Revised Manuscript without Changes Marked (without Author Details) Click here to view linked References

Putting eagle rays on the map by coupling aerial video-surveys and deep learning

Highlights

- Efficient techniques are needed to monitor vulnerable elasmobranchs in space and time.
- Deep learning applied to images is a powerful tool for automated wildlife monitoring.
- Our deep learning model successfully detected 92% of eagle rays on images.
- This study is a step forward for ray monitoring in coral reef ecosystems.

Keywords

Automated species detection, convolutional neural networks, coral reefs, elasmobranchs, New-Caledonia

Abstract

Reliable and efficient techniques are urgently needed to monitor elasmobranch populations that face increasing threats worldwide. Aerial video-surveys provide precise and verifiable observations for the rapid assessment of species distribution and abundance in coral reefs, but the manual processing of videos is a major bottleneck for timely conservation applications. In this study, we applied deep learning for the automated detection and mapping of vulnerable eagle rays from aerial videos. A light aircraft dedicated to touristic flights allowed us to collect 42 hours of aerial video footage over a shallow coral lagoon in New Caledonia (Southwest Pacific). We extracted the videos at a rate of one image per second before annotating them, yielding 314 images with eagle rays. We then trained a convolutional neural network with 80% of the eagle ray images and evaluated its accuracy on the remaining 20% (independent data sets). Our deep learning model detected 92% of the annotated eagle rays in a diversity of habitats and acquisition conditions across the studied coral lagoon. Our study offers a potential breakthrough for the monitoring of ray populations in coral reef ecosystems by providing a fast and accurate alternative to the manual processing of aerial videos. Our deep learning approach can be extended to the detection of other elasmobranchs and applied to systematic aerial surveys to not only detect individuals but also estimate species density in coral reef habitats.

1. Introduction

Elasmobranchs, a subclass of cartilaginous fishes composed of sharks, rays, skates and sawfish, are among the most endangered animal taxa in the oceans (Dulvy et al. 2021). These species are intrinsically sensitive to human activities due to their slow growth rate and limited reproduction capacity, preventing them from quickly recovering from overexploitation (Pacoureau et al., 2021). Elasmobranchs are primarily threatened by targeted fisheries and incidental catches, although habitat degradation is a growing threat for coastal species (Dulvy et al. 2021; Yan et al. 2021). Within the 1,199 species of elasmobranchs assessed by the IUCN in 2021, 10.4% were listed as near-threatened, 15% as vulnerable, 10.1% as endangered (compared to 4.1% in 2010), 7.5% as critically endangered, and 12.9% as data deficient (Dulvy et al. 2021). Rays are even more threatened than sharks with 36% of all species threatened compared to 31.2% (Dulvy et al. 2021). Currently, the limited knowledge and monitoring of elasmobranch abundance and distribution is a major impediment to the implementation of targeted conservation measures (Jabado et al., 2018). To fill these knowledge gaps, new techniques are urgently needed to efficiently and rapidly monitor threatened elasmobranchs in space and time in order to identify their key habitats and provide abundance estimates at the basis of IUCN assessments.

Video-surveys from drones or light aircraft are increasingly used to assess the distribution, behavior and abundance of marine megafauna (Hodgson, Kelly, and Peel 2013; Kelaher et al. 2020; Schofield et al. 2017). Such digital surveys are particularly suited to study sharks and rays in coral lagoons where clear and shallow waters facilitate their detection (Kiszka et al., 2016; Rieucau et al., 2018). Video-surveys offer important advantages over traditional observer-based surveys by generating precise and verifiable observations that are free from observer fatigue and subjectivity (Colefax, Butcher, and Kelaher 2018; Kelaher et al. 2019). However, manual video analysis is a major bottleneck for timely conservation applications, as visualizing hours of footage is both extremely time-consuming and error-prone (Ditria, 2020; Norouzzadeh et al., 2018; Villon et al., 2018). Deep learning algorithms offer great promises to overcome this limitation by allowing the automated identification and detection of species on images (Christin, Hervet, and Lecomte 2019; Norouzzadeh et al. 2018; Torney et al. 2019; Eikelboom et al. 2019). Such models have been successfully applied for the detection of sea turtles (Dujon et al., 2021; Gray et al., 2018), dugongs (Mannocci et al., 2021), pinnipeds (Dujon et al., 2021; Padubidri et al., 2021) and whales (Gray et al. 2019; Guirado et al. 2019). Although there are a few applications for elasmobranchs, these are generally dedicated to monitoring shark risks (Gorkin et al., 2020) rather than conservation objectives requiring abundance and distribution estimates. Accurate deep learning models would drastically increase the efficiency of aerial monitoring for these threatened species.

In this study, we combined aerial video-surveys and deep learning to detect eagle rays and map their distribution throughout a lagoon in New Caledonia, Southwest Pacific. New Caledonia hosts exceptional coral reefs and lagoons, which form one of the three most extensive reef systems in the world (Ceccarelli et al., 2013). Eagle rays are conspicuous rays of the Myliobatidae family that are easily spotted from the surface owing to their relatively large size and characteristic diamond shape (Last, White, and Pogonoski 2010), making them good candidates for automated detection on aerial images. Two species of Myliobatidae are present in New Caledonia, the spotted eagle ray, *Aetobatus narinari* which is common, and the rarer mottled eagle ray, *Aetomylaeus maculatus* (Fricke, Kulbicki, and Wantiez 2011). These species have been classified as globally endangered by the IUCN in 2020 (Dulvy et al. 2020; Rigby et al. 2020), stressing the urgent need to monitor the trends of their populations to feed global indicators like the Living Planet Index (Pacoureau et al., 2021). We trained and evaluated a deep learning model to automatically detect eagle rays on aerial images collected from an ultra-light motor plane (ULM). We then mapped their distribution across the studied lagoon. Our study unravels the potential of deep learning applied to aerial surveys for monitoring the distribution of vulnerable elasmobranchs in coral reefs.

2. Material and methods

2.1. Video data collection

Video-surveys were conducted from an amphibious ULM (AirMax SeaMax) operating touristic flights over the Poé lagoon on the Western coast of New Caledonia (Supplementary Figure A). This lagoon is shallower than 5 m and characterized by shallow reef, seagrass and sandy habitats. The barrier reef includes three deep passes and channels reaching 30 m. Part of the Poé lagoon was declared as a natural reserve (IUCN category IV) in 2006 and is located within the broader South Province Park created in 2009 and the UNESCO World Heritage area established in 2008.

A GoPro Hero Black 7 camera was mounted under the right wing of the ULM, pointing downward. The camera was 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 (2,704 x 1,520 pixels) with integrated image stabilization. The camera was manually triggered by the pilot before each flight. Telemetry data, including GPS coordinates and altitudes, were also recorded by the GoPro along each flight (at a rate of 8 to 12 positions per second). The mean altitude across all flights was 152 m (standard deviation SD= 52 m). At this altitude the image covered a mean surface area of 161 m × 287 m corresponding to a ground sampling distance of 11 cm per pixel. In total, over 42 hours of videos representing 36 fly days were collected from September 2019 to January 2020 in good weather conditions.

