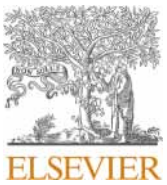
Ecological Informatics 80 (2024) 102499



Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Toward an artificial intelligence-assisted counting of sharks on baited video

Sébastien Villon *, Corina Iovan, Morgan Mangeas, Laurent Vigliola

ENTROPIE, Institut de Recherche pour le Développement (IRD), UR, UNC, CNRS, IFREMER, Centre IRD de Nouméa, 98000 Noumea, New-Caledonia, France

A R T I C L E I N F O

Keywords:
Deep learning
Neural network
Coral reef
Marine ecology
Shark conservation

A B S T R A C T

Given the global biodiversity crisis, there is an urgent need for new tools to monitor populations of endangered marine megafauna, like sharks. To this end, Baited Remote Underwater Video Stations (BRUVS) stand as the most effective tools for estimating shark abundance, measured using the MaxN metric. However, a bottleneck exists in manually computing MaxN from extensive BRUVS video data. Although artificial intelligence methods are capable of solving this problem, their effectiveness is tested using AI metrics such as the F-measure, rather than ecologically informative metrics employed by ecologists, such as MaxN. In this study, we present both an automated and a semi-automated deep learning approach designed to produce the MaxN abundance metric for three distinct reef shark species: the grey reef shark (*Carcharhinus amblyrhynchos*), the blacktip reef shark (*C. melanopterus*), and the whitetip reef shark (*Triaenodon obesus*). Our approach was applied to one-hour baited underwater videos recorded in New Caledonia (South Pacific). Our fully automated model achieved F-measures of 0.85, 0.43, and 0.72 for the respective three species. It also generated MaxN abundance values that showed a high correlation with manually derived data for *C. amblyrhynchos* ($R = 0.88$). For the two other species, correlations were significant but weak ($R = 0.35$–0.44). Our semi-automated method significantly enhanced F-measures to 0.97, 0.86, and 0.82, resulting in high-quality MaxN abundance estimations while drastically reducing the video processing time. To our knowledge, we are the first to estimate MaxN with a deep-learning approach. In our discussion, we explore the implications of this novel tool and underscore its potential to produce innovative metrics for estimating fish abundance in videos, thereby addressing current limitations and paving the way for comprehensive ecological assessments.

1. Introduction

At a time of global biodiversity crisis (Knapp et al., 2021; Lees et al., 2020; Rull, 2022; Tian et al., 2020; Wagner, 2019), top predators are critically endangered by human activity, threatening key ecosystem functions and services (Hammerschlag et al., 2019; Rizzari et al., 2014). In the seas, shark extinction risk is rising with over one third of species endangered globally (Baum and Blanchard, 2010; Boussarie et al., 2018; Davidson and Dulvy, 2017; Edgar et al., 2014; Graham et al., 2010; Jorgensen et al., 2022; Juhel et al., 2018a), mostly due to ever increasing fishing pressure, habitat loss, climate change and pollution (Barone et al., 2022). Although data on shark population abundance is crucial for species conservation planning, monitoring their numbers poses challenges due to their rarity and extensive range. To overcome the issue of data collection, video imagery is becoming increasingly popular as a biomonitoring tool (Ditria et al., 2020; Goodwin et al., 2022; Tian et al., 2020; Whytock et al., 2021) especially since the advent of cheap action cameras. Due to recent advances, artificial intelligence (AI) can solve the problem of video data processing. However, algorithms are usually evaluated on closed sets of still images and AI metrics such as F-measure, rather than on open sets of video streams and metrics used by biologists such as animal abundance. This created a situation where powerful AI methods remain mostly in the realm of emerging tools rather than being used by biologists. Reconciling the outputs of AI algorithms with biology therefore appears to be a priority if certain aspects of the biodiversity crisis are to be addressed in an effective way.

Baited remote underwater video stations (BRUVS) are particularly suitable to observe marine predators such as sharks (Ditria et al., 2021; Goetze et al., 2019; Weinstein, 2018). Current video processing protocols involve a biologist manually annotating sharks on BRUVS to determine their presence and abundance. Technically, biologists only annotate a frame when the number of individuals of a species in that frame exceeds that of previous frames. At the end of the video processing protocol, biologists retain "the maximum number of a particular species

* Corresponding author.
E-mail address: sebastien.villon@ird.fr (S. Villon).

https://doi.org/10.1016/j.ecoinf.2024.102499
Received 22 October 2023; Received in revised form 3 January 2024; Accepted 22 January 2024
Available online 28 January 2024

seen in any one video frame across the duration of the video record" to estimate relative abundance (Langlois et al., 2020; Tian et al., 2020). This abundance metric, called MaxN, is until now the standard and most used method for estimating fish and shark abundance on BRUVS (Cappo et al., 2003; Langlois et al., 2020; Tian et al., 2020; Whitmarsh et al., 2017). Convolutional Neural Networks (CNNs) demonstrate exceptional performance in object detection and identification within images, showcasing the capabilities of artificial intelligence for processing BRUVS videos. Indeed, fish identification and localisation with Deep Learning has been continuously improving since 2015 (Chen et al., 2018; Cui et al., 2020; Jalal et al., 2020; Knausgård et al., 2022; Rathi et al., 2018; Salman et al., 2016; Tian et al., 2020; Villon et al., 2016, 2018).

Regarding sharks, three studies were identified. The short Shark-EYE (Merencilla et al., 2021) study used a YOLOv3 neural network to detect sharks in underwater images. This paper does not classify individuals to species level, but detects sharks with an accuracy of 86%. A second study proposed a pipeline composed of multiple neural networks to localize and identify individuals at the species level (Jenrette et al., 2022). The pipeline includes 1) a shark locator which localizes fish individuals in images 2) a shark identifier which verifies that the detected individuals are sharks 3) a first shark classifier classifies the individuals at the genus level and 4) for each genus, a second classifier identifies the shark at a species level. This 4-stage pipeline allowed the authors to detect and identify 10 species in 7 short videos, with a recall of 69% and a precision of 94%. The third study used two backbones architectures (VGG16 and U-net) to discriminate sharks at an individual level on images (Le et al., 2022). The method was able to predict with 81% accuracy whether individual sharks present in the test dataset had ever been encountered in the training dataset.

In summary, the current literature on studies utilizing CNNs for fish and shark counting in images is promising. However, it indicates that these studies tend to present their findings in terms of accuracy, recall and F-measure rather than comparing model outputs with standard ecological metrics like MaxN abundance. Therefore, it is yet to be determined if AI methods can extract from videos some metrics that are useful for biologists, such as species identification and abundance. It is also unclear whether models exhibiting a high F-measure will produce MaxN abundance values that align with those recorded by biologists.

