Conservation Biology

Leveraging social media and deep learning to detect rare megafauna in video surveys

Laura Mannocci[1,2,3] | Sébastien Villon[2] | Marc Chaumont[3,4] | Nacim Guellati[1] | Nicolas Mouquet[1,5] | Corina Iovan[2] | Laurent Vigliola[2] | David Mouillot[1,6]

[1] MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France

[2] ENTROPIE (IRD, Université de la Réunion, Université de la Nouvelle Calédonie, CNRS, Ifremer), Laboratoire Excellence LABEX Corail, Centre IRD Nouméa, Nouméa, New Caledonia

[3] LIRMM, Univ Montpellier, CNRS, Montpellier, France

[4] University of Nîmes, Nîmes, France

[5] FRB – CESAB, Montpellier, France

[6] Institut Universitaire de France, Paris, France

Correspondence
Laura Mannocci, MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, place Eugène Bataillon, bât 24, CC093, 34095 Montpellier Cedex 5, France.
Email: laura.mannocci@gmail.com

Article impact statement: Our deep-learning method enhances the ability to rapidly detect rare megafauna in the field for monitoring and conservation purposes.

Abstract

Deep learning has become a key tool for the automated monitoring of animal populations with video surveys. However, obtaining large numbers of images to train such models is a major challenge for rare and elusive species because field video surveys provide few sightings. We designed a method that takes advantage of videos accumulated on social media for training deep-learning models to detect rare megafauna species in the field. We trained convolutional neural networks (CNNs) with social media images and tested them on images collected from field surveys. We applied our method to aerial video surveys of dugongs (*Dugong dugon*) in New Caledonia (southwestern Pacific). CNNs trained with 1303 social media images yielded 25% false positives and 38% false negatives when tested on independent field video surveys. Incorporating a small number of images from New Caledonia (equivalent to 12% of social media images) in the training data set resulted in a nearly 50% decrease in false negatives. Our results highlight how and the extent to which images collected on social media can offer a solid basis for training deep-learning models for rare megafauna detection and that the incorporation of a few images from the study site further boosts detection accuracy. Our method provides a new generation of deep-learning models that can be used to rapidly and accurately process field video surveys for the monitoring of rare megafauna.

KEYWORDS
convolutional neural networks, endangered megafauna, internet ecology, monitoring, species detection

Resumen

El aprendizaje profundo se ha convertido en una importante herramienta para el monitoreo automatizado de las poblaciones animales con video-censos. Sin embargo, la obtención de cantidades abundantes de imágenes para preparar dichos modelos es un reto primordial para las especies elusivas e infrecuentes porque los video-censos de campo proporcionan pocos avistamientos. Diseñamos un método que aprovecha los videos acumulados en las redes sociales para preparar a los modelos de aprendizaje profundo para detectar especies infrecuentes de megafauna en el campo. Preparamos algunas redes neurales convolucionales con imágenes tomadas de las redes sociales y las pusimos a prueba con imágenes tomadas en los censos de campo. Aplicamos nuestro método a los censos aéreos en video de dugongos (*Dugong dugon*) en Nueva Caledonia (Pacífico sudoccidental). Las redes neurales convolucionales preparadas con 1,303 imágenes de las redes sociales produjeron 25% de falsos positivos y 38% de falsos negativos cuando las probamos en

video-censos de campo independientes. La incorporación de un número pequeño de imágenes tomadas en Nueva Caledonia (equivalente al 12% de las imágenes de las redes sociales) dentro del conjunto de datos usados en la preparación dio como resultado una disminución de casi el 50% en los falsos negativos. Nuestros resultados destacan cómo y a qué grado las imágenes recolectadas en las redes sociales pueden ofrecer una base sólida para la preparación de modelos de aprendizaje profundo para la detección de megafauna infrecuente. También resaltan que la incorporación de unas cuantas imágenes del sitio de estudio aumenta mucho más la certeza de detección. Nuestro método proporciona una nueva generación de modelos de aprendizaje profundo que pueden usarse para procesar rápida y acertadamente los video-censos de campo para el monitoreo de megafauna infrecuente.

INTRODUCTION

Large animals such as elephants, bears, whales, and sharks (i.e., megafauna) (Moleón et al., 2020) play critical ecological roles in terrestrial and marine environments, such as regulation of food webs, transfer of nutrients and energy, ecosystem engineering, and climate regulation (Enquist et al., 2020; Geremia et al., 2019; Hammerschlag et al., 2019; Mariani et al., 2020). Megafauna includes charismatic species that promote public awareness and empathy (Ducarme et al., 2013), thus providing important socioeconomic benefits from ecotourism (Gregr et al., 2020). Megafauna also comprises some of the most threatened species worldwide (Courchamp et al., 2018) due to cumulative anthropogenic pressures from hunting, habitat loss, pollution, and climate change (Ripple et al., 2019). Together, these pressures have triggered population declines and many megafauna species are now rare or on the brink of extinction (Ceballos et al., 2020; MacNeil et al., 2020; McCauley et al., 2015; Pacoureau et al., 2021). Monitoring any changes in megafauna distribution and abundance is thus critical, but highly demanding in terms of time and money.

Video surveys from piloted and unpiloted aircraft are emerging powerful tools for collecting observations in an automated way over increasingly large spatial and temporal scales (Buckland et al., 2012; Fiori et al., 2017; Hodgson et al., 2018; Lyons et al., 2019). Yet, processing these massive amounts of images creates a major bottleneck in ecology and conservation. To overcome this limitation, deep-learning models applied to video surveys offer great promises for the automated monitoring of megafauna populations (Norouzzadeh et al., 2018; Christin et al., 2019; Eikelboom et al., 2019; Gray et al., 2019).

The strength of deep-learning models lies in their ability to automatically detect and extract features from images based on a large number of labeled examples (LeCun et al., 2015). However, obtaining sufficiently large data sets to train deep-learning models remains a major challenge. Indeed, such algorithms require hundreds of images per species in various contexts to achieve high accuracy (Villon et al., 2018; Christin et al., 2019; Ferreira et al., 2020). Building training databases is even more challenging for rare and threatened marine megafauna because most wild individuals remain in particular and often remote locations (Letessier et al., 2019). As a consequence, successful deep-learning applications for marine megafauna detection have been restricted to seasonally predictable aggregations, such as those of sea turtles in nesting grounds (Gray et al., 2019) and whales in feeding and breeding grounds (Borowicz et al., 2019).

Recently, the joint development of ecotourism, inexpensive digital devices (e.g., GoPro cameras and drones), and high-speed internet has accelerated the sharing of charismatic species images and videos on social media (Toivonen et al., 2019). Contrary to data generated by citizen science or conservation programs, which are most often actively collected, structured, and available online with clear licensing and well-described application programing interfaces (APIs), spontaneously and passively generated social media data are highly heterogeneous in terms of context and quality. The emerging field of iEcology (i.e., internet ecology) is dedicated to the use of such passively or unintentionally collected data to address ecological and environmental issues (Jarić, Correia, et al., 2020; Jarić, Roll, et al., 2020). Yet, until now, iEcology has been applied primarily to explore species occurrences and distributions in the terrestrial realm (Toivonen et al., 2019; Jarić, Correia, et al., 2020). By contrast, the extent to which and how these freely available online resources can provide novel and low-cost support for the training of deep-learning models detecting charismatic megafauna have not been investigated. We designed a new method that couples social media resources and deep-learning models to automatically detect rare megafauna on images. We applied the method to aerial video surveys of dugongs (*Dugong dugon*) in New Caledonia (southwestern Pacific).

