Remote Sensing in Ecology and Conservation
Open Access
ZSL LET'S WORK FOR WILDLIFE

INTERDISCIPLINARY PERSPECTIVES

Machine learning to detect marine animals in UAV imagery: effect of morphology, spacing, behaviour and habitat

Antoine M. Dujon[1], Daniel Ierodiaconou[2], Johanna J. Geeson[3], John P. Y. Arnould[3], Blake M. Allan[2], Kostas A. Katselidis[4] & Gail Schofield[1,5]

[1]Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong Victoria, 3216, Australia
[2]Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Warrnambool Victoria, 3280, Australia
[3]School of Life and Environmental Sciences, Deakin University, Burwood Victoria, 3125, Australia
[4]National Marine Park of Zakynthos, Elefteriou Venizelou Street, Zakynthos GR29100, Greece
[5]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK

Keywords
aerial surveys, demography, satellite imagery, deep learning, artificial intelligence, movement ecology

Correspondence
Antoine M. Dujon and Gail Schofield, Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, Victoria 3216, Australia.
Tel: +61 (0)3 924 45711; +44 (0)20 7882 8774; E-mail: antoine.dujon@yahoo.fr; g.schof@gmail.com

Editor: Kylie Scales
Associate Editor: Phil Bouchet

Received: 28 September 2020; Revised: 25 February 2021; Accepted: 11 March 2021

doi: 10.1002/rse2.205

Remote Sensing in Ecology and Conservation 2021;**7** (3):341–354

Abstract

Machine learning algorithms are being increasingly used to process large volumes of wildlife imagery data from unmanned aerial vehicles (UAVs); however, suitable algorithms to monitor multiple species are required to enhance efficiency. Here, we developed a machine learning algorithm using a low-cost computer. We trained a convolutional neural network and tested its performance in: (1) distinguishing focal organisms of three marine taxa (Australian fur seals, loggerhead sea turtles and Australasian gannets; body size ranges: 0.8–2.5 m, 0.6–1.0 m, and 0.8–0.9 m, respectively); and (2) simultaneously delineating the fine-scale movement trajectories of multiple sea turtles at a fish cleaning station. For all species, the algorithm performed best at detecting individuals of similar body length, displaying consistent behaviour or occupying uniform habitat (proportion of individuals detected, or recall of 0.94, 0.79 and 0.75 for gannets, seals and turtles, respectively). For gannets, performance was impacted by spacing (huddling pairs with offspring) and behaviour (resting vs. flying shapes, overall precision: 0.74). For seals, accuracy was impacted by morphology (sexual dimorphism and pups), spacing (huddling and creches) and habitat complexity (seal sized boulders) (overall precision: 0.27). For sea turtles, performance was impacted by habitat complexity, position in water column, spacing, behaviour (interacting individuals) and turbidity (overall precision: 0.24); body size variation had no impact. For sea turtle trajectories, locations were estimated with a relative positioning error of <50 cm. In conclusion, we demonstrate that, while the same machine learning algorithm can be used to survey multiple species, no single algorithm captures all components optimally within a given site. We recommend that, rather than attempting to fully automate detection of UAV imagery data, semi-automation is implemented (i.e. part automated and part manual, as commonly practised for photo-identification). Approaches to enhance the efficiency of manual detection are required in parallel to the development of effective implementation of machine learning algorithms.

Introduction

Technological advances in unmanned aerial vehicles (UAVs) are facilitating novel ways of detecting, monitoring and assessing wildlife in different ecological settings (Allan et al., 2015, 2019; Chabot et al., 2015; Christiansen et al., 2016; Fu et al., 2018; Mulero-Pázmány et al., 2014; Raoult et al., 2020). However, the use of UAVs as a mainstream tool is currently limited by the extensive time and expertise required to process and extract relevant information (Jones et al., 2006; Wich & Koh, 2018). The potential of UAVs in both long and short-term

monitoring and management actions could be realized by developing ways to extract imagery data in an automated way. This would allow UAVs to be integrated in rapid response tools to mitigate threats or to assess phenological shifts in wildlife with the environment in real time (Anderson & Gaston, 2013; Hazen et al., 2018; Howell et al., 2015; Linchant et al., 2015).

Machine learning algorithms have already been developed to facilitate the annotation and extraction of information from images for commercial, biomedical and security-linked activities (Cruz & Wishart, 2006; Dalal & Triggs, 2005; Ko, 2008; Lecun et al., 2015; Redmon et al., 2015). However, the use of machine learning for detecting animals and plants in ecological settings remains limited, but has great potential (For review see Dujon & Schofield, 2019 or Olden et al., 2008 afor review). A commonly used approach is to train a machine learning algorithm by showing examples of desired inputs and outputs, rather than programming a set of rules for all possible inputs (Jordan & Mitchell, 2015; Lecun et al., 2015). This approach is advantageous as the algorithm is less constrained, and it automatically learns and improves from experience (see Domingos, 2012; Lecun et al., 2015; and Thessen, 2016). The slow integration of machine learning by ecologists to detect plant and animals (Dujon & Schofield, 2019; van Gemert et al., 2015; Gray et al., 2018; Maire et al., 2015) is partly due to algorithms requiring days to train on powerful computers equipped with performant and expensive graphics processing units (GPUs, graphic cards that speeds up matrix-based computation), with no guarantee of success under a given ecological setting (Kaiming et al., 2017; Lamba et al., 2019; Raoult et al., 2020; Redmon et al., 2015; Ren et al., 2015). Consequently, most existing algorithms have limited application beyond the species and sites on which they are originally trained.

Here, we developed a convolutional neural network and validated the performance of this algorithm in: (1) distinguishing the focal organisms (termed regions of interest) of three marine taxa (Australian fur seals *Arctocephalus pusillus*, hereafter referred to as seals, loggerhead sea turtles *Caretta caretta,* hereafter referred to as sea turtles and Australasian gannets *Morus serrator* hereafter referred to as gannets) at different sites, allowing the acquisition of detection, distribution and count data, and (2) simultaneously delineating the fine-scale movement trajectories of multiple sea turtle individuals at focal in-water sites, allowing the acquisition of behaviour, movement and interactions. This algorithm can be implemented on personal computers (i.e. not using dedicated GPU acceleration) using freely available software (i.e. Python, keras and quantum GIS). We provide a readily accessible approach for using machine learning algorithms to optimise use in long-term wildlife monitoring.

Materials and Methods

Data collection

Imagery obtained from UAVs operated at three sites supporting seals, sea turtles and gannets was used. These three species were selected because they are among the most studied taxa in research using UAVs (Dujon & Schofield, 2019; Raoult et al., 2020), in addition to encompassing a wide range of body sizes (0.8–2.5 m) and terrestrial and aquatic habitats, allowing us to demonstrate the potential versatility of the machine learning algorithm for different species and environments. For the seals, a DJI Phantom 4 Professional™ V2 (www.dji.com) was used to take photographs, with continuous forward flight at 35 m altitude along set routes (continuous speed of 4 $m{\cdot}s^{-1}$) over a terrestrial colony at Kanowna Island, Australia (for details, see Supplementary Method 1). For the sea turtles, a DJI Phantom 3 Professional™ was used to collect video with continuous forward flight at 30 and 60 m altitude (providing horizontal field of views of 50 and 100 m, respectively) along set routes (continuous speed of 12 $m{\cdot}s^{-1}$) above the sea at the breeding site of Zakynthos Island, Greece. Hover mode was also used on Zakynthos over a single site, a fish cleaning station to record turtle movement trajectories (for details, see Schofield et al., 2017). For the gannets, a Swellpro Splashdrone quadcoptor (http://www.swellpro.com) was used with continuous forward flight to take photographs at 40 m altitude along set routes (continuous speed of 2 $m{\cdot}s^{-1}$) over a breeding colony at Point Danger, Australia (for details, see Supplementary Method 1). All surveys for the three species were conducted during the daytime and with the camera oriented downward (−90° with respect to the horizon). For each species, altitude was selected to ensure detection while maximising the number of individuals captured to obtain abundance counts and minimising disturbance (see Allan et al., 2019; Raoult et al., 2020; Schofield et al., 2017). In particular, Raoult et al. (2020) showed that target and non-target species are at risk of disturbance when flying at altitudes of <50 m, thus, machine learning algorithms that are effective at these altitudes are required.