Here, we used 185 one-hour BRUVS videos to build an all-in-one CNN-based detector and classifier for reef shark species. Specifically, we assessed its effectiveness to identify and count 3 common South-Pacific species: the grey reef shark *Carcharhinus amblyrhynchos,* the blacktip reef shark *C. melanopterus* and the whitetip reef shark *Triaenodon obesus* in standard ecologically relevant 1-h long underwater BRUVS videos. We then assessed the efficiency of our model through both classic CNN metrics (F-measure, recall, accuracy) and standard ecological abundance metric MaxN. We also explored how such method can lead to semi-automated video processing. Finally, we discussed how Deep Learning processing can deepen our understanding of shark conservation and behavior by making it possible to create new ecological metrics.

1.1. Material

We collected one-hour baited videos (BRUVS) at 185 different stations across New-Caledonia's reefs, a biodiversity hotspot listed world heritage by UNESCO since 2008. Videos were recorded with *Sony hdr cx-7* or *Sony hdr cx-12* cameras, then manually processed by biologists to annotate MaxN (Juhel et al., 2018b). A random division was performed on the 185 one-hour videos, resulting in 163 training videos (80%) and 22 testing videos (20%). Then, 1366 video clips were extracted from the 163 training video dataset, and randomly divided into 1092 training clips (80%) and 274 testing clips (20%). Each clip lasted 15 to 20 s and was centered around species MaxNs. A 5-fold method experimentation was conducted, involving five instances of an 80/20 random selection of clips (refer to Methods for details). Our training data included 8 shark species and 19 non-shark species commonly associated with sharks on baited underwater videos to add diversity in the training. After transforming video clips into still images at a rate of 1 frame per second, we annotated sharks and the 19 non-shark species present in the frames with bounding boxes. This resulted in 26,947 fish annotations (Table 1). Given that the Faster RCNN architecture automatically generates negative samples (Ren et al., 2015), there was no need to annotate any.

It is known that classic deep architectures need numerous image sample per class to work correctly (Lecun et al., 2015; Villon et al., 2022). Hence, shark species that were infrequently captured in our clips (*Carcharhinus albimarginatus, Galeocerdo cuvier, Nebrius ferrugineus, Negaprion acutidens*) were excluded from the testing phase. However, they were included during the training phase to enhance the ability of our CNN to build pertinent feature embeddings. In an effort to improve our model accuracy, we also included 19 non-shark species during the training phase. These species were commonly associated with or exhibited resemblances to the studied shark species. Overall, we assembled two types of datasets

1) a $\text{Video}_{\text{stations}}$ dataset composed of 185 one-hour BRUVS videos. This dataset was divided into the $\text{Training}_{\text{stations}}$ dataset composed of 163 one-hour BRUVS videos (80%) and the $\text{Testing}_{\text{stations}}$ dataset composed of 22 one-hour BRUVS videos (20%);
2) a $\text{Video}_{\text{clips}}$ dataset composed of 1366 video sequences, each lasting 15–20-s. These clips were extracted from the $\text{Training}_{\text{stations}}$ dataset and focused on the MaxN of the studied species. Video clips were cut at one frame per second, generating 23,222 frames. Then, the $\text{Video}_{\text{clips}}$ dataset was randomly divided into 1092 $\text{Training}_{\text{clips}}$ and 274 $\text{Testing}_{\text{clips}}$ datasets (80/20) for each fold, and corresponding frames used for training and testing the CNN models.

Table 1
The annotated dataset with the number of annotations per species, along with details on the number of frames and video clips in which each species appeared. Shark species are indicated with an asterisk (*). Non-shark species show no asterisk. The studied species of sharks are shown in bold.

Species	Number of annotations	Number of frames	Number of clips with the species
*Carcharhinus amblyrhynchos**	**8006**	**4214**	**65**
*Triaenodon obesus**	**1432**	**822**	**40**
*Carcharhinus melanopterus**	**434**	**162**	**32**
*Galeocerdo cuvier**	243	229	7
*Nebrius ferrugineus**	207	190	4
*Negaprions acutidens**	112	112	4
*Carcharhinus albimarginatus**	85	37	3
*Stegostoma fasciatum**	80	80	4
Lutjanus bohar	10,474	4621	90
Plectropomus laevis	2570	1640	62
Epinephelus maculatus	900	615	16
Lethrinus olivaceus	783	430	35
Aprion virescens	454	322	28
Carangoides fulvoguttatus	346	103	10
Carangoides ferdau	267	168	13
Caranx ignobilis	223	77	11
Symphorus nematophorus	132	80	35
Carangoides orthogrammus	107	78	9
Scomberomorus commerson	36	30	5
Chanos Chanos	32	18	4
Lutjanus rivulatus	14	1	1
Grammatorcynus bilineatus	10	8	1

For each fold, the frames from $Training_{clips}$ were used to train the models. The $Testing_{clips}$ were used to assess the robustness of our method on short sequences centered around the presence of fish presence. The $Testing_{stations}$, common to all k-fold were used to assess the robustness of the method on 1-h videos, corresponding to the real use-case scenarios of ecological studies (*Supp.* Fig. 1).

2. Methods

2.1. Deep learning model

To assess the robustness of our method, we performed a 5-fold cross-validation. For each fold, we randomly selected 80% of our dataset $Video_{clips}$ to train our model, and 20% to test it. We could not split directly images from the frame dataset as images from the same videos were very alike which can lead to a model with a low generalization capacity. Using a 80% training/20% testing random split on videos rather than images ensured full independency between images of the training set and testing set.

Our deep-learning models used NASnet architecture (Zoph et al., 2017) with a Faster-rcnn backbone (Ren et al., 2015) implemented in Tensorflow2. The parameters of the model can be found on TensorFlow model zoo[1] under the name "faster_rcnn_nas". All images were resized to 1333 × 800 pixels to match with the pre-training data (COCO dataset (Lin et al., 2014)) used to prepare the first layers of our model and save computing time during the training phase). We used a learning rate of 0.008 with a cosine learning rate decay. For each K-fold, the model was trained on its own version of $Training_{clips}$ through 200,000 iterations, with a batch size of 16 images per iteration. The training was completed in 96 h per model using a GPU-cluster equipped with 4 RTX8000. In order to evaluate the performance of our deep-learning models, we computed the recall, precision, and F-measure for each model. Briefly, a recall of 1 indicates that the model correctly detected all sharks present in the video (False Negative detections are costly), while a recall of 0 indicates that the model missed all sharks. A precision of 1 indicates that detected individuals are all correctly identified, while a precision of 0 indicates that all detected sharks are wrongly identified (False Positives detections are costly). Finally, the F-measure is a harmonic mean of recall and precision. It tends toward 1 when positive detections and correct species identifications outweigh misclassifications and undetected individuals, and toward 0 otherwise. Recall, precision and F-measure are given by the following formulae:

$$Recall_i = \frac{Tp_i}{Tp_i + Fn_i}$$

$$Precision_i = \frac{Tp_i}{Tp_i + Fp_i}$$

$$Fmeasure_i = 2.\frac{Recall_i.Precision_i}{Recall_i + Precision_i}$$

with Tp_i the number of true positives, Fp_i the number of false positives and Fn_i the number of false negatives for species *i*.

