, y_1, x_2, y_2 for each anchor (Figure 3D), makes small adjustments to the original anchors to fit potential objects better, and a smooth loss function is applied. Classification loss and regression loss are calculated at each epoch during training, during which the parameter weights are updated to minimize loss through stochastic gradient descent. Once trained, the model runs inference by selecting anchors with the highest predicted probability, and these anchors are given offset predictions to produce the final bounding box predictions.

A pre-trained version of RetinaNet was used as the starting point of our model weights. This model was trained on Microsoft COCO (Lin et al. 2014), a dataset of over 200,000 labeled images. Using the weights of a model pre-trained on the COCO dataset for our initial training leverages the power of transfer learning (Razavian et al. 2014), which assumes many of the learned convolutional filters will transfer across images from different domains (Kerner et al. 2019). Two CNNs were deployed for detection and enumeration of Black-browed Albatrosses and Southern Rockhopper Penguins using this architecture. The training tiles for the penguin imagery were split smaller to increase visibility for labeling and the tiles had different degrees of overlap, so the training labels were generated separately for both species. Due to these differences, the training labels could not be combined to generate a singular model, although future studies could easily standardize these factors and create more comprehensive models.

The first CNN was trained with 945 tiles containing 12,431 Black-browed Albatrosses from half of the colony

areas. About 104 tiles containing 1,182 albatrosses were used for validation data and 116 tiles with 1,357 albatrosses were set aside for testing data. The second CNN was trained with 2,782 tiles containing 24,670 Southern Rockhopper Penguins from half the colony areas. 308 tiles with 2,610 penguins were used for validation, and 343 tiles with 2,720 penguins were set aside for testing data. The samples for training, validation, and testing were all from separate tiles without overlap. The training, validation, and testing tiles were x by y by 3 tensors (Red, Green, Blue) with the accompanying bounding boxes and labels classifying the examples as an albatross or penguin. All tiles in each dataset were fully labeled with all bird instances.

Both models were trained with a batch size of 2,500 steps, 30 epochs, and took around 3 hr each. We augmented training data via minor rotation, translation, shear, and scaling, as well as flipping in the x - and y directions. Epoch 30 was used to create the model for Black-browed Albatross detection. Epoch 21 was used to create the Southern Rockhopper Penguin detection model. Using the validation data, it was determined that beyond epoch 21 the model began overfitting—when a model memorizes the training data and is not able to generalize new data, which is seen in an increase in validation loss while training loss continues to decrease.

Model Evaluation and Deployment

To determine the performance of a model, the Intersection over Union (IoU), precision, and recall metrics of the testing dataset are used to generate a mean average precision (mAP). IoU is the intersection over the union of the manually created bounding box and the predicted bounding box from the model (Figure 4). The RetinaNet architecture defaults to an IoU threshold of 0.5, and this threshold is widely used in other studies of object detection and instance segmentation (He et al. 2017). In our analysis, we follow this practice, if the IoU is greater than 0.5, the object classification is considered a true positive and if it is less than 0.5, it is a false positive. If a manually created bounding box is present and the model does not detect it, then it is a false negative. Precision and recall are calculated using true positives, false positives, and false negatives, as seen in Equations (1) and (2). Precision is the probability of the predicted bounding box overlapping the ground truth bounding box with $\text{IoU} \geq 0.5$, and high precision means that most detections match ground truth objects. Recall measures the probability of a ground truth object being correctly detected, and high recall means that most ground truth objects were detected. The F1 score, seen in Equation (3), is the weighted average of precision and recall. To optimize the F1 score, the probability threshold for discarding detections was kept at 0.5 for both the albatross and penguin models. Average precision is calculated by sampling the precision-recall curve at all unique recall

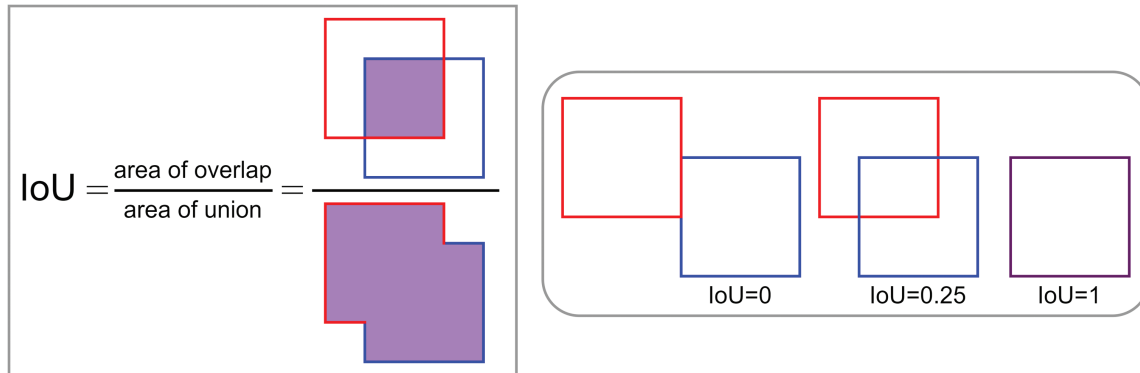


FIGURE 4. IoU metric used to determine overlap between 2 bounding boxes. IoU of 0 indicates no overlap and IoU of 1 indicates complete overlap.