The algorithm was first developed using UAV video footage (24 frames per second, 3840 × 2160 pixels resolution) collected during continuous flight over sea turtles (body size range: 0.8–2.5 m) in a marine setting to obtain position data to determine trajectories. Data from 2 years were used: 2016 ($n = 33$ surveys, April–July; 1980 min total) and 2017 ($n = 18$ surveys, April to July; 1080 min total). For all datasets, the imagery data were manually reviewed by two observers with experience identifying turtles. We then applied the approach to test it in a terrestrial setting for the other two species, representing different body size objects for detection (gannet body size range: 0.8–9 m; seal

body size range: 0.8–2.5 m). The gannet dataset consisted of a geo-referenced orthomosaic, with the whole colony appearing on a single image (see Supplementary Method 1 for details on data collection and pre-processing). The seal dataset consisted of 10 flights collected over 5 days in November and December 2019.

For hovering flight, to record movement patterns of multiple sea turtle individuals at once, we used footage from daytime surveys of a fish cleaning station ($n = 4$ surveys, June-July; 120 min total) frequented by sea turtles, over which the UAV was set to hover at 60 m (providing a horizontal field of view of 100 m).

Selection of a machine learning algorithm

We implemented a relatively simple convolutional neural network (i.e. compared to state-of-the-art object detection algorithms, Kaiming et al., 2017; Redmon et al., 2015; Ren et al., 2015). Specifically, the total number of steps required to train the model and apply it to new data could be implemented on a relatively inexpensive personal computer (i.e. laptop), without a dedicated GPU unit, which would be ideal for use in field settings or under restricted research budgets (the computer used in this study cost <$1150 USD). This contrasts with existing methods that require multiple days to train models using high-technology computers (e.g. with 12 or 16 GB memory costing >$3800 USD; Eikelboom et al., 2019; Gray et al., 2019). Rather than attempting to determine a convolutional neural network architecture from the beginning, which is very inefficient, we used a configuration that was successfully applied to a benchmark dataset previously. This also takes in account that the training step of a convolutional neural network requires considerably more computational resources than the animal detection step. We used a 'convolution-convolution-pooling' architecture that was previously applied to classify small images from a benchmark Canadian Institute for Advanced Research dataset (Krizhevsky, 2009). This dataset contains a mixture of images of common animals and transportation machines. For example, for sea turtles and gannets, the input layer of our model accepted 32 × 32-pixel images (in our case, animals or background) and 150 × 150 pixel images for seals, to account for differences in body size and background (the size of the input image was the same during the training of the model and its application on new data). This approach has the benefit of limiting the number of layers in the convolutional neural network and, hence, the number of parameters, which reduced the time required to train the model. For instance, a few hours was required for our network as opposed to several days, which is required for more complex models (e.g. 2 days for the network created for sea turtles by Gray et al., 2019). As demonstrated in Hasanpour et al. (2016), a well-crafted, yet simple and reasonably deep, architecture can sometimes perform on par with deeper and more complex architectures. The full architecture of the convolutional neural network for sea turtles is presented in Figure 1. Additional information on the functioning of convolutional neural networks is provided in Supplementary Methods 3. The final output of the convolutional neural network is a confidence score for a given window ranging from 0 (background) to 1 (animal) (Figure 1B).

Convolutional neural network training, validation and selection of a confidence threshold

For each species, we created an augmented dataset that was used to train the convolutional neural network (see Supplementary Methods 2 for full details). In brief, each positive and negative sample was duplicated and rotated in a range of orientations and was also flipped horizontally and vertically, generating additional augmented samples for each original sample. This step increases the performance of the convolutional neural network when limited data are available (see Simard, Steinkraus & Platt, 2003 and Zhang et al., 2017 for details on data augmentation). For sea turtles, our training dataset included 5 895 positive samples and 84 805 negative samples once augmented. We then split each augmented dataset into a training dataset (85% of positive and negative samples) that was used to fit the network and a validation dataset (15% of positive and negative samples), which, in turn, was used to estimate the network performance during the training process (Figure 1C).

We trained the network over 20 epochs (one epoch is the input of the entire training dataset through the model during training), with a categorical cross-entropy loss function using the 'adam' optimizer (see Kingma & Ba, 2015), the Rectified Linear Unit activation function and a mini-batch size of 84 (see Ioffe & Szegedy, 2015). A dropout rate of 0.5 was applied to each layer of the network to avoid overfitting (see Hinton, 2014). At the end of the training procedure, the network parameters for the epoch with the best performance on the training dataset were retained. Once the model was trained, we defined a confidence threshold of 0.55 to classify the samples as being part of the background (confidence <0.55) or being an animal (confidence >0.55) based on a sensitivity analysis of the true and false positive rates under varying confidence threshold values (see Supplementary Methods 3 for an example with sea turtles). The threshold was voluntarily set as a low value to avoid missing animals (similar to Gray et al., 2018) because we considered the cost of

