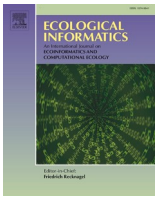
Ecological Informatics 69 (2022) 101673

Contents lists available at ScienceDirect

Ecological Informatics

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



Shark detection and classification with machine learning

J. Jenrette [a,*], Z. Y.-C. Liu [c], P. Chimote [b], T. Hastie [d], E. Fox [b], F. Ferretti [a]

[a] Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA, USA
[b] Department of Computer Science, Virginia Tech, Blacksburg, VA, USA
[c] Hopkins Marine Station, Stanford University, Pacific Grove, CA, USA
[d] Department of Statistics and Department of Biomedical Data Science, Stanford University, CA, USA

ARTICLE INFO

Keywords:
Data-mining
Image classification
Instagram
Machine learning
Remote monitoring
Sharks

ABSTRACT

Suitable shark conservation depends on well-informed population assessments. Direct methods such as scientific surveys and fisheries monitoring are adequate for defining population statuses, but species-specific indices of abundance and distribution coming from these sources are rare for most shark species. We can rapidly fill these information gaps by boosting media-based remote monitoring efforts with machine learning and automation.

We created a database of 53,345 shark images covering 219 species of sharks, and packaged object-detection and image classification models into a Shark Detector bundle. The Shark Detector recognizes and classifies sharks from videos and images using transfer learning and convolutional neural networks (CNNs). We applied these models to common data-generation approaches of sharks: collecting occurrence records from photographs taken by the public or citizen scientists, processing baited remote camera footage and online videos, and data-mining Instagram. We examined the accuracy of each model and tested genus and species prediction correctness as a result of training data quantity.

The Shark Detector can classify 47 species pertaining to 26 genera. It sorted heterogeneous datasets of images sourced from Instagram with 91% accuracy and classified species with 70% accuracy. It located sharks in baited remote footage and YouTube videos with 89% accuracy, and classified located subjects to the species level with 69% accuracy. All data-generation methods were processed without manual interaction.

As media-based remote monitoring appears to dominate methods for observing sharks in nature, we developed an open-source Shark Detector to facilitate common identification applications. Prediction accuracy of the software pipeline increases as more images are added to the training dataset. We provide public access to the software on our GitHub page.

1. Introduction

Sharks are excellent indicators of ocean environmental health however they are constantly challenged by growing fishing pressures as well as poor management and conservation as a result of data paucity, insufficient taxonomic knowledge, and underdeveloped monitoring methods (Jorgensen et al., 2022). Observation data of sharks via surveying and fisheries monitoring is often extremely costly or difficult to collect and is exacerbated when species with larger home ranges are considered (Baum and Blanchard, 2010). Furthermore, classification of sharks is still debated for many species (Serena et al., 2020). The combination of observed global declines and increasing data resolution has resulted in the amount of IUCN-listed threatened species doubling since 2014 (Dulvy et al., 2021). Sharks remain an extremely data deficient group of marine animals and these information gaps contribute to the lack of abundance and distribution indices as well as taxonomic precision needed to properly assess population statistics (Ferretti et al., Unpublished data; Jorgensen et al., 2022).

Image-based biomonitoring is a transformative alternative to expensive and invasive direct observation methods in ecological surveys of marine and terrestrial environments (Siddiqui et al., 2018; Weinstein, 2018; Whytock et al., 2021). With significant advancements in baited underwater remote videos (BRUVs), motion-activated camera traps, and crowdsourced citizen science media, ecological information is being produced at a rate never seen before (Goetze et al., 2019; Weinstein, 2018; Whytock et al., 2021). Importantly, remote monitoring methods generate visual media that can help fill the shark information gaps. These methods are non-invasive and useful for minimizing sampling

* Corresponding author.
E-mail address: jjeremy1@vt.edu (J. Jenrette).

https://doi.org/10.1016/j.ecoinf.2022.101673
Received 17 February 2022; Received in revised form 6 May 2022; Accepted 8 May 2022
Available online 16 May 2022

effort. However they produce large quantities of media to post-process for species identification and analyses, including removing irrelevant images (Swanson et al., 2015). Studies like Tabak et al. (2019) and Malde et al. (2019) stress the importance of using deep learning programs to filter out unrelated content and facilitate rapid sampling.

Deep learning algorithms are highly flexible and well suited for approaching many of these tasks (LeCun et al., 2015; Malde et al., 2019; Siddiqui et al., 2018). They have been used to estimate fish sizes from images (Alvarez Ellacur'ıa et al., 2019; Garcia et al., 2019), identify discarded and processed fish (French et al., 2019), and classify acoustic and movement data (Brautaset et al., 2020; Fallon et al., 2016; Kadar et al., 2020). However, machine-learned detection and image classification of shark species is seldom studied due to a lack of a sufficient amount of training data (Ferretti et al., Unpublished data). Video and photographic documentation of sharks are rarely obtained from commercial fisheries. Such images are more reliably sourced from tourists, social networks, and underwater photographers (Taklis et al., 2020). Consequently, there are few studies that have curated a training dataset of shark images. iSharkFin is a recognition system which can identify 39 species of sharks from pictures of associated dorsal fins (Barone et al., 2022). However this system does not classify whole-body images because it is primarily focused on the tracing of illegal fin trading, and requires users to manually select features and input points that describe fin shape. As a step forward, Seek is a generalist image classifier which sources the iNat2017 archive of 859,000 images to detect and classify over 5000 organisms, including shark species (Horn et al., 2017). This app advances animal classification, but because of its large scope, it remains inaccurate for classifying the 509 species of living sharks.

Because there are many morphologically diverse and data-poor shark species, classifying them is not a straightforward process for machine learning. Here, we approach this challenge by first constructing the largest training dataset of shark images. Second, we combine object-detection and hierarchical classification methods for images and videos to facilitate the creation of biologically relevant data on sharks. Assembling and annotating large amounts of data-mined and user-uploaded media, for the use of conservation, is an emerging approach (Di Minin et al., 2015; Sullivan et al., 2014). Fish species classification with machine learning algorithms has only begun developing within the last two decades (Siddiqui et al., 2018). As a result of interacting with big data and citizen scientists, we have created the largest and most diverse archive of shark images. Few studies have combined object-detection with classification for the purpose of increasing shark taxonomic accuracy (Barone et al., 2022; Horn et al., 2017). Our objective was automatically detecting and classifying, to the species level, any image with perceptible shark features. Because our methods build upon standard recognition and data-mining approaches, we can generate, detect, and classify shark-sourced visual media. We can efficiently post-process BRUVs, camera trap images, remotely operated underwater vehicle (ROV) footage, and shared social media by automatically removing irrelevant content and classifying shark species.

2. Methods

Our shark detection and classification pipeline is composed of several steps and three main components (Fig. 1): 1 - an object-detection model called the Shark Locator (SL), which locates one or several shark subjects in images and draws bounding boxes around them; 2 - a binary sorting model called Shark Identifier (SI) which sorts images of sharks from a pool of heterogeneous images; and 3 - multiclass models called Shark Classifiers (SCs) which classify shark images to the genus and species levels. Combining these three modeling components, we developed a shark identification and classification pipeline called Shark Detector (SD) which can ingest any media containing shark subjects, locate and sort subjects according to relevance, and classify the sharks to the species level. Shark training images for developing these models were mainly sourced from sharkPulse – a crowd-sourcing platform which mines and aggregates shark media from social networks, citizen science projects, user submissions, and other electronic archives (Ferretti et al., Unpublished data).

2.1. Shark locator: object-detection

The identification pipeline starts with the SL. This model is primarily used to inflate an initial training dataset. It crops one or multiple shark subjects from images thereby creating new images. It locates shark subjects in videos (e.g., BRUV footage) and extracts frames with shark subjects. This process had the dual objective of removing any irrelevant subject or noisy background from shark images which challenged the training process, and boosting the training dataset by splitting images with multiple shark subjects into multiple distinct shark training images. Cropping shark features from images and video frames provided better training quality (see Fig. 2).