values whenever the maximum precision value drops. The mAP is then calculated with the area under the precision-recall curve and uses the weighted average of precisions among classes.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{false positive}} \quad (1)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{false negative}} \quad (2)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}} \quad (3)$$

Due to memory constraints on modern graphics processing units (GPUs) detections were run on tiled subsets of the full orthomosaics. These tiles intentionally had a small overlap so that if birds were cut in half by the tiling process, the overlap ensured an uncut bird in one of the images. While it helps to prevent false negatives, this process often results in overlapping bounding boxes around the same bird. A post-processing function for non-max suppression, adapted from Malisiewicz et al. (2011), was applied to each colony area. This non-max suppression function removes duplicate detections based on overlap (IoU = 0.6) and keeps the detection with the highest model confidence score. As penguins were present in pairs, these bounding boxes had a higher degree of overlap. We examined a variety of IoU thresholds to ensure these detections were not removed, although it is likely that some penguin detections were eliminated by the suppression function. Detections were exported as shapefiles with georeferenced coordinates to be overlaid on each orthomosaic. They were reviewed in ArcGIS Pro using a fishnet of rectangular cells covering the extent of detections. Each grid was manually reviewed at a high zoom level (Figure 5).

RESULTS

Model Performance

The mAP for the albatross model was 97.66%, and the mAP for the penguin model was 87.16%. The F1 score for the albatross model at an IoU threshold of 0.5 and model confidence threshold of 0.5 was 0.9162. The F1 score for the same thresholds in the penguin model was 0.8450 (Figure 6).

There were 190,435 albatrosses detected in the Steeple Jason North and South 2018 imagery (colony areas 1–2), 69,989 albatrosses in the 2019 Steeple Jason subsets (colony areas 3–7), and 64,074 albatrosses in the 2019 Grand Jason imagery (colony areas 8–12) (Table 2). As 4 of the sites collected in 2019 at Steeple Jason are overlapping subsets of the 2018 imagery, a more representative count is achieved by summing the results of Steeple Jason 2018 and Steeple Jason South of the Neck (2019) counts. This resulted in 204,690 individual counts at Steeple Jason, and including all the Grand Jason colony areas, a total of 268,764 Black-browed Albatrosses. Although the colony areas on Steeple Jason appeared similar in 2018 and 2019, corroborated to some extent by comparing those colony areas that were photographed in both years, there is likely interannual variation in nest abundance so the summed colony-wide estimates should serve only as a baseline. Furthermore, in this study, we refer to each detected albatross as a “potential breeding individual,” recognizing that transforming individual counts as a proxy for population assessment would require further work including ground-truthing to determine breeding bird numbers from nonbreeding bird numbers as well as accounting for birds that are hidden from aerial view, which were not attempted in this study.

We were not able to photograph all Southern Rockhopper Penguin colony areas in sufficient detail to provide a total count; however, we were able to use CNN detection on those areas photographed in 2019 in which the detected number of penguins at Steeple Jason and Grand Jason was

