Received: 10 January 2019 | Accepted: 14 June 2019

DOI: 10.1111/2041-210X.13246

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry

Patrick C. Gray[1] | Kevin C. Bierlich[1] | Sydney A. Mantell[2] | Ari S. Friedlaender[3] | Jeremy A. Goldbogen[4] | David W. Johnston[1]

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

[2]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

[3]Institute of Marine Sciences, Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California

[4]Department of Biology, Hopkins Marine Station, Stanford University, Monterey, California

Correspondence
Patrick Gray
Email: patrick.c.gray@duke.edu

Funding information
Division of Integrative Organismal Systems, Grant/Award Number: 1656676; Division of Antarctic Sciences, Grant/Award Number: 0823101 and 1440435; Office of Polar Programs, Grant/Award Number: 1644209; North Carolina Space Grant, Grant/Award Number: 1440435,; Stanford University; Microsoft

Handling Editor: Hao Ye

Abstract

1. The flourishing application of drones within marine science provides more opportunity to conduct photogrammetric studies on large and varied populations of many different species. While these new platforms are increasing the size and availability of imagery datasets, established photogrammetry methods require considerable manual input, allowing individual bias in techniques to influence measurements, increasing error and magnifying the time required to apply these techniques.
2. Here, we introduce the next generation of photogrammetry methods utilizing a convolutional neural network to demonstrate the potential of a deep learning-based photogrammetry system for automatic species identification and measurement. We then present the same data analysed using conventional techniques to validate our automatic methods.
3. Our results compare favorably across both techniques, correctly predicting whale species with 98% accuracy (57/58) for humpback whales, minke whales, and blue whales. Ninety percent of automated length measurements were within 5% of manual measurements, providing sufficient resolution to inform morphometric studies and establish size classes of whales automatically.
4. The results of this study indicate that deep learning techniques applied to survey programs that collect large archives of imagery may help researchers and managers move quickly past analytical bottlenecks and provide more time for abundance estimation, distributional research, and ecological assessments.

KEYWORDS
cetaceans, convolutional neural network, deep learning, drones, photogrammetry, population assessments, species identification, unoccupied aerial systems

1 | INTRODUCTION

1.1 | Background on photogrammetry

Accurately measuring the size of animals is essential for wildlife conservation and management, as size can indicate important aspects of life history, such as reproductive status, growth rate, energetic requirements, phenotypic differences between species and populations, and incidents of compromised health related to injury or anthropogenic influences (Blanckenhorn, 2004; Blueweiss et al., 1978; Perryman & Lynn, 1993; Schmidt-Nielsen, 1975). As such, accurate and current morphometric data can help establish the status of, and

support monitoring programs for populations that may be influenced by dynamic environmental factors (Burnett et al., 2018). However, obtaining manual measurements of wild animal populations is logistically challenging, as accessibility can be limited, costly, dangerous, and disruptive to the animal (Gaudioso et al., 2014; Trimble et al., 2011).

Photogrammetry is a non-invasive method for obtaining accurate measurements of animals from photographs. The two main types of photogrammetry methods used in wildlife biology are (a) single camera, where a known scale factor is applied to a single image to measure 2D distances and angles and (b) stereo-photogrammetry, where two or more are used to recreate a 3D model (Gaudioso et al., 2014). These techniques have been used to measure body condition and weight of lactating Mediterranean buffaloes (Negretti, Bianconi, Bartocci, Terramoccia, & Verna, 2008), sexual dimorphism in Western gorillas (Breuer, Robbins, & Boesch, 2007), shoulder heights of elephants (Shrader, Ferreira, & van Aarde, 2006), nutritional status of Japanese macaques (Kurita, Suzumura, Kanchi, & Hamada, 2012), and mass of Weddell seals (Ireland, Garrott, Rotella, & Banfield, 2006).

Aerial photogrammetry has been particularly useful in studying cetaceans, as they spend most of their life beneath the surface and usually reside at great distances from humans (Johnston, 2019). Prior to this technique, measurements of cetaceans were traditionally limited to assessing carcasses collected via whaling (e.g. Ichii, Shinohara, Fujise, Nishiwaki, & Matsuoka, 1998), from bycatch (Read, 1990), or from strandings (Garrigue et al., 2016). Aerial photogrammetric techniques were later developed to calculate morphometric measurements in addition to total length, such as allometric growth and dorsal width measurements to estimate changes in nutritive condition (Ratnaswamy & Winn, 1993). However, occupied aircraft for photogrammetry can be expensive (Arona, Dale, Heaslip, Hammill, & Johnston, 2018), can limit the number of sampling days (Cosens & Blouw, 2003), and present risks for wildlife biologists (Sasse, 2003).

Over the past decade, the affordability and accessibility of small unoccupied aircraft systems (UAS, aka drones) has rapidly increased. These systems are increasingly being used in projects that span the full spectrum of marine science and conservation applications (Johnston, 2019), and are now being used in photogrammetric studies on a variety of odontocete and mysticete cetaceans across polar, temperate, and tropical biomes (Durban, Fearnbach, Perryman, & Leroi, 2015; Christiansen, Dujon, Sprogis, Arnould, & Bejder, 2016). UAS have improved aerial photogrammetry, as they often provide data of similar, if not better, quality than traditional methods (Johnston et al., 2017), and are better suited for ephemeral interactions with marine species that live far from regions that can provide aerial imaging support via occupied aircraft.

1.2 | Deep learning and computer vision in ecological analysis

The growth of ecological digital imagery, ranging from crowdsourced repositories such as iNaturalist (Van Horn et al., 2017), to long duration camera trap studies (Burton et al., 2015), to UAS-based surveys (Johnston, 2019), is providing insight into biological diversity and facilitating a new wave of ecological monitoring. But the expertise and time required to analyse this imagery represents a major bottleneck. Fortunately, this wealth of imagery, along with expert annotation, provides the foundation for accessing the power of modern machine learning algorithms. Large datasets combined with increasing computing power and advancements in artificial intelligence are allowing the technical specialist to automate a wide range of previously labor-intensive ecological analyses. Tasks thought infeasible to automate only a few years ago: animal detection, species identification, and even photogrammetry are now within reach of the ecological research community (Wäldchen & Mäder, 2018; Weinstein, 2017).

Computer vision is an interdisciplinary field concerned with extracting insight from digital images and video, often automating human tasks such as reading handwritten text, identifying individual faces, or informing an autonomous car of its current surroundings for navigation. *Machine learning* (ML) is a subfield of artificial intelligence that uses statistical techniques to "teach" a computer how to do a task by showing it numerous examples of that task being done correctly. Conventional ML approaches to computer vision require considerable image preprocessing and data transformations in a time-consuming process called *feature extraction*.