2.2. Estimating shark abundance on video

Fitted deep-learning models were used to estimate shark abundance (MaxN) on both video clips and complete one-hour videos. For video clips, each frame of the $Testing_{clips}$ dataset was considered independently. For each species, the predicted MaxN value corresponded to the predicted number of individuals in each frame. For full one-hour videos, we first extracted frames from the $Testing_{stations}$ dataset at a rate of one frame per second, then identified and counted sharks on each frame using our deep-learning models. Finally, MaxN was computed as the maximum number of a particular species detected by the models in any one frame throughout the entire duration of the video recording. To mitigate errors resulting from occasional false positives, without increasing the size of the model architecture (e.g., with memory cells), we considered a predicted MaxN to be true if, for a given species, the predicted number of individuals was consistent across at least 2 consecutive frames. Correlation and linear regression analyses were then used to compare MaxN predicted by the models with MaxN observed by biologists. Additionally, the accuracy of MaxN was computed for each species i as follows:

$$MaxN\ accuracy_i = \frac{Correct\ MaxN\ Prediction_i}{Correct\ MaxN\ Prediction_i + Uncorrect\ Maxn\ Prediction_i}$$

A perfect correspondence between predicted and observed MaxN would lead to a correlation coefficient, R, of 1, a regression intercept of 0, a regression slope of 1, and a MaxN accuracy of 1 for all species. Comparisons at the frame level enabled testing the robustness of our method with densely populated information, and associating classic CNN metrics such as recall and precision with ecological metrics. Conversely, comparisons at the station level, involving the processing of 1-h videos, allowed us to assess the efficiency of our method on real field videos with sparse information. Furthermore, we evaluated our methodology of introducing non-targeted species diversity during the training phase.

To evaluate the potential of a semi-automated fish counting method, we utilized the testing dataset of our 5th model. Initially, we processed the testing dataset through our CNN model. Subsequently, we manually reviewed the network's annotations, making corrections where necessary. Finally, we compared the results obtained from both the semi-automated and fully automated methods in terms of recall, precision, and F-measure. Given that all misclassifications were rectified manually with the semi-automated method, the resulting precision was naturally 1, indicating that errors were solely attributable to false detections or missed individuals.

3. Results

3.1. Deep-learning model performance

Models trained on 27 species showed recall values of 0.91, 0.43, and 0.67 for *Carcharhinus amblyrhynchos*, *C. melanopterus*, and *Triaenodon obesus*, respectively. Interestingly, these results closely paralleled those obtained from models trained on only the 3 targeted shark species (0.92, 0.42, and 0.67). Models trained on 27 species demonstrated higher precision, with scores of 0.81, 0.43, and 0.80 compared to 0.72, 0.29, and 0.68 for the 3-species models (Fig. 1). Consequently, the models trained on 27 species exhibited higher F-measures (0.85, 0.43, and 0.72) than those trained on 3 species only (0.80, 0.34, and 0.67). Furthermore, we observed stability across the models for *C. amblyrhynchos*, with a F-measure standard deviation under 0.02 between models. This stability extended to precision and recall, with standard deviations ranging between 0.02 and 0.04. However, the two other species, having less data, displayed a slightly higher F-measure standard deviation, ranging from 0.08 to 0.11 between models. Considering the superior performance of the models trained on 27 species (*Supp.* Table 1), these were selected for subsequent analyses.

3.2. Comparing predicted and observed shark abondance

Significant correlations existed between MaxN shark abundance automatically predicted by our deep-learning models and manually estimated by biologists (Fig. 2). This was true for both video clips