METHODS

Overview of convolutional neural networks

Convolutional neural networks (CNNs) are deep-learning algorithms that are widely used for image classification and object detection (i.e., task of simultaneously localizing and classifying objects in images) (LeCun et al., 2015). CNNs represent by far the most commonly used category of deep-learning algorithms in ecology (Christin et al., 2019) and have been successfully applied to classify, identify, and detect animals on images (e.g., Norouzzadeh et al., 2018; Gray et al., 2019; Ferreira et al., 2020). CNNs consist of stacked groups of convolutional layers

and pooling layers that are particularly suited to process image inputs. Convolutional layers extract local combinations of pixels known as features from images. In the convolution operation, a filter defined by a set of weights computes the local weighted sum of pixels across a given image (LeCun et al., 2015).

In practice, CNNs are fed with large amounts of images in which target objects have been manually annotated so that the CNNs can be trained to associate labels with given objects. During this training phase, the weights of convolution operations are iteratively modified to obtain the desired label by minimizing the error function between the output of the CNN and the correct answer through a process called backpropagation (LeCun et al., 2015). The final output of the CNN is a probability score for each of the learned objects.

With the fast expansion of deep learning, a great number of open-source libraries and associated APIs have been created to facilitate their implementation by nonspecialists, including ecologists (Christin et al., 2019). We relied on the Tensorflow Object Detection API that is specifically designed for training and applying object detection CNNs on images (Abadi et al., 2016).

Existing approaches for dealing with limited data

Limited data are a serious impediment to the widespread use of CNNs for species identification and detection. An efficient approach for dealing with limited training data is transfer learning (i.e., the process of using a pretrained network to perform a new but similar task [West et al., 2007]). In practice, a publicly available CNN with weights pretrained on a large data set is retrained on a smaller data set containing the objects of interest. Artificial data augmentation is another common approach for training CNNs on limited data sets. It consists of applying random transformations, including rotations, translations, and contrast modifications to images, thereby creating more training examples (Zoph et al., 2019; Villon et al., 2018). Reliance on large citizen science programs whereby people upload their own images on dedicated online platforms is also extremely valuable for building species identification databases (e.g., Terry et al., 2020 for ladybirds). Yet, these approaches to overcome the paucity of data to feed deep-learning models are still seldom applied to megafauna, and their potential accuracy when implemented in concert is unknown.

Novel method for detecting rare megafauna

Obtaining large amounts of images is particularly challenging for rare, hidden, or elusive species. For these species, field surveys usually provide only sparse occurrences in a large volume of information. Our novel framework takes advantage of videos accumulated on social media for training CNNs to detect rare megafauna in field video surveys. Although videos posted on social media websites have been used to extract information on species occurrences, phenology, traits, and behavior (Jarić, Correia, et al., 2020), they have been underexploited for training species-detection models. Social media videos have the advan-

tage of focusing on species of interest and saving tedious time of watching field video surveys with scattered sightings. By providing numerous and freely available images in a variety of contexts, social media have the potential to feed a new generation of CNNs robust to the context for detecting the most threatened species.

Our framework for detecting rare megafauna on images has 6 key implementation steps (Figure 1): data collection, image preprocessing, CNN training, CNN application, CNN accuracy assessment, and CNN deployment.

In step 1, social media videos of the species of interest are collected by searching social media websites with appropriate keywords. In parallel and independently, field video surveys are conducted in the study region.

In step 2, images from social media videos are extracted and annotated (i.e., bounding boxes are manually drawn around the species of interest). Annotated images are then partitioned into independent training and testing sets and the training set is artificially augmented. Images from field video surveys are also extracted and annotated.

In step 3, a publicly available, pretrained, object-detection CNN is downloaded and retrained for species detection on the social media data set.

In step 4, the CNN is applied to predict species detections on field survey images.

In step 5, predicted detections are compared with manually annotated bounding boxes on field images and the accuracy of the CNN is assessed. Performance metrics are calculated based on the numbers of true positives (TPs), false positives (FPs), and false negatives (FNs). A TP corresponds to an overlap between a predicted and an annotated box. A predicted bounding box not corresponding to an annotated bounding box is an FP, whereas an annotated bounding box not corresponding to a predicted bounding box is an FN. Precision is the percentage of TP with respect to the predictions (Equation 1). It represents the percentage of predictions that are correct (the closest to 1, the fewest FPs).

$$Precision = TP / (TP + FP).\qquad(1)$$

Recall is the percentage of TP with respect to the annotated objects (Equation 2). It represents the percentage of observations that are actually predicted (the closest to 1, the fewest FNs).

$$Recall = TP / (TP + FN).\qquad(2)$$

Finally, the f1 score evaluates the balance between FPs and FNs. It is an overall measure of the CNN accuracy calculated as the harmonic mean of precision and recall:

$$f1\,score = 2 \times recall \times precision / (recall + precision).\qquad(3)$$