Deep learning is a subfield of ML that uses *neural networks* to automate feature extraction, permitting raw data to be input into a computer and creating high-level abstractions to inform decisions in classification, object detection, or other problems (Lecun, Bengio, & Hinton, 2015). The majority of recent advances in computer vision and object detection have been made with *convolutional neural networks (CNNs)* (He, Zhang, Ren, & Sun, 2016; Krizhevsky, Sutskever, & Hinton, 2012; Long, Shelhamer, & Darrell, 2015). CNNs ingest data in multidimensional arrays (e.g. 1D: text sequences; 2D: imagery or audio; 3D: video) and scan these arrays with a series of windows that transform the raw data into higher level features that represent the original input data through multiple layers of increasing abstraction.

CNN applications within ecology are becoming widespread, including the rapid development of species identification tools (Wäldchen & Mäder, 2018). For example, Norouzzadeha et al. (2017) were able to identify 48 different animal species from camera traps in the 3.2 million image Snapshot Serengeti dataset with 93.8% accuracy, similar to the accuracy of crowdsourced identifications, saving nearly 8.4 years of human labelling effort. Borowicz et al., (2018) successfully counted Adélie penguins in UAS imagery, finding their CNN-based results within 10% of manual counts and requiring manual analysis of only 0.18% of the total area surveyed. More recently, Gray et al. (2018) used a CNN to detect and enumerate olive ridley turtles in the nearshore waters of Ostional, Costa Rica, identifying 8% more turtles in imagery than manual methods with a 66-fold reduction in analyst time.

Applications of deep learning in cetacean studies are few, primarily hindered by small datasets. Several researchers applied CNNs to automate the process of identifying individual right whales

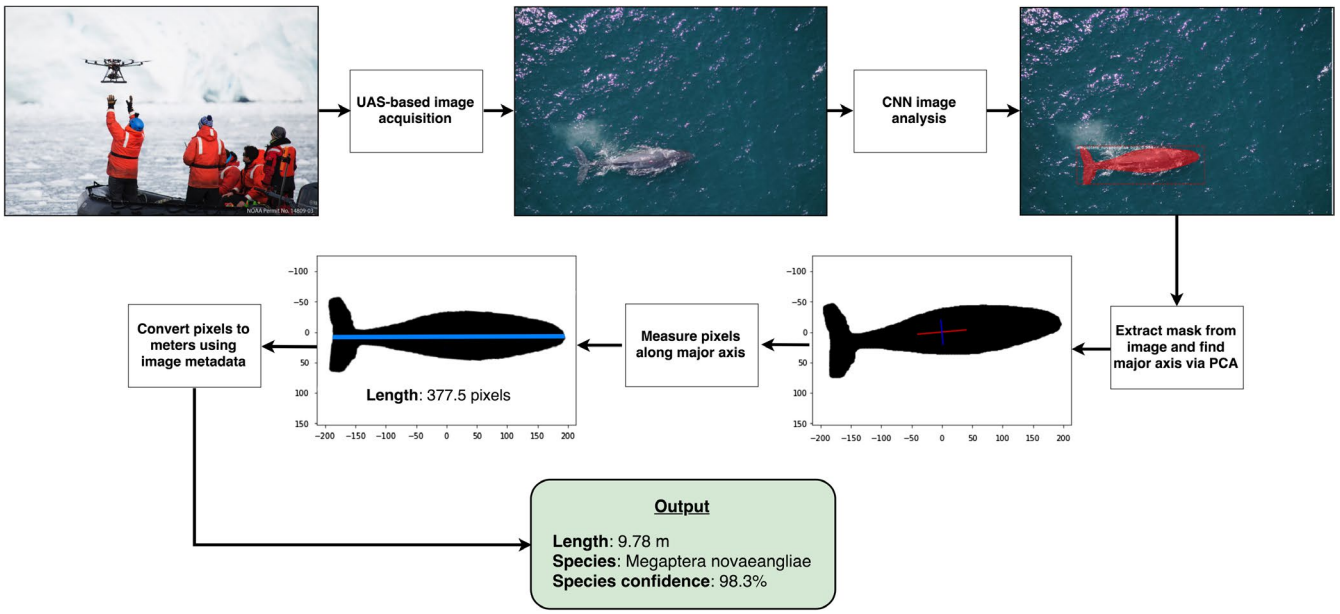

FIGURE 1 Automated photogrammetry workflow. (1) Unoccupied aircraft system (UAS) operations were conducted to collect imagery. (2) A convolutional neural network (CNN) for object detection and instance segmentation was applied to the image, generating a mask outlining the cetacean of interest and a species prediction with a prediction confidence score. (3) Principal component analysis (PCA) was applied to the mask to find its major axis. (4) The cetacean was measured along this axis, obtaining a length in pixels from the anterior tip of the lower jaw to fluke notch and (5) this measurement was converted to meters using the focal length, sensor size, pixel resolution, and altitude. (6) This derived length is then combined with the species and confidence score as the final workflow outputs

Eubalaena glacialis from aerial photos in an online contest (Bogucki et al., 2018), with the most successful group able to identify individual right whales among a population of 447 with 87% accuracy. Despite their promise, the complexity of implementing deep learning approaches, including the burden of obtaining and labelling training datasets, has slowed their widespread adoption in ecology. Particularly for wide ranging ocean species such as cetaceans, common solutions for large scale data collection, such as crowdsourcing or using camera traps, are not feasible.

The purpose of the present study is to leverage the power of CNNs, and the growing capabilities of UAS, to automate species identification and length estimation of whales through the analysis of aerial imagery. Implementing and customizing the Mask R-CNN architecture (He, Gkioxari, Dollar, & Girshick, 2017) we take advantage of transfer learning (Razavian, Azizpour, Sullivan, & Carlsson, 2014) to analyse whale image datasets of moderate size (N = 384) collected in challenging and variable conditions. Transfer learning is the concept that features learned by a neural network for one task can be useful for another unrelated task (Yosinski, Clune, Bengio, & Lipson, 2014) and here we take advantage of this aspect of CNNs, pretraining our model on the 300,000 + image COCO dataset (Lin et al., 2014) which does not contain any whales, yet helps the CNN develop a general ability to classify imagery. Mask R-CNN is a large model and would overfit drastically on a dataset of only 265 images, but leveraging transfer learning we were able to effectively apply our model to new data. Specifically, we trained a CNN to identify humpback whales *Megaptera novaeangliae*, minke whales *Balaenoptera bonaerensis*, and blue whales *Balaenoptera musculus* in UAS imagery and then output a mask of each animal from which we derive length (Figure 1). These CNN-based measurements are then compared with manual UAS-based photogrammetric measurements.