[1] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md

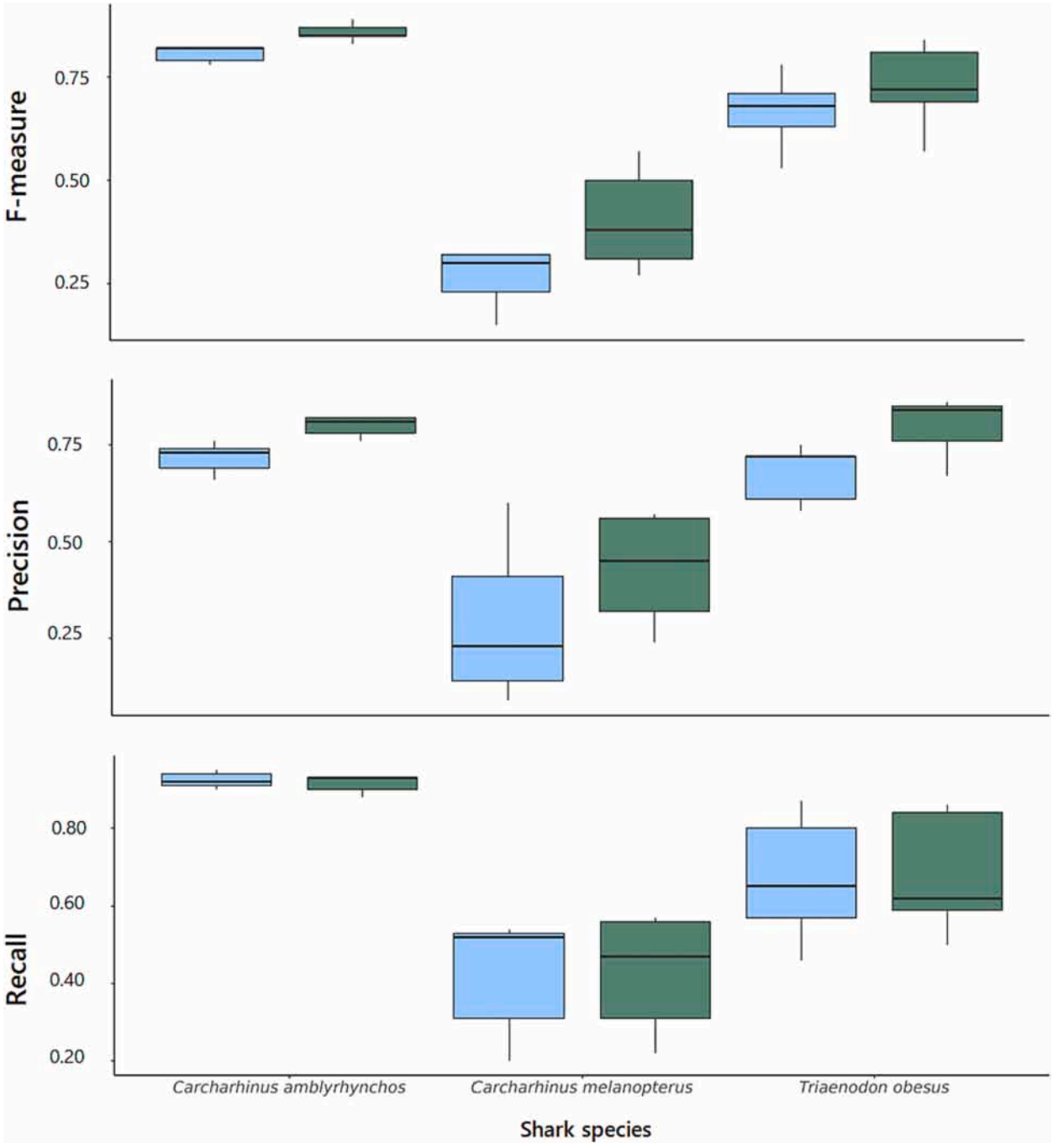


Fig. 1. Boxplot representations of precision, recall and F-measure metrics obtained for the 5 CNN models build through cross-validation. The results obtained on models trained with only the 3 common shark species of interest are in blue, and the results obtained with models trained on 27 species are shown in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

($Testing_{clips}$) and 1-h videos ($Testing_{stations}$) datasets. Pooling all species, the correlation coefficients R were 0.88 and 0.85 for the video clips and 1-h video datasets, respectively. The correlation was highest for *C. amblyrhynchos* ($R = 0.90$), followed by *T. obesus* ($R = 0.72$) and *C. melanopterus* ($R = 0.40$) using the video clips dataset. Lower R values were observed with the full-hour video dataset (0.88 for *C. amblyrhynchos*, 0.35 for *C. melanopterus*, 0.44 for *T. obesus*).

Regression analyses revealed that the models tended to overestimate shark abundance when none were present in the video and underestimate shark abundance when many were present (Fig. 2). While regression intercepts were all statistically significant, the value were close to zero (<0.05 for video clips, <0.7 for full hour videos). This suggests that the models occasionally predicted sharks when none were observed by biologists, resulting in false positive detections especially in data-poor full-hour videos. Regression slopes were consistently lower than one, indicating that the models tended to identify fewer sharks than observed by biologists, particularly when shark abundance was high. Pooling all species, slopes were >0.7, implying that when 6–7 sharks were observed by biologists in a given frame, the deep-learning models tended to miss 1–2 on average. Despite these challenges, the accuracy of our models in predicting MaxNs remained high with the video clip dataset, reaching values of 0.95, 0.99, and 0.94 for *C. amblyrhynchos*, *C. melanopterus*, and *T. obesus*, respectively (Fig. 2 and Fig. 3). Accuracies were relatively lower when using the full-hour video dataset (0.53, 0.77, 0.76 for

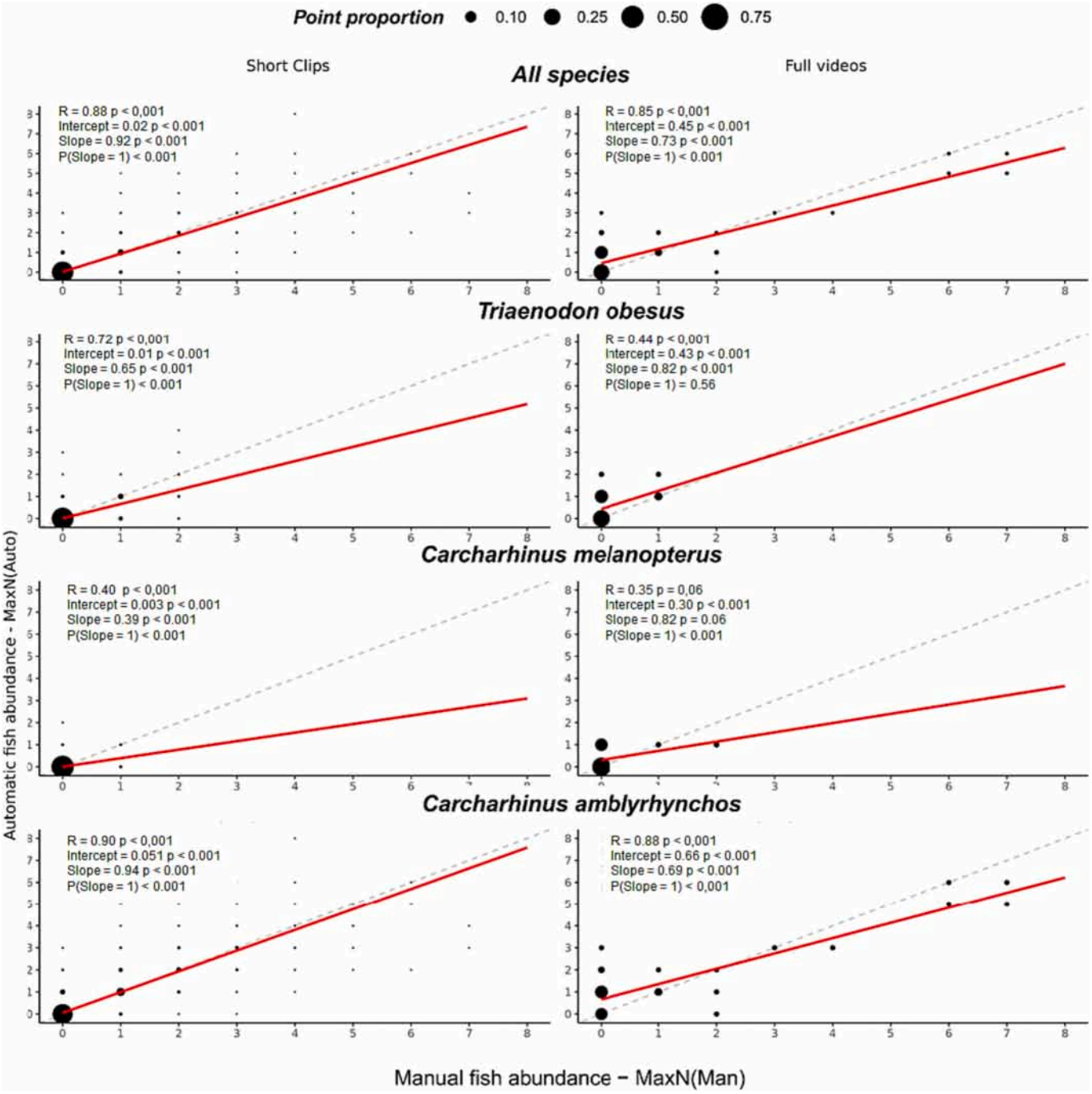


Fig. 2. Linear regressions between MaxN shark abundance manually observed by biologists and MaxN automatically predicted by our deep-learning models. The dotted line represents a perfect prediction of MaxN, e.g. x = y. On the left are the short clips from $Testing_{clips}$, on the right the 60-min videos from $Testing_{stations}$.