2.2. Image annotation

Image annotation is a crucial prerequisite before applying deep learning models (Gray et al., 2018; Norouzzadeh et al., 2018; Villon et al., 2020). All videos were first visualized by a team of students who recorded the times at which they spotted eagle rays (and other megafauna species). Videos that contained at least one eagle ray (representing 114 videos from a total of 228 (Table 1)) were then imported into a custom online application (http://webfish.mbb.univ-montp2.fr/) (Supplementary Figure B). Next, images were extracted from all videos at a rate of one image per second, representing a compromise between image diversity and annotation time.

The annotation procedure consisted in manually drawing rectangle bounding boxes around identified eagle rays and associating labels to these individuals. Only individuals that could be identified without ambiguity as eagle rays, owing to their large size, diamond shape and dark colour, were annotated. Although *Aetomylaeus maculatus* is generally smaller than *Aetobatus narinari*, their color patterns are similar (light spots on a dark disc) so they are easily confused *in situ*. The presence of a long spine near the tail's base of *A. narinari* can help to differentiate it from *A. maculatus* which has a long but spineless tail (Froese and Pauly 2021). Since we could not distinguish one or the other species on aerial videos we built a generic eagle ray (i.e., Myliobatidae) detector, although most sightings were likely of *A. narinari* which is much more common in New Caledonia (Fricke et al., 2011). Each annotation yielded a text file containing the coordinates and label of the bounding box, along with the corresponding image in jpeg format (examples of images are provided in Supplementary Figure C).

2.3. Eagle ray detection model

We used a convolutional neural network (CNN), a class of deep learning models that is widely applied for image classification and object detection, i.e., the task of simultaneously localizing and classifying objects on images (LeCun, Bengio, and Hinton 2015). CNNs represent by far the most commonly used category of deep learning models in ecology (Christin, Hervet, and Lecomte 2019). They are formed by 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 over the three colour channels of a given image (LeCun, Bengio, and Hinton 2015). In practice, CNNs are fed with large amounts of images in which target objects have been manually annotated so they can be trained to associate labels to a given object. During this training phase, the weights are iteratively modified to obtain the desired answer by minimizing the error function between the output of the CNN and the correct answer through a process called backpropagation (LeCun, Bengio, and Hinton 2015). The final output of the CNN is a confidence score for each of the learned objects.

We selected a Faster R-CNN network (Ren et al. 2016) publicly available from the Tensorflow model zoo and tuned it for eagle ray detection on aerial images. The Faster R-CNN is a deep learning algorithm specialized for object detection that consists of two fully-convolutional networks: (1) a region proposal network, which predicts object positions along with their 'objectness' scores and (2) a detection network, which extracts features from the proposed regions and provides class labels for the bounding boxes. 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, leading to better optimization and increased accuracy. Our eagle ray detection framework followed the three main steps detailed below: 1) Image pre-processing, 2) Model training and 3) Model accuracy assessment. The eagle ray detection framework is illustrated in Figure 1.

2.4. Image pre-processing

A total of 314 ULM images containing at least one eagle ray (representing 372 individual encounters) were extracted out of the 79,325 collected images (Table 1)**.** Bounding boxes surrounding eagle rays spanned on average 25 x 25 pixels (pi) on the 2,704 x 1,520 pi images, corresponding to a ratio of 0.0002 between the bounding box area and the image area. To maximize the detection of small eagle rays on ULM images, we split each image into four images with half the original size (i.e., 1,352 x 760 pi). This yielded 308 images with eagle rays (353 individual encounters), as rays located across image boundaries were lost. Image splitting approaches are known to efficiently boost detection accuracy by increasing the relative pixel area of small objects with respect to the entire images, thereby limiting detail losses when images are processed throughout the network (Unel, Ozkalayci, and Cigla 2019) .

Next, images were randomly partitioned, using 80% of images for the training (and validation) subset (corresponding to approximately 250 images) and 20% of images (approx. 60 images) for the testing subset. Full independence between subsets was ensured by selecting images belonging to different videos between the subsets. The training subset was then artificially augmented by applying random transformations to images, including rotations (by -10 to +10 degrees), translations (by -10 to +10 %), scaling (by 80 to 120%), horizontal and vertical flipping, and contrast modification (i.e., multiplying all image pixels with a value ranging from 0.6 to 1.4). Artificial data augmentation is a particularly efficient technique for improving the generalization performance and accuracy of object detection models (Zoph et al., 2019).

2.5 Model training

We initialized our Faster R-CNN with pre-trained weights based on the COCO (Common Objects in Context) dataset (Lin et al., 2015) downloaded from the Tensorflow model zoo. This process of applying previously learned knowledge to solve a new problem, called transfer learning, improves model accuracy and generalization when a limited annotated dataset is available (Chen, Zhang, and Ouyang 2018). We then trained the Faster R-CNN using a stochastic gradient descent optimizer with a momentum of 0.9 for the loss function (Qian, 1999). We applied a learning rate of 10^{-3}, a L2 regularization (with a lambda of 0.004), and a dropout of 50% to mitigate overfitting (Srivastava et al., 2014). The training was stopped after 50,000 iterations to prevent overfitting as indicated by an increasing loss function for the validation subset (Sarle, 1995).

2.6 Model accuracy assessment

The Faster R-CNN was then applied for eagle ray detection on the test subset and its accuracy was evaluated using a 5-fold cross-validation. K-fold cross-validation is a common procedure for evaluating machine learning models while preventing systematic biases due to the partitioning of data subsets (Wong, 2015). The initialized model was trained five times, each time with a different training subset and its accuracy was evaluated five times, each time on an independent test subset.

We applied lenient thresholds of 50% for both the confidence score of predictions and the overlap of predictions with observations, since minimizing false negatives is more crucial than avoiding false positives in the case of rare megafauna species (Villon et al., 2020). As such, a predicted bounding box that was associated with a confidence score of at least 50% and that overlapped at least 50% in surface with an annotated eagle ray was considered a true positive (TP). Predicted bounding boxes not corresponding to an annotated bounding box were false positives (FP), while annotated bounding boxes not corresponding to a predicted bounding box were false negatives (FN). For each cross-validation test subset, the number of TPs, FPs and FNs were computed and performance metrics were calculated as described below.

Precision is the percentage of TPs with respect to all predictions (Equation (1)). It represents the percentage of predictions that are correct (the closest to 1, the fewest false positives):

$$\text{Precision} = TP / (TP + FP) \qquad (1)$$

Recall (or sensitivity) is the percentage of TPs with respect to all annotated objects (Equation (2)). It represents the percentage of positives that are actually predicted (the closest to 1, the fewest false negatives):

$$\text{Recall} = TP / (TP + FN) \qquad (2)$$

Finally, the f1-score evaluates the balance between FPs and FNs. It is an overall measure of accuracy calculated as the harmonic mean of precision and recall (Equation (3)).

$$\text{F1-score} = 2 \times \text{Recall} \times \text{Precision} / (\text{Recall} + \text{Precision}) \qquad (3)$$

Finally, the mean and standard deviation of the performance metrics were computed across the 5-fold cross-validations splits.

