Ecological Informatics 75 (2023) 102036

Contents lists available at ScienceDirect

Ecological Informatics

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

Re-identification of fish individuals of undulate skate via deep learning within a few-shot context

Nuria Gómez-Vargas [a,*], Alexandre Alonso-Fernández [b], Rafael Blanquero [a], Luis T. Antelo [b]

[a] University of Seville, Department of Statistics and Operations Research, Avenida Reina Mercedes s/n, 41012 Seville, Spain
[b] Instituto de Investigacións Mariñas (IIM-CSIC), Spain

ARTICLE INFO

Keywords:
Deep learning
Few-shot learning
Photo-identification
Siamese networks

ABSTRACT

Individual re-identification is critical to track population changes in order to assess status, being particularly relevant in species with conservation concerns and difficult access like marine organisms. For this, we propose photo-identification via deep learning as a non-invasive technique to discriminate between individuals of the undulate skate (*Raja undulata*). Nevertheless, accruing enough training samples might be difficult to achieve in the case of underwater fish images. We develop a novel methodology based on a siamese neural network that incorporates statistical fundamentals as motivation to overcome the few-shot context. Our work provides a hands-on experience and highlights on pitfalls when trying to apply photo-identification in a limited scenario, concerning both data quantity and quality, yet providing remarkable results over the test set including recaptures, where the model is capable of correctly identifying the 70% of the individuals. The findings of this study can be of strong impact for the research teams becoming familiar with deep learning approaches, as it can be easily extended to re-identify individuals of other marine species of interest from a conservation or exploitation point of view.

1. Introduction

In general, poorly understood exploited marine populations seem to be in considerably worse condition than the relatively well-studied ones (Costello et al., 2012). There is a huge potential benefit regarding conservation and sustainable concerns in improving assessment and management of understudied marine species (Hilborn et al., 2020). Therefore, sustainable exploitation and conservation of marine species require an appropriate estimation of population abundances and monitoring trends over time. A relevant example of a species under this situation is the undulate skate, *Raja undulata* Lacepède, 1802. During the last years, the populations of the undulate skate show a negative trend, and the species is globally classified as "endangered" by the International Union for Conservation of Nature (IUCN) Red List (Coelho et al., 2009) and as "near threatened" in Europe (McCully et al., 2015). Besides, this coastal elasmobranch is of high commercial interest for the Galician (NW Spain) artisanal fleet (Alonso-Fernández et al., 2019, 2021). Despite its high vulnerability, knowledge about its biology and ecology remains deficient and, so, it is considered as data-limited population in terms of assessment and management (Alonso-Fernández et al., 2019, 2021).

In such context, several approaches like capture-mark-recapture (CMR) techniques have been previously used in data-limited scenarios to assess and monitor stock status (Jessop, 2000) to properly define management and conservation actions. The development of photo-identification tools that discriminate between individuals by their external characters represents a significant alternative to traditional tagging methods (e.g., subcutaneous chemical markings, transmitters or external colorants), minimising the need for invasive techniques (Marshall and Pierce, 2012; Schneider et al., 2019). The task of photo-identification of fish individuals has been addressed by human observers through pair comparisons (Hirsch and Eckmann, 2015). However, even among experienced researchers, there remains an opportunity for human error and bias (Meek et al., 2013). Moreover, it is a highly time-consuming approach and could be expedited by using computer vision systems, which would also reduce the human biases inherent to the task of re-identifying animals (Schneider et al., 2020). Hence, the development of alternative re-identification processes and putting efforts into enhancing automated photo-identification based on each individual's distinctive natural and time-stable marks (such as skin colour patterns, scars, blotches, etc.) is of remarkably importance.

There has been a breakthrough in computer vision problems due to

* Corresponding author.
E-mail address: ngvargas@us.es (N. Gómez-Vargas).

https://doi.org/10.1016/j.ecoinf.2023.102036
Received 6 September 2022; Received in revised form 20 February 2023; Accepted 20 February 2023
Available online 23 February 2023

the growing computing capability of machines and availability of big data, making it possible to extract high levels of representation of image content (Hassaballah and Hosny, 2019). This field falls into artificial intelligence and, in particular, deep learning, which has recently received large attention from ecologists (Christin et al., 2019). Deep learning is a sub-field of artificial intelligence which focuses on a learning method based on logical structures that closely resemble the architectural characteristics of the brain (the so-called deep artificial neural networks). It consists of processing units within the global system that specialize in detecting certain hidden characteristics in the data. The first record of the application of computer vision in the field of fisheries dates from 1980s, which consisted in a method for sorting species based on shape descriptors derived from binary silhouettes (Tamaya et al., 1982). Later works focused on refining approaches for fish species classification using neural networks (Allken et al., 2019; Antelo et al., 2019; Siddiqui et al., 2018; Tseng et al., 2020).

As cited above, there exist many examples regarding classifying fish species from images, so it could be thought that the problem has been broadly addressed to classify individuals too. However, the pair matching of images to identify fish individuals is generally addressed by the naked eye (Hirsch and Eckmann, 2015) or with the aid of software tools that compare patterns (Dala-Corte et al., 2016; Kristensen et al., 2020). Recent years have witnessed the emergence of deep learning systems, which have demonstrated the accurate re-identification of humans based on image and video data with near perfect accuracy (Schneider et al., 2019). Despite this success, ecologists have yet to utilize these approaches for non-human animal and there are only a few examples in the literature (Bouma et al., 2018; Moskvyak et al., 2021; Nepovinnykh et al., 2020).