In step 6, the CNN is deployed for species detection in real time. Use of small low-power computers for deep-learning applications in real time is becoming increasingly possible. A final expert review of images after the survey can help reduce FPs.
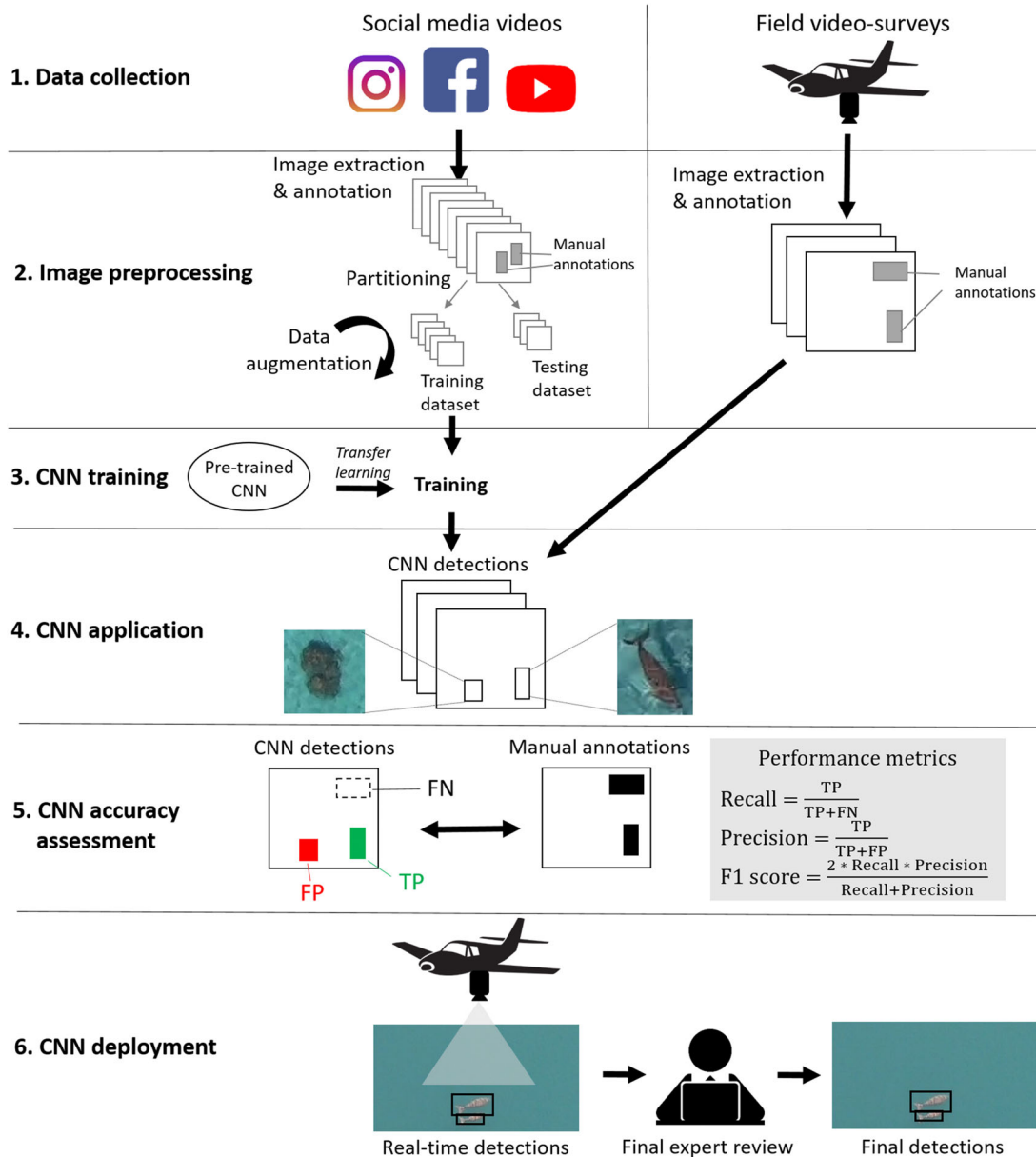
FIGURE 1 Steps in the deep-learning method that detects rare megafauna (CNN, convolutional neural network; TP, true positive; FP, false positive; FN, false negative). Social media images (gray) are used in steps 2 and 3. Field images (black) are used in steps 2, 4, and 5

Case study

We applied steps 1–5 of our framework to aerial video surveys of dugongs in New Caledonia. The dugong is a vulnerable herbivorous marine mammal restricted to coastal seagrass habitat that has become rare due to hunting, habitat degradation, and entanglement in fishing gear (Marsh & Sobtzick, 2019).

Data collection

We searched for aerial videos of dugongs on social media websites (Facebook, Instagram, and YouTube) with the keywords

dugong plus either *drone* or *aerial*. Following ethical standards, we requested the consent of owners (amateurs, nongovernmental organizations, universities, or consultancy companies) to reuse their publicly-shared videos to ensure they were fully aware of our intended use (Ghermandi & Sinclair, 2019). In total, 22 videos were gathered from social media (hereafter WEB videos). Total footage focused on dugongs was of 25 min (Table 1). These videos were acquired with drones in 6 regions throughout the dugong's Indo-Pacific range and were characterized by various resolutions and contexts (details in Appendices S1–S3).

In parallel, we acquired aerial videos from an amphibious ultralight motor plane (ULM) (AirMax SeaMax) operating

TABLE 1 Overview of social-media (WEB)* and field-video (ULM) databases for dugong detection

Database	Number of videos	Mean duration (minutes: seconds)	Total duration (hours: minutes: seconds)	Number of images	Number of images with ≥1 dugong	Mean (SD) image width (pixels)	Mean (SD) image height (pixels)	Mean (SD) bounding box width (pixels)	Mean (SD) bounding box height (pixels)
WEB	22	1: 7	0: 24: 38	1512	1303	1938 (443)	1119 (225)	155 (112)	144 (112)
ULM	57	11: 45	10: 52: 23	42464	161	2704 (0)	1520 (0)	40 (16)	39 (17)

*Details in Appendix S1.

tourist flights over the Poé Lagoon (Figure 2). The Poé Lagoon is a natural reserve (International Union for Conservation of Nature category IV) on the western coast of New Caledonia that hosts a small population of dugongs (Garrigue et al., 2008; Cleguer et al., 2015). A GoPro Hero Black 7 camera was mounted under the right wing of the ULM, pointing downward, and configured to record videos at a rate of 24 frames per second in linear field of view mode at a resolution of 2.7 K (2704 × 1520 pixels). The camera was manually triggered by the pilot before each flight. Because the ULM was not a dedicated scientific platform, its path, speed, and altitude were not standardized. In total, over 42 h of ULM videos were collected from September 2019 to January 2020 in good weather conditions.

Image preprocessing

Image annotation is a crucial prerequisite for training deep-learning models. The WEB videos were imported to a custom online application designed for image annotation (http://webfish.mbb.univ-montp2.fr/). Images were first extracted from the videos at a rate of 1 image per second before annotation. The annotation procedure consisted of manually drawing bounding boxes around identified dugongs and associating labels with these individuals. Only individuals that could be identified with full confidence as dugongs, owing to their size, shape, and color, were annotated and only dugongs for which at least three-quarters of the body was visible were annotated. Each annotation yielded a text file containing the coordinates and label of the annotated bounding box along with the corresponding image in JPEG format.

We also annotated ULM images following the same procedure. To facilitate annotation, ULM videos were visualized and the times at which dugongs were spotted were recorded. Only ULM videos that contained dugong occurrences were imported for annotation (57 ULM videos for a total duration of 11 h [Table 1]).

The annotation step led to the collection of 161 and 1303 images with at least 1 dugong for the ULM and WEB data sets, respectively (Table 1). To maximize the detection of small dugongs, we split each ULM image into 4, 1352 × 760-pixel images, which yielded 172 ULM images (original ULM images may contain more than 1 dugong). Image-splitting approaches efficiently boost detection accuracy by increasing the relative size of small objects with respect to the entire images, thereby

limiting detail losses when images are processed throughout the CNN (Unel et al., 2019).

Next, both ULM and WEB images were randomly partitioned. Eighty percent of the images were used for the training (and validation) and 20% were used for the test. Full independence between training and testing sets was ensured by selecting images belonging to different videos between the sets. We then randomly selected pools of ULM images for incorporation into the training set, equivalent to 2%, 4%, 6%, 8%, 10%, and 12% of training WEB images. These pools represented 24, 50, 76, 101, 121, and 136 ULM images, respectively (136 corresponded to the full ULM training set).

The WEB and ULM training sets were then artificially augmented by applying random transformations to images, including rotations (by −10 to +10 degrees), translations (by −10% to +10%), scaling (from 80% to 120%), horizontal and vertical flipping, and contrast modification (i.e., multiplying all image pixels with a value range of 0.6–1.4). The ULM images were subsequently added to WEB images for the training.