2 | MATERIALS AND METHODS

2.1 | UAS flights and data collection

We collected UAS aerial images of blue and humpback whales off the coast of Santa Barbara and Monterey, California between August – September 2017, and humpback and minke whales along the Western Antarctic Peninsula (WAP) in March 2018. We used two types of UAS hexacopters, the FreeFly Alta 6 (https://freeflysystems.com/alta-6) for data collection in California, and an in-house hexacopter, LemHex-44 (https://sites.nicholas.duke.edu/uas/multirotor-platforms/), for data collection along the WAP. The Alta 6 has a flight time ~20–25 min, while the LemHex-44 has a flight time ~10–15 min. Both aircraft were fitted with a Sony a5100 camera with a 50 mm focal length lens, 23.5 × 15.6 mm sensor size, and 6,000 × 4,000 pixel resolution, as well as a Lightware SF11/C laser altimeter that calculates altitude more accurately than the onboard barometer. The altitude was divided by the focal length of the camera to set the ground sampling distance, or scale, of each photo (see section 2.4). We used similar methods for hand launch and recovery from small boats as described in Durban et al., (2015), with the addition of a first person view screen attached to the flight controller, enabling the pilot to frame the whale and then manually trigger the shutter with a remote connection to collect images. The camera and

laser altimeter were mounted on a gimbal and images were collected at nadir with the animal full frame lengthwise. Images were collected at altitudes between 30–80 m above sea level and at speeds between 0 and 3 m/s to maintain the whale in full frame. Images were collected in bursts as the whale surfaced or was just below the surface.

2.2 | Convolutional neural network architecture

For accurate and automated photogrammetry, with the potential for multiple animals in each image, the CNN must successfully complete instance segmentation – the task of mapping each pixel in an image to a particular class and separating each instance of that class. This allows identification of individual whales within a single image and generates a pixel mask for each. For more detailed information on general CNN architecture and theory, see Lecun et al. (2015). In this study we implemented Mask R-CNN (He et al., 2017) which builds on the foundational work of Faster R-CNN (Ren, He, Girshick, & Sun, 2017) and is capable of ingesting an image and outputting a bounding box around each object of interest, a class for each object (e.g. whale species), and a full pixel mask of the object within each bounding box.

The first layer of a CNN typically creates maps of features such as edges, curves, and color gradients. The feature maps created in deeper layers in a CNN are more abstract and aggregate the previous layer's feature maps, creating combinations of the simple features from the previous layer that in our case may indicate pectoral flippers, flukes, or particular body shapes. Through this process, the CNN extracts the distinguishing features that will permit effective classification (Figure 2b). Typically, a fully connected layer takes the final feature maps, representing useful and high-level image components, and learns a mapping from those feature maps to the output classes.

Faster R-CNN (Ren et al., 2017), the predecessor to our implementation in this study, builds on this typical CNN structure by adding a Region Proposal Network (RPN) at the final feature map step (Figure 2c). This RPN passes a sliding window over the feature maps and generates many bounding box guesses, along with a score estimating how likely the bounding box contains an object that is in a class of interest. The four corners of these proposed regions are then passed to the fully connected layers where they are fine-tuned, and the bounding box is classified (Figure 2d). This architecture, an RPN sitting directly on the CNN feature map, has led to one of the most successful object detection algorithms (Huang et al., 2017).

Mask R-CNN builds on this with a straightforward, yet breakthrough, step for instance segmentation by adding another branch (Figure 2e) that ingests the CNN feature map and runs in parallel to the classification and bounding box fine tuning. This new branch outputs a mask of 1s and 0s for each region proposed by the RPN, indicating object (1) or non-object (0). As its final step these three outputs are combined, resulting in full instance segmentation (Figure 2f). For this study we used Mask R-CNN with ResNet101 (He et al., 2016) as the feature extracting CNN (Figure 2a).

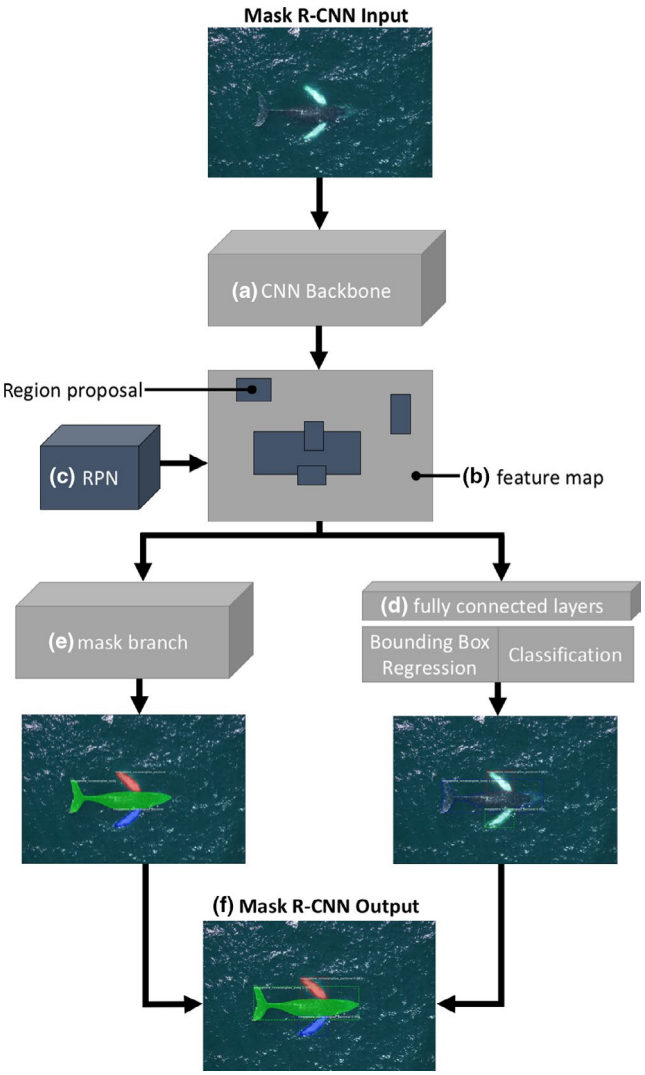

FIGURE 2 Overview of the Mask R-CNN structure. (a) An image is input first into a standard convolutional neural network (CNN) in order to extract meaningful features from the image. This CNN consists of a series of convolutional layers and max pooling layers. The initial layers typically create maps of features such as edges and curves. Feature maps created in deeper layers are more abstract, creating combinations of features (e.g. pectoral flippers, flukes). Through this process, the CNN extracts the distinguishing features that will permit effective classification. This process creates a set of (b) feature maps. (c) A region proposal network (RPN) proposes various bounding boxes from these feature maps that may contain objects of interest and these bounding boxes are passed in parallel to (d) a series of fully connected layers which refines the bounding box corners and makes a class prediction and (e) the mask branch which decides precisely which pixels in the bounding box belong to the object of interest. This leads to (f) predictions for a bounding box, a class, and a mask for pixel-wise segmentation. In this example orange represents the animal's left flipper, blue the right flipper, and green represents the main body and fluke

2.3 | Convolutional neural network training and validation

This CNN was trained on 326 images total, split evenly across species, with 265 for training and 61 for validation (Figure 3).