Regardless of the development of such methods, data collection is usually a limited process. In real conditions, when working with marine wild populations we have to take under consideration the logistical capacity to gather sample images (underwater environment, understaffed research teams, lack of proper protocols to obtain photographs with the required quality, etc.), which inevitably leads to a data scarcity issue. That is the case of the few-shot context, where the problem is to design a supervised learning model to identify new categories from very few tagged samples (Xian, 2020). The focus on realistic scenarios, where perfect conditions cannot be met and small samples are a standard, is what brings out the strong impact on the real application of our model in the wild. We addressed the few-shot learning problem as an *N-way K-shots* classification, where N is the number of individuals (i.e., the number of classes to be identified) and the second term refers to the small amount of K labelled images in each category. To overcome the limitation in the dataset size, we exploit existing deep learning techniques in combination with statistical fundamentals.

Our model is based on a siamese network (Bromley et al., 1993; Chicco, 2021) as artificial neural architecture, which consists of two identical branches that are feed-forward perceptrons (also known as twin networks, this is, they share weights) that create the embeddings of a pair of images joined by an energy function which establishes the similarity between the two inputs. This type of network and its triplet-loss counterpart have gained popularity in the task of animal re-identification (Moskvyak et al., 2021; Nepovinnykh et al., 2020). Concretely, (Nepovinnykh et al., 2020) also chooses a siamese neural network to match pelage patterns of the ringed seals, achieving a 74.6% accuracy after a pattern extraction process and a candidate filtering process.

A convolutional neural network (CNN) was used for the branches to deal with input images, but its training from scratch was infeasible due to our reduced dataset size. Instead, we relied on feature extraction with transfer learning techniques, whose basis is using the learned mapping of the inputs to characteristics (features) by a pre-trained model to extract new mappings from unseen data and solve another problem (Torrey and Shavlik, 2010). Also, we doubled our dataset size with data augmentation techniques (Shorten and Khoshgoftaar, 2019). As the last gear of our methodology, ensemble methods were used in this work as motivation to define the evaluation framework. The way to proceed in ensemble learning is to build a final model by combining the strengths of a collection of simpler base models (Hastie et al., 2009).

Collectedly, we created a methodology that provides robustness, in addition to that is affordable given the small dataset size. Existing deep learning frames and models, such as picking a modern siamese neural network or leveraging feature extraction from pre-trained convolutional architectures, can generate predictors with low biases. But the variability of the real scenario of animal re-identification can affect the performance of these models by increasing the probability of making a poor decision. As a technical novelty, we consider the statistical fundamentals of Machine Learning to create a methodology that encourages the prediction capability and fully exploits our reduced dataset. Moreover, we kept it straightforward so as it can be easily extended to other species use cases within a quality and quantity limited scenario. We believe that this is what brings out the relevance of our study, as it is the standard for most research teams which are becoming familiar with deep learning approaches.

2. Materials and methods

2.1. Target species

The most interesting case studies on which the proposed system could be applied are those corresponding to species with distinctive natural marks (such as skin colour patterns, scars, blotches, etc.) that remain stable through time. We chose as target species the undulate skate *Raja undulata* for both the aforementioned conservation concerns (Alonso-Fernández et al., 2019, 2021; Coelho et al., 2009; McCully et al., 2015) and its photo-identification suitability: it has a recognizable colouring and spot pattern in the back of its disc that is easy to record through photographs (Fig. 1), either in its natural environment or once captured. We focused on a local aggregation that occurs mainly during the summer months in the National Park of Illas Atlánticas de Galicia (Leeb et al., 2021).

2.2. Collecting the database

The images were collected in 2020 and 2021 during field work

Fig. 1. Undulate skate's mosaic pattern.

surveys for acoustic telemetry and traditional external tagging to study the spatial ecology of *Raja undulata*. The images were recorded in variable environments like on board (research or fishing boats) during the tagging sessions or using underwater cameras while diving. Besides, we promoted a citizen science action (see the online campaign, in Spanish, here http://www.iim.csic.es/wp-content/uploads/2021/02/REC_libro_igentac_guiaparticipacion_csic_080221-1.pdf) to increase the number of images while promoting public engagement with marine science and ecosystem conservation.

Given these different sources of images, we have worked with photographs and frames of recorded videos obtained from a variety of digital devices (compact and reflex cameras but also action video cameras). This is a positive feature of the dataset, as it shows that the model is invariant to different data collection conditions. Most of the images were captured using a reflex camera (objective 15-55 mm) and an action camera GoPro Hero 7 (extracting frames from 4 K videos).

Considering the morphology of the target species, the images were taken from a zenith perspective trying to capture the complete back of the disc of the skate (Fig. 1). We only included in the analysis those pictures with a good view of the back of the disc. Pictures from each individual were taken during the same day. However, the shots were separated in time so that we could provide ambiguity in the location and, hence, pictures of the same individual do not share shadows neither occluded body parts. This information is important to show that the model is not relying on superficial features (e.g., shadows) to identify individuals, and instead relies on their visual appearance. In Fig. 2 we included some pictures that exemplify this variability in the dataset.

2.3. Data pre-processing

The images were stored, and the sample was correctly labelled in a category per individual according to the traditional identification via T-bar anchor tags, which have a protective outer sheath around a coloured and individually numbered marking that facilitates the discrimination between known individuals. Available dataset is composed of 1138 images of individuals in unbalanced —with a minimum of 1 picture, a maximum of 31 and, on average, 7 images per individual— categories. The whole database is divided by individual identities into train, validation and test sets. When a recapture is known for its T-bar tag, images of this resighting are kept for the test set; and for once-sighted individuals, we kept in the study those that had at least $n\_eval = 3$ images for their test set so that an odd voting can be done. The rest of the images in each category are 75% – 25% split into train and validation sets, respectively. After discarding those individuals with insufficient data (n_eval images in the test set, at least two in the train set, and some in the validation set), we end up with a train set of 1213 images in 108 categories (i.e., identities). Regarding the test set, we evaluated 370 images, of which 38 belonged to the four known recaptures.