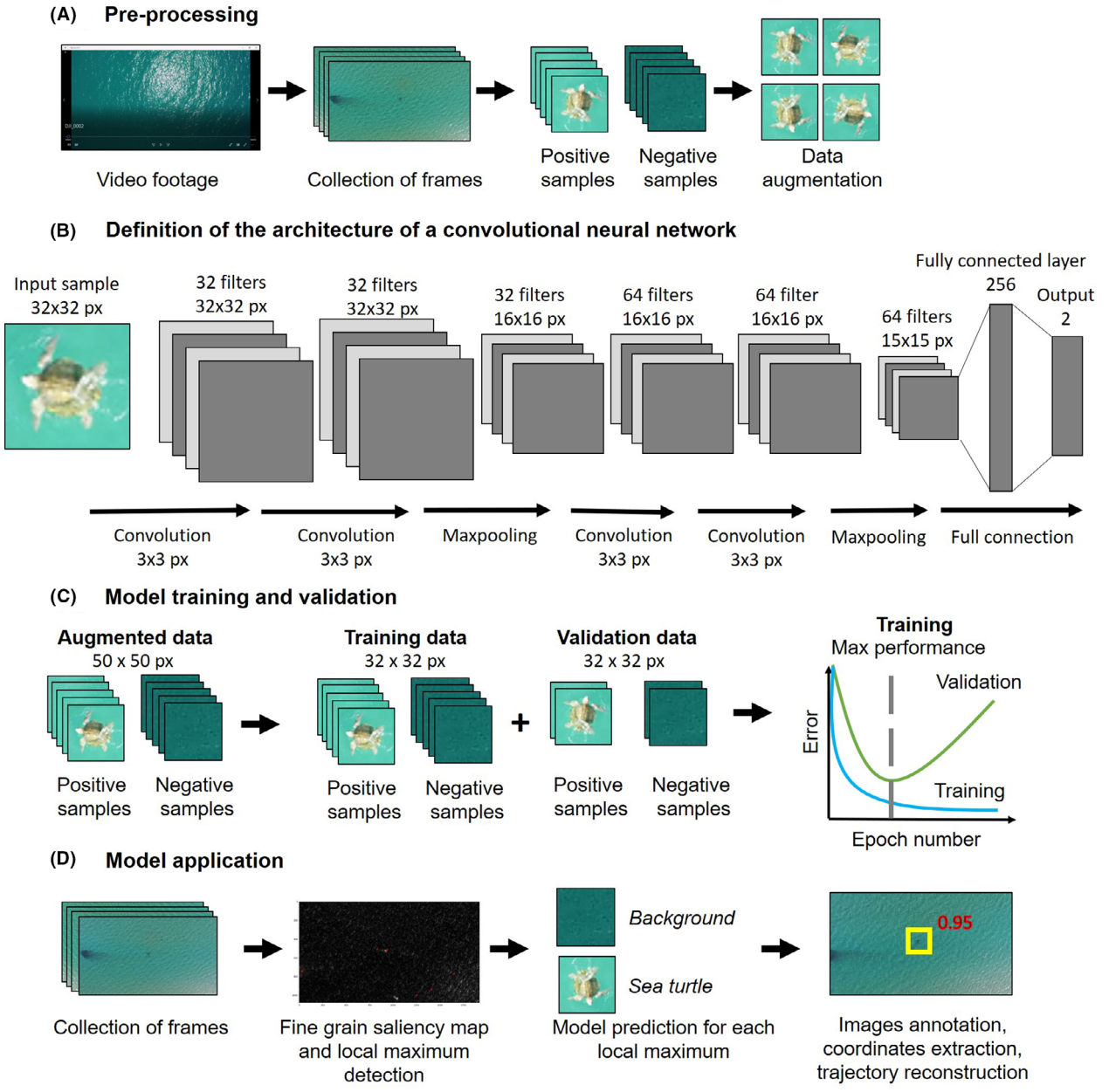

Figure 1. Steps involved in training a convolutional neural network to detect target organisms (in this case sea turtles with 1 m body length): (A) data collection, sample extraction and data augmentation; (B) definition of the architecture of a convolutional neural network showing the architecture used in this study and the operations performed by the network, including the number and resolution of detection filters; (C) training and validation of the convolutional neural network using a training and validation dataset; (D) application of the model to new data and trajectory reconstruction.

missing an animal to be greater than manually reviewing false positives.

Detection of animals

One way of detecting animals on imagery data is to use a sliding window of a given size and stride and to classify each window using a convolutional neural network (for example, see Gray et al., 2018). However, this approach is exhaustive; for a 3840 × 2160 pixels frame, a 50 × 50 pixel and a step size of 25 pixels, over 13 000 windows per frame must be classified, which is computationally prohibitive (requiring 130 s to process a frame using our network and settings). Therefore, we aimed to reduce the

number of windows to process our network, and hence the number of calculations. This was achieved by: (1) reducing the number of pixels that had to be processed (i.e. we downsized each frame to half its original size: 1720 × 1080 pixels, Tanimoto & Pavlidis, 1975); this was implemented for sea turtles and gannets but not seals (because body size is more variable between individuals in seals compared to the two other species) and (2) identifying homogenous background by computing a fine grain saliency map, on which the background appears as low values, while areas noticeably different to the background appear as high values (i.e. a potential animal in our case, Montabone and Soto 2010, Figure 1D). Since the imagery data that we used had very high resolution, reducing their size by 50% was not an issue, nor was the computation of saliency maps, which are well optimized within the OpenCV python library (Bradski, 2000).

We then determined the coordinates of the local maxima for each area with high saliency (each maximum was separated by at least 25 pixels), and extracted a window around each coordinate, which was then classified using the network to determine if it contained a sea turtle or formed part of the background. This approach reduced the number of windows required for processing from 13 000 to 1000–3000 windows per frame (<5–30 s processing time), depending on the complexity of the background and species. Of note, the number of local maxima per image can be easily adapted depending of the performance of the computer used.

Reconstruction of movement trajectory of sea turtles

We first extracted and processed one frame per second of video footage to ensure that consecutive frames overlapped (>50%) to increase the likelihood of detecting animals (i.e. reduce the effect of wind, sun glitter or turbidity). Then, we used the Hungarian algorithm to assign each location returned by the network to a trajectory by minimising a linear cost function based on the distance (in pixels) between locations in consecutive frames (see Kuhn, 1955). This algorithm assigns each location to its most likely trajectory and eliminates potential outliers. Thus, it was possible to estimate the movement trajectory of all animals in a given field of view during hovering.

Analysis of model performance for counting animals applied to seals, gannets and sea turtles

Estimating the performance of a network and its ability to classify newly collected images correctly under a range of scenarios (termed generalization, Zhang et al., 2017) is an important step of training a machine learning algorithm. We used two common metrics of performance for the convolutional neural network that was trained: (1) recall, which quantifies how well the model detected regions of interest (initially sea turtles), which was the proportion of animals detected by the model and (2) precision, which quantifies the proportion of positive identifications returned by the model that are actually correct and is an indication of the number of false positives returned by the model. The two metrics range between 0 (low performance) and 1 (maximum performance), and their equations are respectively defined as:

$$\text{Recall} = \frac{\text{Number of individuals detected by the algorithm in a video or still}}{\text{Number of individuals detected by humans observers in a video or still}} \quad (1)$$

$$\text{Precision} = \frac{\text{Number of true positives}}{\text{Number of true positives and false positives}} \quad (2)$$

For sea turtles, to investigate potential factors affecting the probability of detecting each individual (i.e., recall), we determined: (1) the number of frames in which each sea turtle appeared with no sun glitter (0%) and with sun glitter when overlaying <25%, 25–50% and >50% of individuals, and adapting the classification of Hodgson et al. (2003), (2) the position in the water of each individual (seabed vs. water column), which was validated using their shadows (i.e. sea turtles in the water column cast a shadow on the seabed), (3) the average wind speed based on data recorded by the UAV and data from Zakynthos Airport weather station on the day of the survey and (4) water turbidity at two scales (clear vs. turbid). Turbidity was classified using fixed landmarks in the water (e.g. rocks and anchors) that were positioned at different seabed depths at the study site. If the landmarks were clearly visible, we classified the water as clear; however, if the landmark was partially visible, or not visible at all, the water was classified as turbid.

A Bayesian mixed effects logistic regression model was used to investigate the potential factors influencing recall performance, as this metric is a proportion. We included UAV altitude (30 or 60 m), position of the sea turtle in the water column, water turbidity, average wind speed

and the number of frames with the different categories of sun glitter as fixed effects. We included each flight ID (i.e., 15 min based on battery length) as a random effect, to account for multiple non-independent sea turtle observations during flights (see Supplementary Methods 4 for full detail on the model) (Hodgson et al., 2013; Zuur et al., 2009). We used this model to test whether higher altitude, turbid water, sun glitter and average wind speed reduce the recall of sea turtles, and to detect possible interactions between these factors. In addition, we hypothesized that seabed depth impacts the recall of sea turtles (i.e. that sea turtles resting on the seabed at deeper locations are more difficult to detect). Therefore, we fitted an additional Bayesian mixed effects model between the recall of sea turtles resting on the seabed, water column depth in metres (as a fixed effect) and flight ID (as a random effect). After fitting the model, we used odds ratio (OR) as a measurement of size effect after fitting a Bayesian mixed effects model (Supplementary Methods 4).