C. amblyrhynchos, *C. melanopterus*, and *T. obesus*, respectively). The most challenging issues in our method were images with small/far away individuals in the background (Fig. 4 *A*), individuals with a high degree of torsions when circling around the bait (Fig. 4 *B*) or few identifiable features (Fig. 4 *C*).

3.3. Improving shark abundance estimation with semi-automatic processing

The comparison of fully-automated (*Supp.* Table 2) and semi-automated predictions (*Supp. Table 3*) demonstrated notable enhancements in results. Precision scores increased from 0.76 (*C. amblyrhynchos*), 0.58 (*T. obesus*), and 0.23 (*C. melanopterus*) to 1, eliminating all misclassifications. Similarly, recall improved for all three species, with increases from 0.90 to 0.93, 0.46 to 0.75, and 0.52 to 0.69, respectively. These improvements in precision and recall translated into a substantial enhancement in F-measure, with increments of +0.14, +0.34, and + 0.50, resulting in semi-automated F-scores ranging from 0.82 to 0.97 (Table 2).

4. Discussion

Our study aimed to evaluate the effectiveness of Convolutional Neural Networks (CNNs) in extracting ecologically significant indicators from Baited Remote Underwater Video Stations (BRUVS). Specifically, we focused on assessing their efficiency in extracting MaxN, a widely

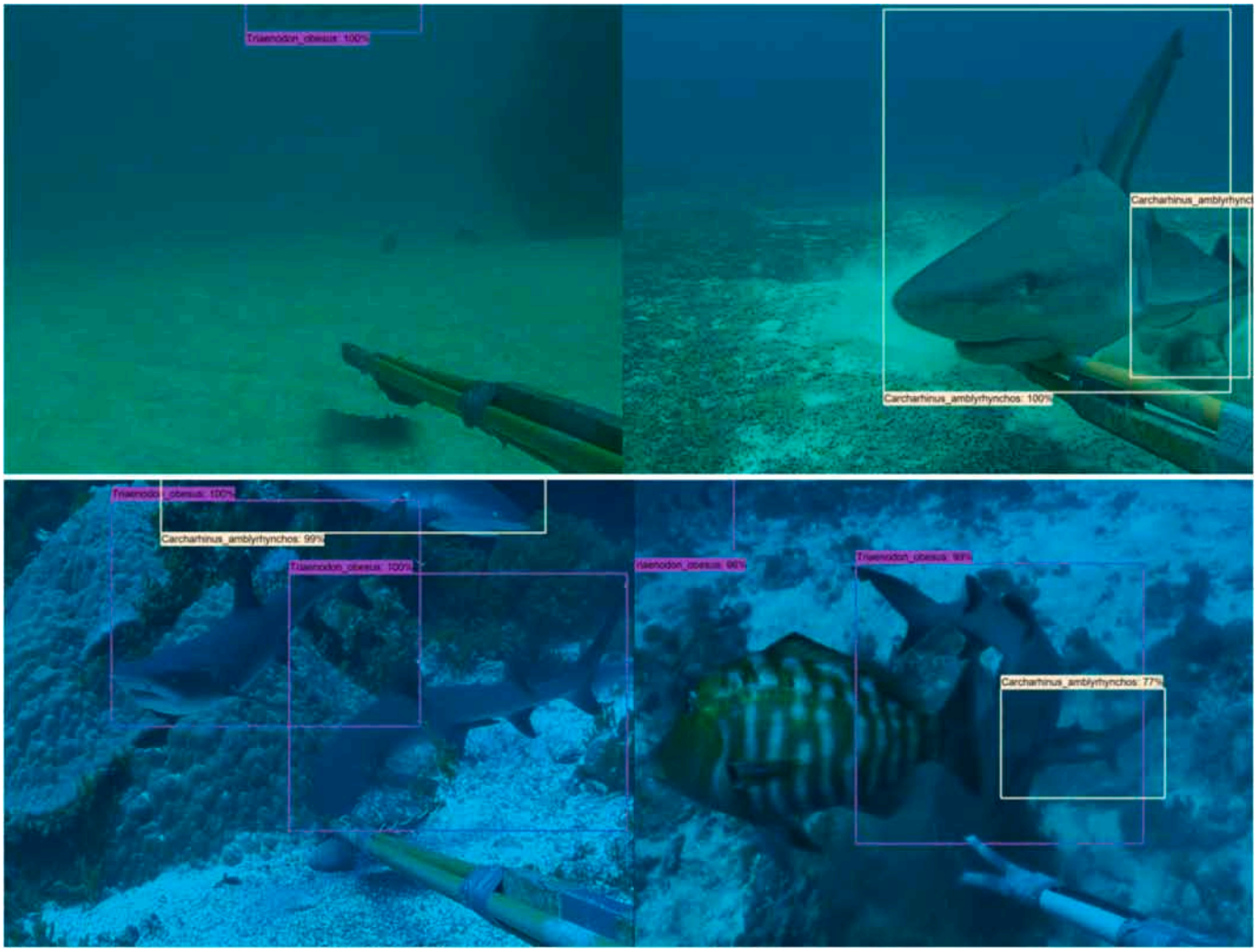


Fig. 3. Shark correctly detected an identified in complexes situations such as high degree of torsions and few visual features due to fish hiding each other and being partially out of the camera recording.

used abundance proxy metric employed by biologists to evaluate fish community structure. We conducted a comparative analysis between a fully automated and a semi-automated method, with manual annotation of MaxN for shark abundance. Our findings indicate that the fully automated method demonstrated high accuracy on dense data, with more varied results on 1-h videos containing sparse data. However, in both cases, strong correlations were observed between predicted MaxNs and human annotations for the three studied shark species. Additionally, the semi-automated method exhibited excellent performances, achieving F-measures up to 0.97 for shark detection and identification. The semi-automated approach holds the potential to significantly reduce the video processing workload for biologists, while still delivering high-quality biodiversity and abundance measurements. We selected the Faster R-CNN as the backbone architecture for our proposed method. To date, this architecture is still the best all-round object detector and classifier. While others, such as Single Shot Detector (SSD) and You Only Look Once (Yolo) are faster, the Faster R-CNN still produced better accuracy results (Bose and Kumar, 2020; Kaarmukilan et al., 2020; Lee et al., 2021). Nevertheless, it's important to highlight that this architecture is adaptable and can be fine-tuned with alternative layers, such as a regression layer, or memory layers. The field of object detection and identification in videos is dynamically evolving, witnessing continuous advancements at a swift pace. To ensure the adaptability and contemporaneity of our proposed method, we aim in the future to incorporate the latest techniques, particularly harnessing state-of-the-art attention mechanisms (Vaswani et al., 2017). Moreover, recent technological strides in time series processing offer promising avenues for refinement. Techniques such as contrastive representation distillation (Tian et al., 2020), federated distillation learning (Xing et al., 2021), long short-term memory (Sepp and Jurgen, 1997), and temporal feature network (Xiao et al., 2021) are among the innovative approaches in this domain. Methodologies such as few-shot/one shot learning and adaptative losses may also improve the models. Such methods may be explored to efficiently discriminate fish with few training data, such as the tiger shark *Galeocerdo cuvier* or the silvertip shark *Carcharhinus albimarginatus*. The integration of these methodologies carries the potential to enhance accuracy metrics and better exploit the temporal dimension of videos.