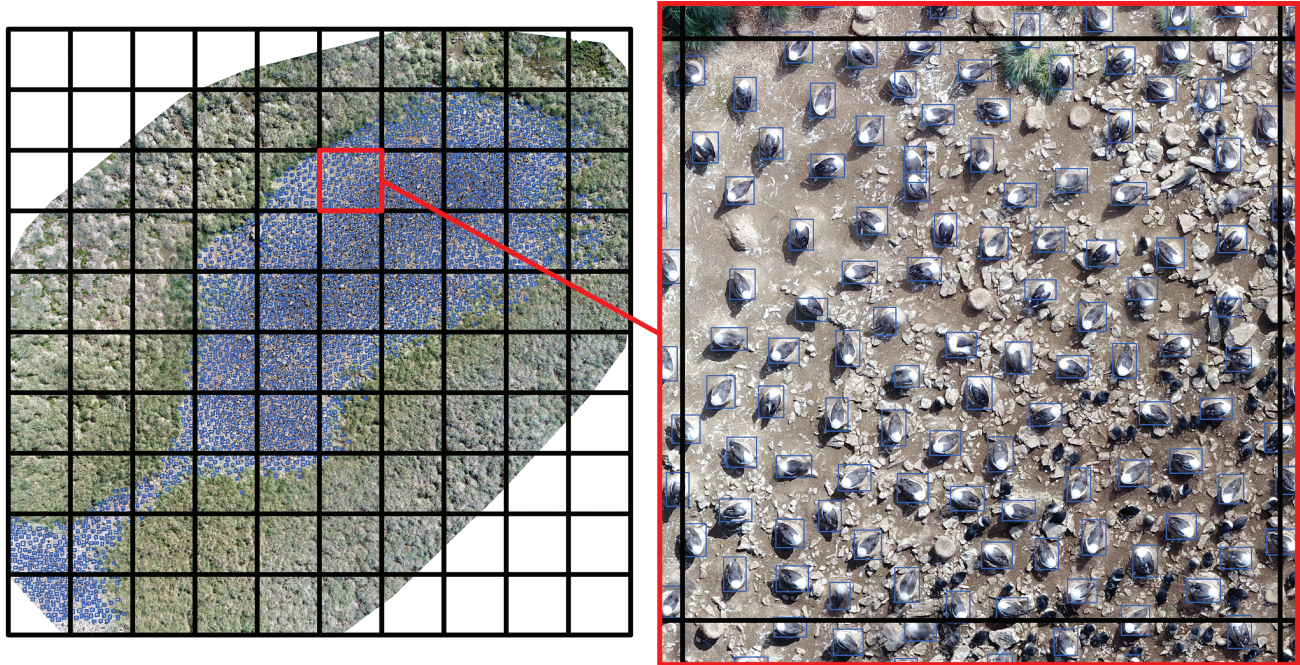


FIGURE 5. Example of visual review process for Steeple Jason “Bubble” area Black-browed Albatross detections at a zoom level of 1:75. Each grid is 10 × 10 m.

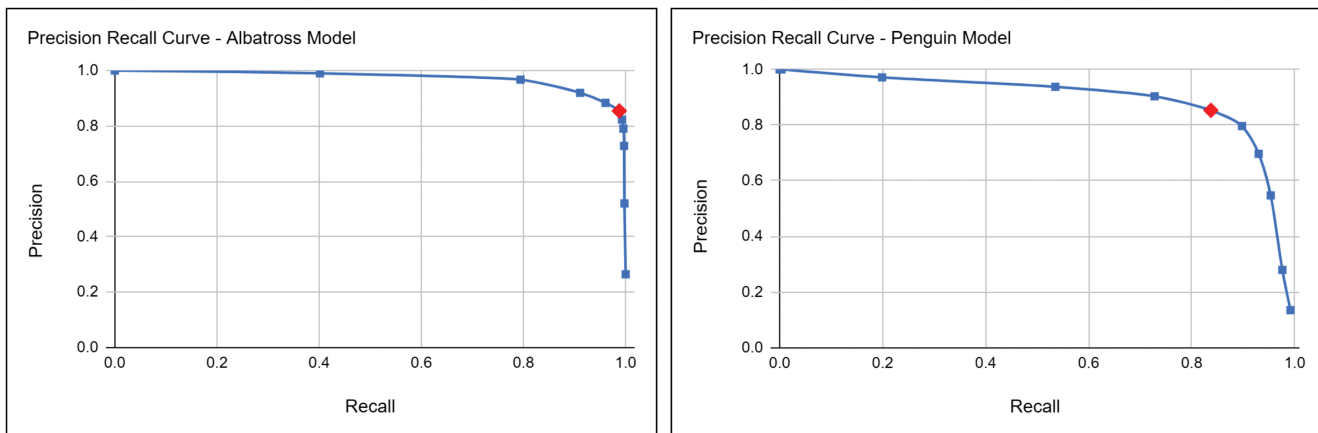


FIGURE 6. Precision recall curves for both albatross and penguin CNNs. Precision vs. recall is plotted as 11 different confidence thresholds from 0 to 1.0. The red diamond indicates the highest F1 score, which is seen at a confidence threshold of 0.5 for both models.

68,779 (Table 3). The Steeple Jason 2018 imagery was not used in these detections as the resolution was too low for the model. This figure is the total number of individual penguins detected.

Comparison with Manual Counts

Compared to the full manual counts from the imagery of Black-browed Albatrosses in Steeple Jason colony areas, there was less than a 5% difference with CNN detections and a 7.4% difference for Grand Jason SE Blob counts (Table 2). For the 2018 Steeple Jason colony areas, the CNN detections were clipped to the same colony area

boundary that was used for the density-based estimation for ease of comparison. The manual density-based estimation for Steeple Jason North had a 2.0% difference with the CNN counts, whereas the estimation for Steeple Jason South was a 9.4% difference (Table 2). Manual counts for Southern Rockhopper Penguins at Steeple Jason Hump, Bubble, and Blob and Grand Jason SE Blob and SW Middle Third were all less than a 5% difference with CNN counts (Table 3).

It took 9.5 hr of analyst time to perform manual counts of albatrosses in 0.043 km² of imagery, while the model took ~20 min to detect a similar number of albatrosses in

TABLE 2. Black-browed Albatross counts of individuals at all sites.

Colony area name	CNN detections	Manual counts	Percent difference
1. Steeple Jason North	133,075	135,774 ^a	2.0
2. Steeple Jason South	57,360	62,990 ^a	9.4
3. Steeple Jason Bubble	2,767	2,834	2.3
4. Steeple Jason Hump	6,419	6,673	3.8
5. Steeple Jason Blob	495	500	1.0
6. Steeple Jason N Tip Gully	46,053	n/a	n/a
7. Steeple Jason South Neck	14,255	n/a	n/a
8. Grand Jason SW Right Third	19,702	n/a	n/a
9. Grand Jason SW Middle Third	8,492	n/a	n/a
10. Grand Jason SW Left Third	11,707	n/a	n/a
11. Grand Jason SE Blob	3,707	3,988	7.4
12. Grand Jason SE Colony	20,466	n/a	n/a

^aDensity-based estimations.