We used the open-source Tensorflow object detection API version 1 (Abadi et al., 2016) in Python version 3 for the training and testing of our model. One training process lasted on average 3 hours on a NVIDIA Quadro P6000 GPU with 64 GB of RAM. The application of the model took on average 5 seconds per image.

2.7 Spatial distribution of eagle rays

Locations of eagle ray occurrences obtained from both manual annotation and the deep learning model were mapped in the study area by retrieving the GPS coordinates of their image identifiers. Locations of all ULM tracks were also mapped by retrieving the GPS coordinates of all video images. To account for the heterogeneous sampling effort, the encounter rate (individuals/km) was mapped throughout the study area. To do so, we created a spatial grid of 0.005° longitude x 0.005° latitude and summed the number of eagle rays and the length of ULM tracks in each cell. The number of individuals was then divided by the length of ULM tracks per cell to obtain the encounter rate. All maps were produced in R (version 4.0.3) with the OpenStreetMap (Fellows, 2019) and ggplot2 packages (Wickham et al., 2020).

3. Results

3.1 Deep learning model accuracy

Our deep learning model trained with 255 images on average (range= 252 - 259 between cross-validations) accurately detected eagle rays on independent images from the same lagoon. The model reached a mean precision of 0.90 on test images (SD= 0.08), meaning that 90% of the model predictions corresponded to a manually annotated eagle ray (i.e., were TPs) (Figure 2). False positives were primarily associated with coral patches. The mean recall was 0.92 (SD= 0.06), meaning that 92% of the annotated eagle rays were detected (Figure 2). The model successfully detected eagle rays in various contexts, as illustrated in Figure 2 and Supplementary Figure D. The mean f1-score balancing FPs and FNs was 0.91 (SD= 0.06). Precision, recall and the f1-score showed little sensitivity to the prediction confidence score (Figure 3).

3.2 Spatial distribution of eagle rays

Eagle rays detected from the deep learning model were distributed throughout the study area, but appeared concentrated in a more intensively surveyed portion of the barrier reef near the easternmost channel (Figure 4-a). The few FPs and FNs were scattered across the lagoon and on the barrier reef (Supplementary Figure E-1). The encounter rate map, accounting for the heterogeneous sampling effort, confirmed the slightly higher occurrence of eagle rays on the barrier reef compared to the lagoon (Figure 4-b). The spatial distributions of detected eagle rays and their encounter rates were similar to that of all annotated eagle rays (Supplementary Figures E-2 and E-3).

4. Discussion

More than one third of all cartilaginous fishes are threatened with extinction, primarily due to overfishing (Dulvy et al. 2021). Rays are no exception as they represent 56.3% of threatened chondrichthyan and 12.3% of ray species are still lacking sufficient data for assessment (Dulvy et al. 2021). As human activities continue to jeopardize these species (Pacoureau et al., 2021; Yan et al., 2021), there is an urgent need for reliable and efficient approaches for monitoring populations. Our study revealed the potential of deep learning for the accurate detection of eagle rays in coral reef ecosystems. Our model trained with fewer than 260 aerial images was able to detect 92% of the eagle rays on independent images from the same lagoon. Our study paves the way towards automated ray population monitoring in coral reefs by providing a fast and accurate alternative to the manual processing of aerial images (Kelaher et al. 2020; Kiszka et al. 2016). While deep learning for elasmobranch aerial detection has been applied in the context of beach surveillance (Gorkin et al. 2020), we present its first implementation towards ecological and conservation applications, including species distribution mapping).

4.1. Eagle ray detection accuracy

Our model achieved a very good detection performance despite the modest size of the training dataset. Obtaining large amounts of images for training deep learning models is a major bottleneck for ecological and conservation applications (Christin, Hervet, and Lecomte 2019). To overcome this limitation, we relied on transfer learning and artificial data augmentation, two efficient techniques that are widely used for training models in data-limited situations (Schneider et al., 2020). The model was successful at avoiding missed occurrences (false negatives), which is most critical when the objective is to detect vulnerable species that occur in low numbers such as rays and sharks (Villon et al., 2020). Eagle rays were consistently detected across the diversity of habitats (e.g., soft bottom and barrier reef) and acquisition conditions (e.g., luminosity, altitude and camera angle) in our study area. The robustness of the model at detecting eagle rays in more contexts and its generalization to new data could be further increased by expanding the size of both the training and the test datasets and the contextual variety at new sites in New-Caledonia and beyond. Moreover, there is a need to test the model’s generalizability to a larger dataset in the future, as the size of the test dataset is also limited. The model was equally successful at avoiding false positives, with few misdetections primarily associated with coral patches. To eliminate these false positives, coral patch annotations could be incorporated into the training dataset so that the model explicitly learns this class. As deep learning algorithms rapidly improve, we could further enhance our eagle ray detection method by using most up-to-date object detection CNNs such as the YOLOv3 that achieved a high performance on fish detection (Jalal et al., 2020).

4.2. Comparison with other monitoring methods

Effective conservation requires up-to-date and high quality data collected with limited monetary and human costs over repeated periods (Fust and Loos 2020). Previous studies on the distribution and movements of eagle rays have relied on acoustic (DeGroot et al., 2020) and satellite telemetry (Ajemian and Powers 2014). Active acoustic telemetry implies following the individuals in order to determine their movements in the water column, but is generally restricted to few individuals and necessitates a large array of hydrophones (DeGroot et al., 2020). Satellite telemetry allows tracking rays over potentially large spatial scales, but is constrained by the frequency and precision of GPS data and associated costs (Ajemian and Powers 2014). Both methods are intrusive as they require catching and manipulating individuals to attach the tags properly. Surveys from scuba divers and baited remote underwater videos (Rizzari, Frisch, and Magnenat 2014; Ward-Paige 2017) are also used for elasmobranch censuses, especially for species that live further from the surface . However, these underwater surveys are limited in their spatial extent and may fail to detect the most elusive species (Juhel et al., 2017). Moreover, observations may not be precisely located and are not verifiable, unlike those derived from video footage.

In this study, aerial images collected from an off-the-shelf camera and processed with a deep learning algorithm allowed us to precisely locate eagle rays in a coral lagoon at low financial and operational costs. The opportunistic use of an aircraft dedicated to touristic flights led to an heterogeneous survey effort, preventing the estimation of abundance from the traditional strip transect methodology (Kiszka et al., 2016; Sykora-Bodie et al., 2017). Nevertheless, our accurate algorithm will be applicable to images collected along systematically-designed transects for abundance estimation in the future. Despite the heterogeneous survey effort, the current method suggests a widespread distribution of eagle rays across a variety of coral reef habitats, which is in accordance with previous study (Ajemian and Powers 2014). Future studies should seek to quantify habitat preferences of eagle rays by linking effort-corrected encounter rates to local habitat information (Ajemian, Powers, and Murdoch 2012; DeGroot et al. 2020).