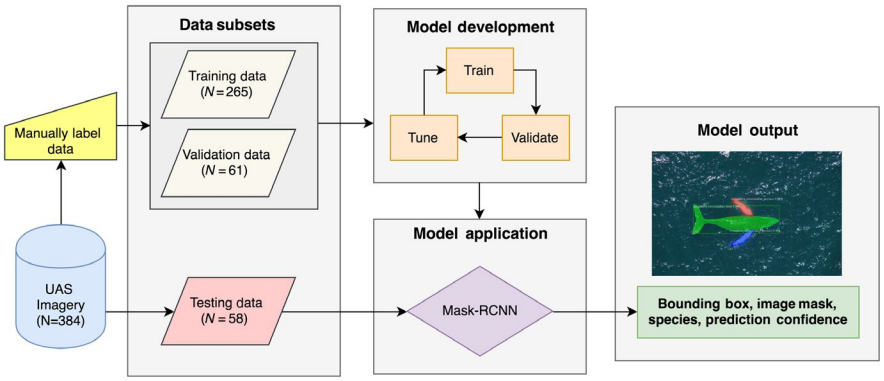

FIGURE 3 Overview of Mask R-CNN model development, application and output as applied to unoccupied aircraft system (UAS) imagery of cetaceans

Fifty-eight additional images (17 blue whale, 30 humpback whale, and 11 minke whale), all of different individuals, were then used for testing species ID and comparing the automated measurement method versus manual methods on this previously unseen data. Using a species ID label and masks manually drawn to delineate the body and pectorals in each image, the CNN was trained to optimize the performance of the model on a multi-task loss function minimizing the errors in RPN class, RPN bounding box, final class, final bounding box, and final mask with respective ratios of 1, 2, 2, 2, 5 in order to prioritize the mask itself above all other components. This loss function was optimized using stochastic gradient descent (SGD), with a learning rate of 0.001, momentum of 0.9, and weight decay of 0.001. Training was initialized with the Microsoft COCO dataset (Lin et al., 2014), after which training of the heads with the rest of the network frozen was done for 69 epochs, followed by training of ResNet101 layers 5 and up for 16 epochs, followed by training of ResNet101 layers 3 and up for 61 epochs, and finally training of the entire model for 34 epochs. Model parameters were selected from epoch 174 which presented the lowest loss on the validation set of $N = 61$. Training took 2 days using an NVIDIA Tesla K80 with 12GB of memory.

2.4 | Measurement method

2.4.1 | Manual

UAS images were manually selected to measure the total length of individual whales if the animal appeared straight with minimal curvature, was at the surface or just below, and the outer edge of the lower jaw and fluke notch were both clearly visible. All manual measurements were performed using ImageJ 1.5i. The segmented or straight-line tool was used to draw a line from the tip of the lower jaw to the fluke notch to measure the distance in number of pixels. The total length of the whale was then calculated using similar methods as Fearnbach, Durban, Ellifrit, and Balcomb (2011), by multiplying the number of pixels and the ground sampling distance (GSD) (Equation 1). GSD was calculated using Equation 2.

$$\text{Total Length } (m) = \#\text{pixels} * \text{GSD} \quad (1)$$

$$\text{GSD} = (a/f) * (\text{Sw}/\text{Pw}) \quad (2)$$

where a = altitude (m), f = focal length (mm), and Sw = width of sensor size (mm), and Pw = the width of the image resolution in pixels. The width was used for the sensor size and image resolution because the whale was captured full frame widthwise.

In addition, masks were manually drawn to delineate whales for training the CNN and used to compare measurements based on manual masks to measurements based on masks predicted by the CNN. All manual masks were drawn using the VGG VIA software (http://www.robots.ox.ac.uk/~vgg/software/via/) (Figure S1).

2.4.2 | Automated

CNN-based measurements were generated using a five-step workflow (Figure 1). Following UAS operations the trained Mask R-CNN model was run on the image, generating a mask delineating the cetacean of interest (Figure 4), a species prediction, and a confidence score from 0.0 to 1.0 in the species prediction. Principal component analysis (PCA) was conducted on this mask to find the major axis (first eigenvector) in order to measure the full length. The mask was measured along that axis, obtaining a length in pixels from lower jaw to fluke notch. This measurement was converted to meters (Equations 1, and 2) resulting in a length along with the species and prediction confidence score.

3 | RESULTS

3.1 | Measurements

Measurement results were similar between CNN and manual methods (Figure 5). A comparison of CNN-based length versus conventional manually measured length for all three species is presented in Figure 6a. There is a mean difference of 0.31 m between CNN-based lengths and manual lengths across all three species. Two comparisons were identified as extreme outliers for the model, one of a blue whale and one of a humpback whale. The outlier blue whale comparison was conducted on mask derived from a picture that contains a boat approaching a whale to tag it (Figure 6b). The humpback outlier

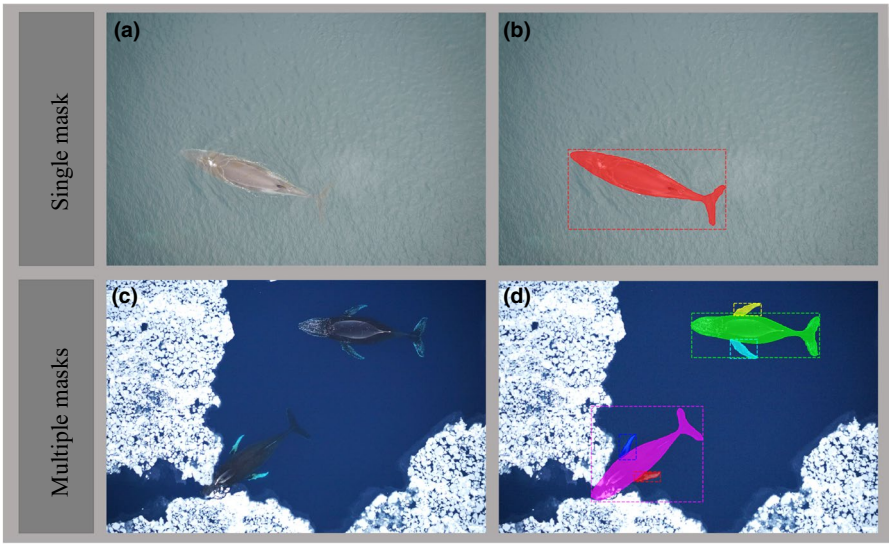

FIGURE 4 Examples of mask and bounding box outputs generated from the trained Mask R-CNN model. (a) An unoccupied aircraft system (UAS) image of a minke whale and its (b) single mask output and (c) a UAS image of two humpback whales and its (d) multiple mask output