To build the SL, we sourced Tensorflow's Model Garden (Yu et al., 2020) and used a Faster Region-based Convolutional Neural Network (Faster-RCNN) algorithm. The model was trained with the Common Objects in Context (COCO) datasets (consisting of 236 shark images) to detect and draw boxes around sharks (Lin et al., 2014; Ren et al., 2016). Faster-RCNN can detect more than one object within a frame, allowing

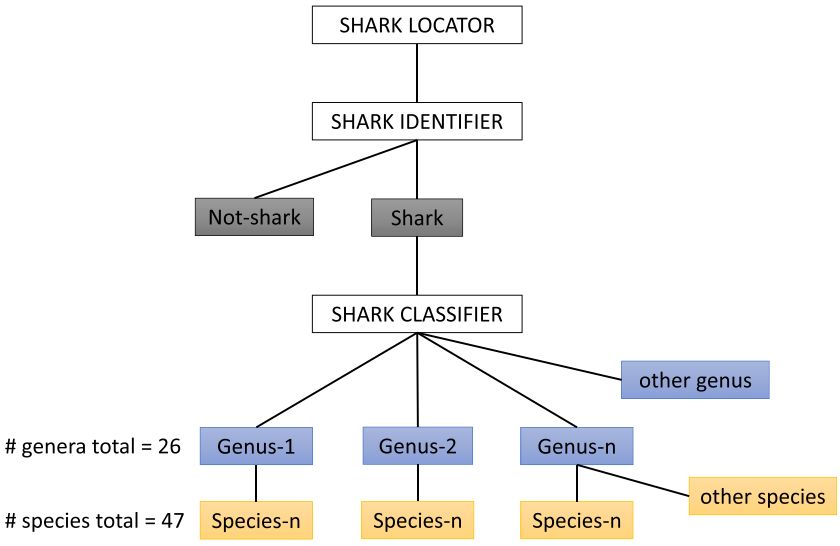


Fig. 1. The Shark Detector system is composed of object-detection and classification packages that work best in a step-wise procedure. Additionally, by detecting shark subjects, the Shark Locator synthetically supplements the sharkPulse archive with cropped shark images available to Shark Identifier and Shark Classifier models as new training data. Videos are processed in the order of locating, identifying, and then classifying. Heterogeneous data-mined datasets are processed in the order of identifying and then classifying.

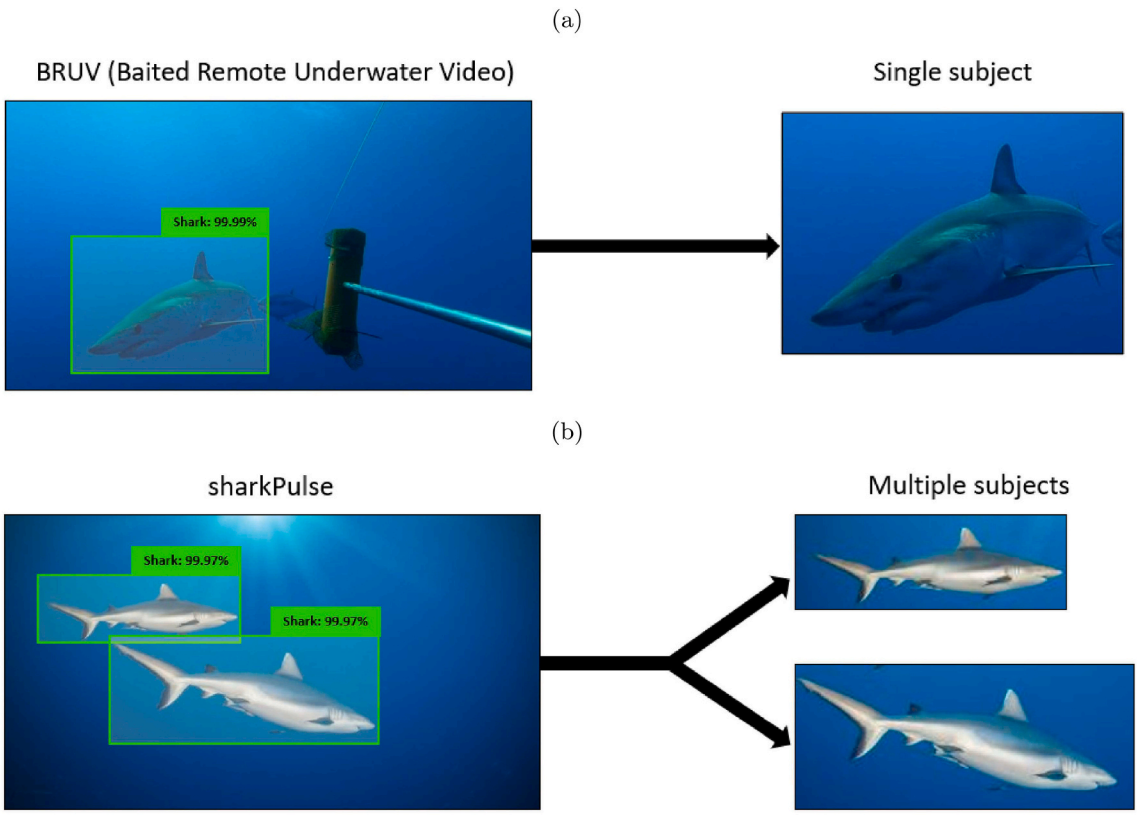


Fig. 2. The SL object-detection model draws boxes corresponding to confidence levels of shark presence. (a) A juvenile shortfin mako is detected and a single autocropped image is processed, removing irrelevant objects such as the bait canister and bluefin tuna. (b) Multiple *Carcharhinidae* species are detected and two images are cropped from a single image.

multiple boxes to be drawn. To reduce processing time, we set a limit of 10 boxes that could be drawn within a single frame. The SL boosted our classification dataset from 24,546 images to 53,345 images.

2.2. Shark Identifier: binary model

Second, we developed a binary sorting model. The SI identifies shark vs. non-shark subjects in images and is used to filter out non-shark images, before the remaining images are taxonomically classified. We sourced 53,345 shark images from Instagram and sharkPulse, and additionally, we sourced 50,260 non-shark images from Instagram (Table 1). First, we constructed the SI to learn key shark features from training images by optimization of the binary cross-entropy loss function (LeCun and Bengio, 1998). We incorporated a pre-trained model to reduce the number of training steps. This is known as transfer learning. The SI was pre-trained with the VGG16 network which is trained on 1.28 million images with 1000 categories. VGG16 achieves 92.7% test accuracy on the ImageNet dataset (Simonyan and Zisserman, 2015). CNN algorithms perform best when the categories of interest are well represented and balanced (Liu et al., 2019). Non-shark images were sourced entirely from Instagram. We resized images to 150 × 150 pixels to reduce memory consumption.

Table 1
SD packages trained with images sourced from various social networks and online archives.

	# of models	Training images	Test images	training source
Shark Locator	1	236		COCO dataset
Shark Identifier	1	93,244	10,361	sharkPulse, Flickr, iNaturalist, Instagram, YouTube
Genus-specific Classifier	1	33,050	3672	sharkPulse, Flickr, iNaturalist, Instagram, YouTube
Species-specific Classifier	18	17,319	1924	sharkPulse, Flickr, iNaturalist, Instagram, YouTube

To increase training accuracy, we used position image augmentation techniques, i.e., we artificially augmented images with transformations such as width and height shifting, shearing, zooming, and rotations (Tabak et al., 2019; Taylor and Nitschke, 2017). Then, we constructed convolutional-pooling layers which act as checkpoints for summarizing features the model has learned (LeCun and Bengio, 1998). When shark features are learned from trained images, the CNN generates parameters called weights. Weights were first initialized when we pre-trained networks on the ImageNet dataset. We froze the bottom pre-trained layers to prevent weights from being modified while we trained the top layers for shark features. VGG16 contains 16 pre-trained layers and we added four layers to train for shark features. We trained the CNN to accept augmented and regular training images as raw pixels and gradually learn output predictions as they passed through convolutional-pooling layers. To facilitate adaptive learning, we adjusted the algorithm's learning rate to 5e-4 with the Adam optimizer (Kingma and Ba, 2014). To avoid vanishing gradients and improve training speed, we incorporated Rectified Linear Units (i.e., ReLU activation function) into the CNN's fully connected layers (Nair and Hinton, 2010). The output layer was composed of two neurons for classification, corresponding to the number of classes being trained: *shark* and *non-shark*. These neurons were normalized with a softmax activation function (Bridle, 1990). We trained the model with 90% of the training set, and validated with the remaining images over 10 epochs. One epoch represents one full cycle where the algorithm has processed the entire training dataset. To prevent the model from overfitting, we incorporated dropout and regularization parameters. Dropout effectively removes a percentage of

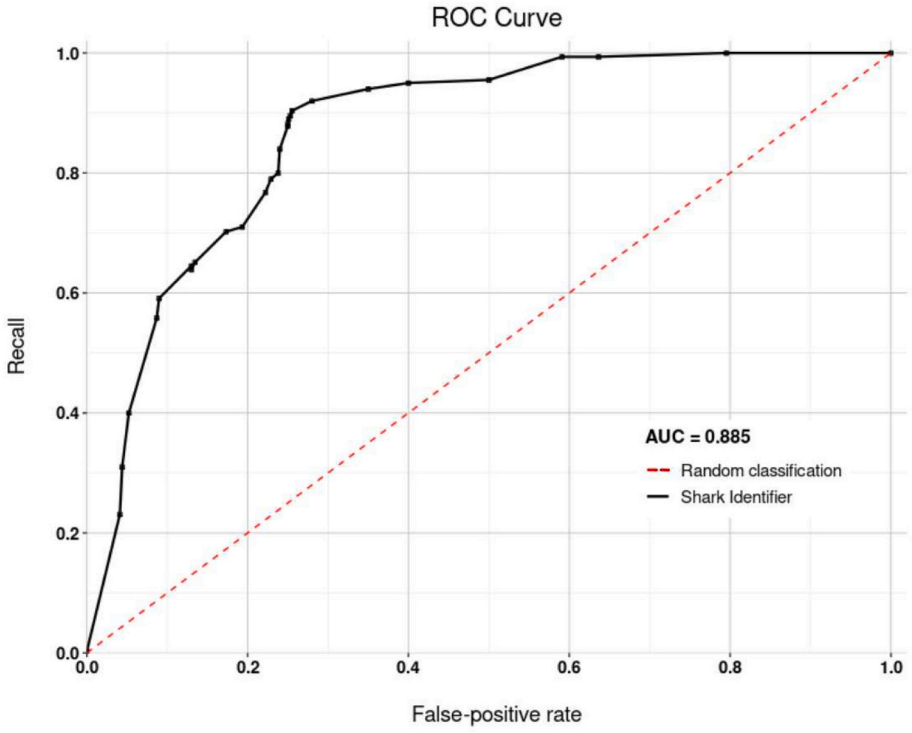


Fig. 3. Receiver operating characteristic (ROC) curve of the SI binary classification scheme. The Area Under the Curve (AUC) conveys a probability measure of how likely the model is to separate between positive and negative classes. A straight diagonal line indicates a no-skill classification model which discriminates randomly.