TABLE 3. Southern Rockhopper Penguin counts of individuals.

Colony area name	CNN detections	Manual counts	Percent difference
3. Steeple Jason Bubble	2,015	2,069	2.6
4. Steeple Jason Hump	1,911	1,914	0.2
5. Steeple Jason Blob	1,459	1,483	1.6
6. Steeple Jason N Tip Gully	29,114	n/a	n/a
7. Steeple Jason South Neck	4,660	n/a	n/a
8. Grand Jason SW Right Third	1,781	n/a	n/a
9. Grand Jason SW Middle Third	2,516	2,476	1.6
10. Grand Jason SW Left Third	3,848	n/a	n/a
11. Grand Jason SE Blob	7,282	7,557	3.7
12. Grand Jason SE Colony	14,194	n/a	n/a

the same area. 10.5 hours of analyst time were required for manual counts of penguins in the same imagery and the model took ~50 min to run detections. Using these comparisons, both models cut down detection time to <10% of analyst time for individual counts. For the density-based method on the Steeple Jason 2018 imagery, it took ~50 hr of analyst time to estimate 198,764 albatrosses in 0.29 km² of imagery, while it took the albatross model 10 hr to detect 190,435 albatrosses in that area. Using the deep learning model reduced manual analyst time to 20% for density-based estimations.

Common Errors

Common errors made by the Black-browed Albatross model were typically false positives, where the model picked up on other seabirds, rocks, and shadows as albatross (Figure 7). Albatross were also missed across sites, but false positives were much more common. Common errors in the Southern Rockhopper Penguin model were also false positives, including rocks, shadows, and albatross. Due to their smaller size, feather markings that are harder to distinguish, and lack of regular nest configuration, Southern Rockhopper Penguins are far harder to distinguish than Black-browed Albatrosses, therefore the model missed penguins in many areas. Additionally, it is hard for the human eye to pick up on both species in the

lower resolution imagery, so this is likely also a limitation for the CNN.

DISCUSSION

This is one of the first studies to successfully use CNNs for seabird counts in drone imagery at large mixed colonies. Our method spanned a large area, was effective and accurate, achieving 97.66% and 87.16% accuracy detecting and counting Black-browed Albatrosses and Southern Rockhopper Penguins, respectively, on Steeple Jason and Grand Jason Islands, sites of global importance to both species. Many studies focus on CNN-based detection in individual images (Hodgson et al. 2018, Hong et al. 2019) or the use of other computer vision techniques for detection in orthomosaics (Rush et al. 2018, Lyons et al. 2019). This study builds on previous work to run a deep learning algorithm for object detection on large orthomosaics. This is unique and particularly enabling when investigating large colonies such as those found on the Falkland (Malvinas) Islands because it provides opportunities to link focal bird locations with broad habitat features, allowing researchers to explore patterns and processes of individual organisms and how they interact with the larger ecosystem. Many of these patterns could be missed when