CNN training

We used a Faster R-CNN (Ren et al., 2016) pretrained on the COCO (common objects in context) data set (Lin et al., 2015) that was publicly available from the Tensorflow model zoo. The Faster R-CNN is a deep-learning model specialized for object detection that consists of 2 fully convolutional networks: a region proposal network, which predicts object positions along with their objectness scores, and a detection network, which extracts features from the proposed regions and outputs the bounding boxes and class labels (Ren et al., 2016). We specifically used a Faster-RCNN with a ResNet-101 backbone, a deep architecture in which layers have been reformulated as residual functions of input layers, so as to improve optimization and accuracy (He et al., 2015).

To fulfil our objectives, we performed baseline and mixed runs. The baseline run (R0) trained the Faster-RCNN with WEB images only. The objective of this run was to evaluate the accuracy of a deep-learning model trained with WEB images exclusively for dugong detection on ULM images. The mixed runs (R2–R12) trained the Faster-RCNN with WEB images mixed with a small number of randomly selected ULM images (2–12% of WEB images). The objective of the mixed runs was to assess the extent to which incorporating ULM images into
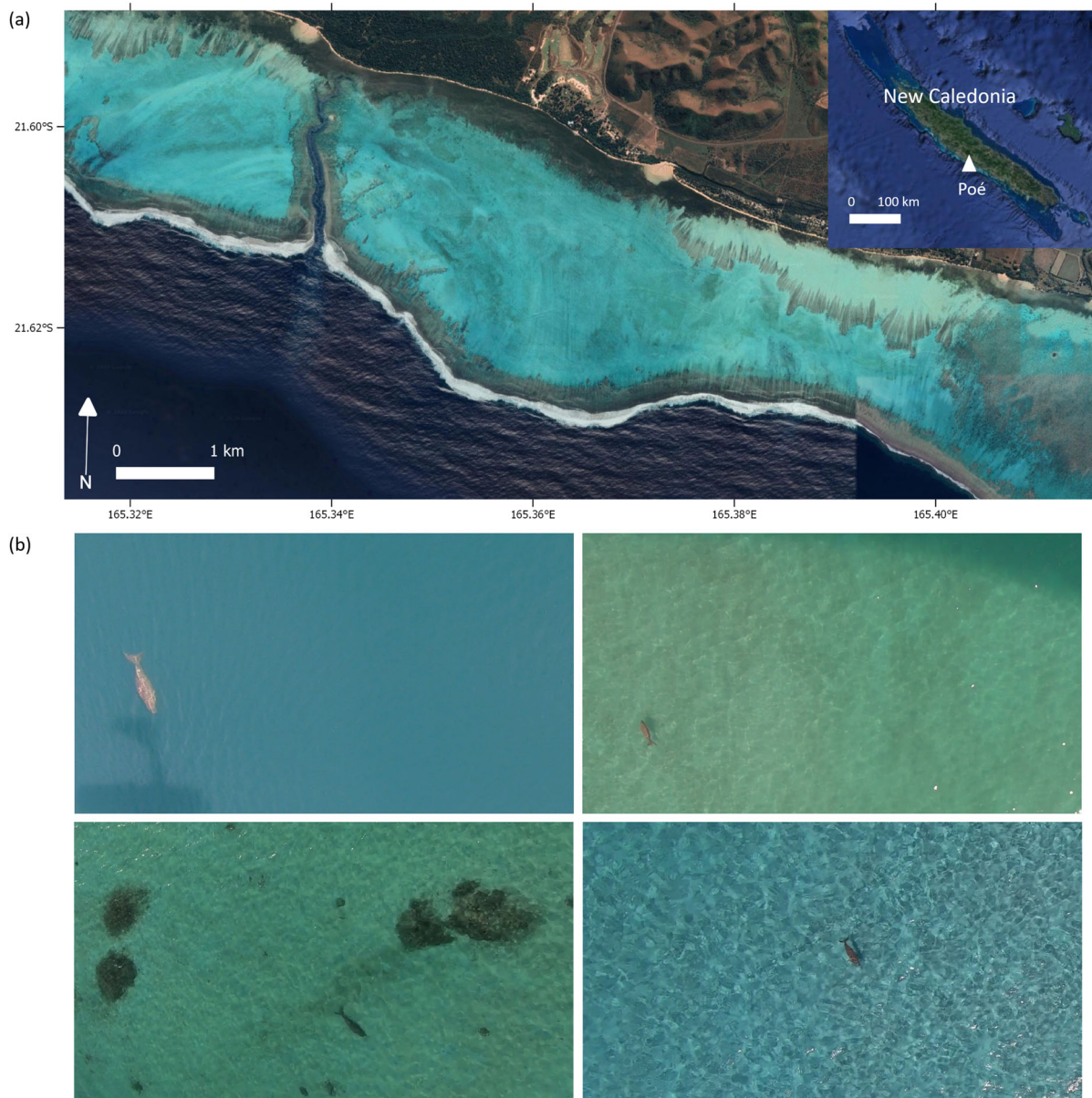
(a)

(b)

FIGURE 2 (a) Map of the Poé Lagoon study area in New Caledonia and (b) examples of dugong images collected by ULM. Imagery source for the map: Google Earth

the training data set improves capacity of dugong detection on field images. Dugong detection frameworks for the baseline and mixed runs are illustrated in Appendix S4.

We fine-tuned the pretrained Faster R-CNN (with the same network in the baseline and mixed runs). A stochastic gradient descent optimizer with a momentum of 0.9 for the loss function was applied (Qian, 1999). A learning rate of 10^{-4}, L2-regularization (lambda of 0.004), and a dropout of 50% were used to mitigate overfitting (Srivastava et al., 2014). The training was stopped after 12,000 iterations to prevent overfitting, as would be indicated by an increasing loss for the validation set (Sarle, 1995).

We used the open-source Tensorflow object detection API version 1 (Abadi et al., 2016) in Python 3 to train our CNN. The

training process lasted on average 4 h on a NVIDIA Quadro P6000 GPU with 64 GB of RAM and 24 GB of GPU memory.