neurons from the model that have learned features, and helps the model generalize when predicting new images. We set dropout to 30%. Image augmentation and early-stopping are forms of regularization, and are used to minimize validation error (Srivastava et al., 2014). We incorporated early-stopping of training when test accuracy decreased for three consecutive epochs. Then we measured the ROC curve in Fig. 3. We built the CNN in Python using the Keras and Tensorflow packages (Abadi et al., 2016; Chollet, 2015).

2.3. Shark classifier: genus and species classification

We developed the SC as a hierarchical classification framework for classifying the identified shark images taxonomically. We trained one genus-specific model and a series of local species-specific models - one for each genus (Fig. 1). The SC ingests the filtered shark images and classifies them at the genus level with the genus-specific classifier (GSC). Then, depending on the genus, a species-specific classifier (SSCg) would predict the most likely species.

We trained the SC with the sharkPulse database, images cropped with the SL, and Instagram images. In total, the SC contains 74 genera and 219 species of sharks with an average of 167 images per species (Table 1). The GSC was trained with 36,722 images and the SSCg was trained with 19,243 images. We evaluated the recall of a genus class vs. its training data quantity in Fig. 4a which revealed an average of 433 ± 47 images were needed to produce >50% recall. Recall measures the proportion of shark images that were correctly classified. Most genera (>64%) did not contain this many images to produce adequate training quality. Once a genus was classified, we looked at the same relationship for species classes in Fig. 4b and discovered an average of 161 ± 41 images were needed to produce >50% recall. We used these averages as training data quantity thresholds for the SC (see Fig. 4c and d). Next, to better understand why models confused classes with each other, we examined two metrics. We used Pielou's evenness index, usually employed to assess whether and to what extent species's abundances are uniform in a community, to quantify how balanced training datasets were (Pielou, 1966).

Then we evaluated the difference in morphology by calculating the Euclidean distance between species. This was done by collecting a common set of morphometric measurements for 124 species that were available in the rfishbase R package and that represented our dataset (Boettiger et al., 2012; Norouzzadeh et al., 2018). We used total length, standard length, fork length, and head length. Then we calculated the average for each measurement to create a morphometric centroid of all species. We compared each species to this centroid to assess morphological homogeneity.

For the GSC, we trained 36,722 images across 26 genus classes that met the training data threshold of 433 images (Table 1). We trained a 27*th* class with 2593 images to represent the >64% of genera that did not meet the training data threshold. This class was labeled *other genus*. Since there were 18 shark genera containing two or more species, we developed 18 SSCg models having a variable number of classes. SSCg species classes that contained <161 images were added to an *other species* class. The exception to this rule occurred when a genus contained exactly two species (e.g., *Echinorhinus* and *Negaprion*). In this case, regardless of training data quantity, both species were trained with their respective labels and the *other species* class was not incorporated. Regardless of species-specific dataset size, if their parent genus did not meet the threshold, a SSCg local model was not trained. We trained 18 SSCg models with 19,243 images. The SC is capable of classifying 47 species.

We optimized the models with the categorical cross-entropy function to learn shark features that are specific to genus and species classes (LeCun and Bengio, 1998). To prevent redundant feature training and incorporate fewer parameters, we used DenseNet201 as our pre-trained network for multiclass classification (Huang et al., 2016). We used image augmentations and passed our training dataset through 20 convolutional-pooling layers to generate feature maps at each layer, creating weights. We adjusted the SC's learning rate to 9e-3 with the Adagrad optimizer (Duchi et al., 2011). We activated layers with the *sigmoid* function (Narayan, 1997) which, in the case of the SC, facilitated

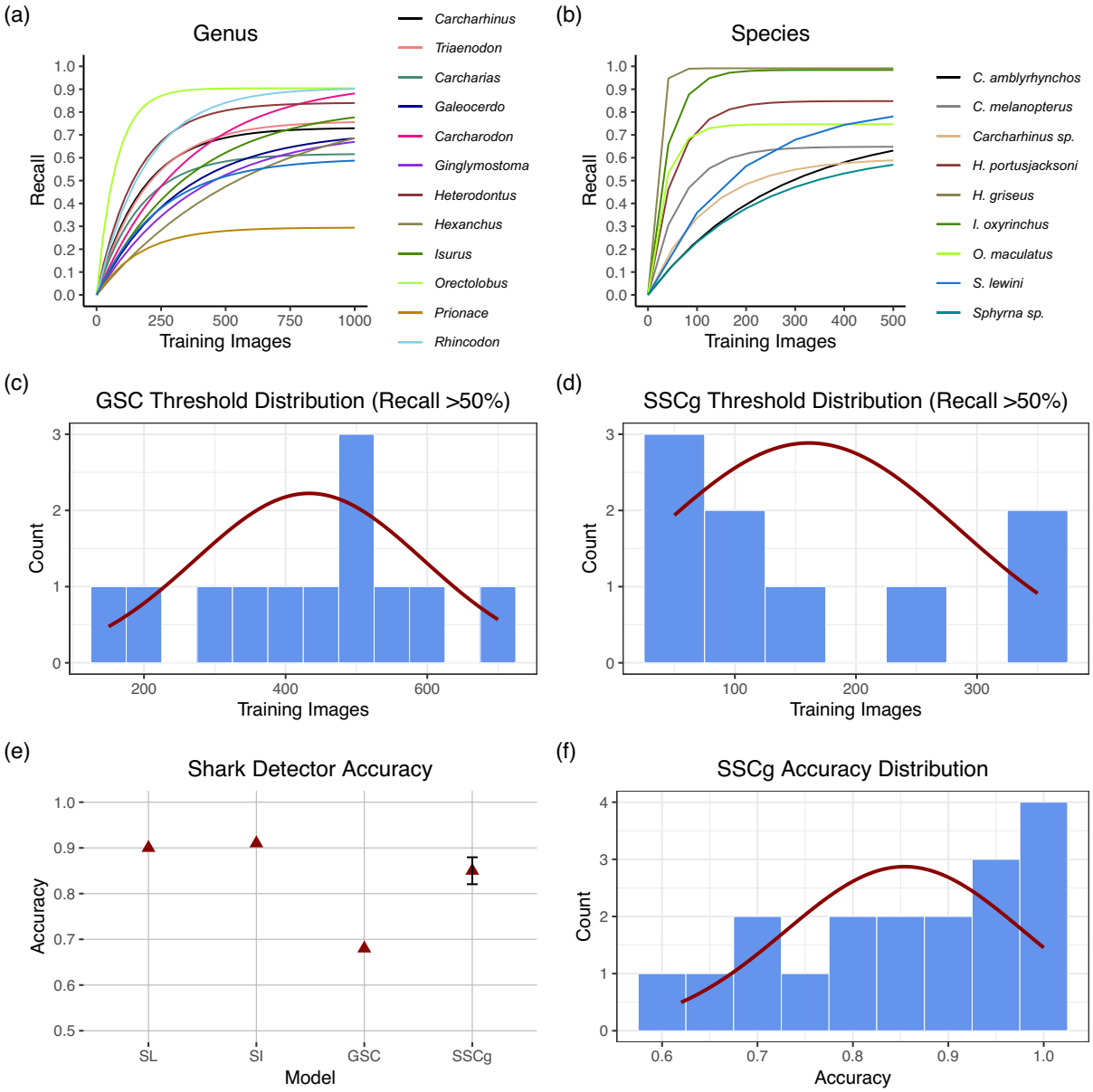


Fig. 4. Measured performance of SD components. (a) GSC accuracy for 13 genus classes as a result of training dataset size fit with a 2-parameter asymptotic model. The asymptotic curves represent the maximum recall a class can achieve within the model. (b) SSCg accuracy of seven species classes and two classes that contain a data-poor *Carcharhinus* sp. and *Sphyrna* sp. (c) Distribution of dataset size threshold for 12 GSC classes (*Prionace* was excluded due to recall never reaching 50%). Curves represent the density of normal distribution. (d) Distribution of dataset size threshold for nine SSCg classes whose parent genera contain more than two species. (e) Performance of all SD components with a standard error interval for SSCg models. (f) Accuracy distribution of all 18 SSCg models.

higher test accuracy than ReLU activation units. We normalized classes with the softmax activation function (Bridle, 1990). We tuned dropout to 15% and trained the models over 15 epochs while incorporating early-stopping if validation loss increased for 5 consecutive epochs (Srivastava et al., 2014).