Our approach mimicked the way humans estimate fish biodiversity and abundance from BRUVS video stations, which involves manually detecting, identifying, and counting fish. Manual annotation allows ecologists to calculate MaxN standard abundance measurements. However, MaxN measurements not only take time, they also produce biased abundance estimates when the number of fish is >20 and the screen is saturated (Conn, 2011; MacNeil et al., 2020). On the contrary, automated methods not only offer instantaneous and effortless measurements but also have the capacity to provide unbiased fish abundance estimates, such as MeanCount. Unlike MaxN, MeanCount represents the mean number of fish per frame, eliminating the saturation bias. However, calculating MeanCount necessitates numerous observations for each species in a video, escalating the cost of annotation and

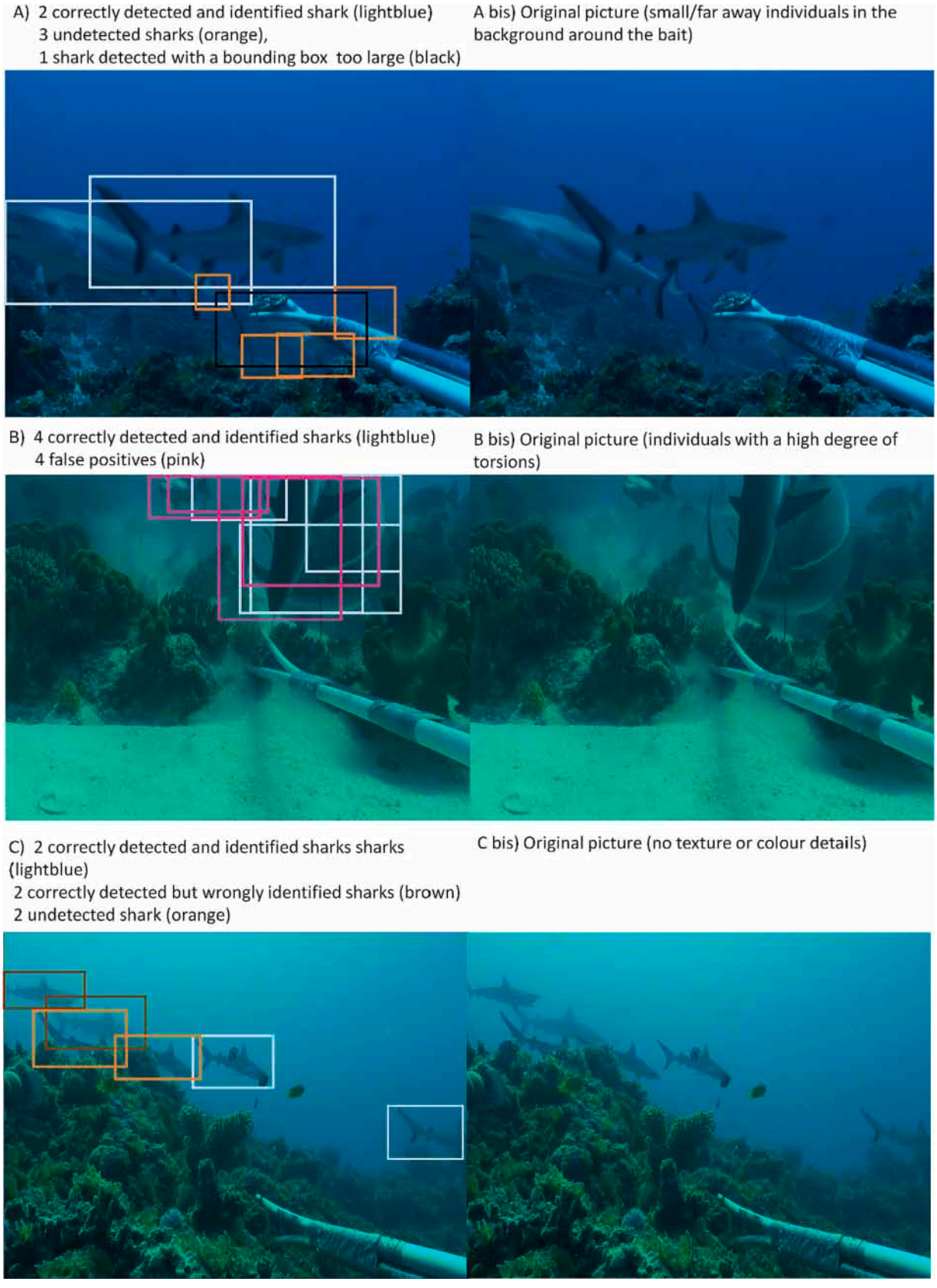


Fig. 4. Examples of shark detection or identification issues resulting in important differences between predicted MaxN and human observation.

Table 2
Precision, recall and F-measure through automatic and semi-automatic processing.

Reef shark species	Automated			Semi-Automated		
	Precision	Recall	F-measure	Precision	Recall	F-measure
C. amblyrhynchos	0.76	0.9	0.83	1	0.93	**0.97**
T. obesus	0.58	0.46	0.52	1	0.75	**0.86**
C. melanopterus	0.23	0.51	0.32	1	0.69	**0.82**

underscoring the importance of an automatic method capable of counting each species in each frame (a task nearly impossible to accomplish by human annotators). In addition to existing metrics, CNN detection and counting open avenues for novel indicators like species co-occurrence (i.e. species appearing together in a frame), species avoidance (i.e. species never appearing in the same frame), order of species appearance on videos, time spent by species on screen, etc. Time series, particularly, exemplify information that cannot be manually extracted given the time required to identify each fish of each species in each frame over hours of videos. Without a doubt, these novel AI-generated indicators would be immensely valuable to biologists, potentially revealing new insights into ecosystem functioning. Furthermore, such methods are applicable to all types of videos, fixed or moving, such as transects and diver-operated BRUVS.

5. Conclusion

Our study proved the effectiveness of Convolutional Neural Networks (CNNs) in extracting ecologically significant indicators from Baited Remote Underwater Video Stations (BRUVS). As for now, semi-automated methods, as well as automated MeanCount, could be used transitionally until fully automated methods accounting for the specificities of marine ecosystems can provide usable results for ecologists.