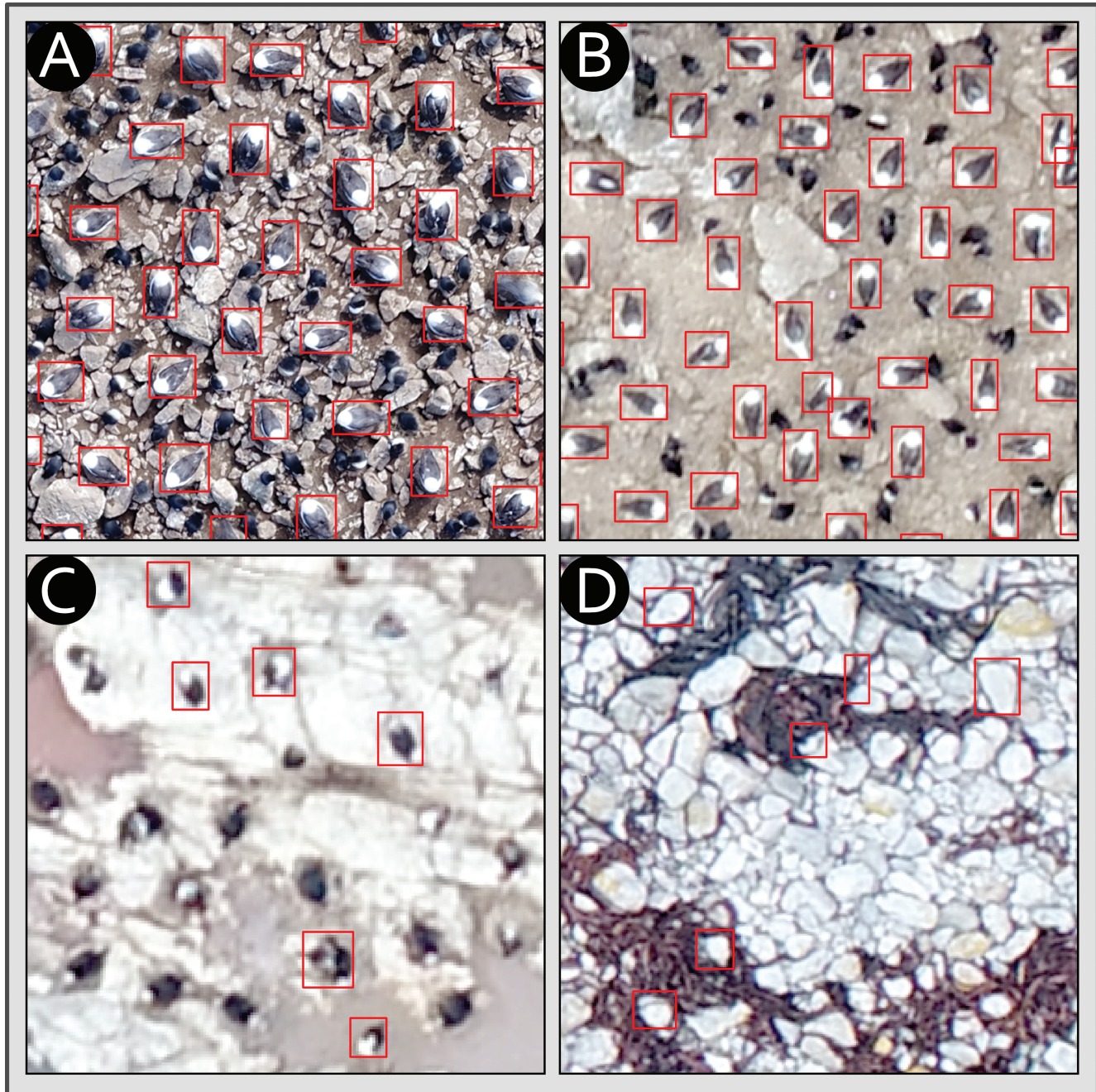


FIGURE 7. CNN generated detections of albatrosses overlaid on orthomosaics at (A) Steeple Jason Bubble (2019) and (B) Steeple Jason North (2018). Common false positives seen at Steeple Jason North are seen in (C) other seabirds and (D) rocks.

using a static density-based metric or when constrained to a small area.

The results of the present CNN analysis were within 5% of manual individual counts for both Black-browed Albatrosses and Southern Rockhopper Penguins for all but one site and all manual counts were higher than CNN counts except for one site. This may be partially due to the possibility of manually counting the same bird in different

tiles, as overlap was not considered while generating training data. Duplicate training bounding boxes could be removed with a non-max suppression function, and further research is required to assess how overlap may bias counts. The density-based Black-browed Albatross estimations based on manual counts at Steeple Jason North and South were both higher than CNN counts, which may be because the CNN counts individuals across the whole area

whereas the manual estimations counted individuals in smaller specific quadrats.

Our study provides important considerations for future flight planning for seabird colony surveys. The albatross model had difficulty in the Steeple Jason 2018 lower resolution imagery, suggesting that the 5 cm/pixel resolution may be approaching a detection threshold. The 2018 imagery was also not sufficient to detect penguins. The penguin model had difficulty even in the 1.80 cm pixel⁻¹ imagery, suggesting that this may be the maximum threshold for accurate detection using this approach. Generally, collecting imagery at a high resolution requires flying lower to the ground, which covers less area per flight but may be a necessary tradeoff for more accurate abundance estimation using this type of drone sensor combination. Flying lower to the ground may also introduce disturbance for more mobile species. In the highest resolution imagery at 0.5 cm pixel⁻¹, there were very few ghosting artifacts generated by the photogrammetry software. Ghosting artifacts occur when the animal moves during flight lines and the software has trouble stitching the object together. In the present study, the lack of ghosting artifacts in the highest resolution imagery of both albatrosses and penguins suggests that there was little disturbance and flying at this height did not significantly increase error. Ideally, to produce one single model that can accurately detect both species of birds, it is recommended to obtain imagery lower than 1.5 cm pixel⁻¹ resolution.

There are also key integration considerations to ensure data continuity. For both species, there is significant variability in the trajectory and magnitude of population change amongst annually monitored sites and sources of error in density-based ground counts come from natural variation in breeding density and sampling error in colony area measurement (Huin and Reid 2006, Baylis 2012). Adding sites to annual monitoring can reduce variability and the use of drones to estimate colony size has been proven to reduce cumulative variance when compared to traditional approaches. Yet, duplicate counts using the new drone count method and the previous methods will be required to determine the true ratio of the magnitude of traditional counts to drone counts (Hodgson et al. 2016). Once the ratios are determined with corresponding error metrics, the new automated drone counts can be compared to historical counts. The upfront cost to implement drone monitoring and automated counting is substantial and these methods may only be applicable to long-term monitoring of large, complex colonies where the cost tradeoff is deemed worth it.