2.4. Shark detector performance

To demonstrate the potential of this approach, we applied the SD to three common cases of data generation methods involving sharks, and measured performance. In the first method, we used the SL to locate shark subjects in the sharkPulse dataset. We used a detection threshold of 0.9. Before locating shark subjects, we resized images to 512 × 512. We calculated the proportion of shark images that were located by the SL.

Second, we evaluated a data pipeline we developed for sharkPulse, where global sighting records of sharks are generated from the social network Instagram (Jenrette et al., Unpublished data). We extracted images from Instagram, identified shark images (Table 2) and classified them (Table 3). In this case, we excluded the SL when processing Instagram images because it was more computationally expensive and did not significantly impact classification accuracy.

In the third method, we post-processed two BRUV videos and 5 YouTube (YT) videos with the goal of locating, identifying, and classifying all sharks present (Fig. 2 and Table 4). Videos were chosen to represent varying habitat types, conventional ecological surveys, and data-mining methods focused on sharks. We processed all videos one after the other to evaluate total processing time in addition to individual processing time. All videos were recorded at 30 frames per second. We extracted 1 frame per second and resized frames to 512 × 512 to reduce

Table 2
Hashtags relevant to specific shark species that were data-mined from Instagram. The result was heterogeneous datasets of images. We measured the SI's sorting accuracy and error rate at a confidence threshold = 0.5.

Hashtag of shark group	Test images	TP	FP	TN	FN	Recall	Precision	Specificity	FPR	FNR	F1 score
#tigershark	1590	299	19	1261	11	0.96	0.94	0.99	0.01	0.04	0.95
#blueshark	1269	343	22	889	15	0.96	0.94	0.98	0.02	0.04	0.95
#whaleshark	1144	309	21	800	14	0.96	0.94	0.97	0.03	0.04	0.95
#makoshark	988	228	15	730	15	0.94	0.94	0.98	0.02	0.06	0.94
#scallopedhammerhead	871	149	18	701	3	0.98	0.89	0.97	0.03	0.02	0.93
#sandtigershark	1011	220	21	753	17	0.93	0.91	0.97	0.03	0.07	0.92
#nurseshark	1370	180	23	1156	11	0.94	0.89	0.98	0.02	0.06	0.91
#greatwhites	849	251	24	550	24	0.91	0.91	0.96	0.04	0.09	0.91
#blacktipshark	955	209	40	700	6	0.97	0.84	0.95	0.05	0.03	0.90
#portjacksonshark	1012	191	28	971	13	0.94	0.87	0.97	0.03	0.06	0.90
#sixgillshark	287	107	19	150	11	0.91	0.85	0.89	0.11	0.09	0.88
#spottedwobbegong	180	97	18	55	10	0.91	0.84	0.75	0.25	0.09	0.87
#whitetipreefshark	890	250	43	561	36	0.87	0.85	0.93	0.07	0.13	0.86
#greyreefshark	901	203	55	600	43	0.83	0.79	0.92	0.08	0.17	0.81
total	13,317	3036	366	9877	229	0.93	0.89	0.96	0.04	0.07	0.91

Table 3
SC classification of data-mined images from Instagram. Recall was measured for the SC's top species prediction as well as the top three predictions.

Species	Scientific Name	Training images	Test images	Recall	Top-3 Recall
Whale shark	*Rhincodon typus*	1602	309	0.95	0.99
Port jackson shark	*Heterodontus portusjacksoni*	1172	191	0.87	0.98
White shark	*Carcharodon carcharias*	2290	251	0.87	0.9
Whitetip reef shark	*Triaenodon obesus*	1786	250	0.79	0.92
Blacktip reef shark	*Carcharhinus melanopterus*	829	209	0.77	0.91
Shortfin mako	*Isurus oxyrinchus*	1360	228	0.76	0.95
Spotted wobbegong	*Orectolobus maculatus*	1019	97	0.74	0.99
Nurse shark	*Ginglymostoma cirratum*	821	180	0.7	0.9
Tiger shark	*Galeocerdo cuvier*	1117	299	0.68	0.91
Grey reef shark	*Carcharhinus amblyrhynchos*	550	203	0.68	0.88
Bluntnose six-gill shark	*Hexanchus griseus*	792	107	0.68	0.71
Sand tiger shark	*Carcharias taurus*	2405	220	0.67	0.89
Scalloped hammerhead	*Sphyrna lewini*	274	149	0.6	0.84
other species		1086	88	0.5	0.71
Blue shark	*Prionace glauca*	990	343	0.29	0.76
total		**18,093**	**3124**	**0.7**	**0.9**

Table 4
Performance metrics of SD components to locate, identify, and classify sharks from two baited remote videos and five YouTube videos that collectively depict eight species of sharks. SL threshold = 0.99, SI threshold = 0.5.

	Sicilian Channel	Palau archipelago	YT 1	YT 2	YT 3	YT 4	YT 5
Video length (min)	35.4	17.7	49.2	10.5	4.3	14	5.6
Processing time (min)	37.1	18.2	55	13.5	6.2	17.2	8
Frames extracted	2121	1055	2951	630	255	841	333
# of shark images	812	152	855	0	120	256	82
# of non-shark images	1309	903	2096	630	135	585	251
Shark species	*I. oxyrinchus*	*G. cuvier, C. amblyrhynchos*	*C. carcharias, G. cuvier, S. mokarran*	None	*G. cirratum*	*C. taurus, S. mokarran*	*C. melanopterus*
SL Recall	0.9	0.88	0.89		0.8	0.9	0.91
SL Precision	0.91	0.84	0.9		0.93	0.86	0.87
SL Specificity	0.92	0.85	0.87	0.93	0.89	0.84	0.84
SI Specificity	0.97	0.94	0.9		0.95	0.96	0.96
SC Recall	0.62	0.79	0.69		0.78	0.82	0.73
SC Top-3 Recall	0.7	0.86	0.76		0.84	0.88	0.81

memory consumption and decrease processing time. The extracted frames were then screened with the SL. The SL threshold was increased to 0.99 to minimize the false-positive rate. To test the SL's specificity on a video that did not contain any shark subjects, we processed a YT Video that exclusively depicted typical coral reef habitat and numerous fish species, but did not contain any sharks (see YT Video 2 in Table 4). We annotated shark-located frames with the time-stamp in the video they were extracted from. This was done so we could check the video at the exact time a shark was located. We sorted shark-located frames with the SI (threshold = 0.5). Finally, the identified shark images were classified with the SC. We calculated the SC's recall for its top guess and its top three guesses.

For the Instagram pipeline, we data-mined 14 separate datasets of images, pertaining to 14 target species, to first assess the SI's performance. Media was easier to extract when we identified the hashtags delineated in Table 2 (Kim et al., 2016). These species represented morphologically diverse groups of sharks and showed the highest validation accuracies when training the SC. Therefore, this analysis would demonstrate the SI's ability to recognize general shark features as well as the SC's ability to classify species that were best trained. We collected an average of 951 images per hashtag, so the task of manually validating classified species was not time-consuming. We collected images along with their metadata from Instagram with the open-source program InstaCrawlR (Schroeder, 2018). We removed duplicate images by

comparing their pixel content and retaining only one copy (Lundrigan, 2018). To evaluate the relationship between false-positive rate and accuracy of the SI at various thresholds (ranging from 5e-6 to 0.999), we generated a receiver operating characteristic (ROC) curve (Fig. 3). We used a threshold of 0.5 to identify 3533 data-mined shark images from Instagram. After image identification, we combined all images labeled as sharks and classified them with the SC (Table 3). We analysed the relationship between recall and training dataset size of 13 genera and 14 species to identify a target training size for classes with not-yet enough information (Fig. 4). We fit a 2-parametric asymptotic model $y = a(1 - e^{-bx})$ and identified the asymptote value for each class.

Finally, we compared the performances of the iNaturalist's classifier Seek with the SD. We created a test dataset by sourcing 400 new images from Instagram. We collected Instagram images using the methods above, except we specified different hashtags: #shark and #sharkfishing to collect 13 shark species. We then removed identical images. From each hashtag we pooled a total of 200 images that represented shark subjects commonly photographed above and below water. Then we processed the dataset with both models and calculated their accuracy.

For evaluating performance of the SD packages, we calculated recall, precision, specificity, and F_1 score as follows:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (1)$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (2)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \quad (3)$$

$$\text{F}_\beta = \frac{1 + \beta^2}{\beta^2 (recall^{-1} + precision^{-1})} \quad (4)$$

TP = true positive, FP = false positive, TN = true negative, FN = false negative.

Recall, also known as sensitivity, measures the proportion of positives that are correctly recognized. Precision measures the proportion of retrieved individuals that are relevant. Specificity measures the proportion of correctly identified negatives. F_1 score measures the tradeoff between recall and precision. We provide these models and their training data in an instructional GitHub repository (https://github.com/JeremyFJ/Shark-Detector).