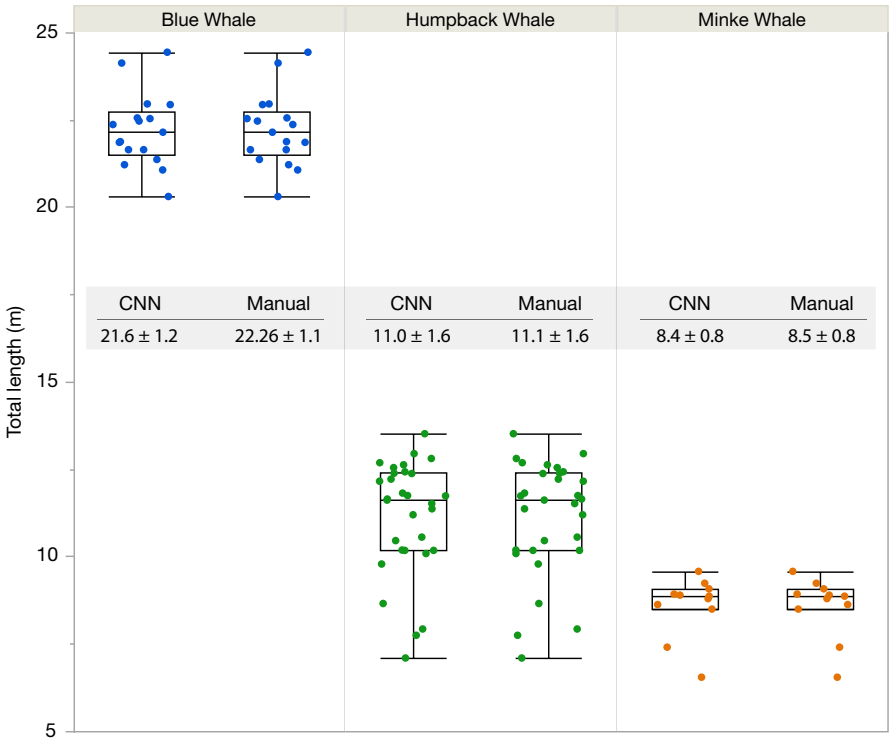

FIGURE 5 Overall results for convolutional neural network (CNN)-based measurements are similar to manual methods. In addition to the box plots showing the distribution of lengths, the center data bar presents mean measurements and standard deviation in meters. Minke whales had the least difference in mean total length between both methods, followed by humpbacks, and then blues. Humpbacks had the widest size range compared to the other baleen whales

stems from mask derived from a photo that has a whale producing a large blow that obscures part of the back of the animal (Figure 6c). Figure 7 shows the distribution of differences between the CNN-based measurements and the manual lengths, and while blue whales appear to have a much larger difference it is worth noting that proportionally the differences are similar, −2.8% for blue whales, −1.5% for humpback whales, and −2.4% for minke whales (Figure 6d).

3.2 | Species identification

The CNN correctly predicted whale species in 57 of 58 (98%) images and 95% of automated measurements were within 5% of manual measurements with a maximum difference in one example of 13%. All species prediction confidence scores output from the CNN were above 80%, except the one misclassified individual which had a prediction confidence score of 63%.

3.3 | Comparison to manual masks

Comparisons between manual masks and CNN-based masks were assessed through an intersection over union (IoU) approach. This approach is a common evaluation metric for object detection and instance segmentation and reports the area of overlap between two masks divided by their area of union (He et al., 2017). Intersection

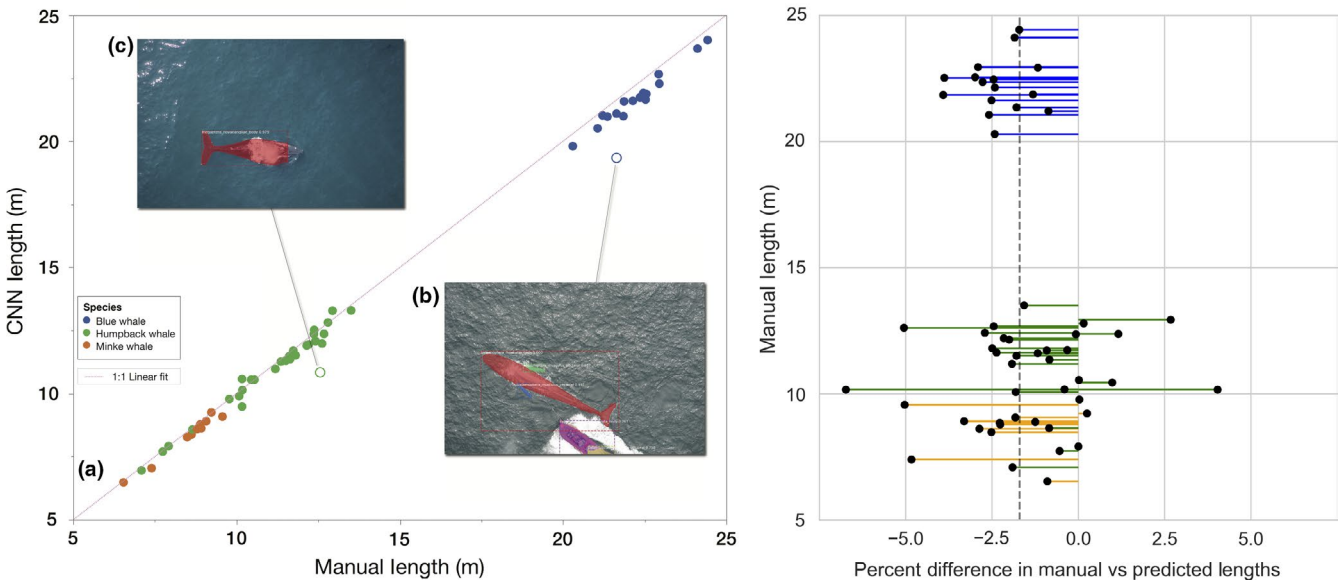

FIGURE 6 (a) Comparison of CNN-based length versus conventional manually measured length for all three species. The 1:1 line indicates that automated measurements tracked manual measurements closely but slightly underestimated length, particularly for blue whales. Two comparisons were identified as outliers for the model, one of a blue whale (b) and one of a humpback whale (c). The outlier blue whale image contains a boat approaching a whale to tag it and the humpback outlier stems from an image that has a whale producing a large blow that obscures part of the back of the animal. (d) The difference in manual versus predicted lengths across scale with the two outliers removed. The dashed line indicates the mean difference across all measurements (−1.7%)