It is also crucial to balance the risks of Type I (false positive) vs. Type II (false negative) errors within the deep learning method. Deep learning models can be tuned to internally filter fewer detections, outputting more low

confidence detections and minimizing Type II errors at the cost of more Type I. In the case of some endangered species, there is less concern in committing Type I errors because the objective is to count as many individuals as possible (Shrader-Frechette 1994). In situations where detections absolutely cannot be missed, but the amount of data to review is intractably large, models with low confidence thresholds followed by analyst verification could still save substantial manual time with few missed detections. The appropriate confidence threshold will be situation-dependent, but the benefit of deep learning models is seen in the ability to easily fine-tune these systems based on desired outcome.

The benefits of the methods presented here for improving long-term seabird monitoring could be substantial. Monitoring the world's largest Black-browed Albatross and second largest Southern Rockhopper Penguin colonies is crucial for determining thresholds of concern, and therefore impact management of both species. Breeding abundance and productivity can be intricately linked to the physical properties of the marine ecosystem, indicating changes in marine health (Diamond and Devlin 2003). Southern Rockhopper Penguins are particularly vulnerable to the increasingly stochastic marine environment, and mass mortality events may increase due to climate change (Baylis 2012, Dehnhard et al. 2013). With more frequent and accurate population monitoring researchers should be more likely to detect short-term and long-term variability, although further research is required to automatically count chicks and determine breeding success, determine nonbreeding from breeding individuals, and to determine the appropriate method to count breeding pairs of Southern Rockhopper Penguins instead of individuals.

The results of this study indicate that drone imagery coupled with deep learning is a viable tool for population monitoring at the Falkland Islands and beyond. Although there was a significant investment of time in development, both models can be applied to future drone imagery. Collecting drone imagery of the same sites across multiple years will allow for consistent comparisons and analysis of population trends and patterns at whole colony scales and provide details on the ecological context of colony areas and how they may be changing over time. Reducing the time spent manually counting seabirds may free up time for more complex research questions and patterns to be explored through both in situ sampling and remote sensing methods. While monitoring large, hard-to-access seabird colonies proves to be challenging, the push toward automated methods can significantly increase capabilities without sacrificing data quality and accuracy.

ACKNOWLEDGMENTS

We would like to thank Walter Sedgwick, Wildlife Conservation Society trustee, for guidance and support throughout the project. We also thank Rob and Lorraine McGill, local managers and administrators of Grand Jason and Steeple Jason islands, for support and logistics, Mike and Nikki McRae, captain and crew of the “Seaquest,” for transport, and Falklands Conservation for providing assistance with permits and research insights.

Funding statement: This research was funded with support from the Wildlife Conservation Society, the Island Foundation, and the Francis Goelet Charitable Trust, and Robert G. Goelet.

Ethics statement: Fieldwork was conducted under Research License No: R28/2019 issued by the Environmental Office of the Falkland Islands Government, with UAV flight permit certificate No: 19_OTO_P0538 issued by the Falkland Islands Civil Aviation Department.

Conflict of interest statement: The authors declare no conflicts of interest.

Author contributions: D.W.J., G.H., and M.C.H. conceived the ideas and designed the methodology; W.C.S., V.D.C., and G.H. led the expeditions and collected the data; M.C.H. and P.C.G. led software development; M.C.H. and N.C. analyzed the data; M.C.H. led the writing of the manuscript and S.C., P.C.G., G.H., and D.W.J. edited the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data availability: The open-source convolutional neural network code and all relevant imagery, including training, validation, and testing tiles and labels, is available on the Duke Research Data Repository: <https://doi.org/10.7924/r4dn45v9g> (Hayes et al. 2020).