3. Results

3.1. Boosting training data

The SD components performed well individually and as a stepwise process. Overall, the SL performed at 89% accuracy, the SI at 91% accuracy, and the SC at 69% accuracy. The SL located 90% of shark images from the sharkPulse data archive (n = 24,546 shark images) and generated novel training data by extracting only shark features. By locating one or multiple subjects in shark images (Fig. 2), the SL cropped

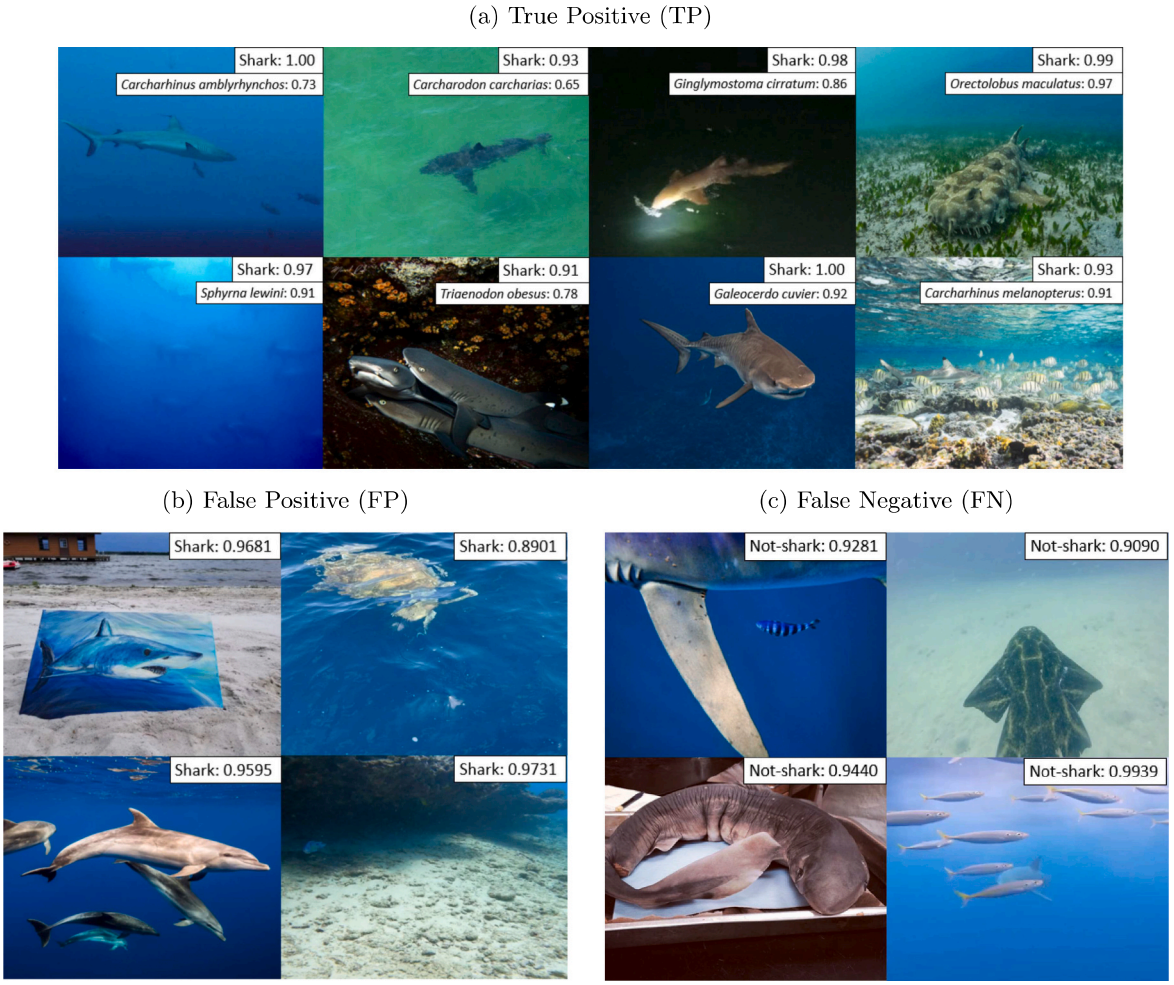


Fig. 5. Images identified by the SI and subsequent classification by the SC. (a) The SI and SC correctly identify a diverse collection of shark images by classifying underwater photographs, images with foreground and background noise, images with hardly discernible shark features, and eight different species. (b) Common subjects that were misclassified by the SI such as cetaceans (and other marine and terrestrial animals), empty foregrounds, inscrutable objects, and *fake* sharks. (c) The SI misses shark presence due to partially concealed features.

28,799 additional images to inflate the original training dataset. This approach appended 14,888 cropped images to GSC's training dataset and 9979 images to the SSCg's training datasets. All subsequent SD models dramatically increased their accuracy as a result of ingesting this training data. We observed an average 3.5% increase in test accuracy of all models that used training datasets inflated by the SL. In Table S1, we tabulate all genus and species classes that are available in sharkPulse, the amount of training data, and their observed validation accuracy (if the training data threshold was met).

3.2. SC training and overall performance

By examining the training data threshold distribution of GSC and SSCg classes, it was revealed that 433 ± 47 images are needed to achieve >50% recall among genus classes and 161 ± 41 images are needed to achieve the same recall among SSCg classes (see Fig. 4c and d, and Table S1). However variability across genera and species is high in Fig. 4a and b, and the relationship between the two variables depends on morphological distinctiveness as well as the level of training data balance. Furthermore, to obtain the same recall, we observed highly distinct species such as *Rhincodon typus* (morphological Euclidean distance to centroid = 19.0) and *Orectolobus* spp. (distance = 25.6) required significantly less training images than species with common physical attributes such as *Carcharhinus* spp. (distance = 8.3) and *Prionace glauca* (distance = 9.0). We calculated the Pielou diversity index of the GSC to be 0.94 (scale 0–1), meaning genus training datasets were overall well-balanced. The average Pielou diversity index of SSCg models was 0.77. Lastly, we compared the classification accuracy of the iNaturalist model Seek and the SD on 400 random shark images sourced from Instagram. There were 13 species to classify. Seek performed at 62% top classification recall while the SD performed at 73% top recall.

3.3. Instagram

By data-mining images from Instagram, we created 14 datasets. The SI removed non-shark images and retained shark images with 91% overall accuracy (Table 2). About 5% of actual shark images were not related to the hashtag they were scraped from (i.e., they were other species). The area under the correlation between recall and false-positive rate of the SI represented a successful classification probability of 0.885 (Fig. 3). We noticed the SI displayed lower recall when sorting video frames that had not yet been cropped for shark subjects, but performed well when sorting heterogeneous data-mined images. The SI displayed a very low false-positive rate and false-negative rate. However, we noticed images like those in Fig. 5b and c represented commonly misclassified images. These misclassifications occurred 9% of