For gannets, the algorithm was trained using the same architecture and protocol that we used on sea turtles. We used 30 individuals that were randomly selected from the colony (5% of total colony size) to generate 266 positive samples and 2184 negative samples to train the model after augmenting the dataset. It is often recommended that roughly the same number of positive and negative samples are used when training a machine learning algorithm (Buda et al., 2018). However, it is not always possible to balance the two types of samples, especially if the amount of imagery data is limited, which was the case for the gannets. Here, the difference between the number of positive and negative samples for the gannets was not an issue because the animals were clearly differentiated from the background, which facilitated the training of the convolutional neural network. For seals, the architecture was modified to accept larger input samples; however, the number of layers and general architecture of the network remained the same as for the other two species. In total, we used 9853 positive samples and 10 255 negative samples of seals to train the model after augmenting the dataset. For both gannets and seals, we quantified the recall and precision of the model by screening the whole colony (which fit in a single image by generating orthomosaics) with the trained model, and compared the detections returned by the model to the locations of the animals manually extracted from ImageJ software by placing a point midway on the head-tail axis of the animals.

Analysis of model performance for recording the movement trajectories of sea turtles

Using the neural network and Hungarian algorithm, we reconstructed the movement trajectories of all sea turtles visiting the fish cleaning station on the video footage. We quantified the recall for each trajectory, along with the relative positioning error of locations. The locations of all turtles within all frames processed by the algorithm were manually extracted (centre of the carapace was recorded as the turtle location and defined as the middle of a snippet). The relative positioning error was defined as the distance (in cm) between the location returned by the algorithm and the location manually recorded. We then determined the mean relative positioning error for each trajectory.

Reporting of statistics

We reported all of the parameters estimated from the Bayesian mixed effects models, followed by their 95% credible intervals between square brackets. The equations and details of the prior distributions are provided in Supplementary Methods 4. An explanatory variable was considered to have a significant effect on recall if the 95% credible intervals of the OR did not contain 1. Similarly, two proportions or odd ratios were considered significantly different if their 95% credible intervals did not overlap.

Reporting of hardware and software

All calculations and video processing were completed on a Dell Inspiron 15 5000 Series laptop equipped with a 64-bit Windows 10© operating system, an Intel® Core™ I7-6500 CPU two cores processor cadenced at 2.5 GHz and 8 Gig of RAM. No GPU acceleration was used.

Manually cropped samples and coordinates were extracted using ImageJ software version 1.51k (Eliceiri et al., 2012). The general manipulation of frames, data augmentation procedures and computation of fine grained saliency maps were performed using the OpenCV version 3.4.0 python library (Bradski, 2000). The convolutional neural networks were trained using Keras python library version 2.4.3 (Chollet, 2015). The Hungarian algorithm was computed using scikit-learn version 0.18.1 python library (Pedregosa et al., 2012).

All Bayesian statistical analyses were performed using the 'MCMCglmm' package (Hadfield, 2010) within the R software version 3.3.2. (R Development Core Team 2013). Geographical data were assimilated using Quantum-GIS software version 2.18 (QGIS Development Team 2015).

Results

Counts of individual animals

The success of the machine learning algorithm varied depending on species and parameters, including

morphology, spacing, behaviour and habitat uniformity. For the gannet colony (terrestrial; $n = 587$ individuals), the overall precision of the model was 0.74, with an overall recall of 0.94. Individuals ($n = 56$, 10% of the colony size) that were not detected by the algorithm were mostly juveniles that were partially hidden (i.e. overlapping) by their parents (Figure 2A). The habitat was uniform (sandy substrate), and all birds in the image were resting (not flying).

In comparison, for the colony of seals (also terrestrial-haul out site), the overall precision of the model was 0.27, with an overall recall of 0.79. This low result was driven by multiple factors, including habitat (the presence of seal-sized boulders), sexual dimorphism of adults, the small size of pups and the presence of creches in which pups aggregated (Figure 2B). The model was not able to distinguish between multiple accolated seals pups.

The precision and recall for the sea turtles in the sea was 0.20 and 0.75, when only using individuals in the sea surface layer (i.e. comparable to the conditions existing machine learning algorithms were trialled for sea turtles). At our study site, sea turtles were well-spaced (>10 m), and were all adults with no sexual dimorphism (both adult males and females measured between 0.6 and 1 m body length). However, the marine environment generated different issues in detection compared to terrestrial study sites, as individuals were distributed across a range of seabed depths from shore to 5 m. The logistic regression model indicated that the probability of detecting a sea turtle resting on the seabed was lower than that (recall 0.44; [95% CI: 0.40, 0.47]) of a turtle basking in the surface layer or in the water column (recall 0.74; [95% CI: 0.70, 0.77]). Detection decreased significantly with each additional meter in depth (OR: 0.51, [95% CI: 0.41, 0.62]), and was close to zero at a depth of 5 m (Figure 3B). In addition, when water was turbid, significantly fewer turtles resting on the seabed were detected when compared to clear conditions (OR: 0.23 [95% CI: 0.12, 0.40]) and those in the water column (OR: 0.84 [95% CI: 0.49, 1.55]). The altitude (either 30 or 60 m) at which the UAV was flown did not affect the recall of sea turtles, regardless of water turbidity (OR: 1.00 [95% CI: 0.97, 1.03]) or the position of turtles in the water column (OR: 1.01 [95% CI: 0.98, 1.03], Figure 3A). Similarly, the average wind speed (max 24.1 $km \cdot h^{-1}$) did not significantly affect the detection of sea turtles (OR: 1.02 [95% CI: 0.99, 1.04]). When the number of frames with no glint overlapping an individual processed by the algorithm increased, so did the probability of detecting an individual (OR: 1.33 [95% CI: 1.13, 1.53], Figure 3C). Frames where sun glint overlaid turtles reduced the detection rate (OR: 0.86, [95% CI: 0.67, 1.12], Figure 3C). The overall precision of the model when not accounting for any of the variables was 0.20.