Shark and ray monitoring requires detection and census methods that are adapted to the studied habitats. While aerial surveys are very efficient in coral reefs with clear and shallow waters, open water or turbid waters (e.g., estuaries, mangroves) require non-visual methods such as acoustic telemetry. Environmental DNA is also an innovative method at the species level that can be notably used to detect rare species including elasmobranchs (Boussarie et al., 2018). Coral lagoons are major habitats for eagle rays (DeGroot et al. 2020; Ajemian, Powers, and Murdoch 2012) and our aerial approach proved efficient for monitoring populations in these habitats. Our approach can be complemented by other methods (e.g., eDNA, acoustic telemetry) in habitats where eagle rays can occur (Ajemian and Powers 2014; Sellas et al. 2015) but waters are deep or turbid.

4.3. Implications for elasmobranchs monitoring in coral reefs

Data on population trends and distributions of rays and sharks are difficult to collect; yet, such information is critical to establish appropriate conservation and management actions (Dwyer et al., 2020; MacNeil et al., 2020; Pacoureau et al., 2021). Our approach combining video-surveys and deep learning offers a potential breakthrough for the automated monitoring of eagle rays in coral reef ecosystems. The ability of our model at detecting eagle rays in the variety of habitats and conditions encompassed by our data highlights its potential robustness in a broad range of contexts. Future work should assess the model transferability to other coral lagoons in New Caledonia and beyond. Robust detection models would be particularly beneficial for ray monitoring in the Indo-Pacific biodiversity triangle where conservation efforts are most urgent due to pronounced levels of human threats (Dulvy et al. 2021). Our deep learning approach could be applied to other distinctive elasmobranchs, provided sufficient images of these species are available for training the model.

Finally, our approach could be extended to systematic video-surveys from manned aircraft or drones to not only detect individuals, but also count them to derive abundance estimates and species density maps in a study area. Drones make a viable alternative to manned aircraft for marine megafauna surveys (Gray et al., 2018; Hodgson et al., 2013; Kelaher et al., 2020; Kiszka et al., 2016), alleviating safety risks, monetary costs and carbon emissions (Hodgson et al., 2013). However, the use of drones is subject to strict airspace regulations, and legislation in many areas necessitates the pilot to maintain visual-line-of-sight with the drone (Raoult et al., 2020). The platform choice will ultimately depend on the study question and the required imagery characteristics. Using an aircraft dedicated to touristic flights allowed us to achieve greater spatial and temporal coverage than would have been possible with a single drone and with no need to acquire permits. This method could be implemented in other touristic locations (e.g., Australia, French Polynesia, the Caribbean) where local companies operate scenic, low altitude flights over coastal areas.

Overall, our cost-effective approach succeeded in collecting high-quality images for training a deep learning model able to detect 92% of eagle rays in coral reefs. This new eagle ray detector will be critical for deriving abundance estimates in order to closely monitor these vulnerable populations in the future.

Author contributions

LM, DM, LV and MC conceived the ideas and designed the methodology; LM, LV and DM collected the data; LM and LD analysed the data and developed scripts; LD and LM led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data statement

The eagle ray image database will be made available on Zenodo.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv:1603.04467 [Cs]*. http://arxiv.org/abs/1603.04467

Ajemian, M. J., & Powers, S. P. (2014). Towed-float satellite telemetry tracks large-scale movement and habitat connectivity of myliobatid stingrays. *Environ Biol Fish*, 15.

Ajemian, M. J., Powers, S. P., & Murdoch, T. J. T. (2012). Estimating the Potential Impacts of Large Mesopredators on Benthic Resources : Integrative Assessment of Spotted Eagle Ray Foraging Ecology in Bermuda. *PLoS ONE*, *7*(7), e40227. https://doi.org/10.1371/journal.pone.0040227

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J.-B., Kiszka, J. J., Kulbicki, M., Manel, S., Robbins, W. D., Vigliola, L., & Mouillot, D. (2018). Environmental DNA illuminates the dark diversity of sharks. *Science Advances*, *4*(5), eaap9661. https://doi.org/10.1126/sciadv.aap9661

Ceccarelli, D. M., McKinnon, A. D., Andréfouët, S., Allain, V., Young, J., Gledhill, D. C., Flynn, A., Bax, N. J., Beaman, R., Borsa, P., Brinkman, R., Bustamante, R. H., Campbell, R., Cappo, M., Cravatte, S., D'Agata, S., Dichmont, C. M., Dunstan, P. K., Dupouy, C., … Richardson,

A. J. (2013). The Coral Sea. In *Advances in Marine Biology* (Vol. 66, p. 213- 290). Elsevier. https://doi.org/10.1016/B978-0-12-408096-6.00004-3

Chen, Z., Zhang, T., & Ouyang, C. (2018). End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images. *Remote Sensing*, *10*(1), 139. https://doi.org/10.3390/rs10010139

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*(10), 1632- 1644. https://doi.org/10.1111/2041-210X.13256

Colefax, A. P., Butcher, P. A., & Kelaher, B. P. (2018). The potential for unmanned aerial vehicles (UAVs) to conduct marine fauna surveys in place of manned aircraft. *ICES Journal of Marine Science*, *75*(1), 1- 8. https://doi.org/10.1093/icesjms/fsx100

DeGroot, B., Roskar, G., Brewster, L., & Ajemian, M. (2020). Fine-scale movement and habitat use of whitespotted eagle rays Aetobatus narinari in the Indian River Lagoon, Florida, USA. *Endangered Species Research*, *42*, 109- 124. https://doi.org/10.3354/esr01047

Ditria, E. M. (2020). Automating the Analysis of Fish Abundance Using Object Detection : Optimizing Animal Ecology With Deep Learning. *Frontiers in Marine Science*, *7*, 9.

Dujon, A. M., Ierodiaconou, D., Geeson, J. J., Arnould, J. P. Y., Allan, B. M., Katselidis, K. A., & Schofield, G. (2021). Machine learning to detect marine animals in UAV imagery : Effect of morphology, spacing, behaviour and habitat. *Remote Sensing in Ecology and Conservation*, *7*(3), 341- 354. https://doi.org/10.1002/rse2.205

Dulvy, N. K., Carlson, J., Charvet, P., Ajemian, M. J., Bassos-Hull, K., Blanco-Parra, M., Chartrain, E., Derrick, D., Dia, M., Diop, M., Doherty, P., Dossa, J., De Bruyne, G., Herman, K., Leurs, G. H. L., Mejía-Falla, P. A., Navia, A. F., Pacoureau, N., Pérez Jiménez, J. C., … Williams, A. B. (2020). *Aetobatus narinari : The IUCN Red List of Threatened Species 2021* [Data set]. International Union for Conservation of Nature. https://doi.org/10.2305/IUCN.UK.2021-2.RLTS.T42564343A201613657.en

Dulvy, N. K., Pacoureau, N., Rigby, C. L., Pollom, R. A., Jabado, R. W., Ebert, D. A., Finucci, B., Pollock, C. M., Cheok, J., Derrick, D. H., Herman, K. B., Sherman, C. S., VanderWright, W. J., Lawson, J. M., Walls, R. H. L., Carlson, J. K., Charvet, P., Bineesh, K. K., Fernando, D.,

… Simpfendorfer, C. A. (2021). Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. *Current Biology*, *31*(21), 4773-4787.e8. https://doi.org/10.1016/j.cub.2021.08.062

Dwyer, R. G., Krueck, N. C., Udyawer, V., Heupel, M. R., Chapman, D., Pratt, H. L., Garla, R., & Simpfendorfer, C. A. (2020). Individual and Population Benefits of Marine Reserves for Reef Sharks. *Current Biology*, *30*(3), 480-489.e5. https://doi.org/10.1016/j.cub.2019.12.005

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*(11), 1875‑ 1887. https://doi.org/10.1111/2041-210X.13277

Fellows, I. (2019). *OpenStreetMap : Access to Open Street Map Raster Images*. https://CRAN.R-project.org/package=OpenStreetMap

Fricke, R., Kulbicki, M., & Wantiez, L. (2011). Checklist of the fishes of New Caledonia, and their distribution in the Southwest Pacific Ocean (Pisces). *STUTTGARTER BEITRÄGE ZUR NATURKUNDE A*, 123.