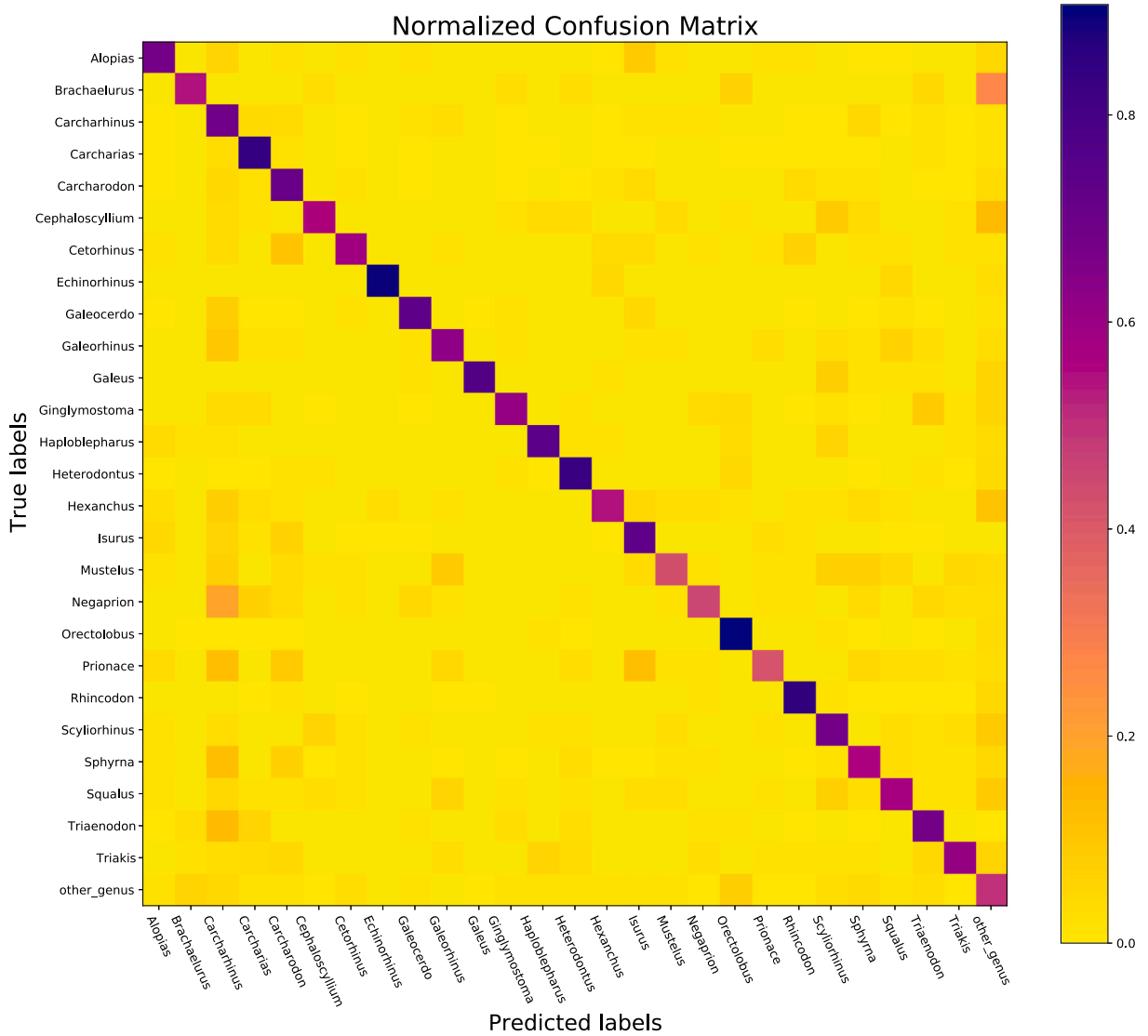


Fig. 6. GSC normalized confusion matrix of 26 shark genera classes. A 27*th* class *other genus* represents 48 data-deficient genera.

Table 5
List of species, and number of training images, for which we could infer a taxonomic identification at the genus level (with the GSC model) and species level (with the SSCg models). Acc = classification accuracy. Genera are included if the training data >433 images and species were included if the training data >161 images (please see results for these thresholds). Species with fewer images than the threshold were grouped in the *Genus* spp. classes. If a genus contained exactly two species, both species were trained regardless of training data quantity. A complete list of species and their available training images are given in Table S1.

Species	Images(Acc)	Species	Images(Acc)	Species	Images(Acc)
Alopias	1185(0.65)	**Galeorhinus**	791(0.53)	**Orectolobus**	2021(0.92)
A. vulpinus	353(0.81)	*G. galeus*	791(1)	*O. maculatus*	1019(0.82)
Alopias spp.	174(0.72)	**Galeus**	575(0.72)	*O. halei*	542(0.62)
Brachaelurus	479(0.65)	*G. melastomus*	376(1)	*O. ornatus*	281(0.34)
B. waddi	299(0.96)	**Ginglymostoma**	945(0.61)	*Orectolobus* spp.	97(0.6)
B. colcloughi	162(1)	*G. cirratum*	821(1)	**Prionace**	990(0.42)
Carcharhinus	4963(0.71)	*G. unami*	124(0.92)	*P. glauca*	990(1)
C. melanopterus	829(0.7)	**Haploblepharus**	680(0.61)	**Rhincodon**	1602(0.88)
C. amblyrhynchos	550(0.67)	*H. fuscus*	271(0.96)	*R. typus*	1602(1)
C. limbatus	488(0.41)	*H. edwardsii*	215(0.23)	**Scyliorhinus**	964(0.63)
C. leucas	402(0.66)	*Haploblepharus* spp.	194(0.75)	*S. canicula*	378(0.97)
C. obscurus	259(0.85)	**Heterodontus**	2180(0.82)	*Scyliorhinus* spp.	94(0.5)
C. perezi	245(0.24)	*H. portusjacksoni*	1172(0.94)	**Sphyrna**	1591(0.54)
C. plumbeus	212(0.45)	*H. galeatus*	343(0.89)	*S. tiburo*	377(0.79)
Carcharhinus spp.	815(0.74)	*H. francisci*	337(0.74)	*S. lewini*	274(0.82)
Carcharias	2405(0.84)	*H. japonicus*	306(1)	*S. mokarran*	165(0.47)
C. taurus	2405(1)	*Heterodontus* spp.	22(0.84)	*Sphyrna* spp.	140(0.79)
Carcharodon	2290(0.72)	**Hexanchus**	971(0.67)	**Squalus**	1044(0.52)
C. carcharias	2290(1)	*H. griseus*	792(1)	*S. acanthias*	182(1)
Cephaloscyllium	663(0.56)	*Hexanchus* spp.	8(0)	*Squalus* spp.	130(0.77)
C. isabellum	323(1)	**Isurus**	1636(0.72)	**Triaenodon**	1786(0.69)
C. laticeps	264(1)	*I. oxyrinchus*	1360(0.99)	*T. obesus*	1786(1)
Cephaloscyllium spp.	76(0.8)	*I. paucus*	62(0.25)	**Triakis**	1060(0.59)
Cetorhinus	642(0.57)	**Mustelus**	677(0.43)	*T. semifasciata*	673(1)
C. maximus	642(1)	*M. canis*	187(0.68)	*T. megalopterus*	213(0.86)
Echinorhinus	516(0.87)	*Mustelus* spp.	335(0.88)	*T. scyllium*	161(1)
E. cookei	452(1)	**Negaprion**	910(0.38)	*Triakis* spp.	12(0)
E. brucus	6(0)	*N. brevirostris*	171(1)		
Galeocerdo	1117(0.72)	*N. acutidens*	69(0)		
G. cuvier	1117(1)				

the time.

The SC performed at 70% accuracy when classifying Instagram images of 17 species (see Table 3). The SC classified 3124 shark images to 14 species and, in addition, three species to the *other species* label (two *Carcharhinus* spp. and one *Sphyrna* sp.). When testing recall, we noticed that, as the test class was trained with increasingly less images, the model would proportionally mistake that class for physically similar genera. We examined the recall asymptotes for all classified genera as well (Fig. 4a). Many classes never met their asymptotic values, meaning that inflating their training datasets would improve recall. The classes *Orectolobus*, *Rhincodon*, and *Heterodontus* reached their asymptotic values with <1000 training images, and at relatively high recall (92%, 90%, and 83% respectively). Some classes exhibited low peak recall, such as blue sharks (*Prionace glauca*). From the confusion matrix (Fig. 6), blue sharks were commonly mistaken for *Carcharhinus* species. The low asymptotic curve indicates that, with the same GSC structure, class recall for blue sharks cannot significantly improve unless training datasets are balanced and boosted.

Similarly, we assessed the recall of species classes within SSCg models (see Fig. 4b). Both the *Hexanchus* and *Isurus* models contained two unbalanced classes (see Table 5 for training datasets). So, we anticipated that the recall of these models' dominant classes would peak even if they were trained with fewer images. The models *Carcharhinus*, *Heterodontus*, *Orectolobus*, and *Sphyrna* contained mostly balanced training datasets with four or more classes. Interestingly, port jackson shark (*Heterodontus portusjacksoni*), spotted wobbegong (*Orectolobus maculatus*), and blacktip reef shark (*Carcharhinus melanopterus*) classes reached their maximum recall while being trained with <200 training images. Grey reef shark (*Carcharhinus amblyrhynchos*), other *Carcharhinus* sp., scalloped hammerhead (*Sphyrna lewini*), and other *Sphyrna* sp. classes did not meet their maximum recall with <500 training images.

3.4. BRUV surveys and online videos

We allowed the boosted SD to locate and classify eight shark species from seven videos without manual interaction. We processed two BRUV recordings and five YT videos that made up 136 min of total video footage, which contained eight species of sharks. We spent 6.2 h manually validating all of the extracted frames (n = 8185 frames). It took the SD 2.6 h to process all videos in succession. The SL located 89% of available shark frames (n = 2277 frames) and the SI filtered out false-positive images with 94% specificity (Table 4). The SC classified all species with an average top recall of 69% and top-3 recall of 76%. The SL showed 93% specificity and a false-positive rate of 7% when processing YT Video 2, which did not contain sharks.

4. Discussion

Historically hampered by problems of data paucity, shark research is transitioning toward a time with ubiquitous big data. Embracing this movement requires being able to capture and structure the increasing amount of information that is available online and generated by modern scientific monitoring. On this line, we developed a modular software package targeted at identifying and classifying shark images from unstructured and unlabeled media. In this package, location and identification models were able to detect sharks with 90% and 91% recall, respectively. Further, a pseudo-hierarchical classification structure classified 26 genera and 47 shark species, at 69% and an average of 85% recall, respectively. Trained on the largest and most diverse shark image dataset compiled so far, this software facilitates rapid data collection on sharks and generation of biologically relevant data, including boosting information for data-poor species.

Expressing the full potential of this approach would be supporting a completely automated data analysis pipeline. We have shown that surveys and online archives can be automatically processed for shark

classification, although with human review needed. While the SD is going toward complete automation, there is still room for improvement. Our shark detector is currently the most efficient software for locating, identifying and classifying sharks from unlabeled media. The SD top species predictions were 11% more accurate than iNaturalist's Seek (currently the best general-purpose biodiversity classifier available) (Horn et al., 2017). Yet model accuracy can still be improved, especially for the GSC model.