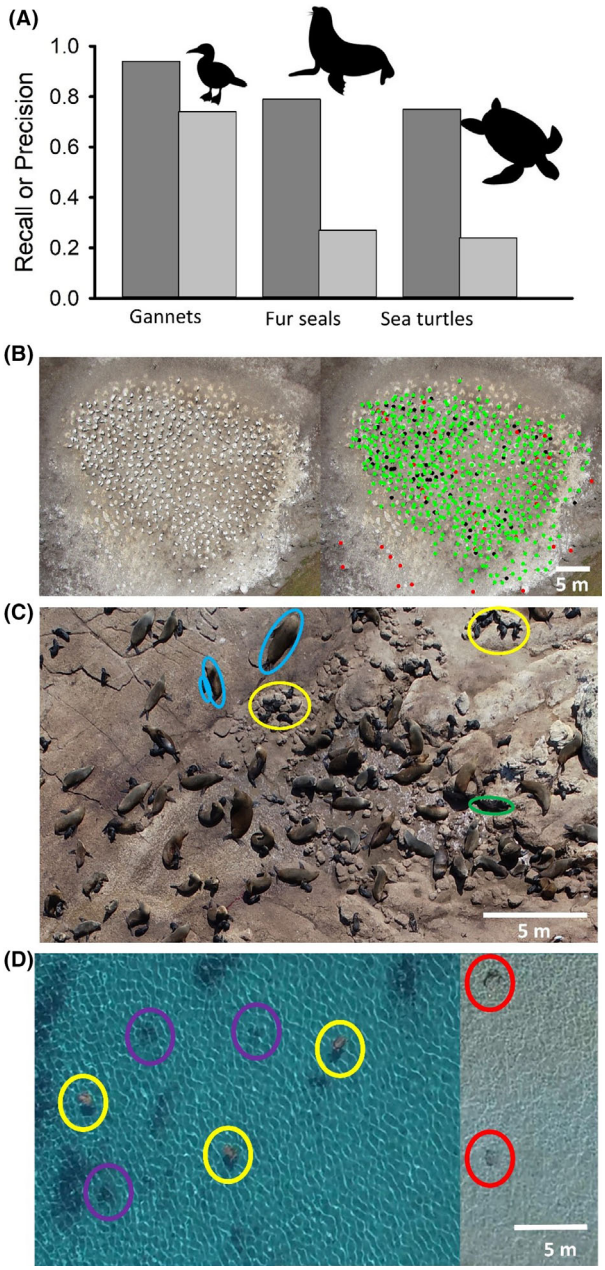

Figure 2. (A) Recall (dark grey bars) and precision (light grey bars) of the convolutional network for gannets, seals and sea turtles. (B) Example of detection accuracy of the convolutional neural network for the gannet colony. Green dots represent individuals detected by both the human observer and convolutional neural network, black dot indicates individuals only detected by the human observer (mostly juveniles huddled with their parent), and red dots indicate false positives. Examples of factors impacting detection in (C) seals: body size (blue ovals: adult male, female and pup), spacing and behaviour (yellow ovals: pups huddling in creches) and habitat (green oval: seal-sized, shaped and coloured boulders); and (D) sea turtles: habitat complexity (purple ovals: seagrass) hindering detection in deeper waters (yellow ovals: *c.* 4 m seabed depth) and low contrast in shallower waters (red ovals: <1 m seabed depth).

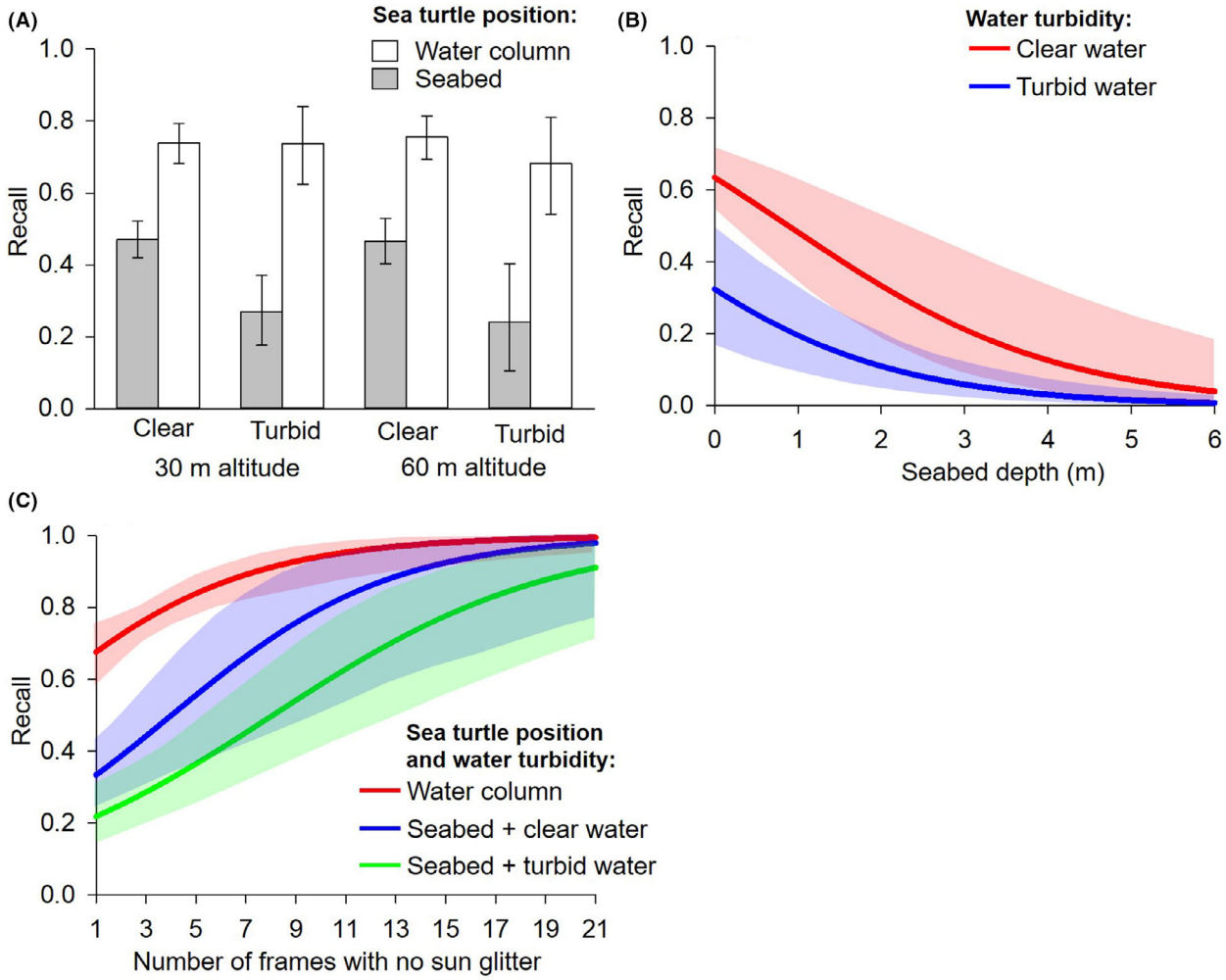

Figure 3. (A) Overall recall of sea turtles as a function of UAV flight altitude, position of individuals in the water column and water turbidity. Error bars represent 95% credible intervals. (B) Effect of depth and water turbidity on the recall of sea turtles located on the seabed. (C) Effect of the number of frames in which sun glitter did not overlay sea turtles on recall, taking into account turtle position in the water column and water turbidity. Note, detection of sea turtles located in the water column was not affected by turbidity. In (B and C), the solid bold lines represent average recall, while shaded bands represent 95% credible intervals.

Focal movement of sea turtles

We recorded the movement of 20 sea turtles from five processed videos (Figure 4, Supplementary Table 1). Overall, the average detection rate was 63 ± 24% (range: 9–99%).

The average duration of the movement records was 250 ± 255 s (range: 18–689 s), which equated to an average number of 38 ± 14 locations per minute (range: 5–58). The overall relative average positioning error was 23 ± 13 cm (range: 0–183, see Supplementary Table 1 for details of each trajectory), with 99% of locations having a relative positioning error of <55 cm. As the sampling rate was high, the overall movement pattern of all individuals was captured well. Five trajectories had a detection rate of <50% (Figure 4), primarily because these individuals were positioned on the seabed (at ca 2.5 m deep, see Figure 2C).