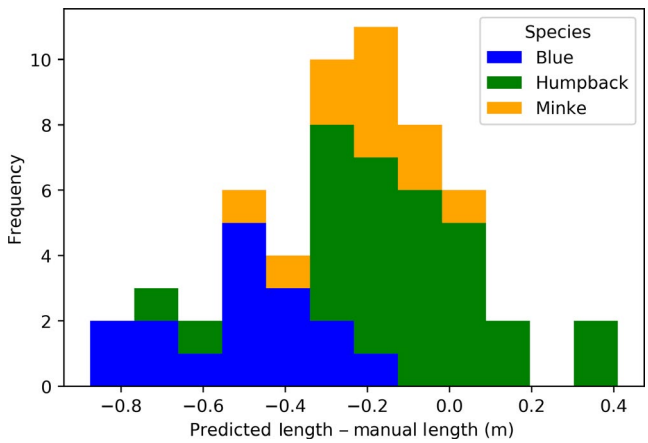

FIGURE 7 Histogram showing the differences between the predicted length from the convolutional neural network and the manually measured lengths in meters. This distribution shows the slight underestimation in general and increased underestimation of blue whales

over union (IoU) in this study compares very favorably with other studies in instance segmentation (where >0.5 is typically considered a good detection) with a mean IoU of 0.85 and standard deviation of 0.05 when comparing manually drawn masks to predicted masks. Detection rate is also typically presented in terms of precision (proportion of detections were true positives) and recall (proportion of true positives were detected). Since the whales were large and typically easy to detect, no false negatives were generated in the test imagery and the only false positive was the tagging boat, leading to a precision of 0.983 and a recall of 1.0.

To isolate the PCA-based approach used to generate lengths from masks from the error in the CNN predictions and assess, it on its own we used it to derive lengths from manually drawn masks and compared these with the manual linear method, leading to a mean difference of 0.06 m, considerably under the mean difference of 0.31 m for CNN-based lengths. This demonstrates that the error in estimated length is primarily driven by the automated mask generation, and not with the PCA-based estimation of length from that mask.

4 | DISCUSSION

4.1 | Historical context and ecological insights

The results of the present study provide the first example of how CNNs can be applied to automatically identify marine mega vertebrate species and subsequently estimate their length. The CNN performed well in both tasks, and with further training could be applied confidently in studies seeking to rapidly assess imagery datasets of other species. Our results are consistent with other total length ranges of whales reported by other studies (excluding calf sizes) using UAS photogrammetry on blue whales (Durban et al., 2016) and humpback whales (Christiansen et al., 2016). We also present the first photogrammetry results for Antarctic minke whales. Our results yield a slightly wider size range of minke whales encompassing smaller animals compared to historical reports from harvests

(Armstrong & Siegfried, 1991). This may be a result of bias towards sampling larger individuals during scientific harvests, morphological differences between coastal and pelagic Antarctic minke whales, or a bias in our imagery of finding whales that were approachable.

The present study builds upon previous work (see Karnowski, Johnson, & Hutchins, 2016 for a brief review) by providing species identification and length estimation in a single automated pipeline. Photogrammetric measurements of whales have not been previously automated, and still require considerable analyst effort even with streamlined computer workflows for image preparation and presentation. The model additionally outputs a confidence score along with its species prediction, this can be used as a simple filtering mechanism by using an empirically determined confidence threshold (80% in our model) as a flag for manual verification.

In previous work, reported measurement errors ranged from 1%–5%, providing enough precision and accuracy to detect relatively minor changes in overall body condition of animals related to breeding activities (Christiansen et al., 2016). In our current CNN, measurement errors much greater than 5% of body length were restricted to outliers arising from novel and complicated images being presented to the CNN (Figure 6). In spite of these outliers, our CNN-based approach provides sufficient accuracy to establish different age classes in baleen whales (adult vs. juvenile or calves) based on length. If these error proportions are similar for girth measurements there should be sufficient resolution to assess smaller changes in body size, including body condition measurements (Christiansen et al., 2016).

4.2 | Caveats and considerations

Nearly all differences in length measurements when comparing the automated method to the manual are due to a slight underestimation of length in the automated method. Much of the training data was from different images of the same individuals, thus while it was capturing some variation in positioning and body morphology, it was not truly capturing a huge variation in animals and conditions. Humpback and minke whales have similar coloration and more contrast with the water, potentially accounting for the increased accuracy over blue whale measurements.

Some of the largest errors in measurements appear to stem from novel phenomena in the input imagery. For example, one image of a blue whale was mis-measured by the CNN by 2.3 m, well above the idealized 5% error threshold for length measurement. In this case, the input image has an approaching zodiac from a tagging event, which washed the caudal portion of the whale with the wake (Figure 6). Similarly, the humpback outlier had a large respiratory blow, which resulted in the CNN truncating the mask. It is likely that the lack of boats and blows in the training imagery resulted in these errors. It is our assumption that given more examples of these phenomena in the training data, as long as they do not actually obscure the image, they would be relatively easily handled by the CNN, as evidenced by other applications of Mask R-CNN that gracefully handle partially obscured classes of interest (He et al., 2017). Interestingly, these novel phenomena did not influence the correct identification of whale species. The CNN was able to deal with multiple animals and naturally occurring environmental parameters (e.g. sea ice) without degradation in accuracy, likely due to the prevalence of these conditions in training imagery. Despite overall accuracy in length and species predictions the present CNN appears to have trouble resolving the tips of flukes and fins (e.g. the minke pectoral fins in Figure 4b), and as such would be unlikely to provide robust morphometrics of control surfaces. While there were some non-whale objects in our imagery (e.g. sea ice, boats) they were quite limited compared to what one may find in aerial imagery of many terrestrial or near-shore habitats. The additional variability in other habitats, and thus added potential for mis-classifications, will likely call for more training data, sufficiently representing these other object classes, to reach acceptable accuracies. While precise estimates aren't feasible, training sample sizes in the high hundreds to low thousands, combined with transfer learning, will likely lead to these acceptable accuracies (Guo, Liu, Georgiou, & Lew, 2018).

While the manual method can currently be considered the most accurate approach to length estimation in photogrammetric studies of whales, it still suffers from variation associated with the analyst making the measurements. Indeed, manual measurements made by multiple individuals can vary, and has been shown to have a coefficient of variation (CV) < 1% (Christiansen et al., 2018). We added a second researcher to manually measure each whale in each image to test this variation and had a mean CV of 0.38%, with an average difference of 7 cm between measurements. The automated method, once trained and set, has negligible variation in each run, and once trained with more data, may provide a better tool for comparing fine differences in length of individuals.