CNN application and accuracy assessment

The CNN was applied to dugong detection on the test set, and its accuracy was evaluated using a 5-fold cross-validation, a common procedure for evaluating machine learning models (Wong, 2015). Specifically, the pretrained CNN was trained 5 times, each time with a different training subset, and its accuracy was evaluated 5 times, each time on an independent test subset, before averaging the results. Because minimizing FNs is more crucial than avoiding FPs for the detection of rare species (Villon et al.,
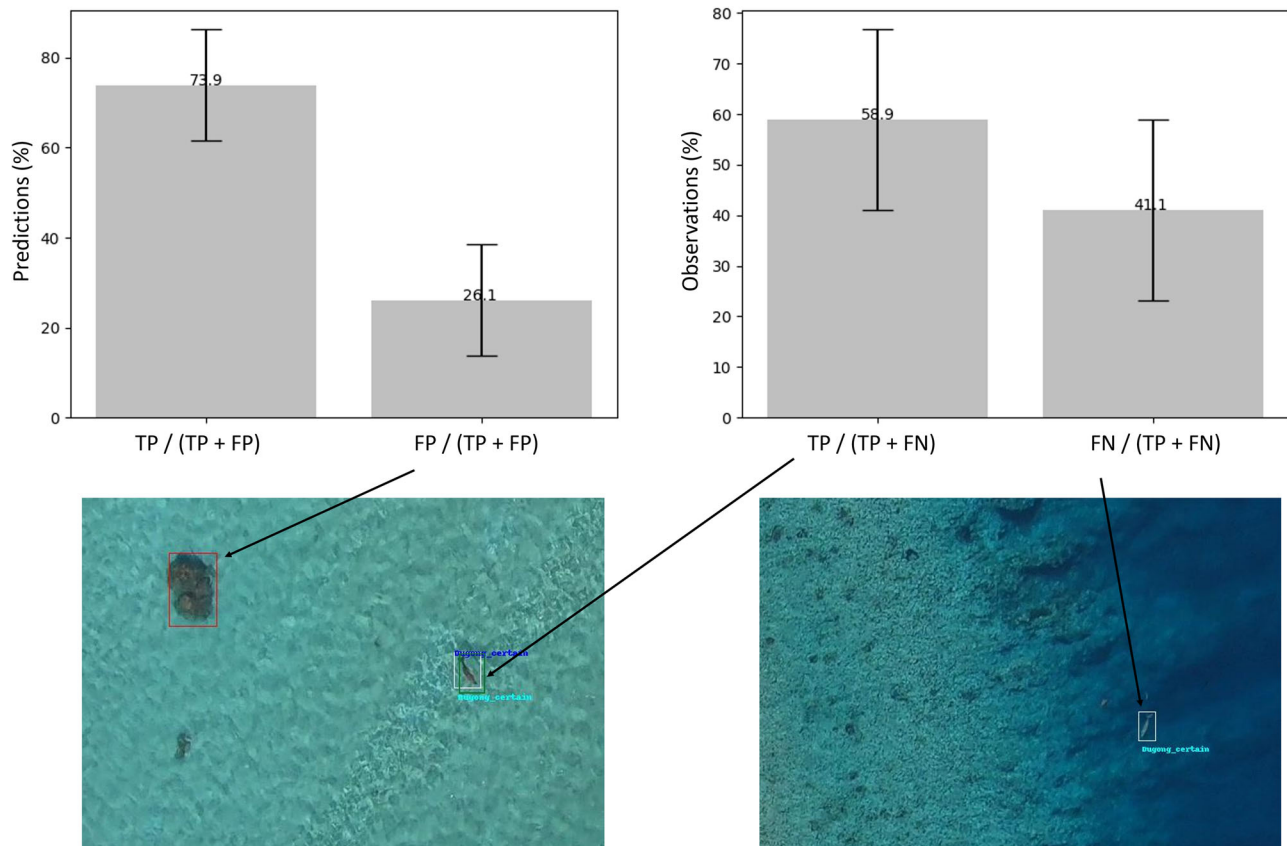
FIGURE 3 Results of dugong detection for the baseline run (trained with social media images only) applied to test field images: left graph, mean percentage of true positives (TPs) and false positives (FPs) in the predictions; right graph, mean percentage of TPs and FNs (false negatives) in observations (error bars, SD). Images are examples for a TP (green) (predicted bounding box associated with an annotated bounding box [white]), an FP (red) (predicted bounding box not corresponding to an annotated bounding box; here a coral patch), and an FN (annotated bounding box not corresponding to a predicted bounding box)

2020), we used lenient thresholds of 50% for both the confidence score of predictions and the overlap of predictions with observations. This meant that a dugong detection that was associated with a confidence score of at least 50% and that overlapped at least 50% in surface with a dugong annotation was considered a TP. For each cross-validation test subset, we calculated the number of TPs, FPs, and FNs and derived the precision, recall, and f1 score (defined above). We then computed the mean and standard deviation (SD) of these metrics across the 5 cross-validation test sets.

RESULTS

The CNNs trained with WEB images only (R0) successfully detected dugongs on ULM images. We found that 73.9% of the predictions corresponded to a manually annotated dugong (i.e., a TP), and the remaining were FPs (primarily coral patches and sun glint on the water) (Figure 3). Of the dugong annotations (i.e., observations), 41.1% did not correspond to a prediction, so were FNs (Figure 3). The baseline run yielded a mean precision of 0.75 on test ULM images (SD 0.12) and a mean recall of 0.62 (SD 0.18), corresponding to 25% and 38% of FPs and

FNs, respectively. The mean f1 score, balancing FPs and FNs, was 0.66 (SD 0.13) (Appendix S5).

The CNNs trained with both WEB and ULM images (R2-R12) showed improved dugong detection accuracy on ULM images. The mean recall increased from 0.62 (for R0) to 0.79 (for R12), corresponding to near a 50% drop in FNs (Figure 4, continuous line). The mean precision decreased slightly in the mixed runs compared with the baseline run, ranging from 0.68 (for R6) to 0.73 (for R10). The f1 score increased from R0 to R4, reaching 0.72 before stabilizing. The large SDs around the mean (e.g., 0.12–0.19 for precision) highlighted the variability between the 5 cross-validation tests.

For comparison, the accuracy of CNNs was also evaluated on WEB images. Mean performance metrics were slightly higher than when evaluated on ULM images; precision was 0.79, recall was 0.82, and the f1 score was 0.80 for R0 (Figure 4, dashed line). For the mixed runs, the mean recall increased slightly (0.88 for R12) and the precision ranged from 0.74 to 0.79, with some variability among the cross-validation sets.