Then, in order to exclude those non-informative elements, we proceeded with a manual preprocessing of the images consisting of image cropping and removal of background using the lasso tool of Adobe Photoshop®. Despite the background removal, the photo-identification procedure regarding marine species is more complex since angling and underwater photos provide an ambiguous scenario not only in the background but also in potential existing shadows over the animal pattern together with possible occluded parts of the body. An example

Fig. 2. Variability in images regarding the environment, cameras, times of the day and shadows.

where pelvic fins and pterygopodes are not visible is shown in Supplementary Fig. S1.

2.4. Methodology

In this section we provide sufficient details so the work can be replicated for re-identification of individuals based on natural markings. The code to reproduce the analysis and guidelines are available at https://github.com/nuriagomv/Application-of-Deep-Learning-techniques-for-the-photo-identification-of-fish-individuals.

2.4.1. Data augmentation process

The most common approach to deal with a dataset of insufficient size is implementing a data augmentation process. The data augmentation techniques must be carefully selected so as not to generate images that could never actually be found. For instance, mirror flipping would be a wrong selection for studying the pattern in the skin of individuals of the undulate skate since this technique would change it completely, invalidating the generated image. Therefore, the selected data augmentation techniques were: i) affine transformations (scaling, translation, rotation, and shear); ii) pixel sum transformation; iii) hue, saturation, and contrast modifications; and iv) adding Gaussian noise to images. Fig. 3 shows an original photo and an example of the resulting output of applying each selected data augmentation technique. The next step is to define a sequence of transformations where a new augmented photo will be the result of a combination of all the aforementioned child augmenters, each of them randomly applied with a probability $p = 0.5$.

An interesting question that arises at this point is whether data augmentation should be done before or after splitting the sample into the training, validation, and test sets. It could be thought that if the data is increased before dividing the sample, there would be a high probability of passing practically the same image (just adding some noise) to the training and evaluation sets, thus falling into overfitting. However, even though the original pictures were provided with some diversity, by augmenting with such multiple combinations in the transformations before splitting, what we are doing is encouraging the variability of images of the same individual. Therefore, we are actually helping the model to learn and to fine-tune the parameters of the algorithm.

2.4.2. Network architecture

In this work, siamese neural networks were chosen among the metric learning methods because of their learning paradigm in two steps. Metric learning is an approach based directly on a metric that compares the relationship between images of the same or different classes. The first step consists in training a neural network that can distinguish between the equality/difference of classes of a pair of inputs. This so-called verification model is trained to give the probability that a particular pair belongs to the same class, by randomly sampling pairs of inputs belonging to the same and different categories. In a second stage —known as one-shot task—, we select the support set of categories in which an individual could be assigned, and the model is used to evaluate new test samples in a pairwise manner against one input per possible class, choosing the one with the highest score. Fig. 4 depicts these two steps for the case of individual identification.

For the convolutional twin networks, the choice was to import a Keras image classification model, *InceptionResNetV2* (Szegedy et al., 2017), loaded with weights pre-trained on ImageNet (Deng et al., 2010). The weights of the network have been frozen and the fully connected top layer (in charge of the classification task) was not included. Instead, the units of the last of the convolutional layers (in charge of feature extraction) are flattened into a single vector, which is the embedding of the input image. This is, we rely on the embedding that the convolutional base provides.

What is done next is to define the comparison layer, choosing the elementwise L1-distance as the energy function that joins the pair of outputs. Finally, this layer is given to a single dense output unit, obtaining a weighted distance. In this way, we define a model that will train which parts of the embeddings are more relevant for the proposed problem. The final output unit uses as activation function the sigmoid function, which gives the same class probability prediction $p = \sigma(\sum_i w_i \cdot |x_{1i} - x_{2i}|)$, where w_i are the parameters that are learned by the model during training, weighting the importance of the component-wise distance. The final architecture of the defined siamese network is depicted in Fig. 5.

2.4.3. Ensemble-inspired evaluation framework

When using ensemble learning techniques, the final decision for each

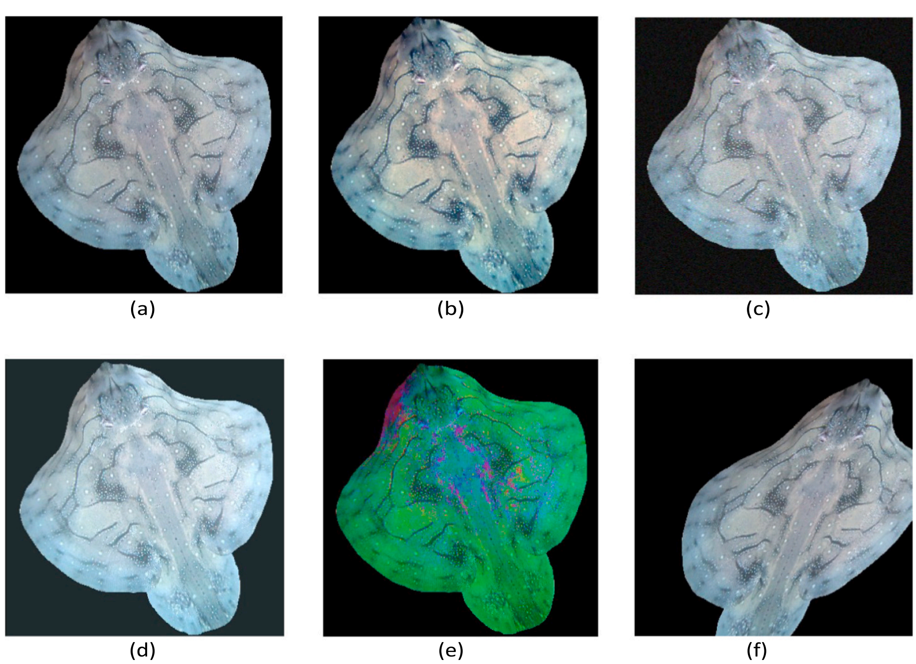

Fig. 3. Considered data augmentation techniques. (a) Original image; (b) Contrast transformation; (c) Additive Gaussian noise transformation; (d) Pixel sum transformation; (e) Hue and saturation transformation; (f) Affine transformations.