The mean error in our study (0.31 m), may not be sufficient resolution to determine age and sex class of other smaller animals such as dolphins, porpoises, turtles, and smaller terrestrial animals, but errors are based on sensor size and altitude, not the true size of the objects. Thus, if you can capture images closer to the animals, error will be reduced and similar error proportions as demonstrated in this work (2.1% of total length) can be achieved for much smaller target organisms. Ongoing improvements in sensor size and UAS altitude error will further facilitate analysis of smaller organisms.

4.3 | Future work

While adoption is still early, deep learning techniques are beginning to have a major impact in ecology and the environmental sciences where new methods of data collection, (e.g. UAS, satellites, camera traps) generate vast quantities of information. As deep learning continues to mature, it is critical that it be integrated into the ecology and conservation communities as a powerful new tool to understand our natural world. Future work within cetacean photogrammetry includes extending this system to additional species, adding other morphometric and allometric measurements such as girth, fluke width, rostrum to blowhole, etc., and eventually adding keypoint detection to conduct full behavioral analysis in video. If cetacean surveys are routinely conducted by UAS rather than ships and planes

these automated capabilities will facilitate analysis and allow rapid management and ecological insights.

5 | CONCLUSIONS

The results of this study indicate that deep learning techniques can enhance photogrammetric workflows aiming to identify and accurately measure baleen whales (and likely other species) in aerial imagery rapidly. The techniques described above, if applied to aerial survey programs that collect large archives of imagery, may help researchers move quickly past analytical bottlenecks and provide more time for abundance estimation, distributional research and ecological assessments. As UAS platforms evolve towards longer flight times and better sensor packages, automated workflows like the one presented here will be crucial for moving from data collection to inference and knowledge.

ACKNOWLEDGEMENTS

We thank Julian Dale for logistical support, UAS testing, and maintenance. We thank Clara Bird for data processing and imagery analysis support. Funding provided in part by the North Carolina Space Grant Graduate Research Fellowship, NSF IOS-1656676, NSF OPP-1644209, NSF-ANT 0823101 and 1440435, and Terman Fellowship from Stanford University. Cloud-based computational portions of this project were funded by the Microsoft AI for the Earth program and local computing was supported by the NVIDIA Corporation through the donation of a Titan Xp GPU.

AUTHORS' CONTRIBUTIONS

P.C.G., K.C.B., and D.W.J. conceived the ideas and designed the methodology; D.W.J., A.S.F., and J.A.G. led the expeditions and managed the projects; K.C.B., D.W.J., A.S.F., and J.A.G. collected the data; P.C.G. and S.A.M. led software development; P.C.G., S.A.M., K.C.B., and D.W.J. analysed the data; P.C.G. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

All processing and convolutional neural network code is available on Github.com at url: https://github.com/patrickcgray/cetacean_photogram or https://doi.org/10.5281/zenodo.3255030. All relevant imagery, including training and testing labels, is stored on the Dryad Data Repository at Gray et al., 2019. https://doi.org/10.5061/dryad.7482v2n.

ORCID

Patrick C. Gray https://orcid.org/0000-0002-8997-5255

David W. Johnston https://orcid.org/0000-0003-2424-036X

REFERENCES

Armstrong, A. J., & Siegfried, W. R. (1991). Consumption of antarctic krill by minke whales. *Antarctic Science*, *3*(1), 13–18. https://doi.org/10.1017/S0954102091000044

Arona, L., Dale, J., Heaslip, S. G., Hammill, M. O., & Johnston, D. W. (2018). Assessing the disturbance potential of small unoccupied aircraft systems (UAS) on gray seals (*Halichoerus grypus*) at breeding colonies in Nova Scotia. *Canada. Peerj*, *6*, e4467. https://doi.org/10.7717/peerj.4467

Blanckenhorn, W. U. (2004). The evolution of body size: what keeps organisms small? *The Quarterly Review of Biology*, *75*(4), 385–407. https://doi.org/10.1086/393620

Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R., & Sams, S. (1978). Relationships between body size and some life history parameters. *Oecologia*, *37*(2), 257–272. https://doi.org/10.1007/BF00344996

Bogucki, R., Cygan, M., Khan, C. B., Klimek, M., Milczek, J. K., & Mucha, M. (2018). Applying deep learning to right whale photo identification. *Conservation Biology*, *33*(3), 676–684. https://doi.org/10.1111/cobi.13226

Borowicz, A., McDowall, P., Youngflesh, C., Sayre-McCord, T., Clucas, G., Herman, R., … Lynch, H. J. (2018). Multi-modal survey of Adélie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-22313-w

Breuer, T., Robbins, M. M., & Boesch, C. (2007). Using photogrammetry and color scoring to assess sexual dimorphism in wild western gorillas (Gorilla gorilla). *American Journal of Physical Anthropology*, *134*(3), 369–382. https://doi.org/10.1002/ajpa.20678

Burnett, J. D., Lemos, L., Barlow, D., Wing, M. G., Chandler, T., & Torres, L. G. (2018). Estimating morphometric attributes of baleen whales with photogrammetry from small UASs: A case study with blue and gray whales. *Marine Mammal Science*, *35*(1), 108–139. https://doi.org/10.1111/mms.12527

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

Christiansen, F., Dujon, A. M., Sprogis, K. R., Arnould, J. P. Y., & Bejder, L. (2016). Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales. *Ecosphere*, *7*(10), e01468. https://doi.org/10.1002/ecs2.1468

Christiansen, F., Vivier, F., Charlton, C., Ward, R., Amerson, A., Burnell, S., & Bejder, L. (2018). Maternal body size and condition determine calf growth rates in southern right whales. *Marine Ecology Progress Series*, *592*, 267–281. https://doi.org/10.3354/meps12522

Cosens, S. E., & Blouw, A. (2003). Size- and age-class segregation of bowhead whales summering in northern Foxe Basin: A photogrammetric analysis. *Marine Mammal Science*, *19*(2), 284–296. https://doi.org/10.1111/j.1748-7692.2003.tb01109.x

Durban, J. W., Fearnbach, H., Barrett-Lennard, L. G., Perryman, W. L., & Leroi, D. J. (2015). Photogrammetry of killer whales using a small hexacopter launched at sea. *Journal of Unmanned Vehicle Systems*, *3*(June), 1–5. https://doi.org/10.1139/juvs-2015-0020

Durban, J. W., Moore, M. J., Chiang, G., Hickmott, L. S., Bocconcelli, A., Howes, G., … LeRoi, D. J. (2016). Photogrammetry of blue whales with an unmanned hexacopter. *Marine Mammal Science*, *32*(4), 1510–1515. https://doi.org/10.1111/mms.12328