Froese, R., & Pauly, D. (2021). Froese, R. and D. Pauly. Editors. 2021.FishBase. World Wide Web electronic publication. Www.fishbase.org, (06/2021). *FishBase. World Wide Web electronic publication. www.fishbase.org, (06/2021)*.

Fust, P., & Loos, J. (2020). Development perspectives for the application of autonomous, unmanned aerial systems (UASs) in wildlife conservation. *Biological Conservation*, *241*, 108380. https://doi.org/10.1016/j.biocon.2019.108380

Gorkin, R., Adams, K., Berryman, M. J., Aubin, S., Li, W., Davis, A. R., & Barthelemy, J. (2020). Sharkeye : Real-Time Autonomous Personal Shark Alerting via Aerial Surveillance. *Drones*, *4*(2), 18. https://doi.org/10.3390/drones4020018

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*(9), 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., & Johnston, D. W. (2018). A Convolutional Neural Network for Detecting Sea Turtles in Drone Imagery. *Methods in Ecology and Evolution*, 2041-210X.13132. https://doi.org/10.1111/2041-210X.13132

Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., & Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, *9*(1), 14259. https://doi.org/10.1038/s41598-019-50795-9

Hodgson, A., Kelly, N., & Peel, D. (2013). Unmanned Aerial Vehicles (UAVs) for Surveying Marine Fauna : A Dugong Case Study. *PLOS ONE*, *8*(11), e79556. https://doi.org/10.1371/journal.pone.0079556

Jabado, R. W., Kyne, P. M., Pollom, R. A., Ebert, D. A., Simpfendorfer, C. A., Ralph, G. M., Dhaheri, S. S. A., Akhilesh, K. V., Ali, K., Ali, M. H., Mamari, T. M. S. A., Bineesh, K. K., Hassan, I. S. E., Fernando, D., Grandcourt, E. M., Khan, M. M., Moore, A. B. M., Owfi, F., Robinson, D. P., … Dulvy, N. K. (2018). Troubled waters : Threats and extinction risk of the sharks, rays and chimaeras of the Arabian Sea and adjacent waters. *Fish and Fisheries*, *19*(6), 1043‐ 1062. https://doi.org/10.1111/faf.12311

Jalal, A., Salman, A., Mian, A., Shortis, M., & Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, *57*, 101088. https://doi.org/10.1016/j.ecoinf.2020.101088

Juhel, J.-B., Vigliola, L., Mouillot, D., Kulbicki, M., Letessier, T. B., Meeuwig, J. J., & Wantiez, L. (2017). *Reef accessibility impairs the protection of sharks*. 11.

Kelaher, B. P., Colefax, A. P., Tagliafico, A., Bishop, M. J., Giles, A., & Butcher, P. A. (2020). Assessing variation in assemblages of large marine fauna off ocean beaches using drones. *Marine and Freshwater Research*, *71*(1), 68. https://doi.org/10.1071/MF18375

Kelaher, B., Peddemors, V., Hoade, B., Colefax, A., & Butcher, P. (2019). Comparison of sampling precision for nearshore marine wildlife using unmanned and manned aerial surveys. *Journal of Unmanned Vehicle Systems*, 32.

Kiszka, J., Mourier, J., Gastrich, K., & Heithaus, M. (2016). Using unmanned aerial vehicles

(UAVs) to investigate shark and ray densities in a shallow coral lagoon. *Marine Ecology Progress Series*, *560*, 237‑ 242. https://doi.org/10.3354/meps11945

Last, P. R., White, W. T., & Pogonoski, J. J. (2010). *Descriptions of new sharks and rays from Borneo*. CSIRO Marine & Atmospheric Research.

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

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

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:1405.0312 [Cs]*. http://arxiv.org/abs/1405.0312

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, J. E. (2020). Global status and conservation potential of reef sharks. *Nature*, *583*(7818), 801‑ 806. https://doi.org/10.1038/s41586-020-2519-y

Mannocci, L., Villon, S., Chaumont, M., Guellati, N., Mouquet, N., Iovan, C., Vigliola, L., & Mouillot, D. (2021). Leveraging social media and deep learning to detect rare megafauna in video surveys. *Conservation Biology*, cobi.13798. https://doi.org/10.1111/cobi.13798

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*(25), E5716‑ E5725. https://doi.org/10.1073/pnas.1719367115

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*(7843), 567‑ 571. https://doi.org/10.1038/s41586-020-03173-9

Padubidri, C., Kamilaris, A., Karatsiolis, S., & Kamminga, J. (2021). Counting sea lions and elephants from aerial photography using deep learning with density maps. *Animal Biotelemetry*, *9*(1), 27. https://doi.org/10.1186/s40317-021-00247-x

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, *12*(1), 145‑151. https://doi.org/10.1016/S0893-6080(98)00116-6

Raoult, V., Colefax, A. P., Allan, B. M., Cagnazzi, D., Castelblanco-Martínez, N., Ierodiaconou, D., Johnston, D. W., Landeo-Yauri, S., Lyons, M., Pirotta, V., Schofield, G., & Butcher, P. A. (2020). Operational Protocols for the Use of Drones in Marine Animal Research. *Drones*, *4*(4), 64. https://doi.org/10.3390/drones4040064

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks. *ArXiv:1506.01497 [Cs]*. http://arxiv.org/abs/1506.01497

Rieucau, G., Kiszka, J. J., Castillo, J. C., Mourier, J., Boswell, K. M., & Heithaus, M. R. (2018). Using unmanned aerial vehicle (UAV) surveys and image analysis in the study of large surface-associated marine species : A case study on reef sharks *Carcharhinus melanopterus* shoaling behaviour. *Journal of Fish Biology*, *93*(1), 119‑127. https://doi.org/10.1111/jfb.13645

Rigby, C. L., Bin Ali, A., Chen, X., Derrick, C., Dharmadi Ebert, D. A., Fahmi Fernando, D., Gautama, D. A., Haque, A. B., Herman, K., Ho, H., Hsu, H., Krajangdara, T., Maung, A., Seyha, L., Sianipar, A., Tanay, D., Utzurrum, J.-A. T., Vo, V. Q., Yuneni, R. R., & Zhang, J. (2020). *Aetomylaeus maculatus : The IUCN Red List of Threatened Species 2020*. International Union for Conservation of Nature. https://www.iucnredlist.org/species/60120/124440727

Rizzari, J. R., Frisch, A. J., & Magnenat, K. A. (2014). Diversity, abundance, and distribution of reef sharks on outer-shelf reefs of the Great Barrier Reef, Australia. *Marine Biology*, *161*(12), 2847‑2855. https://doi.org/10.1007/s00227-014-2550-3

Sarle, W. S. (1995). *Stopped training and other remedies for overfitting*. Computer Science.