The GSC acts as the parent node for multiclass classification among the SD components and is, therefore, most challenged in the pipeline (see Fig. 4e). The GSC typically displayed lower classification error with more training samples. However, misclassification also depended on physical distinctiveness and training data balance and content. For example, carpet sharks (*Orectolobus* spp.) are easily identified because they are physically unique. When comparing morphological Euclidean distances with all species represented, *Orectolobus* species exhibited one of the highest distances (25.6) meaning they are among the most physically dissimilar taxa. We also noticed the content of *Orectolobus* species' image and video archives were homogenous because they are strictly bottom-dwelling sharks and are almost exclusively observed in benthic habitats. Therefore, the class achieved a high recall (92%) with <1000 training images (see Fig. 4a). Conversely, blue sharks (*Prionace glauca*) exhibit similar morphometric measurements to the centroid of the training dataset (Euclidean distance = 9.0). They are frequently observed in various marine habitats by photographers, divers, and recreational and commercial fishermen (Campana et al., 2009), resulting in heterogeneous image and video archives. As we continue to capture this heterogeneity and physical distinctiveness by gathering more images, we expect classification accuracy to increase. But currently, *P. glauca* experiences low recall (29%) with <1000 training images.

During training of the GSC, 15% of test images ($n = 539$ images) were mistaken for the *Carcharhinus* genus (trained with 4963 images and represents 24 species) and the *other genus* class (trained with 2593 images and represents 48 genera and 172 species). Because the *Carcharhinus* and *other genus* classes describe 196 species, their training datasets are heterogeneous and variable in morphology, imbalanced relative to other smaller genera, and represent 21% of the entire training dataset. While the GSC training dataset was shown to be well-balanced with Pielou's diversity index of 0.94, we can minimize confusion and improve overall classification accuracy by continuing to balance data-poor genera. Furthermore, boosting genera that do not reach the training threshold would remove them from the *other genus* label, reduce confusion with the label, and increase the SC's taxonomic range. The SC will gain a new classifiable genus capable of achieving >50% recall. However, morphological diversity will still affect the GSC's overall training accuracy.

SSCg models are composed of child nodes that utilize previous taxonomic information from the GSC. As expected, average SSCg classification accuracy (85% with 3.5% standard error) was higher than GSC accuracy (Fig. 4e and f). Nonetheless, even SSCg models were challenged by imbalanced datasets and class similarity. The *Hexanchus* and *Isurus* models each contain two classes, where the dominant class was trained with an average of 50 times more images than the non-dominant class. Recall was perfect for *Hexanchus griseus* and *Isurus oxyrinchus* (Fig. 4b) because the model did not learn the misrepresented class. This affected our SSCg threshold distribution (see Fig. 4d) by indicating less training images were needed to reach >50% recall, without taking into account that the classes are imbalanced. Further, fitting asymptotic recall functions of different SSCg model classes was useful for gauging future data boosting efforts. For instance, we noticed a pattern where dominant classes attained their maximum recall (<500 training images) while non-dominant classes did not. This would suggest that maximum recall values are useful as benchmarks, but will change as species are boosted and classes are increasingly represented. And so, to best increase overall SC top recall and species coverage (Table 3), we must grow the number of taxonomically labeled images while prioritizing data-poor species, balancing training datasets, and increasing image diversity.

SI/SL misclassifications are only $<10\%$ frequent. The Faster-RCNN model allowed the SL to achieve high recall (89%), precision (88%), and specificity (93%) (Ren et al., 2016). Varying habitat types and methods of videotaping, and presence of non-shark fauna did not significantly impact these metrics. VGG16 allowed the SI to achieve a F_1 score of 91% (Simonyan and Zisserman, 2015). Performance of the SI and SL, like the SC, can be boosted by inflating training datasets.

The training and validation datasets are substantial considering the scarcity of visual information repositories for most shark species. SharkPulse contains the largest repository of shark images and provides a consistent influx of shark-specific media by combining several data collection approaches: data scraping from online archives, user submissions, and synthetic image generation techniques. Our training data is high quality due to crowdsourcing validation of taxonomic and spatiotemporal information. This facilitates continuous data collection and classification accuracy and is slowly being adopted for conservation (Ferretti et al., Unpublished data; Horn et al., 2017; Mart'ın et al., 2021). For example, iNaturalist's Seek was trained on a massive database of crowdsourced images that were validated by the application's users. Effectively, iNaturalist and iSharkFin grow with user submissions which can improve the models' classification accuracy (Barone et al., 2022; Horn et al., 2017). We adopted this approach and combined it with automated data scraping and synthetic image generation techniques, making the Shark Detector a novel instrument for collecting visual media of sharks. While the Shark Detector excels at classification accuracy and taxonomic range compared to other methods, there are still objectives to strive for. Seek is available on smartphones and as a result, they can equip everyone with intelligent monitoring capabilities. Increasing citizen science interactions and validation effort with mobile applications would continue to improve the quality of data that is sourced from sharkPulse and the Shark Detector.

The largest limitation of the Shark Detector is classification accuracy for data-poor species. Boosting natural and synthetic image generation techniques will inflate the training datasets of these species considerably, and subsequently increase classification accuracy. We showed how data-mining (Table 2), object-detected cropping (Fig. 2), and image augmentations are effective data generation approaches. Social networks like Instagram offer an inexhaustible source of shark and non-shark images (Jenrette et al., Unpublished data). Furthermore, synthetic image generation can be significantly improved. We can extract cropped images of fish and paste them onto randomly selected backgrounds while incorporating transformations. This approach will effectively generate thousands of new images from a handful of genuine images (Allken et al., 2018). As new shark images are ingested and validated, the Shark Detector will immediately use them, automatically funneling those images into the appropriate training datasets. The SD will be a rapidly evolving AI, automatically collecting and generating new shark images, training models, and growing smarter with each step.

Utilizing unsupervised models for shark detection and species identification has multiple applications, including processing online videos, survey footage, and big data (Siddiqui et al., 2018). This allows us to expand possibilities for filling information gaps in shark populations, even beyond traditional fisheries monitoring techniques. Instagram is a massive data cloud that offers tremendous opportunities for generating biologically relevant data. However, it contains a daunting amount of irrelevant content that would be unrealistically filtered with manual validation (Migliaccio et al., 2019). Plus, even targeted shark images often lack taxonomic and spatiotemporal information that need to be inferred with post processing. When 91% of noisy data is removed (which is the current SI capability), and the remaining content is taxonomically classified, validation suddenly becomes practical. Furthermore, filtering and classifying facilitate the development of geoparsing and time-stamping programs (Migliaccio et al., 2019). Preliminary investigations suggest that Instagram posts of sharks can be effectively

transformed into occurrence records with these taxonomic and spatio-temporal identifiers (Jenrette et al., Unpublished data).

We compiled object-detection and classification models into a self-standing package that utilizes transfer learning, deep learning, and multiple CNNs. To our knowledge, this is the most reliable identification software for general-purpose shark recognition trained with the largest and most diverse shark image dataset available today. This package is available to all researchers willing to use and customize these models for accommodating their own experimental avenues (https://github.com/JeremyFJ/Shark-Detector). Our aim is facilitating rapid data collection for boosting data-poor species, expanding the SD taxonomic range, and increasing its accuracy. The Shark Detector is valuable for transitioning into the big data revolution for filling information gaps while supplementing traditional fisheries monitoring techniques. Identification without time-consuming validation can fundamentally change the design and quality of studies focusing on shark ecology, biology, and conservation.

Authors' contribution

FF and JJ conceived the idea and, along with EF and TH, contributed to designing the methodology. ZL developed the object-detection software and, along with PC and JJ, developed the training architecture for identification and classification software. FF and JJ collected the data. JJ and PC analysed the data and JJ led the manuscript writing. All authors contributed to drafts and gave final approval for publication submission.

Data availability

Code for building, training, testing, and customizing all models are available on GitHub at https://github.com/JeremyFJ/Shark-Detector. Model weights will be available in a Google Drive folder at https://drive.google.com/drive/folders/1KdVkSn4avPCa4iGjLp6Lf8IVSEAURQqs?usp=sharing.

Declaration of Competing Interest

There are no conflicts to report.

Acknowledgments

We thank the Global Change Center and the Bertarelli Foundation for their financial support. We thank our undergraduate student Lauren Morris for collecting and labeling training images. We thank the undergraduate students of the CS 4624 (Multimedia, Hypertext, and Information Access) and ECE 4994 (Undergraduate Research) classes – Steven Gordon, Gregory Chang, Mason Mulgrew, Hunter Debay, Aman Kothari, Ashutosh Tiwari, Feneel Patel, Guanang Su, Hang Lyu, Ray Raya, and Tirth Shroff – for researching classification model architectures and constructing a web page within sharkPulse for validating shark sightings and integrating the Shark Detector as a species suggestion assistant.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2022.101673.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Zheng, X., 2016. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. CoRR, abs/1603.04467. https://doi.org/10.48550/arXiv.1603.04467.

Allken, V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiout, T., Malde, K., 2018. Fish species identification using a convolutional neural network trained on synthetic data. ICES J. Mar. Sci. 76 (1), 342–349. https://doi.org/10.1093/icesjms/fsy147.