LITERATURE CITED

- Abd-Elrahman, A., L. Pearlstine, and F. Percival (2005). Development of pattern recognition algorithm for automatic bird detection from unmanned aerial vehicle imagery. *Surveying and Land Information Science* 65:37–45.
- Akçay, H. G., B. Kabasakal, D. Aksu, N. Demir, M. Öz, and A. Erdoğan (2020). Automated bird counting with deep learning for regional bird distribution mapping. *Animals* 10:1207.
- Aniceto, A. S., M. Biuw, U. Lindstrøm, S. A. Solbø, F. Broms, and J. Carroll (2018). Monitoring marine mammals using unmanned aerial vehicles: Quantifying detection certainty. *Ecosphere* 9:e02122.
- Barbraud, C., and H. Weimerskirch (2001). Emperor Penguins and climate change. *Nature* 411:183–186.
- Baylis, A. M. M. (2012). 2010 Archipelago-Wide Census Gentoos and Rockhopper Penguins. Falklands Conservation, Stanley, Falkland Islands.
- Baylis, A. M. M., A. C. Wolfaardt, S. Crofts, P. A. Pistorius, and N. Ratcliffe (2013). Increasing trend in the number of Southern Rockhopper Penguins (*Eudyptes c. chrysocome*) breeding at the Falkland Islands. *Polar Biology* 36:1007–1018.
- Borowicz, A., P. McDowall, C. Youngflesh, T. Sayre-McCord, G. Clucas, R. Herman, S. Forrest, M. Rider, M. Schwaller, T. Hart, et al. (2018). Multi-modal survey of Adélie Penguin megacolonies reveals the Danger Islands as a seabird hotspot. *Scientific Reports* 8:3926.
- Bost, C. A., and Y. LeMaho (1993). Seabirds as bio-indicators of changing marine ecosystems: New perspectives. *Acta Oecologica* 14:463–470.
- Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. Proceedings of the 19th International Conference on Computational Statistics (COMPSTAT'2010), Paris, France, August 22–27, 177–186.
- Brodrick, P. G., A. B. Davies, and G. P. Asner (2019). Uncovering ecological patterns with convolutional neural networks. *Trends in Ecology & Evolution* 34:734–745.
- Chabot, D., S. R. Craik, and D. M. Bird (2015). Population census of a large Common Tern colony with a small unmanned aircraft. *PLoS One* 10:e0122588.
- Chabot, D., and C. M. Francis (2016). Computer-automated bird detection and counts in high-resolution aerial images: A review. *Journal of Field Ornithology* 87:343–359.
- Christin, S., É. Hervet, and N. Lecomte (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10:1632–1644.
- Crofts, S., and A. Stanworth (2017). Falkland Islands Seabird Monitoring Programme Annual Report 2016/2017 (SMP24). Falklands Conservation, Stanley, Falkland Islands.
- Crofts, S., and A. Stanworth (2019). Falkland Islands Seabird Monitoring Programme Annual Report 2018/2019 (SMP26). Falklands Conservation, Stanley, Falkland Islands.
- Croxall, J. P., and P. A. Prince (1979). Antarctic seabird and seal monitoring studies. *Polar Record* 19:573–595.
- Croxall, J. P., S. H. Butchart, B. Lascelles, A. J. Stattersfield, B. Sullivan, A. Symes, and P. Taylor (2012). Seabird conservation status, threats and priority actions: A global assessment. *Bird Conservation International* 22:1–34.
- Dehnhard, N., M. Poisbleau, L. Demongin, K. Ludynia, M. Lecoq, and J. F. Masello (2013). Survival of rockhopper penguins in times of global climate change. *Aquatic Conservation: Marine and Freshwater Ecosystems* 23:777–789.
- Descamps, S., A. Béchet, X. Descombes, A. Arnaud, and J. Zerubia (2011). An automatic counter for aerial images of aggregations of large birds. *Bird Study* 58:302–308.
- Diamond, A. W., and C. M. Devlin (2003). Seabirds as indicators of changes in marine ecosystems: Ecological monitoring on Machias Seal Island. *Environmental Monitoring and Assessment* 88:153–175.
- Gray, P. C., K. C. Bierlich, S. A. Mantell, A. S. Friedlaender, J. A. Goldbodem, and D. W. Johnston (2019). Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods in Ecology and Evolution* 10:1490–1500.
- Gray, P. C., A. B. Fleishman, D. J. Klein, M. W. McKown, V. S. Bézy, K. J. Lohmann, and D. W. Johnston (2018). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution* 10:345–355.
- Groom, G., I. K. Petersen, M. D. Anderson, and A. D. Fox (2011). Using object-based analysis of image data to count birds: Mapping of Lesser Flamingos at Kamfers Dam, Northern Cape, South Africa. *International Journal of Remote Sensing* 32:4611–4639.