Fearnbach, H., Durban, J. W., Ellifrit, D. K., & Balcomb, K. C. (2011). Size and long-term growth trends of endangered fish-eating killer whales. *Endangered Species Research*, *13*(3), 173–180. https://doi.org/10.3354/esr00330

Garrigue, C., Oremus, M., Dodémont, R., Bustamante, P., Kwiatek, O., Libeau, G., … Dalebout, M. L. (2016). A mass stranding of seven Longman's beaked whales (Indopacetus pacificus) in New Caledonia, South Pacific. *Marine Mammal Science*, *32*(3), 884–910. https://doi.org/10.1111/mms.12304

Gaudioso, V., Sanz-Ablanedo, E., Lomillos, J. M., Alonso, M. E., Javares-Morillo, L., & Rodríguez, P. (2014). 'Photozoometer': A new photogrammetric system for obtaining morphometric measurements of elusive animals. *Livestock Science*, *165*, 147–156. https://doi.org/10.1016/j.livsci.2014.03.028

Gray, P. C., Bierlich, K. C., Mantell, S. A., Friedlaender, A. S., Goldbogen, J. A., & Johnston, D. W. (2019). Data from: Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Dryad Digital Repository*, https://doi.org/doi.org/10.5061/dryad.7482v2n.

Gray, P. C., Fleishman, A. B., Klein, D. J., McKown, M. W., Bézy, V. S., Lohmann, K. J., & Johnston, D. W. (2018). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, *12*, 0–2, https://doi.org/10.1111/2041-210X.13132

Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*, *7*(2), 87–93. https://doi.org/10.1007/s13735-017-0141-z

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob, 2980–2988. doi:https://doi.org/10.1109/ICCV.2017.322.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Retrieved from, https://arxiv.org/abs/1512.03385.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua, 3296–3305. https://doi.org/10.1109/CVPR.2017.351.

Ichii, T., Shinohara, N., Fujise, Y., Nishiwaki, S., & Matsuoka, K. (1998). Interannual changes in body fat condition index of minke whales in the Antarctic. *Marine Ecology Progress Series*, *175*, 1–12. https://doi.org/10.3354/meps175001

Ireland, D., Garrott, R. A., Rotella, J., & Banfield, J. (2006). Development and application of a mass-estimation method for Weddell seals. *Marine Mammal Science*, *22*(2), 361–378. https://doi.org/10.1111/j.1748-7692.2006.00039.x

Johnston, D. W. (2019). Unoccupied aircraft systems in marine science and conservation. *Annual Review of Marine Science*, *11*(1), 439–463. https://doi.org/10.1146/annurev-marine-010318-095323

Johnston, D. W., Dale, J., Murray, K. T., Josephson, E., Newton, E., & Wood, S. (2017). Comparing occupied and unoccupied aircraft surveys of wildlife populations: Assessing the gray seal (Halichoerus gryus) breeding colony on Muskeget Island, USA. *Journal of Unmanned Vehicle Systems*, *5*, 178–191. https://doi.org/10.1139/juvs-2017-0012

Karnowski, J., Johnson, C., & Hutchins, E. (2016). Automated video surveillance for the study of marine mammal behavior and cognition. *Animal Behavior and Cognition*, *3*(4), 255–264. https://doi.org/10.12966/abc.05.11.2016

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1–9, https://doi.org/10.1016/j.protcy.2014.09.007

Kurita, H., Suzumura, T., Kanchi, F., & Hamada, Y. (2012). A photogrammetric method to evaluate nutritional status without capture in habituated free-ranging Japanese macaques (*Macaca fuscata*): A pilot study. *Primates*, *53*(1), 7–11. https://doi.org/10.1007/s10329-011-0280-4

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

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., …Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8693.LNCS(PART, 5),740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07–12-June, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.

Negretti, P., Bianconi, G., Bartocci, S., Terramoccia, S., & Verna, M. (2008). Determination of live weight and body condition score in lactating Mediterranean buffalo by Visual Image Analysis. *Livestock Science*, *113*(1), 1–7. https://doi.org/10.1016/j.livsci.2007.05.018

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M., Packer, C., & Clune, J. (2017). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, *115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Perryman, W. L., & Lynn, M. S. (1993). Identification of geographic forms of common dolphin (Delphinus delphis) from aerial photogrammetry. *Marine Mammal Science*, *9*(2), 119–137. https://doi.org/10.1111/j.1748-7692.1993.tb00438.x

Ratnaswamy, M. J., & Winn, H. E. (1993). Photogrammetric estimates of allometry and calf production in fin whales, balaenoptera physalus. *Journal of Mammalogy*, *74*(2), 323–330. https://doi.org/10.2307/1382387

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 512–519, https://doi.org/10.1109/CVPRW.2014.131

Read, J. (1990). Estimation of body condition in harbour porpoises, Phocoena phocoena. *Canadian Journal of Zoology*, *68*(9), 1962–1966. https://doi.org/10.1139/z90-276

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Sasse, B. (2003). Job-related mortality of wildlife workers in the United State, 1937–2000. *Wildlife Society Bulletin*, *31*(4), 1015–1020. https://doi.org/10.2307/3784446

Schmidt-Nielsen, K. (1975). Scaling in biology: The consequences of size. *Journal of Experimental Zoology*, *194*(1), 287–307. https://doi.org/10.1002/jez.1401940120

Shrader, A. M., Ferreira, S. M., & van Aarde, R. J. (2006). Digital photogrammetry and laser rangefinder techniques to measure African elephants. *South African Journal of Wildlife Research*, *36*(1), 1–7.

Trimble, M. J., van Aarde, R. J., Ferreira, S. M., Nørgaard, C. F., Fourie, J., Lee, P. C., & Moss, C. J. (2011). Age determination by back length for African Savanna elephants: Extending age assessment techniques for aerial-based surveys. *PLoS ONE*, *6*(10), e26614. https://doi.org/10.1371/journal.pone.0026614

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., …Belongie, S. (2017). The iNaturalist species classification and detection dataset. Retrieved from http://arxiv.org/abs/1707.06642.

Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, *9*(11), 2216–2225. https://doi.org/10.1111/2041-210X.13075

Weinstein, B. G. (2017). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*(3), 533–545. https://doi.org/10.1111/1365-2656.12780

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 1–9). https://doi.org/10.1109/IJCNN.2016.7727519.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Gray PC, Bierlich KC, Mantell SA, Friedlaender AS, Goldbogen JA, Johnston DW. Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods Ecol Evol*. 2019;00:1–11. https://doi.org/10.1111/2041-210X.13246