Precision–recall curves calculated for various prediction confidence thresholds illustrated the critical trade-off between FNs and FPs (Figure 5). The closer this threshold was to 50%
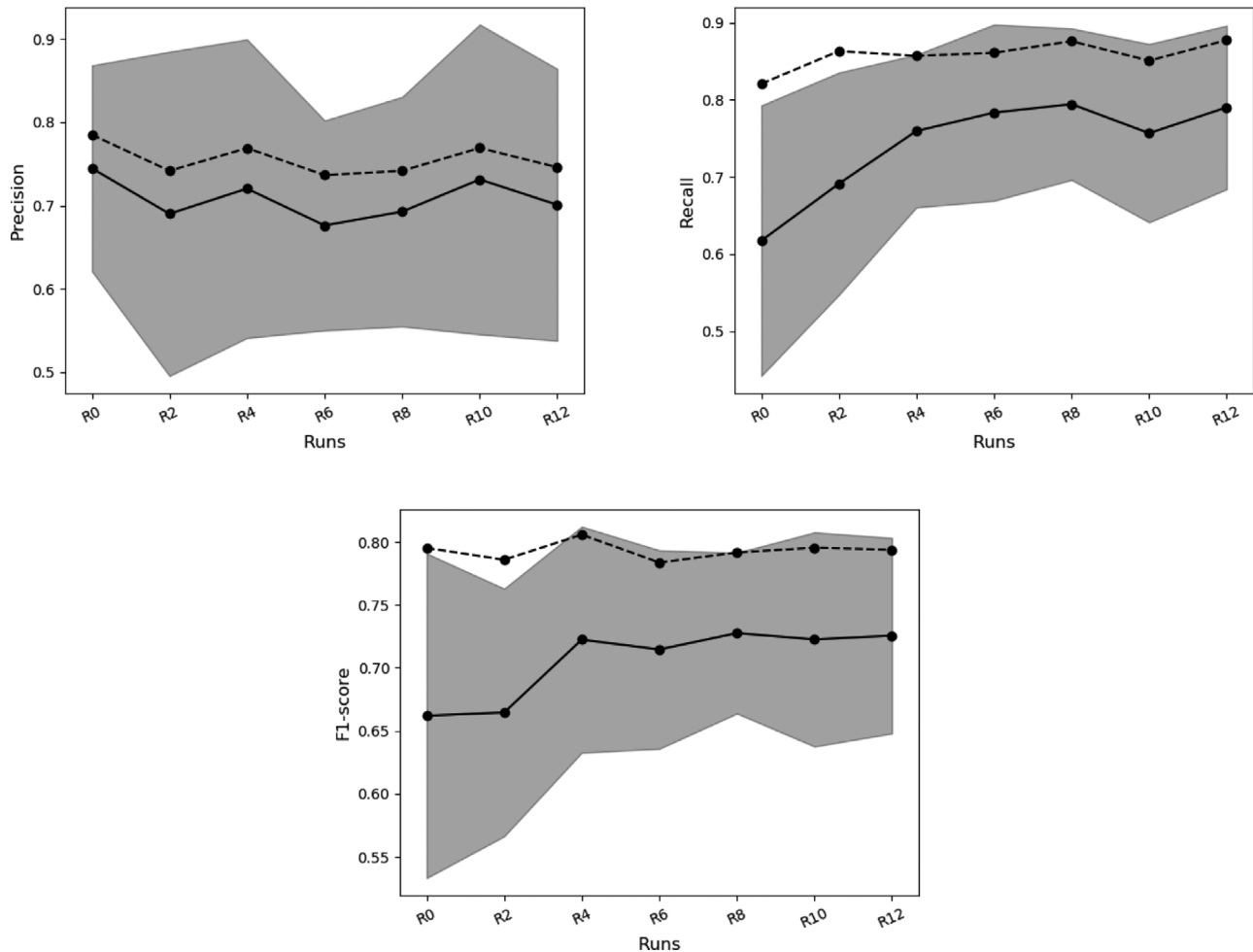
FIGURE 4 Mean precision, recall, and f1 score (balance between false positives and false negatives) of convolutional neural networks detecting dugongs in test field (ULM) images (continuous line) versus test social media (WEB) images (dashed line) for all runs (R0, training with WEB images only; R2–R12, training with WEB images mixed with a number of ULM images equivalent to 2–12% of WEB images) (shading, SD of metrics evaluated on test ULM images). Values of all performance metrics are in Appendix S5
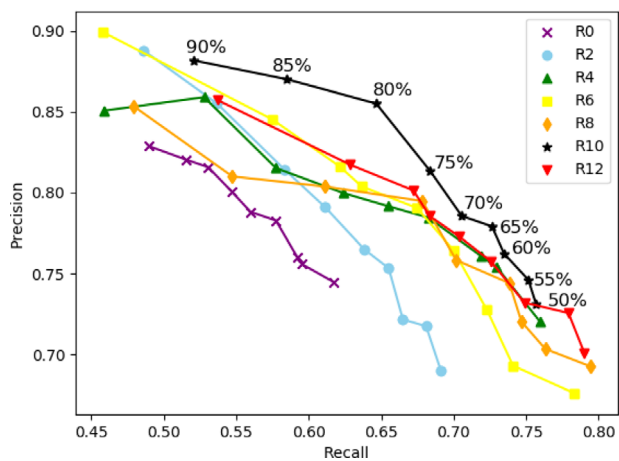

FIGURE 5 Mean precision–recall curves for all runs (R0–R12) of deep-learning models calculated based on the test field images of dugongs collected at Poé Lagoon (dot, threshold value from 50% to 90%). Recall is the metric to maximize when detecting rare species, such as the dugong (lower right corner of graph)

(selected in this study), the more dugongs were detected, increasing the recall at the cost of precision. Figure 5 highlights the increased recall in the mixed runs relative to the baseline run.

DISCUSSION

Social media potential for training rare megafauna detection models

Obtaining large image data sets is a major bottleneck for the automated monitoring of species with deep-learning models (Christin et al., 2019). This issue is exacerbated for marine megafauna species that occur in low numbers, range over vast areas, and spend most of their time underwater, hence offering few sightings. Our novel framework leverages social media resources for building training data sets and CNNs that can be used to detect rare megafauna on videos and images. Our approach proved effective when applied to aerial video surveys

of dugongs. The CNNs trained with 1303 images collected from social media yielded only 25% FPs and 38% FNs when tested on ULM images collected in a novel region (New Caledonia), with another recording device (a GoPro), and in a different context (Poé Lagoon). The addition of 136 ULM images from New Caledonia to the training data set led to a nearly 50% reduction in the number of FNs, which is critical when surveying rare species. The CNNs still yielded imperfect results when tested on independent WEB image sets, highlighting the need for intrinsic CNN improvements to maximize detection accuracy across image sources. These results show that social media offer a solid basis to train CNNs for the automated detection of rare species and that the addition of few field images into the training data set can further boost the detection capacity.

At a time where digital technologies shape contemporary society, social media have become an unprecedented source of information for environmental research (Ghermandi & Sinclair, 2019). Images and videos accumulated on social media are increasingly exploited to address biodiversity conservation challenges (Toivonen et al., 2019) and generate ecological knowledge within the broader iEcology approach (Jarić, Correia, et al., 2020; Jarić, Roll, et al., 2020). This process is referred to as passive or opportunistic crowdsourcing because materials spontaneously generated by humans are posted online and shared independent of formal citizen science programs (Ghermandi & Sinclair, 2019). To our knowledge, we are the first to exploit the potential of iEcology for training algorithms, as most deep-learning applications have relied on active crowdsourcing through structured citizen science programs (Van Horn et al., 2015, 2018; Terry et al., 2020).

Megafauna species are particularly suited to this purpose because they are extremely popular among the public, eliciting abundant social media activity. Leveraging social media resources allowed the collection of an unprecedented training database for dugong, owing to its sample size (1303 images) and broad geographical reach (6 regions spanning the dugong's range) (map in Appendix S3). Because of their opportunistic nature and lack of precise geolocation information, social media videos inherently have limited value for species monitoring, but our results highlight their great potential for training CNNs that can detect species even in different contexts.