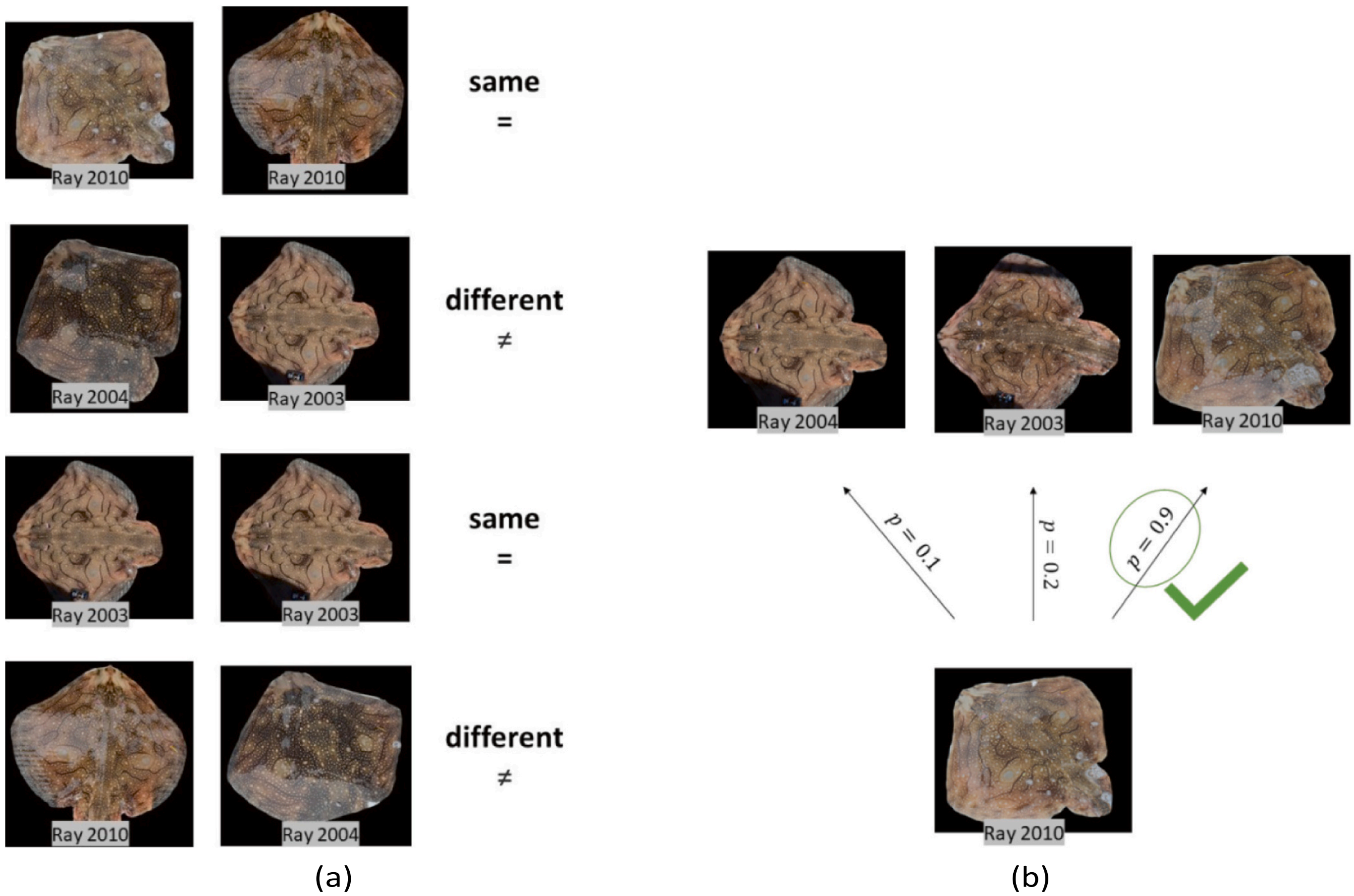

Fig. 4. Learning paradigm of siamese networks in two steps. (a) Verification model, which gives the probability that a particular pair belongs to the same class; (b) One-shot task, which evaluates a new test sample in a pairwise manner against one input per possible class, choosing the one with the highest score.