Author contributions

SV, LV, CI and MM designed the study. SV, LV collected the data. SV, LV and MM performed analyses. SV wrote the first draft. All authors contributed to writing, reviewing, editing, and gave final approval.

Funding

The study was funded by grant ANR "SEAMOUNTS" #ANR-18-CE02-0016, the French Oceanographic Fleet, and IRD core funding.

CRediT authorship contribution statement

Sébastien Villon: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Corina Iovan:** Investigation, Methodology, Project administration, Writing – review & editing. **Morgan Mangeas:** Formal analysis, Investigation, Methodology, Writing – review & editing. **Laurent Vigliola:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Visualization, Writing – review & editing.

Declaration of competing interest

All authors declare no competing interests.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102499.

References

Barone, M., Mollen, F.H., Giles, J.L., Marshall, L.J., Villate-Moreno, M., Mazzoldi, C., Pérez-Costas, E., Heine, J., Guisande, C., 2022. Performance of iSharkFin in the identification of wet dorsal fins from priority shark species. Eco. Inform. 68 https://doi.org/10.1016/j.ecoinf.2021.101514.

Baum, J.K., Blanchard, W., 2010. Inferring shark population trends from generalized linear mixed models of pelagic longline catch and effort data. Fish. Res. 102 (3), 229–239. https://doi.org/10.1016/j.fishres.2009.11.006.

Bose, S.R., Kumar, V.S., 2020. Efficient inception V2 based deep convolutional neural network for real-time hand action recognition. IET Image Process. 14 (4), 688–696. https://doi.org/10.1049/iet-ipr.2019.0985.

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. http://advances.sciencemag.org/.

Cappo, M., Harvey, B.E., Malcolm, H., Speare, P., 2003. Potential of video techniques to monitor diversity, abundance and size of fish in studies of marine protected areas. www.geomsoft.com.au.

Chen, G., Sun, P., Shang, Y., 2018. Automatic fish classification system using deep learning. In: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2017-November, pp. 24–29. https://doi.org/10.1109/ICTAI.2017.00016.

Conn, P.B., 2011. An evaluation and power analysis of fishery independent reef fish sampling in the gulf of mexico and U.S. South Atlantic.

Cui, S., Zhou, Y., Wang, Y., Zhai, L., 2020. Fish detection using deep learning. Appl. Comp. Intellig. Soft Comp. 2020 https://doi.org/10.1155/2020/3738108.

Davidson, L.N.K., Dulvy, N.K., 2017. Global marine protected areas to prevent extinctions. Nat. Ecol. Evol. 1 (2) https://doi.org/10.1038/s41559-016-0040.

Ditria, E.M., Lopez-Marcano, S., Sievers, M., Jinks, E.L., Brown, C.J., Connolly, R.M., 2020. Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning. Front. Mar. Sci. 7 https://doi.org/10.3389/fmars.2020.00429.

Ditria, E.M., Jinks, E.L., Connolly, R.M., 2021. Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. In: Animal Behaviour, vol. 177. Academic Press, pp. 31–37. https://doi.org/10.1016/j.anbehav.2021.04.018.

Edgar, G.J., Stuart-Smith, R.D., Willis, T.J., Kininmonth, S., Baker, S.C., Banks, S., Barrett, N.S., Becerro, M.A., Bernard, A.T.F., Berkhout, J., Buxton, C.D., Campbell, S. J., Cooper, A.T., Davey, M., Edgar, S.C., Försterra, G., Galván, D.E., Irigoyen, A.J., Kushner, D.J., Thomson, R.J., 2014. Global conservation outcomes depend on marine protected areas with five key features. Nature 506 (7487), 216–220. https://doi.org/10.1038/nature13022.

Goetze, J.S., Bond, T., McLean, D.L., Saunders, B.J., Langlois, T.J., Lindfield, S., Fullwood, L.A.F., Driessen, D., Shedrawi, G., Harvey, E.S., 2019. A field and video analysis guide for diver operated stereo-video. Methods Ecol. Evol. 10 (7), 1083–1090. https://doi.org/10.1111/2041-210X.13189.

Goodwin, M., Halvorsen, K.T., Jiao, L., Knausgård, K.M., Martin, A.H., Moyano, M., Oomen, R.A., Rasmussen, J.H., Sørdalen, T.K., Thorbjørnsen, S.H., 2022. Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. ICES J. Mar. Sci. 79 (2), 319–336. https://doi.org/10.1093/icesjms/fsab255.

Graham, N.A.J., Spalding, M.D., Sheppard, C.R.C., 2010. Reef shark declines in remote atolls highlight the need for multi-faceted conservation action. Aquat. Conserv. Mar. Freshwat. Ecosyst. 20 (5), 543–548. https://doi.org/10.1002/aqc.1116.

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. In: Trends in Ecology and Evolution, 34. Elsevier Ltd., pp. 369–383. https://doi.org/10.1016/j.tree.2019.01.005. Issue 4.

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. Eco. Inform. 57 https://doi.org/10.1016/j.ecoinf.2020.101088.

Jenrette, J., Liu, Z.Y.C., Chimote, P., Hastie, T., Fox, E., Ferretti, F., 2022. Shark detection and classification with machine learning. Eco. Inform. 69 https://doi.org/10.1016/j.ecoinf.2022.101673.

Jorgensen, S., Micheli, F., White, T., Van Houtan, K., Alfaro-Shigueto, J., Andrzejaczek, S., Arnoldi, N., Baum, J., Block, B., Britten, G., Butner, C., Caballero, S., Cardeñosa, D., Chapple, T., Clarke, S., Cortés, E., Dulvy, N., Fowler, S., Gallagher, A., Ferretti, F., 2022. Emergent research and priorities for shark and ray