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution*, *10*(7), 3503‐3517. https://doi.org/10.1002/ece3.6147

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. https://doi.org/10.3354/meps12193

Sellas, A. B., Bassos-Hull, K., Perez-Jimenez, J. C., Angulo-Valdes, J. A., Bernal, M. A., & Hueter, R. E. (2015). Population Structure and Seasonal Migration of the Spotted Eagle Ray, Aetobatus narinari. *Journal of Heredity*, *106*(3), 266‐275. https://doi.org/10.1093/jhered/esv011

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*, 30.

Sykora-Bodie, S. T., Bezy, V., Johnston, D. W., Newton, E., & Lohmann, K. J. (2017). Quantifying Nearshore Sea Turtle Densities : Applications of Unmanned Aerial Systems for Population Assessments. *Scientific Reports*, *7*(1), 17690. https://doi.org/10.1038/s41598-017-17719-x

Torney, C. J., Lloyd‐ Jones, D. J., Chevallier, M., Moyer, D. C., Maliti, H. T., Mwita, M., Kohi, E. M., & Hopcraft, G. C. (2019). A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, *10*(6), 779‐787. https://doi.org/10.1111/2041-210X.13165

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)*, 582‐591. https://doi.org/10.1109/CVPRW.2019.00084

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. https://doi.org/10.1016/j.ecoinf.2018.09.007

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*(1), 10972. https://doi.org/10.1038/s41598-020-67573-7

Ward-Paige, C. A. (2017). Global evaluation of shark sanctuaries. *Global Environmental Change*, 16.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dewey, D. (2020). *ggplot2 : Create Elegant Data Visualisations Using the Grammar of Graphics*. https://cran.r-project.org/web/packages/ggplot2/index.html

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839‑ 2846. https://doi.org/10.1016/j.patcog.2015.03.009

Yan, H. F., Kyne, P. M., Jabado, R. W., Leeney, R. H., Davidson, L. N. K., Derrick, D. H., Finucci, B., Freckleton, R. P., Fordham, S. V., & Dulvy, N. K. (2021). Overfishing and habitat loss drive range contraction of iconic marine fishes to near extinction. *Science Advances*, *7*(7), eabb6026. https://doi.org/10.1126/sciadv.abb6026

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., & Le, Q. V. (2019). Learning Data Augmentation Strategies for Object Detection. *ArXiv:1906.11172 [Cs]*. http://arxiv.org/abs/1906.11172

Tables

Table 1: Overview of the New Caledonia video database. Abbreviations: SD= standard deviation, pi= pixels.

Number of videos	Mean video duration	Total video duration	Total number of images	Number of images with ≥ 1 eagle ray		Number of individual encounters	
				2,704 x 1,520 pi images	**1,352 x 760 pi images**	**2,704 x 1,520 pi images**	**1,352 x 760 pi images**
114	11.70 min (SD= 0.93 min) equivalent to 11 min 42 s	22 h 14 min 19 s	79,325	314	308	372	353

Figures

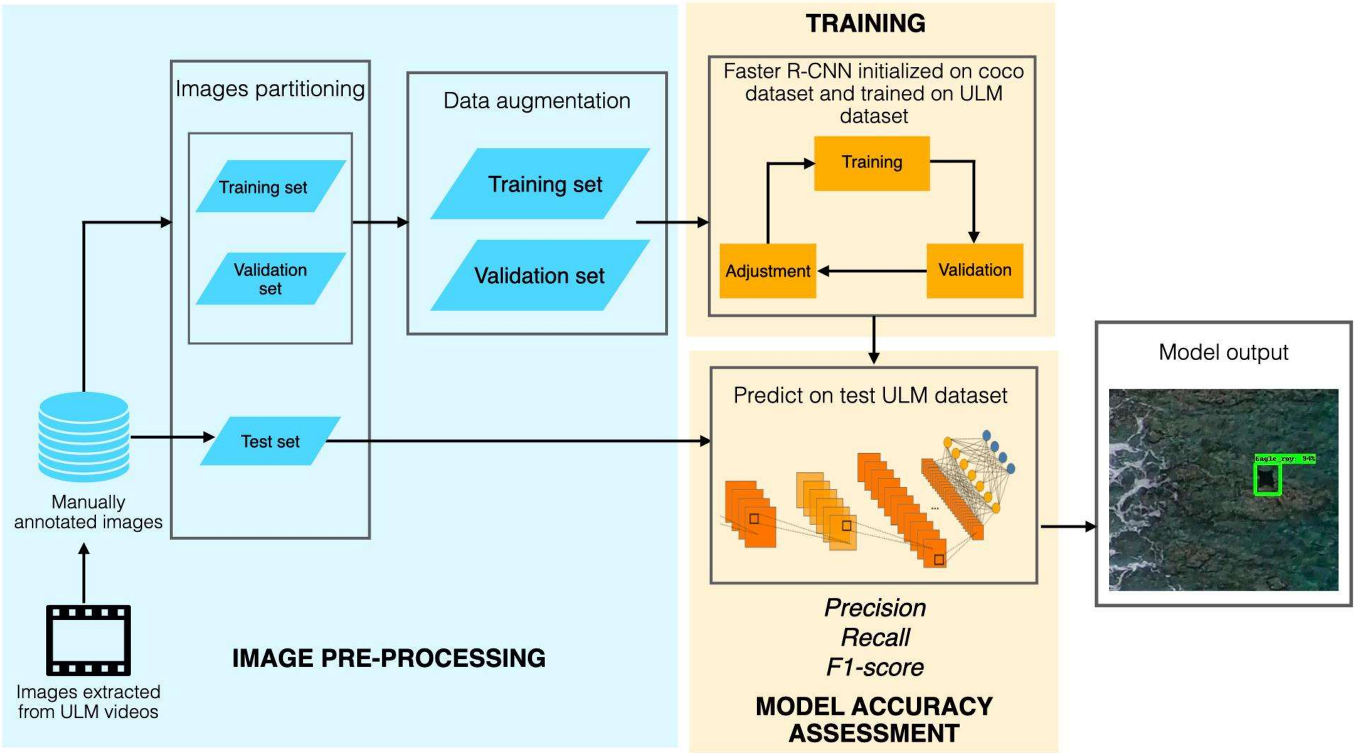

Figure 1: Eagle ray detection framework with three main steps. 1) Image pre-processing: Images are extracted from the ULM videos and manually annotated. These images are then partitioned into independant training, validation and test sets. Training and validation sets are augmented by applying random transformations such as rotations and translations to images. 2) Training: A Faster R-CNN with weights pre-trained on the COCO dataset is downloaded from the Tensorflow model zoo and trained on the training set. The training is stopped before overfitting as indicated by an increasing loss function for the validation set. 3) Model accuracy assessment: The trained Faster R-CNN is applied for eagle ray detection on the test set. Precision, recall and the f1-score are then derived to evaluate the model accuracy. The final output is a detected bounding box with an associated confidence score for each of the detected eagle rays. These steps are detailed in sections 2.4., 2.5. and 2.6.