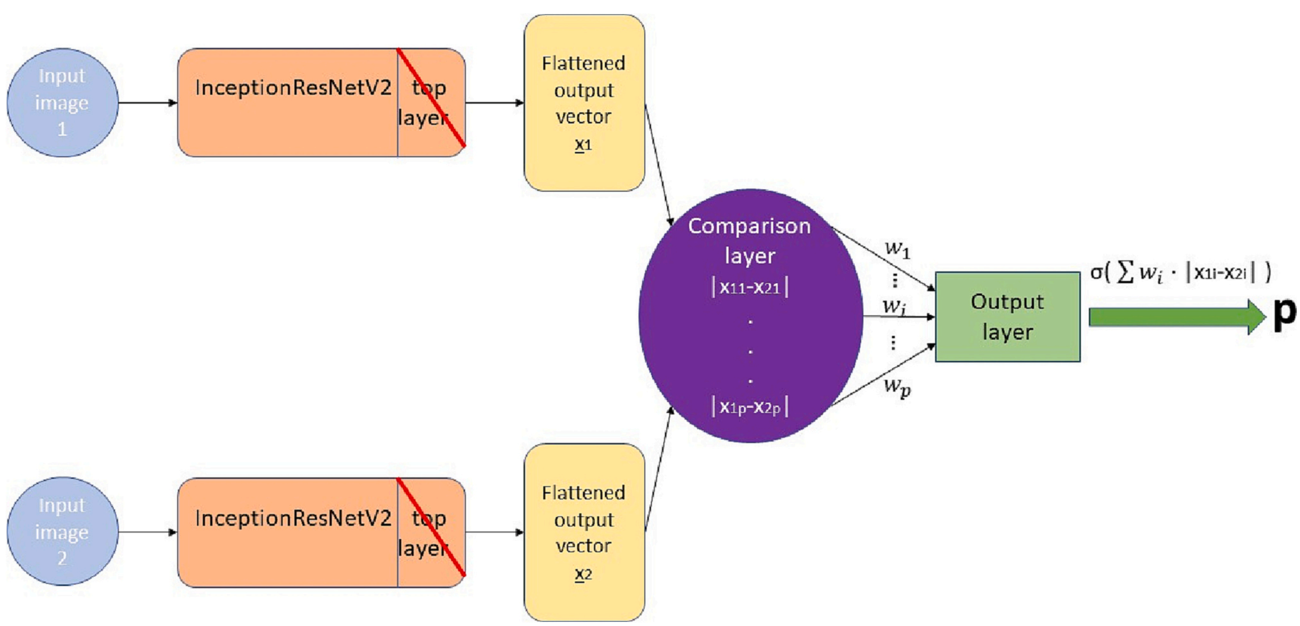

Fig. 5. Siamese network defined for our model.

new test sample instance is made based on a vote of the predictions obtained from a collection of base models by averaging them if the output is numerical or by considering majority voting if it is a classification decision. In this work, we took advantage of the statistical motivation for ensemble methods, which get to minimize the prediction error by tackling the bias-variance trade-off in the expected prediction error. The goal of ensemble systems is to create several classifiers with relatively fixed (or similar) bias and then combining their outputs, say by averaging, to reduce the variance and smooth the model (Zhang and Ma, 2012). Our problem domain also makes it suitable to average and vote: it is well-known that a certain small set of new images belongs to a unique individual, either because they were taken at one time by the same person or because the individual was previously marked.

In contrast to ensemble learning, we did not consider several base models. Instead, our single final model (hence, trained and with a fixed small bias) was used to evaluate all the images (where we know that there exists a high variability). If we consider our support set which consists of N individuals, each with K_n images, $n = 1, \ldots, N$; and a new unknown individual in the query set with samples $\{\text{image}_i\}_{i=1}^{n_{eval}}$. First, we defined the score that a single image belongs to one of the categories in the support set as the mean of the probabilities of being similar to each of the samples of that category, thus smoothing the prediction. This is,

$$p(image_i, n) = \frac{1}{K_n}\sum_{j=1}^{K_n} p\left(image_i, image_j\right)\forall n = 1, \ldots, N$$

As is natural, the final prediction will be the identity with the highest score.

$$d_i = arg \max_{n=1,\ldots,N} p(image_i, n)\forall i = 1, \ldots, n_{eval}$$

Then, we built the final decision by considering the prediction that our model gives for every image in that certain set of new samples and performed majority voting. In this way, the probability of making a poor decision is reduced and we give a more robust one. This is,

$$n^* = arg \max_{n=1,\ldots,N} \sum_{i=1}^{n_{eval}} \mathbb{1}_{(d_i=n)}$$

This proposed methodology is depicted in Supplementary Fig. S2.

2.5. Training the model

This system was built under Python 3.9.5 using TensorFlow 2.4.1 and training was performed on an Intel(R) Core(TM) i9-9900K CPU @ 3.60GHz processor and 32GB RAM memory. Learning was performed with Stochastic Gradient Descendent (SGD) method, optimizing its parameters (learning rate and momentum) and binary cross-entropy as a loss function. The algorithm was also asked to exceed a certain threshold of variability in its predictions, i.e., that the standard deviation of the predicted probabilities exceeds a certain threshold —to be tuned— so that the output of the classifier is not almost the same for all images that feed the model. Moreover, the input image size was optimized, since there is a trade-off between the information provided by large resolutions and the number of weights of the network that need to be trained.

The parameters were tuned with Bayesian optimization (Snoek et al., 2012). This selection was made because it does not require to study every possible combination of the parameters in a grid search, it works by incorporating the information that was learned in previous function evaluations to choose an optimal set of coordinates of the search space for the next evaluation. This is made by calculating the posterior predictive distribution for the function's value at each point. Table 1 gathers the search space and the selected parameters at the end of the optimization process.

Table 1
Model training optimized parameters.

Parameter	Nomenclature	Search interval	Optimized value
Learning rate	μ	$[10^{-5}, 10^{-2}]$	0.096
Momentum	β	[0,1]	0.845
SD threshold	σ	[0.05, 0.25]	0.05
Input size	s	{(75, 125, 3), (100, 150, 3), (200, 250, 3)}	(200, 250, 3)

Each optimization run consisted of 3000 train iterations with a batch size of 25 individuals to pick a pair of similar and a pair of dissimilar photos (i.e., a batch size of 100 images), and validating each 300 iterations. TensorBoard —a set of visualization tools included in the open-source library for machine learning TensorFlow— was used to study the performance of the runs. The selected parameters correspond to the run shown in Supplementary Fig. S3.