Alvarez Ellacur'ıa, A., Palmer, M., Catal'an, I.A., Lisani, J.-L., 2019. Image-based, unsupervised estimation of fish size from commercial landings using deep learning. ICES J. Mar. Sci. 77 (4), 1330–1339. https://doi.org/10.1093/icesjms/fsz216.

Barone, M., Mollen, F.H., Giles, J.L., Marshall, L.J., Villate-Moreno, M., Mazzoldi, C., Guisande, C., 2022. Performance of iSharkFin in the identification of wet dorsal fins from priority shark species. Eco. Inform. 68, 101514 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.

Boettiger, C., Lang, D.T., Wainwright, P.C., 2012. rfishbase: exploring, manipulating and visualizing FishBase data from R. J. Fish Biol. 81 (6), 2030–2039. https://doi.org/10.1111/j.1095-8649.2012.03464.x.

Brautaset, O., Waldeland, A.U., Johnsen, E., Malde, K., Eikvil, L., Salberg, A.-B., Handegard, N.O., 2020. Acoustic classification in multifrequency echosounder data using deep convolutional neural networks. ICES J. Mar. Sci. 77 (4), 1391–1400. https://doi.org/10.1093/icesjms/fsz235.

Bridle, J.S., 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Souli'e, F.F., H'erault, J. (Eds.), Neurocomputing. Springer, Berlin Heidelberg, pp. 227–236. https://doi.org/10.1007/978-3-642-76153-9.

Campana, S., Joyce, W., Manning, M., 2009. Bycatch and discard mortality in commercially caught blue sharks prionace glauca assessed using archival satellite pop-up tags. Marine Ecol. Progr. Ser. 387, 241–253. https://doi.org/10.3354/meps08109.

Chollet, F., 2015. Keras. April 2022. https://keras.io/.

Di Minin, E., Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. Front. Environ. Sci. 3, 63. https://doi.org/10.3389/fenvs.2015.00063.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12 (61), 2121–2159. January 2022. http://jmlr.org/papers/v12/duchi11a.html.

Dulvy, N.K., Pacoureau, N., Rigby, C.L., Pollom, R.A., Jabado, R.W., Ebert, D.A., Simpfendorfer, C. A., 2021. Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. Curr. Biol. 31 (21), 4773–4787. https://doi.org/10.1016/j.cub.2021.08.062.

Fallon, N., Fielding, S., Fernandes, P., 2016. Classification of Southern Ocean krill and icefish echoes using random forests. ICES J. Mar. Sci. 73 (8), 1998–2008. https://doi.org/10.1093/icesjms/fsw057.

French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N., Needle, C., 2019. Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. ICES J. Mar. Sci. 77 (4), 1340–1353. https://doi.org/10.1093/icesjms/fsz149.

Garcia, R., Prados, R., Quintana, J., Tempelaar, A., Gracias, N., Rosen, S., Løvall, K., 2019. Automatic segmentation of fish using deep learning with application to fish size measurement. ICES J. Mar. Sci. 77 (4), 1354–1366. https://doi.org/10.1093/icesjms/fsz186.

Goetze, J.S., Bond, T., McLean, D.L., Saunders, B.J., Langlois, T.J., Lindfield, S., 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.

Horn, G.V., Aodha, O.M., Song, Y., Shepard, A., Adam, H., Perona, P., Belongie, S.J., 2017. The iNaturalist challenge 2017 dataset. CoRR, abs/1707.06642. https://doi.org/10.48550/arXiv.1707.06642.

Huang, G., Liu, Z., Weinberger, K.Q., 2016. Densely connected convolutional networks. CoRR, abs/1608.06993. https://doi.org/10.48550/arXiv.1608.06993.

Jorgensen, J., Micheli, F., White, T.D., Houtan, K.S.V., Alfaro-Shigueto, J., Andrzejaczek, S., Ferretti, F., 2022. Emergent research and priorities for elasmobranch conservation. Endanger. Species Res. https://doi.org/10.3354/esr01169.

Kadar, J.P., Ladds, M.A., Day, J., Lyall, B., Brown, C., 2020. Assessment of machine learning models to identify port jackson shark behaviours using tri-axial accelerometers. Sensors 20 (24). https://doi.org/10.3390/s20247096.

Kim, Y., Huang, J., Emery, S., 2016. Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. J. Med. Internet Res. 18 (2), e41. Feb 26. https://doi.org/10.2196/jmir.4738.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv. https://doi.org/10.48550/arXiv.1412.6980.

LeCun, Y., Bengio, Y., 1998. Convolutional Networks for Images, Speech, and Time Series. In the Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge, MA, USA, pp. 255–258. May 2022. https://dl.acm.org/doi/10.5555/303568.303704.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. May 2022. http://www.nature.com/articles/nature14539.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., 2014. Microsoft CoCo: common objects in context. European conference on computer vision. https://doi.org/10.48550/arXiv.1405.0312.

Liu, Z.Y.C., Chamberlin, A.J., Shome, P., Jones, I.J., Riveau, G., Ndione, R.A., De Leo, G. A., 2019. Identification of snails and parasites of medical importance via convolutional neural network: an application for human schistosomiasis. bioRxiv. https://doi.org/10.1101/713727.

Lundrigan, P., 2018. Duplicate image finder. GitHub. May 2021. https://github.com/philipbl/duplicate-images.

Malde, K., Handegard, N.O., Eikvil, L., Salberg, A.-B., 2019. Machine intelligence and the data-driven future of marine science. ICES J. Mar. Sci. 77 (4), 1274–1285. https://doi.org/10.1093/icesjms/fsz057.

Mart'ın, B., Gonz'alez-Arias, J., Vicente-V'ırseda, J., 2021. Machine learning as a successful approach for predicting complex spatio–temporal patterns in animal species abundance. Mach. Learn. 44, 289–301. https://doi.org/10.32800/abc.2021.44.0289.

Migliaccio, F., Carri'on, D., Ferrario, F., 2019. Semantic validation of social media geographic information: A case study on instagram data for expo Milano 2015. ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. 4213, 1321–1326. https://doi.org/10.5194/isprs-archives-XLII-2-W13-1321-2019.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. Omnipress, pp. 807–814. April 2022. https://dl.acm.org/doi/10.5555/3104322.3104425.

Narayan, S., 1997. The generalized sigmoid activation function: competitive supervised learning. Inf. Sci. 99 (1), 69–82. https://doi.org/10.1016/S0020-0255(96)00200-9.

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. Proc. Natl. Acad. Sci. 115 (25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115.

Pielou, E., 1966. The measurement of diversity in different types of biological collections. J. Theor. Biol. 13, 131–144. https://doi.org/10.1016/0022-5193(66)90013-0.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. CoRR. https://doi.org/10.48550/arXiv.1506.01497.

Schroeder, J., 2018. Crawl public instagram data using R scripts without API access token. Github. February 2021. https://github.com/JonasSchroeder/InstaCrawlR.

Serena, F., Abella, A.J., Bargnesi, F., Barone, M., Colloca, F., Ferretti, F., Moro, S., 2020. Species diversity, taxonomy and distribution of chondrichthyes in the Mediterranean and Black sea. Eur. Zool. J. 87 (1), 497–536. https://doi.org/10.1080/24750263.2020.1805518.

Siddiqui, S.A., Salman, A., Malik, M.I., Shafait, F., Mian, A., Shortis, M.R., Harvey, E.S., 2018. Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES J. Mar. Sci. 75 (1), 374–389. https://doi.org/10.1093/icesjms/fsx109.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. arXiv. https://doi.org/10.48550/arXiv.1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (56), 1929–1958. February 2022. http://jmlr.org/papers/v15/srivastava14a.html.

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Kelling, S., 2014. The eBird enterprise: an integrated approach to development and application of citizen science. Biol. Conserv. 169, 31–40. https://doi.org/10.1016/j.biocon.2013.11.003.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific Data 2 (1), 150026. Jun 09. https://doi.org/10.1038/sdata.2015.26.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Miller, R. S., 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods Ecol. Evol. 10 (4), 585–590. https://doi.org/10.1111/2041-210X.13120.

Taklis, C., Giovos, I., Karamanlidis, A., 2020. Social media: a valuable tool to inform shark conservation in Greece. Mediterranean Marine. Science. https://doi.org/10.12681/mms.22165.

Taylor, L., Nitschke, G., 2017. Improving deep learning using generic data augmentation. CoRR. abs/1708.06020. https://doi.org/10.48550/arXiv.1708.06020.

Weinstein, B.G., 2018. Scene-specific convolutional neural networks for video-based biodiversity detection. Methods Ecol. Evol. 9 (6), 1435–1441. https://doi.org/10.1111/2041-210X.13011.

Whytock, R.C., Swiezewski, J., Zwerts, J.A., Bara-Slupski, T., Koumba Pambo, A.F., Rogala, M., 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.

Yu, H., Chen, C., Du, X., Li, Y., Rashwan, A., Hou, L., Li, J., 2020. TensorFlow Model Garden. December 2021. https://github.com/tensorflow/models.