Enhanced detection robustness

Species identification and detection are largely influenced by the individual's surroundings, so that deep-learning models are mostly valid locally (Villon et al., 2018; Ferreira et al., 2020). In the marine environment, weather conditions and habitats (e.g., coral, sand, seagrass), but also acquisition characteristics (e.g., altitude), strongly influence the image quality and contents. The 2 fundamental differences between ULM and WEB videos were their acquisition platform (a GoPro attached to a manned aircraft vs. drones) and location (New Caledonia vs. 6 other regions). This resulted in WEB images being more heterogeneous than ULM images, both in terms of their acquisition characteristics and conditions. This heterogeneity of WEB

images enhanced the CNN robustness by allowing the detection of dugongs in various contexts.

The addition of modest amounts of ULM images to the training data set (equivalent to up to 12% of WEB images) allowed a decrease in the number of FNs to 21% (Figure 4). This improvement in the dugong detection capacity was likely due to the incorporation of contextual elements specific to the Poé Lagoon, such as distinct light and habitat features. The addition of ULM images also increased the variability around precision because FPs were detected on unseen elements of the background (e.g., sun glint on the water). Overall, these results suggest that CNNs trained with various sources of social media images can successfully detect species in new locations where conditions potentially differ from those of the training data set. This ability of CNNs to shift domains (i.e., their transferability) is particularly valued in ecological applications of deep learning (Schneider et al., 2020; Terry et al., 2020) and brings hope for the development of models at the global scale.

Precision–recall trade-offs for rare species detection

Avoiding missed detections (i.e., FNs) is critical when studying species with scarce occurrences, such as threatened marine megafauna. The dugong is characterized by low abundance and declining populations worldwide and is currently recognized as vulnerable to extinction (Marsh & Sobtzick, 2019). As recommended for rare species, our approach aimed to minimize the number of FNs at the expense of FPs (Villon et al., 2020). To do so, we applied lenient thresholds of 50% for both the confidence of predictions and their overlap with observations, meaning that a combination of a 50% prediction confidence score and a 50% overlap of the prediction with the observation was sufficient to assign a TP. Indeed, when the objective is to detect rare species on images, recall is the metric to maximize (bottom-right corner of Figure 5). Similarly, Gray et al. (2019) tuned their sea turtle detection model to maximize recall at the cost of precision. The drawback of this approach is that detections in a novel unannotated image data set need to be reviewed by humans in order to exclude potential FPs (e.g., coral patches).

Model improvement perspectives

Training CNNs with limited data sets of rare species incontestably results in lower performances than with very large data sets of common species. However, standards are different when studying abundant, gregarious, and accessible animals (e.g., many African mammals) versus rare marine mammals that mostly occur in low numbers, small groups, and remote places. To reduce FPs, images of corals could be incorporated into the training data set so that the model explicitly learns this class. However, this would require additional annotation and processing time that, in practice, may be incompatible with the pressing monitoring needs for species with conservation concerns.

Moreover, the reduction of FPs is not the main challenge in conservation because experts can verify the detections and correct them if necessary. By contrast, reducing FNs to avoid missed occurrences is of high priority. Toward this aim, we suggest more images from social media or field surveys be included in the training data set. Thresholds could also be decreased below 50%, but at the cost of FPs.

Implications for rare megafauna monitoring

Together, deep learning and social media provide a potential breakthrough for video-based monitoring of rare megafauna populations. First, social media can provide many times more images than aerial surveys in a study area (here, 1303 WEB v. 161 ULM images) in different contexts, boosting the size and diversity of training data sets. Our findings show that incorporating local field images in the training data set is still required to decrease FNs and that time needs to be devoted to annotate some of these field images to improve overall detection accuracy. Second, social media offer the possibility to gather images from many more locations than would be achieved with conventional surveys, expanding the environments in which the species may be found (e.g., seagrass, barrier reefs, open water), thereby increasing the transferability of models. As such, training data sets derived from social media are paving the way toward global deep-learning models capable of detecting a given species in any location. Importantly, harnessing social media data collected by nature enthusiasts residing or traveling near biodiversity hotspot locations also saves time, labor, and money (Ghermandi & Sinclair, 2019). Social-media trained models also represent an unparalleled opportunity to get the public involved in ecological research applications and engaged in conservation.

We found that social media provided a rich, underexplored, and ever-increasing source of information for training deep-learning models able to successfully detect a rare, charismatic species. Our approach is generally applicable to other large marine and terrestrial vertebrates that elicit social media activity. Combining deep-learning models and social media not only helps build robust species detection tools that are applicable in various contexts, but also has the potential to save substantial survey resources. Our method extends the value of iEcology for the production of a new generation of accurate and global models able to continuously process video surveys from a wide variety of sources to allow monitoring rapid biodiversity changes in near real time.

ORCID
Laura Mannocci https://orcid.org/0000-0001-8147-8644

LITERATURE CITED

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, A., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., … Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv.* http://arxiv.org/abs/1603.04467

Borowicz, A., Le, H., Humphries, G., Nehls, G., Höschle, C., Kosarev, V., & Lynch, H. J. (2019). Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLoS ONE*, *14*, e0212532.

Buckland, S. T., Burt, M. L., Rexstad, E. A., Mellor, M., Williams, A. E., & Woodward, R. (2012). Aerial surveys of seabirds: the advent of digital methods. *Journal of Applied Ecology*, *49*, 960–967.

Ceballos, G., Ehrlich, P. R., & Raven, P. H. (2020). Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences*, *117*, 13596–13602.

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*, 1632–1644.

Cleguer, C., Grech, A., Garrigue, C., & Marsh, H. (2015). Spatial mismatch between marine protected areas and dugongs in New Caledonia. *Biological Conservation*, *184*, 154–162.

Courchamp, F., Jaric, I., Albert, C., Meinard, Y., Ripple, W. J., & Chapron, G. (2018). The paradoxical extinction of the most charismatic animals. *PLoS Biology*, *16*, e2003997.

Ducarme, F., Luque, G. M., & Courchamp, F. (2013). What are "charismatic species" for conservation biologists? *BioSciences Master Reviews*, *10*, 1–8.

Eikelboom, J. A. J., Wind, J., Van De Ven, E., Kenana, L. M., Schroder, B., De Knegt, H. J., Van Langevelde, F., & Prins, H. H. T. (2019). Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, *10*, 1875–1887.

Enquist, B. J., Abraham, A. J., Harfoot, M. B. J., Malhi, Y., & Doughty, C. E. (2020). The megabiota are disproportionately important for biosphere functioning. *Nature Communications*, *11*, 699.

Ferreira, A. C., Silva, L. R., Renna, F., Brandl, H. B., Renoult, J. P., Farine, D. R., Covas, R., & Doutrelant, C. (2020). Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, *11*, 1072–1085.

Fiori, L., Doshi, A., Martinez, E., Orams, M. B., & Bollard-Breen, B. (2017). The use of unmanned aerial systems in marine mammal research. *Remote Sensing*, *9*, 543.

Garrigue, C., Patenaude, N., & Marsh, H. (2008). Distribution and abundance of the dugong in New Caledonia, southwest Pacific. *Marine Mammal Science*, *24*, 81–90.

Geremia, C., Merkle, J. A., Eacker, D. R., Wallen, R. L., White, P. J., Hebblewhite, M., & Kauffman, M. J. (2019). Migrating bison engineer the green wave. *Proceedings of the National Academy of Sciences*, *116*, 25707–25713

Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*, *55*, 36–47.