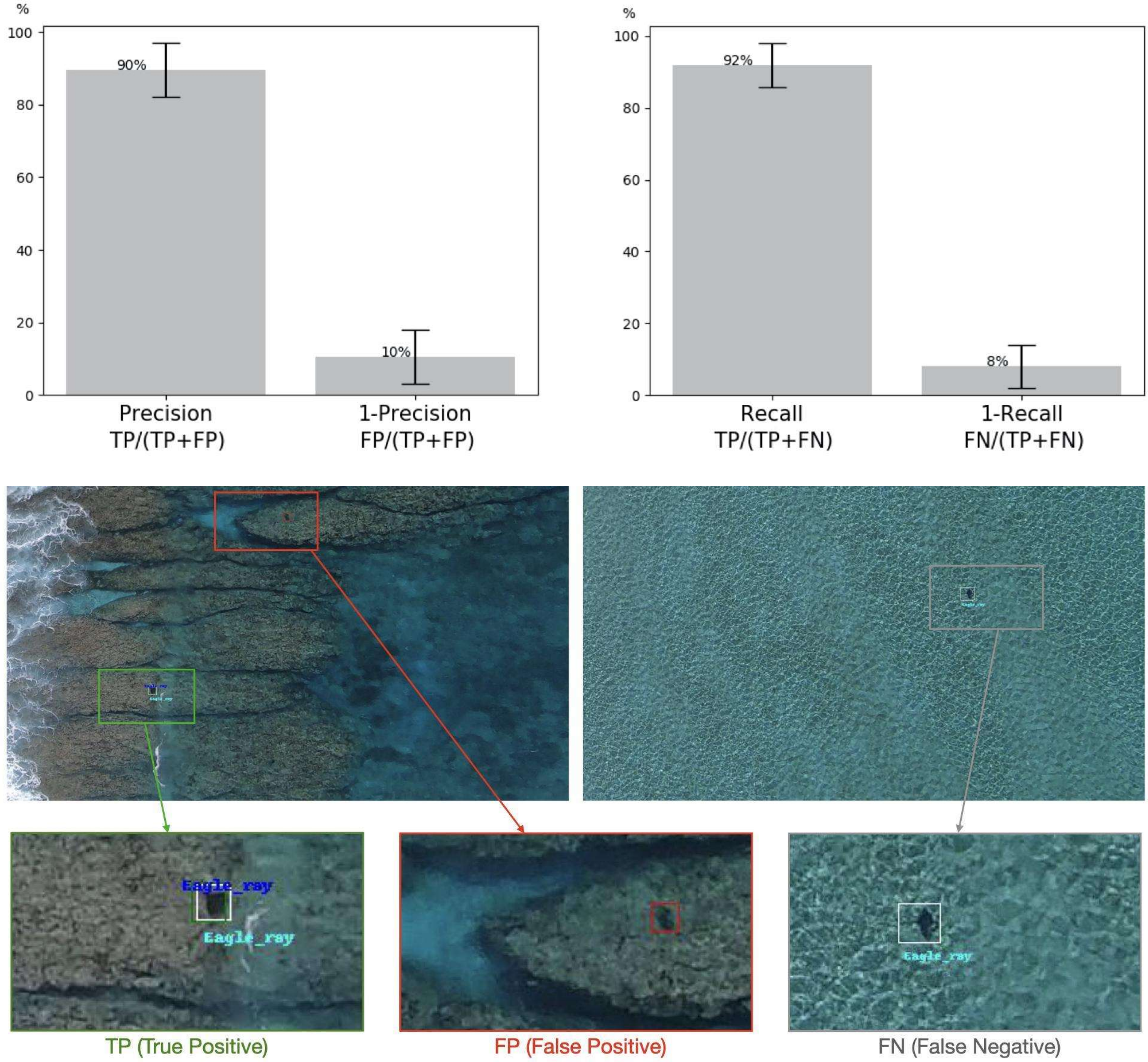

Figure 2: Results of eagle ray detection on test images for a prediction confidence score of 50%. The left graph shows the mean percentage of true positives (TPs) and false positives (FPs) with respect to all predictions. The right graph shows the mean percentage of TPs and false negatives (FNs) in the observations. The error bars are the standard deviations from the means. Examples are provided below the graphs for a TP in green (prediction associated with an annotation shown in white), a FP in red (prediction not corresponding to an annotation; here a coral patch) and a FN (annotation not corresponding to a prediction). Further examples of detection results are provided in Appendix D.