Discussion

This study demonstrated that it is possible to train a small convolutional neural network to detect wildlife of <1 to >2 m in body size across aquatic and terrestrial systems, that is both user-friendly and can be run using open-source software libraries and low cost computing resources, which is vital for organizations with limited resources (ImageJ, Python and Keras, see Chollet, 2015;

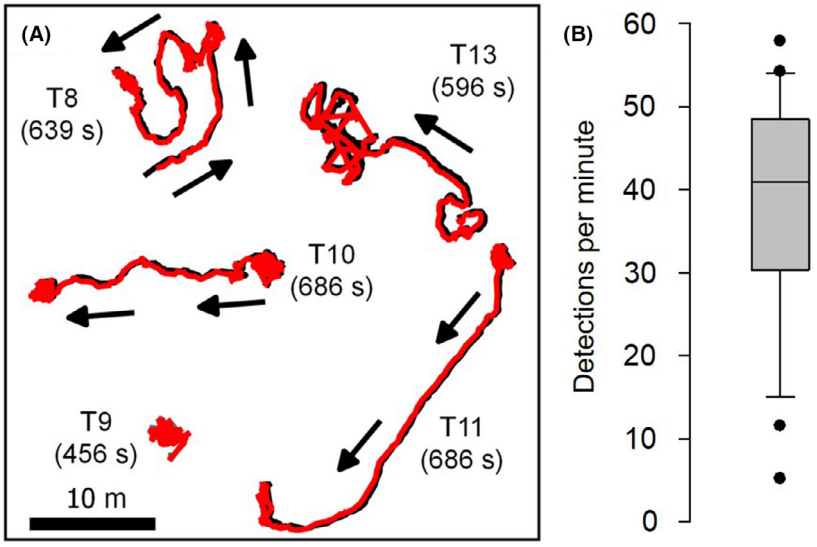

Figure 4. (A) Trajectories of five sea turtles (four of which were moving, and one that remained stationary) reconstructed from processing video footage during hovering mode using a convolutional neural network. Black lines represent the trajectories that were manually obtained from the video frames, while red lines represent the trajectories obtained using the machine learning algorithm. Black arrows indicate the direction of movement. The number in parentheses under each sea turtle ID represents the trajectory record duration in seconds. (B) Boxplot of the number of locations per minute obtained for the 20 sea turtles for which a trajectory was reconstructed using the machine learning algorithm.

Eliceiri et al., 2012). Our approach is operable in the field, with a low-cost computing device that could be applied for real-time scenarios (i.e. as shark detection). However, this study also shows that the precision of the network varies with respect to species, morphology, spacing, behaviour and habitat. We recommend that these factors are considered when planning to use convolutional neural networks, or any other algorithm, to process UAV data of animals. A potential approach would be to use a machine learning algorithm for components of the data in which all four criteria are met and processing the remainder manually (such as the pup creches of seals). We also demonstrated the potential of using this network to reconstruct the movement trajectories of individual animals, providing ecological information on how individuals move in relation to conspecifics or other organisms (predators or prey) and complementing remote tracking studies (Raoult et al., 2018; Schofield et al., 2017, 2019). Overall, automating the detection of wildlife represents a potentially useful tool for both researchers and managers to enhance monitoring effort and science-based approaches; however, our study also demonstrates that, in some circumstances, precision might require correction when conducting censuses. This could be partially achieved by accounting for imperfect detection of groups and individuals when estimating abundance (e.g. Clement, Converse & Royle, 2017). However, we demonstrate that the application of a machine learning approach provides enormous potential for semi-automating a previously tedious manual approach that would expedite image processing efficiency, especially with the increasing size of datasets and time series collections available.

While not state of the art, the performance of our network was comparable to other machine learning networks trained to detect and count birds, marine mammals and sea turtles (Gray et al., 2018; Hodgson et al., 2018; Maire et al., 2015) from UAV and aerial imagery data. Our model also outperformed older methods developed for terrestrial ecosystems (e.g. van Gemert et al., 2015) with the added advantage of users being able to fully train and run it on standard and older generation laptops (without a relatively expensive dedicated GPU), facilitating its use by organizations that potentially have limited access to funds, such as non-governmental organisations in developing countries. Furthermore, because we trialled our machine learning algorithm on three taxa (birds, reptiles and mammals) occupying two contrasting environments (terrestrial and marine), we were able to identify key parameters impacting detection and hence, performance. Performance might also be impacted by UAV model and mode of data collection (e.g. photographs vs. video); however, there was no noticeable impact on our analyses, as both video and photographic data were of very high resolution. Our algorithm partially overcame the challenge of detecting relatively small animals with high resolution imagery data (e.g. see Hu & Ramanan, 2017; Lalonde et al., 2017), being able to detect animals representing percentages as low as 0.0006% (1 m^2 out of

1750 m^2) and 0.0002% (1 m^2 out of 6500 m^2 of the surface of a 4K resolution frame, as for sea turtles). Furthermore, our algorithm was not impacted by ranges of 50 cm in body size for sea turtles (i.e. adult sea turtles, 0.6–1 m variation across both males and females; Schofield et al., 2017). However, sexual dimorphism in seals (adult males vs. females) and the small size of pups meant that the model must be trained separately for each group category. This approach could be used if enough data were available (which was not the case in this study), because the network only required a few hours to train each species compared to a deep learning algorithm that has hundreds of layers that would typically require days to train (Lecun et al., 2015). Overall, more samples are required to train a model on complex backgrounds compared to simpler backgrounds when target species are elusive or when large variation in body size exists among individuals. We recommend including several thousand initial samples when training a convolutional neural network with an architecture like the one used in the current study. In addition, convolutional neural networks trained to solve an initial detection task can be reused as a starting point to train a model to solve a second detection task (termed 'transfer learning', Yosinski, Clune, Bengio, & Lipson, 2014). This could prove highly useful for implementing long-term monitoring programs where important parameters might not be detected initially, requiring the dataset to be re-evaluated. As the amount of data collected using UAVs increases, the implementation of machine learning models to process large amounts of imagery data will become an important step for long-term monitoring programs (e.g. Tabak et al., 2019; Villon et al., 2018).

Other parameters also impact detection by algorithms leading to false positive or negative estimates of animal numbers, with this issue potentially being exacerbated by machine learning tools (Domingos, 2012; Kampichler et al., 2010; Olden et al., 2008; Raoult et al., 2020). The uniformity of the habitat in which animals occur is a key issue influencing detection (Dujon & Schofield, 2019). For instance, the bird colony had a highly uniform background with high detection rates. In comparison, the seal colony was more heterogenous, with the presence of seal-sized boulders and cracks generating false positives. For sea turtles, detection was equivalent to previous studies (e.g. Gray et al., 2018) when the same conditions were used (e.g. surface layer individuals); however, when different sea depths were explored, including turtles occupying different positions in the water (seabed and water column), detection levels noticeably dropped. Interestingly, while previous studies have identified turbidity and glare/glitter of the sun as major issues in aquatic systems (Brack, Kindel, & Oliveira, 2018; Hodgson et al., 2013), we found that the longer a sea turtle appeared on the footage, the more likely it would be detected, thus overcoming these issues. Therefore, UAV speed could be adjusted to increase the time an individual is in a frame and the likelihood of detection. In parallel, by flying UAVs at 60 m altitude of 60 m, coverage is enhanced, with detection of turtles to depths of at least 5 m being possible.

In addition, we had issues with distinguishing pairs or groups of animals. Examples included juvenile gannets huddled with their parents (i.e. appearing as a single unit) or seal pups forming tight aggregations in creches. This issue has also been reported in previous studies of birds and mammals (Brack et al., 2018; Chabot & Francis, 2016). Other detection issues also include cases where an animal is partially hidden beneath an object (e.g. vegetation or cliff) or are only partially visible under water (Bonnin et al., 2018; Brack et al., 2018; Koh & Wich, 2012). In such instances, adjusting UAV altitude might not be sufficient; however, other approaches could be implemented, including measuring areas or segmenting an area considered to contain an animal by the network, for example by grouping pixels (Gonzalez et al., 2016; Maire et al., 2015). Alternatively, a convolutional neural network could be trained on specific animal body parts (i.e. head, flipper or tail). Specific approaches could be employed to overcome this issue, but would incur increased computational costs, which would require more expensive computers (see for example Hu & Ramanan, 2017; Lamba et al., 2019; Lecun et al., 2015).