Gray, P. C., Fleishman, A. B., Klein, D. J., Mckown, M. W., Bézy, V. S., Lohmann, K. J., & Johnston, D. W. (2019). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, *10*, 345–355.

Gregr, E. J., Christensen, V., Nichol, L., Martone, R. G., Markel, R. W., Watson, J. C., Harley, C. D. G., Pakhomov, E. A., Shurin, J. B., & Chan, K. M. A. (2020).

Cascading social-ecological costs and benefits triggered by a recovering keystone predator. *Science*, *368*, 1243–1247.

Hammerschlag, N., Schmitz, O. J., Flecker, A. S., Lafferty, K. D., Sih, A., Atwood, T. B., Gallagher, A. J., Irschick, D. J., Skubel, R., & Cooke, S. J. (2019). Ecosystem function and services of aquatic predators in the anthropocene. *Trends in Ecology & Evolution*, *34*, 369–383.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv*. http://arxiv.org/abs/1512.03385

Hodgson, J. C., Mott, R., Baylis, S. M., Pham, T. T., Wotherspoon, S., Kilpatrick, A. D., Raja Segaran, R., Reid, I., Terauds, A., & Koh, L. P. (2018). Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, *9*, 1160–1167.

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., Firth, J. A., Gaston, K. J., Jepson, P., Kalinkat, G., Ladle, R., Soriano-Redondo, A., Souza, A. T., & Roll, U. (2020). iEcology: Harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*, 630–639.

Jarić, I., Roll, U., Arlinghaus, R., Belmaker, J., Chen, Y., China, V., Douda, K., Essl, F., Jähnig, S. C., Jeschke, J. M., Kalinkat, G., Kalous, L., Ladle, R., Lennox, R. J., Rosa, R., Sbragaglia, V., Sherren, K., Šmejkal, M., Soriano-Redondo, A., … Correia, R. A. (2020). Expanding conservation culturomics and iEcology from terrestrial to aquatic realms. *PLOS Biology*, *18*, e3000935.

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

Letessier, T. B., Mouillot, D., Bouchet, P. J., Vigliola, L., Fernandes, M. C., Thompson, C., Boussarie, G., Turner, J., Juhel, J.-B., Maire, E., Caley, M. J., Koldewey, H. J., Friedlander, A., Sala, E., & Meeuwig, J. J. (2019). Remote reefs and seamounts are the last refuges for marine predators across the Indo-Pacific. *PLOS Biology*, *17*, e3000366.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: common objects in context. *arXiv*. http://arxiv.org/abs/1405.0312

Lyons, M. B., Brandis, K. J., Murray, N. J., Wilshire, J. H., Mccann, J. A., Kingsford, R. T., & Callaghan, C. T. (2019). Monitoring large and complex wildlife aggregations with drones. *Methods in Ecology and Evolution*, *10*, 1024–1035.

Macneil, M. A., Chapman, D. D., Heupel, M., Simpfendorfer, C. A., Heithaus, M., Meekan, M., Harvey, E., Goetze, J., Kiszka, J., Bond, M. E., Currey-Randall, L. M., Speed, C. W., Sherman, C. S., Rees, M. J., Udyawer, V., Flowers, K. I., Clementi, G., Valentin-Albanese, J., Gorham, T., … Cinner-Show, J. E. (2020). Global status and conservation potential of reef sharks. *Nature*, *583*, 801–806.

Mariani, G., Cheung, W. W. L., Lyet, A., Sala, E., Mayorga, J., Velez, L., Gaines, S. D., Dejean, T., Troussellier, M., & Mouillot, D. (2020). Let more big fish sink: Fisheries prevent blue carbon sequestration—Half in unprofitable areas. *Science Advances*, *6*, eabb4848.

Marsh, H., & Sobtzick, S. (2019). *Dugong dugon (amended version of 2015 assessment)*. The IUCN red list of threatened species 2019. https://www.iucnredlist.org/en

Mccauley, D. J., Pinsky, M. L., Palumbi, S. R., Estes, J. A., Joyce, F. H., & Warner, R. R. (2015). Marine defaunation: Animal loss in the global ocean. *Science*, *347*, 1255641.

Moleón, M., Sánchez-Zapata, J. A., Donázar, J. A., Revilla, E., Martín-López, B., Gutiérrez-Cánovas, C., Getz, W. M., Morales-Reyes, Z., Campos-Arceiz, A., Crowder, L. B., Galetti, M., González-Suárez, M., He, F., Jordano, P., Lewison, R., Naidoo, R., Owen-Smith, N., Selva, N., Svenning, J.-C., … Tockner, K. (2020). Rethinking megafauna. *Proceedings of the Royal Society B: Biological Sciences*, *287*, 20192643.

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

Pacoureau, N., Rigby, C. L., Kyne, P. M., Sherley, R. B., Winker, H., Carlson, J. K., Fordham, S. V., Barreto, R., Fernando, D., Francis, M. P., Jabado, R. W., Herman, K. B., Liu, K.-M., Marshall, A. D., Pollom, R. A., Romanov, E. V., Simpfendorfer, C. A., Yin, J. S., Kindsvater, H. K., & Dulvy, N. K. (2021). Half a century of global decline in oceanic sharks and rays. *Nature*, *589*, 567–571.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, *12*, 145–151.

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: towards real-time object detection with region proposal networks. *arXiv*. http://arxiv.org/abs/1506.01497

Ripple, W. J., Wolf, C., Newsome, T. M., Betts, M. G., Ceballos, G., Courchamp, F., Hayward, M. W., Valkenburgh, B., Wallach, A. D., & Worm, B. (2019). Are we eating the world's megafauna to extinction? *Conservation Letters*, *12*, e12627.

Sarle, W. (1995). *Stopped training and other remedies for overfitting*. Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics.

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. *Ecology & Evolution*, *10*, 3503–3517.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Terry, J. C. D., Roy, H. E., & August, T. A. (2020). Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods in Ecology and Evolution*, *11*, 303–315.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298–315.

Unel, F. O., Ozkalayci, B. O., & Cigla, C. (2019). *The power of tiling for small object detection*. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA. https://ieeexplore.ieee.org/document/9025422/

Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeirotis, P., Perona, P., & Belongie, S. (2015). *Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., & Belongie, S. (2018). The iNaturalist species classification and detection dataset. *arXiv*. http://arxiv.org/abs/1707.06642

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

Villon, S., Mouillot, D., Chaumont, M., Subsol, G., Claverie, T., & Villéger, S. (2020). A new method to control error rates in automated species identification with deep learning algorithms. *Scientific Reports*, *10*, 10972.

West, J., Ventura, D., & Warnick, S. (2007). *Spring research presentation: A theoretical foundation for inductive transfer*. Brigham Young University, College of Physical and Mathematical Sciences. https://web.archive.org/web/20070801120743/http://cpms.byu.edu/springresearch/abstract-entry?id=861

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*, 2839–2846.

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., & Le, Q. V. (2019). Learning data augmentation strategies for object detection. *arXiv*. http://arxiv.org/abs/1906.11172

SUPPORTING INFORMATION

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