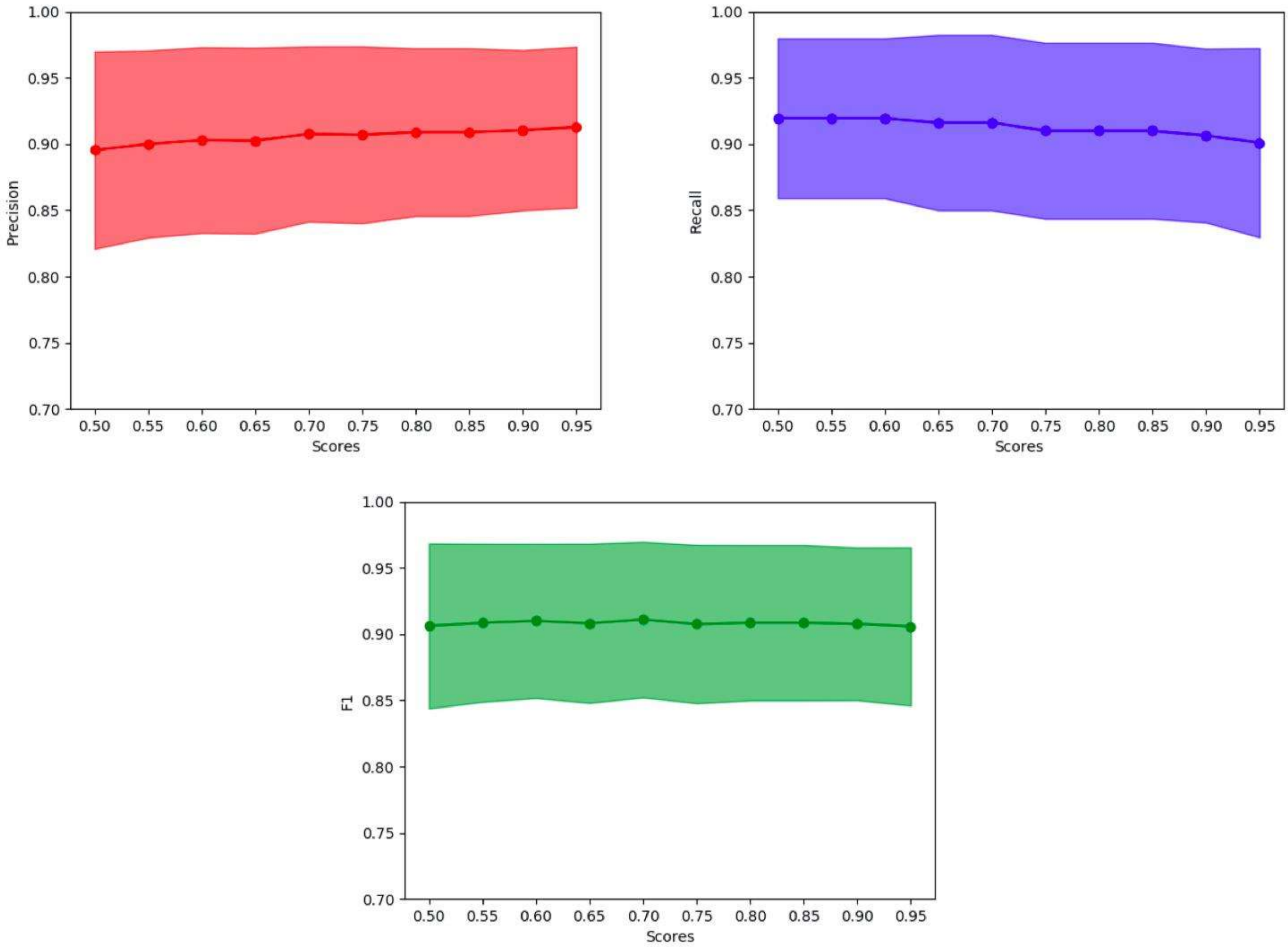

Figure 3: Mean precision, recall and f1-score on the test images for varying prediction confidence scores. The standard deviation is represented by the shaded area.

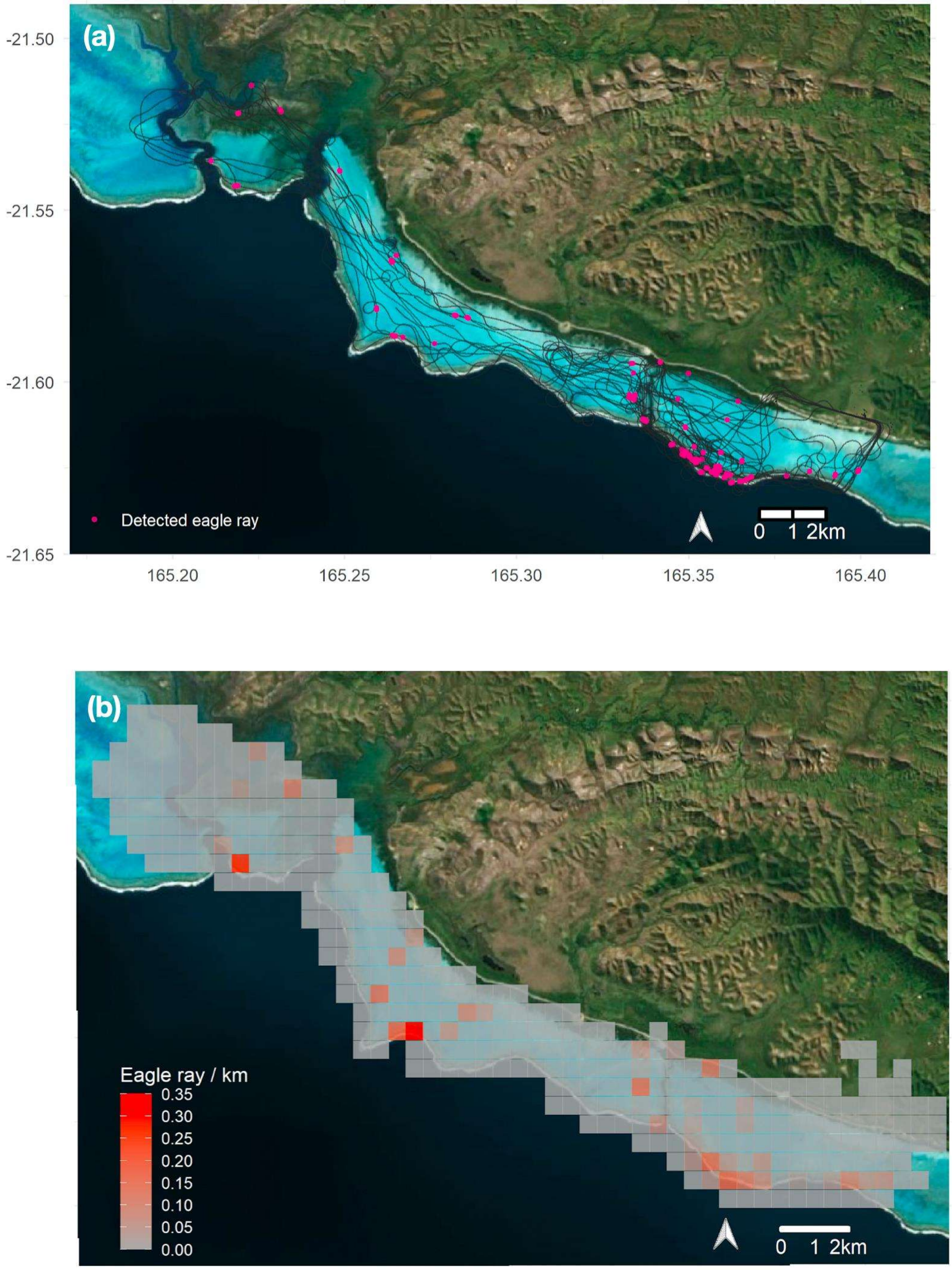

Figure 4: Spatial distribution of (a) eagle ray detections (dots) from the deep learning model mapped by retrieving the GPS coordinates of their image identifiers and the corresponding ULM flight tracks (black lines) and (b) the encounter rate (individuals/km) of detected eagle rays calculated on a spatial grid of 0.005° longitude x 0.005° latitude (the calculation is detailed in section 2.7).

Putting eagle rays on the map by coupling aerial video-surveys and deep learning

Authors: Desgarnier L.[1], Mouillot D.[1,2], Vigliola L.[3], Chaumont M.[4,5], Mannocci L.[1,3,4]

[1] MARBEC (Univ Montpellier, CNRS, Ifremer, IRD), Montpellier, France.

[2] Institut Universitaire de France, Paris, France.

[3] ENTROPIE (IRD, Université de la Réunion, Université de la Nouvelle Calédonie, CNRS, Ifremer), Noumea, New Caledonia, France.

[4] LIRMM (Université de Montpellier, CNRS), Montpellier, France.

[5] Université de Nîmes, Nîmes, France.

Corresponding author: Lila Desgarnier
liladesgarnier@gmail.com
MARBEC Laboratory, 93 Place Eugène Bataillon, 34090 Montpellier, France

Acknowledgements

We are grateful to Jugurtha Ifticen, Lola Romant, Nacim Guellati, Gwendal Quimbre and Matis Toitot for their help with visualisation of ULM videos and annotation of images. We thank Air Paradise for their collaboration in collecting ULM video sequences in New Caledonia. This project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 845178 ('MEGAFAUNA').