3. Results

Our approach based on a siamese artificial neural architecture obtained a best validation accuracy of 94.5% after 900 train iterations. The weights of the aforementioned best validation run were saved and made up the final model. When the model was used to finally predict the identity of the unseen individuals in the test set, the model correctly identified the 70% of the skates. As the test set included re-sightings of recaptures that happened about a year after the first shots, that loss of 20 percentage points shows the generalization when the model is deployed to real life processing.

By correctly identifying an individual, we mean that the individual is among the selection that the algorithm proposes by majority voting. For instance, for a single image, it may happen that the score of belonging to a category is the highest and the same for more than one category of the support set. Hence, an example of a difficult identification instance might be: if we have 3 pictures of an individual in the test set and the predicted identities are *{[Ray 1, Ray 3, Ray 6], [Ray 1], [Ray 2, Ray 6]}*, our model proposes as identity both *Ray 1* and *Ray 6*. We emphasize that, unlike a softmax classifier, our model did not have to forcibly choose an individual from the support set. Despite the requested variability in the predictions, the majority vote procedure may select more than one individual. To depict this behavior, Fig. 6 illustrates a heatmap of a square matrix that represents the "intensity of assignment" in the prediction for each individual. This is, $A = (a_{ij})_{i,\, j=1,\, \ldots,\, 108}$ with

$$a_{ij} = \begin{cases} \dfrac{1}{\#\{\text{identities predicted for individual } i\}} & \text{if individual } j \text{ was predicted as } i \\ 0 & \text{otherwise} \end{cases}$$

We can appreciate a clearly marked diagonal with darker colours, corresponding to that high percentage of times that the prediction is

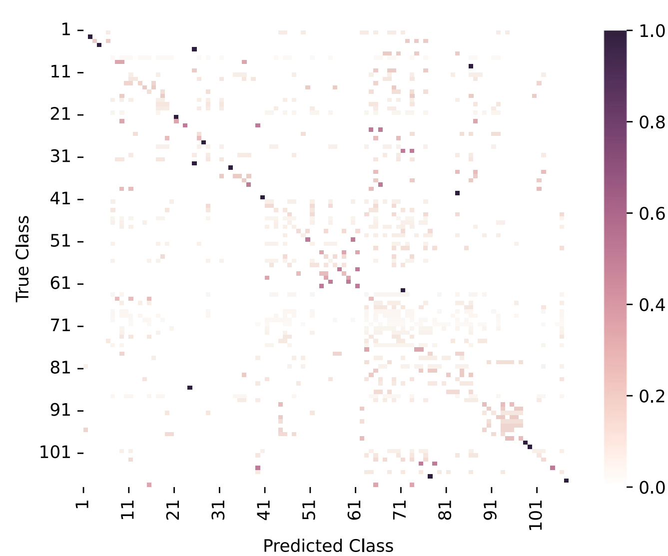

Fig. 6. Heatmap of individual assignments in the predictions. It shows coloured cells (i,j) if individual in column j was predicted as identity for individual in row i, and white otherwise. The strongest colour is for the case where only one individual was proposed as identity and becomes clearer when the number of identities proposed in the majority vote procedure is greater.

correct on the test set.

We would have a one-to-one identification if the model predicted only one individual as its output, but it is not set to forcedly choose one as in a softmax classification. Even if the model may give more than one individual as prediction, the correctness of the identification does not depend on that. Fig. 7 is a bar chart that depicts how many individuals were proposed by the majority vote prediction and whether it was a correct/wrong prediction in that case. We can visualize a chart clearly shifted to the left with a median value of 6.5. Purple coloured bars, corresponding to the correct classification predictions, also predominate when only a few identities are predicted.

Finally, the algorithm can be used to give not only the majority vote predictions but also to order a list of the top most voted identities in the prediction for each individual, understanding a "top detection" as an identity that has been voted a number of times that exceeds half of the number of photos that the individual has in the support set. We cannot speak in terms of a "top-*k* accuracy" as in (Nepovinnykh et al., 2020) because we do not follow any sort of *k*-nearest neighbours (where *k* is fixed) approach but have an ordered list of the top most voted individuals for each prediction that meet some criteria instead. In this sense, the decision is left to the discretion of the researcher in charge, considerably narrowing down the set of individuals for the pairwise comparison. With this procedure, the individuals in the test set are correctly detected the 89% of the times. In this case, the median is of 26 individuals, which narrows down the possibilities by a quarter. Supplementary Fig. S4 provides a similar bar chart as before but showing the frequency of the number of individuals proposed in the top detections.

4. Discussion

Deep learning is a rising tool for ecologists, covering a wide range of topics (Christin et al., 2019), such as the identification of animal species (Norouzzadeh et al., 2021) or the estimation of biodiversity in large data sets (James and Bradshaw, 2020). Among the current applications of deep learning techniques in ecology, identification and classification of individuals based on photographs have been proposed as a useful tool for monitoring wild populations (Goodwin et al., 2022; Schneider et al., 2019). However, marine ecosystems still represent a challenge, particularly for data collection, and the examples in marine taxa are still scarce. The studied species, *R. undulata*, is considered as data limited population in terms of assessment and management (Alonso-Fernández et al., 2019, 2021). Therefore, our case of study falls into the field of few-shot learning, which has become a very important research problem in the field of deep learning in recent years since a data-limited scenario is the standard in the deployment of real applications.