We also demonstrated the utility of our network to record the fine scale movement of multiple animals at once, which is a potential advantage of UAVs, where the trajectories of individuals were assimilated separately and then overlaid (Gaspar et al., 2011; Raoult et al., 2018). This function provides a way of documenting how individuals move in relation to conspecifics, other organisms (prey and predators) and their environment (e.g. substrate/habitat type) without the need to capture and attach units (Schofield et al., 2019). In particular, our convolutional neural network generated more locations (on average one location every 2 s) with extremely high accuracy (<0.5 m) compared to commonly used tracking technologies for aquatic animals at a similar spatial scale; specifically, 70 locations per day of <70 m accuracy using Fastloc-GPS (Dujon et al., 2014, 2017, 2018) and 180 locations per day of <40 m accuracy for acoustic tracking using Vemco Positioning System (Stieglitz & Dujon, 2017). However, UAVs are limited compared to these technologies with respect to duration of monitoring (20–30 min battery life per flight for most commercial UAVs) and distance from user (Christie et al., 2016; Gonzalez et al., 2016; Koh & Wich, 2012). Therefore, using UAVs in combination with remote technologies could help advance ecological studies,

particularly of marine vertebrates (Hays et al., 2019), from tracking individuals to monitoring populations and communities (Schofield et al., 2019).

In conclusion, we demonstrated that while the same machine learning algorithm can be used with high accuracy within and across different species, no single algorithm can capture all components optimally, even within sites, due to the variable effects of morphology, spacing behaviour and habitat. We recommend that rather than attempting to fully automate detection of UAV imagery data, semi-automation is implemented (i.e. part automated and part manual, as with photo-identification). For instance, 'regions' of data where animals meet the criteria for optimal detection are automated, while 'regions' where this is not possible are processed manually, possibly using GIS-based orthomosiacs; however, the current need for powerful computers must be resolved first. Thus, approaches to enhance the efficiency of manual detecting are required in parallel to the development of semi-automated machine learning algorithms. Ultimately, this study provides a way of reducing the current gap between the field of machine learning and ecology, including conservation and management. We anticipate machine learning algorithm to be increasingly implemented in long-term monitoring programs, facilitating evidence-based conservation programs.

Author Contributions

AMD and GS conceived the study; KAK, GS, JJG, DI, BA conducted the sea turtle fieldwork and assimilated the data; DI, BA and JYPA collected the Gannet data. JPY and JJG collected the seals colony data. DI and BA undertook the image processing. AMD prepared the augmented datasets, coded and trained the machine learning algorithms and performed the statistical analyses; AMD and GS led the writing of the manuscript with contributions from all authors.

References

Allan, B.M., Ierodiaconou, D., Hoskins, A.J. & Arnould, J.P.Y. (2019) A rapid UAV method for assessing body condition in fur seals. *Drones*, **3**, 24.

Allan, B.M., Ierodiaconou, D., Nimmo, D.G., Herbert, M. & Ritchie, E.G. (2015) Free as a drone: ecologists can add UAVs to their toolbox. *Frontiers in Ecology and the Environment*, **13**, 354–355.

Anderson, K. & Gaston, K.J. (2013) Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, **11**, 138–146.

Bonnin, N., Van Andel, A., Kerby, J., Piel, A., Pintea, L. & Wich, S. (2018) Assessment of chimpanzee nest detectability in drone-acquired images. *Drones*, **2**, 17.

Brack, I.V., Kindel, A. & Oliveira, L.F.B. (2018) Detection errors in wildlife abundance estimates from Unmanned Aerial Systems (UAS) surveys: synthesis, solutions, and challenges. *Methods in Ecology and Evolution*, **9**, 1864–1873.

Bradski, G. (2000) The OpenCV library. *Dr. Dobb's Journal of Software Tools*, **120**, 122–125.

Buda, M., Maki, A. & Mazurowski, M.A. (2018) A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, **106**, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011.

Chabot, D., Craik, S.R. & Bird, D.M. (2015) Population census of a large common tern colony with a small unmanned aircraft. *PLoS One*, **10**, 1–14.

Chabot, D. & Francis, C.M. (2016) Computer-automated bird detection and counts in high-resolution aerial images: a review. *Journal of Field Ornithology*, **87**, 343–359.

Chollet, F. (2015) 'Keras. Github.' GitHub. Available at: https://github.com/fchollet/keras [Accessed 27th March 2021].

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**, e01468.

Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M. & Hanson, L. (2016) Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology. *Frontiers in Ecology and the Environment*, **14**, 241–251.

Clement, M.J., Converse, S.J. & Royle, J.A. (2017) Accounting for imperfect detection of groups and individuals when estimating abundance. *Ecology and Evolution*, **7**, 7304–7310. https://doi.org/10.1002/ece3.3284.

Cruz, J.A. & Wishart, D.S. (2006) Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, **2**, 59–77.

Dalal, N. & Triggs, B. (2005) Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, pp. 886–893.

Domingos, P. (2012) A few useful things to know about machine learning. *Communications of the ACM*, **55**, 78.

Dujon, A.M., Lindstrom, R.T. & Hays, G.C. (2014) The accuracy of Fastloc-GPS locations and implications for animal tracking. *Methods in Ecology and Evolution*, **5**, 1162–1169.

Dujon, A.M. & Schofield, G. (2019) Importance of machine learning for enhancing ecological studies using information-rich imagery. *Endangered Species Research*, **39**, 91–104.

Dujon, A.M., Schofield, G., Esteban, N. & Hays, G.C. (2017) Fastloc-GPS reveals daytime departure and arrival during long-distance migration and the use of different resting. *Marine Biology*, **164**, 187. https://doi.org/10.1007/s00227-017-3216-8.

Dujon, A.M., Schofield, G., Lester, R.E., Papafitsoros, K. & Hays, G.C. (2018) Complex movement patterns by foraging

loggerhead sea turtles outside the breeding season identified using Argos-linked Fastloc-Global Positioning System. *Marine Ecology*, **39**, e12489.

Eikelboom, J.A.J., Wind, J., van de Ven, E., Kenana, L.M., Schroder, B., de Knegt, H.J. et al. (2019) Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, **10**, 1875–1887. https://doi.org/10.1111/2041-210X.13277.

Eliceiri, K., Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. (2012) NIH Image to ImageJ : 25 years of image analysis. *Nature Methods*, **9**, 671–675.

Fu, Y., Kinniry, M. & Kloepper, L.N. (2018) The Chirocopter: a UAV for recording sound and video of bats at altitude. *Methods in Ecology and Evolution*, **9**, 1531–1535.

Gaspar, T., Oliveira, P. & Silvestre, C. (2011) UAV-based marine mammals positioning and tracking system. In: Proceedings of the 2011 IEEE International Conference on Mechatronics and Automation. Beijing, China, pp. 1050–1055.

Gonzalez, L.F., Montes, G.A., Puig, E., Johnson, S., Mengersen, K. & Gaston, K.J. (2016) Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors*, **16**, 97. https://doi.org/10.3390/s16010097.

Gray, P.C., Bierlich, K.C., Mantell, S.A., Friedlaender, A.S., Goldbogen, J.A. & Johnston, D.W. (2019) Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods in Ecology and Evolution*, **10**, 1490–1500. https://doi.org/10.1111/2041-210X.13246.

Gray, P.C., Fleishman, A.B., Klein, D.J., McKown, M.W., Bézy, V.S., Lohmann, K.J. et al. (2018) A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, **10**, 345–355.

Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *Journal of Statistical Software*, **33**, 1–22.

Hasanpour, S.H., Rouhani, M., Fayyaz, M. & Sabokrou, M. (2016) Lets keep it simple, using simple architectures to outperform deeper and more complex architectures. arXiv.org arXiv:1608, 1–18.