- Hayes, M. C., P. C. Gray, G. Harris, W. C. Sedgwick, V. D. Crawford, N. Chazal, S. Crofts, and D. W. Johnston (2020). Data from: Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. Duke Research Repository. doi:10.7924/r4dn45v9g
- Hazen, E. L., B. Abrahams, S. Brodie, G. Carroll, M. G. Jacox, M. S. Savoca, K. L. Scales, W. J. Sydeman, and S. J. Bograd (2019). Marine top predators as climate and ecosystem sentinels. *Frontiers in Ecology and the Environment* 17:565–574.
- He, K., G. Gkioxari, P. Dollár, and R. Girshick (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22–29, 2980–2988.
- Hodgson, J. C., S. M. Baylis, R. Mott, A. Herrod, and R. H. Clarke (2016). Precision wildlife monitoring using unmanned aerial vehicles. *Scientific Reports* 6:22574.
- Hodgson, A., N. Kelly, and D. Peel (2013). Unmanned aerial vehicles (UAVs) for surveying marine fauna: A dugong case study. *PLoS One* 8:e79556.
- Hodgson, J. C., R. Mott, S. M. Baylis, T. T. Pham, S. Wotherspoon, A. D. Kilpatrick, R. Raja Segaran, I. Reid, A. Terauds, and L. P. Koh (2018). Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution* 9:1160–1167.
- Hong, S.-J., Y. Han, S.-Y. Kim, A.-Y. Lee, and G. Kim (2019). Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors* 19:1651.
- Huang, J., V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21–26, 3296–3297.
- Huin, N., and T. Reid (2006). Census of the Black-Browed Albatross Population of the Falkland Islands 2000 and 2005. Falklands Conservation, Stanley, Falkland Islands.
- Johnston, D. W. (2019). Unoccupied aircraft systems in marine science and conservation. *Annual Review of Marine Science* 11:439–463.
- Kerner, H. R., K. L. Wagstaff, B. D. Bue, P. C. Gray, J. F. Bell, and H. B. Amor (2019). Toward generalized change detection on planetary surfaces with convolutional autoencoders and transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12:3900–3918.
- LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep learning. *Nature* 521:436–444.
- Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie (2017b). Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21–26, 936–944.
- Lin, T.-Y., P. Goyal, R. Girshick, K. He, and P. Dollár (2017a). Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22–29, 2999–3007.
- Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Lawrence Zitnick (2014). Microsoft COCO: Common objects in context. *Computer Vision – ECCV* (13th European Conference on Computer Vision) 2014, Zurich, Switzerland, September 6–12, 740–755.
- Linchant, J., J. Lisein, J. Semeki, P. Lejeune, and C. Vermeulen (2015). Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges: A review of UASs in wildlife monitoring. *Mammal Review* 45:239–252.
- Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg (2016). SSD: Single shot multibox detector. *Computer Vision – ECCV* (14th European Conference on Computer Vision) 2016, Amsterdam, the Netherlands, October 11–14, 9905:21–37.
- Lyons, M. B., K. J. Brandis, N. J. Murray, J. H. Wilshire, J. A. McCann, R. T. Kingsford, and C. T. Callaghan (2019). Monitoring large and complex wildlife aggregations with drones. *Methods in Ecology and Evolution* 10:1024–1035.
- Malisiewicz, T., A. Gupta, and A. A. Efros (2011). Ensemble of exemplar-SVMs for object detection and beyond. 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, November 6–13, 89–96.
- Moreno, C. A., R. Castro, L. J. Mujica, and P. Reyes (2008). Significant conservation benefits obtained from the use of a new fishing gear in the Chilean Patagonian Toothfish Fishery. *CCAMLR Science* 15:79–91.
- Morghenthaler, A., E. Frere, A. Raya Rey, C. Torlaschi, P. V. Cedrola, E. Tiberi, R. Lopez, E. Mendieta, M. L. Carranza, S. Acardi, et al. (2018). Unusual number of Southern Rockhopper Penguins, *Eudyptes chrysocome*, molting and dying along the Southern Patagonian coast of Argentina: Pre-molting dispersion event related to adverse oceanographic conditions? *Polar Biology* 41:1041–1047.
- Pütz, K., J. G. Smith, R. J. Ingham, and B. H. Lüthi (2002). Winter dispersal of Rockhopper Penguins *Eudyptes chrysocome* from the Falkland Islands and its implications for conservation. *Marine Ecology Progress Series* 240:273–284.
- Ratcliffe, N., D. Guihen, J. Robst, S. Crofts, A. Stanworth, and P. Enderlein (2015). A protocol for the aerial survey of penguin colonies using UAVs. *Journal of Unmanned Vehicle Systems* 3:95–101.
- Razavian, A. S., H. Azizpour, J. Sullivan, and S. Carlsson (2014). CNN features off-the-shelf: An astounding baseline for recognition. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, June 23–28, 512–519.
- Redmon, J., S. Divvala, R. Girshick, and A. Farhadi (2016). You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27–30, 779–788.
- Ren, S., K. He, R. Girshick, and J. Sun (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39:11137–1149.
- Rush, G. P., L. E. Clarke, M. Stone, and M. J. Wood (2018). Can drones count gulls? Minimal disturbance and semiautomated image processing with an unmanned aerial vehicle for colony-nesting seabirds. *Ecology and Evolution* 8:12322–12334.
- Seymour, A. C., J. Dale, O. Hammill, P. N. Halpin, and D. W. Johnston (2017). Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Scientific Reports* 7:45127.
- Shrader-Frechette, K. (1994). *Ethics of Scientific Research*. Rowman & Littlefield Publishers, Lanham, MD, USA.
- Sun, C., A. Shrivastava, S. Singh, and A. Gupta (2017). Revisiting unreasonable effectiveness of data in deep learning era. 2017 Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, October 22–29, 843–852.
- Thompson, K. R., and P. Rothery (1991). A census of the Black-browed Albatross *Diomedea melanophrys* population on

- Steeple Jason Island, Falkland Islands. *Biological Conservation* 56:39–48.
- Weimerskirch, H., P. Inchausti, C. Guinet, and C. Barbraud (2003). Trends in bird and seal populations as indicators of a system shift in the Southern Ocean. *Antarctic Science* 15:249–256.
- Wolfaardt, A. (2012). An Assessment of the Population Trends and Conservation Status of Black-Browed Albatrosses in the Falkland Islands. JNCC, Peterborough, UK.
- Zacharias, M. A., and J. C. Roff (2001). Use of focal species in marine conservation and management: A review and critique. *Aquatic Conservation: Marine and Freshwater Ecosystems* 11:59–76.
- Zlocha, M., Q. Dou, and B. Glocker (2019). Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels. 22nd International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI 2019, Shenzhen, China, October 13–17, 402–410.