This study is one of the first attempts to apply deep learning techniques to photo-identification within a fish monitoring program. Traditional monitoring systems rely on artificial markers to differentiate each individual, which may pose a threat to animal survival or cause behavioural alteration and stress due to the combined effect of capture, handling and tagging. But also tag loss could affect performance of mark-recapture designs (Sackett and Catalano, 2017). For this reason, the application of similarity comparison networks, such as a siamese network and its triplet-loss counterpart, has recently emerged for the re-identification of marine individuals (Moskvyak et al., 2021; Nepovinnykh et al., 2020).

In the experimental part of the work, the model achieved a high accuracy for a straightforward matching process. Our accuracy of 0.7 means that the majority voting procedure proposed the correct individual the 70% of the times. A methodology consisting not in providing a one-to-one identification but to the most similar individuals is the chosen option if a conventional classification, where the number of categories is fixed, is to be avoided. (Moskvyak et al., 2021; Nepovinnykh et al., 2020) have strategies based on the *k*-nearest neighbours, and analogously in (Bouma et al., 2018) an algorithm is developed to match an individual to existing images in a catalogue of known individuals and return the top *k* identities, ranked by similarity. It must be mentioned that the achieved classification success is lower when compared with these works. However, we consider that our study is a straightforward methodology, and it is developed within a much more limited scenario, which is the standard for most research teams which are becoming familiar with deep learning approaches.

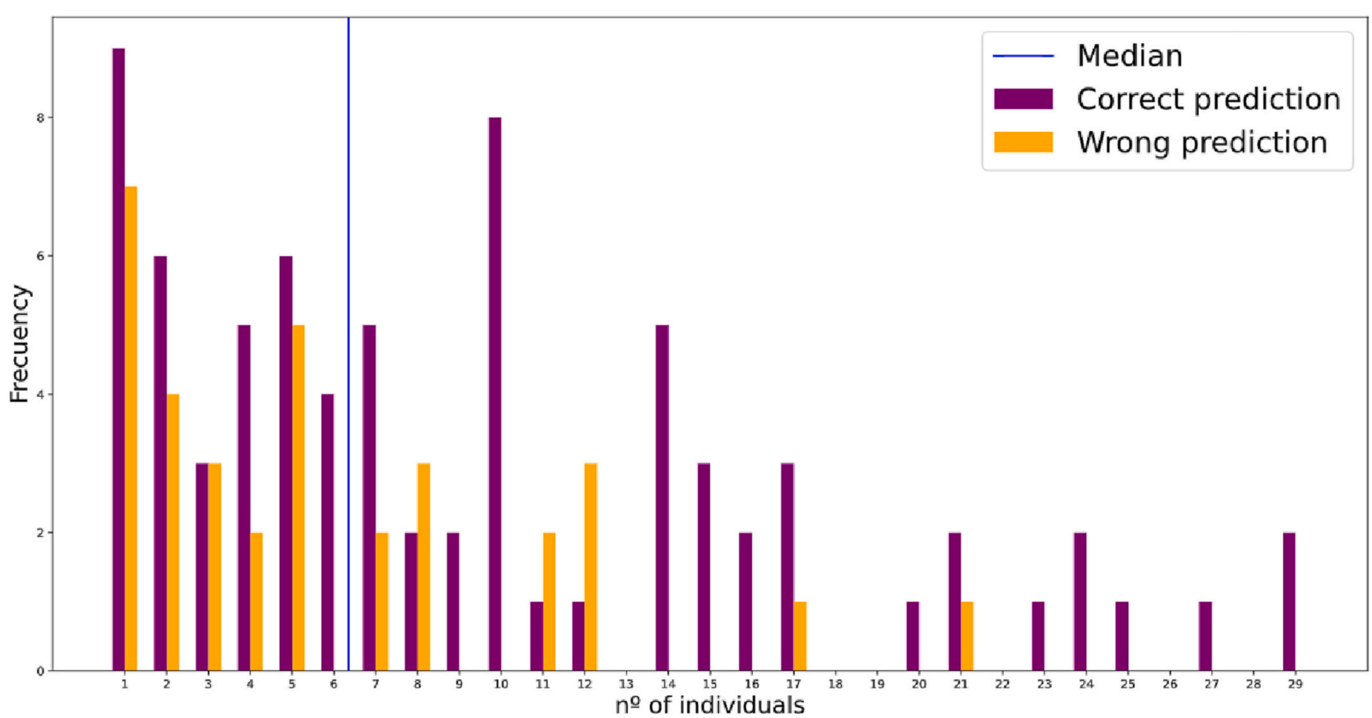

Fig. 7. Bar chart of the quantity of individuals voted in the majority vote procedure. Purple and orange bars correspond to the count of correct and wrong predictions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In (Nepovinnykh et al., 2020) we also find the choice of a siamese network. However, the first big difference that we come across is that their dataset was composed of nearly 100,000 images, compared to our few-shot learning approach. This allows them to also train the convolutional part of the network. They obtain a 74.6% accuracy after a feature engineering process that deleted all the noise —lighting conditions, shadows, etc.— from the patterns. When using the original patches, their accuracy decreases in 10 percentage points. Lastly, they develop a candidate filtering and ranking sequence based on topology-preserving projections. Whilst our work has been a successful attempt of individual recognition —the algorithm was able to correctly classify images of recaptured individuals— in a straightaway procedure: using the statistical properties of the bias-variance trade-off allows us to propose the most similar identities in a robust way without the need for filtering.