Hays, G.C., Bailey, H., Bograd, S.J., Bowen, W.D., Campagna, C., Carmichael, R.H. et al. (2019) Translating marine animal tracking data into conservation policy and management. *Trends in Ecology & Evolution*, **34**, 459–473. https://doi.org/10.1016/j.tree.2019.01.009.

Hazen, E.L., Scales, K.L., Maxwell, S.M., Briscoe, D.K., Welch, H., Bograd, S.J. et al. (2018) A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. *Science Advances*, **4**, eaar3001.

Hinton, G. (2014) Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, **15**, 1929–1958.

Hodgson, A., Kelly, N. & Peel, D. (2013) Unmanned aerial vehicles (UAVs) for surveying marine fauna: a dugong case study. *PLoS One*, **8**, 1–15.

Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspoon, S., Kilpatrick, A.D. et al. (2018) Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, **9**, 1160–1167.

Howell, E.A., Hoover, A., Benson, S.R., Bailey, H., Polovina, J.J., Jeffrey, A. & et al. (2015) Enhancing the TurtleWatch product for leatherback sea turtles, a dynamic habitat model for ecosystem-based management. *Fisheries Ocenanography*, **24**, 57–68. https://doi.org/10.1111/fog.12092.

Hu, P. & Ramanan, D. (2017) Finding tiny faces. In: The IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, pp. 951–959.

Ioffe, S. & Szegedy, C. (2015) Batch normalization : accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France, pp. 448–456.

Jones, G.P., Pearlstine, L.G. & Percival, H.F. (2006) An assessment of small unmanned aerial vehicles for wildlife research. *Wildlife Society Bulletin*, **34**, 750–758.

Jordan, M.I. & Mitchell, T.M. (2015) Machine learning: trends, perspectives, and prospects. *Science*, **349**, 255–260.

Kaiming, H., Georgia, G., Doll, P. & Girshick, R. (2017) Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Las Condes, Chile, pp. 1–12.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H. & Arriaga-Weiss, S. (2010) Classification in conservation biology: a comparison of five machine-learning methods. *Ecological Informatics*, **5**, 441–450.

Kingma, D.P. & Ba, J.L. (2015) Adam: a method for stochastic optimization. In: International Conference on Learning Representations 2015. San Diego, CA, pp. 1–15.

Ko, T. (2008) A survey on behavior analysis in video surveillance for homeland security applications. In: *Applied imagery pattern recognition workshop, 2008. AIPR '08. 37th IEEE*. Washington, DC, pp. 1–8.

Koh, L.P. & Wich, S.A. (2012) Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation. *Tropical Conservation Science*, **5**, 121–132.

Krizhevsky, A. (2009) Learning multiple layers of features from tiny images.

Kuhn, H.W. (1955) The Hungarian method for the assignement problem. *Naval Research Logistics Quarterly*, **2**, 83–87.

Lalonde, R., Zhang, D. & Shah, M. (2017) ClusterNet: detecting small objects in large scenes by exploiting spatio-temporal information. *Preprint arXiv*, **1704**, 1–19.

Lamba, A., Cassey, P., Segaran, R.R. & Koh, L.P. (2019) Deep learning for environmental conservation. *Current Biology*, **29**, R977–R982.

Lecun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, **521**, 436–444.

Linchant, J., Lisein, J., Semeki, J., Lejeune, P. & Vermeulen, C. (2015) Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges. *Mammal Review*, **45**, 239–252.

Maire, F., Mejias Alvarez, L. & Hodgson, A. (2015) Automating marine mammal detection in aerial images captured during wildlife surveys: a deep learning approach. In: *Advances in Artificial Intelligence: 28th Australasian Joint Conference*. Canberra, ACT, Australia, pp. 1–13.

Montabone, S., and A. Soto.2010. Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image and Vision Computing* **28**, 391–402.

Mulero-Pázmány, M., Stolper, R., Van Essen, L.D., Negro, J.J. & Sassen, T. (2014) Remotely piloted aircraft systems as a rhinoceros anti-poaching tool in Africa. *PLoS One*, **9**, 1–10.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*, **83**, 171–193.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2012) Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

QGIS Development Team. 2015. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available from: http://qgis.osgeo.org. [Accessed 27th March 2021].

R Development Core Team. (2013) *R: a language and environment for statistical computing*. Available from: https://cran.r-project.org/ Accessed 27th March 2021.

Raoult, V., Colefax, A.P., Allan, B.M., Cagnazzi, D., Castelblanco-mart, N., Ierodiaconou, D. et al. (2020) Operational protocols for the use of drones in marine animal research. *Drones*, **4**, 64. https://doi.org/10.3390/drones4040064.

Raoult, V., Tosetto, L. & Williamson, J. (2018) Drone-based high-resolution tracking of aquatic vertebrates. *Drones*, **2**, 37.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2015) You only look once: unified, real-time object detection. arXiv preprint arXiv:1506.02640.

Ren, S., He, K. & Girshick, R. (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems 28 (NIPS 2015). Montréal, Canada, pp. 91–99.

Schofield, G., Esteban, N., Katselidis, K.A. & Hays, G.C. (2019) Drones for research on sea turtles and other marine vertebrates – a review. *Biological Conservation*, **238**, 108214. https://doi.org/10.1016/j.biocon.2019.108214.

Schofield, G., Papafitsoros, K., Haughey, R. & Katselidis, K. (2017) Aerial and underwater surveys reveal temporal variation in cleaning-station use by sea turtles at a temperate breeding area. *Marine Ecology Progress Series*, **575**, 153–164.

Simard, P.Y.Y., Steinkraus, D. & Platt, J.C.C. (2003) Best practices for convolutional neural networks applied to visual document analysis. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition. Edinburgh, UK, pp. 958–963.

Stieglitz, T.C. & Dujon, A.M. (2017) A groundwater-fed coastal inlet as habitat for the Caribbean queen conch *Lobatus gigas* – an acoustic telemetry and space use analysis. *Marine Ecology Progress Series*, **571**, 139–152.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., VerCauteren, K.C., Snow, N.P. et al. (2019) Machine learning to classify animal species in camera trap images: applications in ecology. *Methods in Ecology and Evolution*, **10**, 585–590.

Tanimoto, S. & Pavlidis, T. (1975) A hierarchical data structure for picture processing. *Computer Graphics and Image Processing*, **119**, 104–119.

Thessen, A. (2016) Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, **1**, e8621.

van Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P. & Wich, S. (2015) Nature conservation drones for automatic localization and counting of animals. In: Agapito, L., Bronstein, M.M. & Rother, C. (Eds.) *Computer vision – ECCV 2014 workshops*. Zurich, Switzerland, pp. 255–270.

Villon, S., Mouillot, D., Chaumont, M., Emily, S., Subsol, G., Claverie, T. et al. (2018) A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological Informatics*, **48**, 238–244.

Wich, S.A. & Koh, P.L. (2018) *Conservation drones: mapping and monitoring biodiversity*. Oxford, UK: Oxford University Press.

Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014) How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, **27**, 3320–3328.

Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. (2017) Understanding deep learning requires rethinking generalization. In: International Conference on Learning Representations. Toulon, France, pp. 1–15.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Berlin, Germany: Springer Berlin Heidelberg.

Supporting Information

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

Supplementary Method 1. Seals and Gannet colony data collection and pre-processing.

Supplementary Methods 2. Creation of an augmented dataset required to train the convolutional neural network.

Supplementary Methods 3. Performance of the neural network on the test dataset and definition of the confidence threshold.

Supplementary Methods 4. Fitting and validation of the fix effect models.

Supplementary Results 1. Main statistics of the 20 trajectories of sea turtles visiting the fish cleaning station recorded using the combination of a convolutional neural network and the Hungarian algorithm.