conservation. Endanger. Species Res. 47, 171–203. https://doi.org/10.3354/esr01169.
Juhel, J.B., Vigliola, L., Mouillot, D., Kulbicki, M., Letessier, T.B., Meeuwig, J.J., Wantiez, L., 2018a. Reef accessibility impairs the protection of sharks. J. Appl. Ecol. 55 (2), 673–683. https://doi.org/10.1111/1365-2664.13007.
Juhel, J.B., Vigliola, L., Mouillot, D., Kulbicki, M., Letessier, T.B., Meeuwig, J.J., Wantiez, L., 2018b. Reef accessibility impairs the protection of sharks. J. Appl. Ecol. 55 (2), 673–683. https://doi.org/10.1111/1365-2664.13007.
Kaarmukilan, S.P., Poddar, S., Thomas, A.K., 2020. FPGA based deep learning models for object detection and recognition comparison of object detection: Comparison of object detection models using FPGA. In: Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020, pp. 471–474. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00088.
Knapp, S., Aronson, M.F.J., Carpenter, E., Herrera-Montes, A., Jung, K., Kotze, D.J., La Sorte, F.A., Lepczyk, C.A., Macgregor-Fors, I., Macivor, J.S., Moretti, M., Nilon, C.H., Piana, M.R., Rega-Brodsky, C.C., Salisbury, A., Threlfall, C.G., Trisos, C., Williams, N. S.G., Hahs, A.K., 2021. A research agenda for urban biodiversity in the global extinction crisis. In: BioScience, 71. Oxford University Press, pp. 268–279. https://doi.org/10.1093/biosci/biaa141. Issue 3.
Knausgård, K.M., Wiklund, A., Sørdalen, T.K., Halvorsen, K.T., Kleiven, A.R., Jiao, L., Goodwin, M., 2022. Temperate fish detection and classification: a deep learning based approach. Appl. Intell. 52 (6), 6988–7001. https://doi.org/10.1007/s10489-020-02154-9.
Langlois, T., Goetze, J., Bond, T., Monk, J., Abesamis, R.A., Asher, J., Barrett, N., Bernard, A.T.F., Bouchet, P.J., Birt, M.J., Cappo, M., Currey-Randall, L.M., Driessen, D., Fairclough, D.V., Fullwood, L.A.F., Gibbons, B.A., Harasti, D., Heupel, M.R., Hicks, J., Harvey, E.S., 2020. A field and video annotation guide for baited remote underwater stereo-video surveys of demersal fish assemblages. Methods Ecol. Evol. 11 (11), 1401–1409. https://doi.org/10.1111/2041-210X.13470.
Le, N.A., Moon, J., Lowe, C.G., Kim, H.-I., Choi, S.-I., 2022. An automated framework based on deep learning for shark recognition. J. Mar. Sci. Eng. 10 (7), 942. https://doi.org/10.3390/jmse10070942.
Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. In: Nature, 521. Nature Publishing Group, pp. 436–444. https://doi.org/10.1038/nature14539. Issue 7553.
Lee, J., Wang, P., Xu, R., Dasari, V., Weston, N., Li, Y., Bagchi, S., Chaterji, S., 2021. Benchmarking video object detection systems on embedded devices under resource contention. In: EMDL 2021 - Proceedings of the 2021 5th International Workshop on Embedded and Mobile Deep Learning, Part of MobiSys, 2021, pp. 19–24. https://doi.org/10.1145/3469116.3470010.
Lees, A.C., Attwood, S., Barlow, J., Phalan, B., 2020. Biodiversity scientists must fight the creeping rise of extinction denial. In: Nature Ecology and Evolution, 4. Nature Research, pp. 1440–1443. https://doi.org/10.1038/s41559-020-01285-z. Issue 11.
Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2014. Microsoft COCO: Common Objects in Context. 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.
Merencilla, N.E., Sarraga Alon, A., Fernando, G.J.O., Cepe, E.M., Malunao, D.C., 2021. Shark-EYE: A deep inference convolutional neural network of shark detection for underwater diving surveillance. In: Proceedings of 2nd IEEE International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2021, pp. 384–388. https://doi.org/10.1109/ICCIKE51210.2021.9410715.
Rathi, D., Jain, S., Indu, S., 2018. Underwater Fish Species Classification using Convolutional Neural Network and Deep Learning (arXiv:1805.10106v1 [cs.CV]). June. http://arxiv.org/abs/1805.10106.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. http://arxiv.org/abs/1506.01497.
Rizzari, J.R., Frisch, A.J., Connolly, S.R., 2014. How robust are estimates of coral reef shark depletion? Biol. Conserv. 176, 39–47. https://doi.org/10.1016/j.biocon.2014.05.003.
Rull, V., 2022. Biodiversity crisis or sixth mass extinction? Does the current anthropogenic biodiversity crisis really qualify as a mass extinction? EMBO reports 23 (1), e54193. https://www.eol.org/.
Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., Harvey, E., 2016. Fish species classification in unconstrained underwater environments based on deep learning. Limnol. Oceanogr. Methods 14 (9), 570–585. https://doi.org/10.1002/lom3.10113.
Sepp, Hochreiter, Jurgen, Schmidhuber, 1997. Long short term memory. Neural Comput. 9, 1735–1780.
Tian, Y., Krishnan, D., Research, G., Isola, P., 2020. Contrastive Representation Distillation. http://github.com/HobbitLong/RepDistiller.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. http://arxiv.org/abs/1706.03762.
Villon, S., Chaumont, M., Subsol, G., Villéger, S., 2016. Coral reef fish detection and recognition in underwater videos by supervised machine learning. In: Advanced Concepts for Intelligent Vision Systems: 17th International Conference, ACIVS 2016.
Villon, S., Mouillot, D., Chaumont, M., Darling, E.S., Subsol, G., Claverie, T., Villéger, S., Fr, V., 2018. A deep learning algorithm for accurate and fast identification of coral reef fishes in underwater videos. Eco. Inform. Ecological informatics. 48, 238–244.
Villon, S., Iovan, C., Mangeas, M., Vigliola, L., 2022. Confronting deep-learning and biodiversity challenges for automatic video-monitoring of marine ecosystems. Sensors 22 (2). https://doi.org/10.3390/s22020497.
Wagner, D.L., 2019. Insect declines in the Anthropocene. Annu. Rev. Entomol. https://doi.org/10.1146/annurev-ento-011019.
Weinstein, B.G., 2018. A computer vision for animal ecology. J. Anim. Ecol. 87 (3), 533–545. Blackwell Publishing Ltd. https://doi.org/10.1111/1365-2656.12780.
Whitmarsh, S.K., Fairweather, P.G., Huveneers, C., 2017. What is big BRUVver up to? Methods and uses of baited underwater video. In: Reviews in Fish Biology and Fisheries, 27. Springer International Publishing, pp. 53–73. https://doi.org/10.1007/s11160-016-9450-1. Issue 1.
Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Słupski, T., Koumba Pambo, A.F., Rogala, M., Bahaa-el-din, L., Boekee, K., Brittain, S., Cardoso, A.W., Henschel, P., Lehmann, D., Momboua, B., Kiebou Opepa, C., Orbell, C., Pitman, R.T., Robinson, H. S., Abernethy, K.A., 2021. Robust ecological analysis of camera trap data labelled by a machine learning model. Methods Ecol. Evol. 12 (6), 1080–1092. https://doi.org/10.1111/2041-210X.13576.
Xiao, Z., Xu, X., Xing, H., Luo, S., Dai, P., Zhan, D., 2021. RTFN: a robust temporal feature network for time series classification. Inf. Sci. 571, 65–86. https://doi.org/10.1016/j.ins.2021.04.053.
Xing, H., Xiao, Z., Qu, R., Zhu, Z., Zhao, B., 2021. An Efficient Federated Distillation Learning System for Multi-task Time Series Classification. http://arxiv.org/abs/2201.00011.
Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V., 2017. Learning Transferable Architectures for Scalable Image Recognition. http://arxiv.org/abs/1707.07012.