Similar works are (Bouma et al., 2018; Moskvyak et al., 2021). It is true that (Bouma et al., 2018) considers the data scarcity problem, but both have bigger datasets compared to ours. This allows them to train fine-tuned CNNs and develop a methodology based on extracting particular embeddings of the species in question, specializing in capturing their natural markings quite well. In contrast, considering a realistic limited scenario forced us to explore only feature extraction by defining an approach that specializes in how to weight the relevance in the comparison of the extracted embeddings for our problem. This supposes an advantage with respect to those previous works which cannot just use straightforward feature extraction but continue training the CNN to obtain particular embeddings for their models to correctly identify the individuals.

The field of Machine Learning is where Artificial Intelligence and Statistics converge. Our study clearly highlights the potential of applying modern siamese neural networks in combination with traditional statistical results to individual photo-identification within this framework. The application of these deep networks is a promising way to revolutionize the way in which individuals are re-identified and managed owing to its multiple advantages: it solves the problem of few-shot categories, which is the most common scenario in the marine environment; and its architecture makes the solution easily extended to other species with a unique recognizable pattern. Moreover, siamese neural networks provides a flexible framework in two steps that adapts both to the problem of identifying the current individuals in the available database as well as to address the later arrival of new specimens (Goodwin et al., 2022). The first step is very useful in our problem domain because the verification task could be trained even with pairs of images of individuals of the same species but with databases from other spatial domains. The reason is that, at this first stage, this model only aims at learning which parts of the images are discriminating since the identification categories are not defined yet. In the following step, the flexibility in choosing which individuals are in the support would allow to broaden the existing set when new individuals are detected.

As indicated previously, when a recapture was known, images of this re-sighting were kept for the test set. One of these recaptures was detected by our model with eight of ten votes with the maximum probability. *Re*-sightings occurred about a year after the first shots of these individuals, meaning that our model is able to attribute the same identity to two pictures of the same animal taken on different days. This long time period between the pictures shows the strong impact of this work, that indeed enables the re-identification of known/marked individuals.

It is important to emphasize that our database was collected from several projects with a variety of objectives different than the individual recognition, resulting in a quite varied sample due to environmental and light conditions. We believe that our lesser success in terms of accuracy is related to this variation of the database. Nevertheless, when thinking about the application of these type of models in wildlife research, one should aim to train models that are invariant to ambient conditions. This is because most of the times researchers will not have the perfect standard conditions in the wild. Therefore, putting too much focus on obtaining high accuracies and not necessarily on the real application of these models in the wild is not always right.

Collectedly, the findings of the proposed model suppose a strong impact and have several implications on the real applications of wildlife research, concerning that it means a feasible application for most of research teams: we face the problem of dealing with a small amount of photographs per individual to train the model, altogether with considering a model invariant to different data collection techniques, and a scenario with non-perfect conditions in the wild. Furthermore, our model is very low time-consuming and can even be trained on CPU since it has been oriented to train the fewest number of weights possible (always taking into account the trade-off with the information provided to the model). This implied that our architecture was already trained within about a day, allowing us to run many iterations of Bayesian optimization and hence explore multiple options for the tuning parameters in promising regions.

Moreover, our deep learning-based method can be easily extended to identify individuals from other species due to the fact that photo identification has been proven to be a potential method for individual recognition when applied to fish individuals (Benjamins et al., 2018; Marshall and Pierce, 2012). We already identified species in the study area, within the same national park, that fit with the minimum requirements to apply this photo-identification approach, like *Labrus bergylta* (Mucientes et al., 2019). For this reason, the defined procedure of photo-identification of fish individuals based on deep learning techniques and applied in a data-limited scenario, represents a meaningful contribution to this neglected topic. Its implementation will improve the current monitoring program of the target species using non-invasive tracking techniques, and so, reducing the impact on the studied population.

We came across certain limitations in this work, and further research lines are needed. Initially, we proceeded with a manual preprocessing of image (cropping and removal of background). We tested how would the algorithm perform with no preprocessing, as building an image set by manually delimiting the representative part of the pattern to study is highly time consuming —it took about a minute to delimitate each image in a dataset of size 1138—. The absence of pre-processing is a pitfall of the algorithm, as to obtain the same results we had to downsize the number of identities by more than a third. However, this limitation could be solved by developing and applying automatic object instance segmentation techniques, such as training a Mask R-CNN (He et al., 2017) to extract the bodies from the background (Álvarez-Ellacuría et al., 2020; Tseng et al., 2020). Regardless of the way to differentiate the body of the individuals, the preprocessing step is of high importance as can reduce the number of images needed and, for instance, if we have ambiguity in the seabed this can confuse the algorithm and result in a bad classification. This is necessary to show that our model is not relying on superficial features (e.g., shadows) to identify animals, and instead rely on their visual appearance, being invariant to different data collection conditions. Also, explainability analysis —such the ones in (Angelov and Soares, 2020) using prototypes or by using gradient-based methods (Selvaraju et al., 2017)— could be subject of future exploration in order to know what patterns are learned by the deep learning system.

On the other hand, we rely on the embedding that the pre-trained convolutional base provides. This is, we trust on that the CNN was exposed to a sufficiently large and varied original dataset that the learned weights for the feature extraction will be able to generalize well to our problem. This hypothesis can be assumed since ImageNet is much larger in scale and diversity and much more accurate than the current image datasets (Deng et al., 2010). InceptionResNetV2 (Szegedy et al., 2017) is a variation of the earlier *InceptionV3* model which borrows some ideas from Microsoft's ResNet papers (He et al., 2016b, 2016a). This CNN was chosen because it achieved a new state of the art in terms of accuracy on the ImageNet image classification benchmark (https://ai.googleblog.com/2016/08/improving-inception-and-image.html, access date 07/04/2022). However, a comparison of architectures for transfer learning is not the goal of this paper, but to state that it is feasible to directly use feature extraction to make up a model that only trains which parts of the extracted embeddings are more relevant for the proposed problem. We do not deny that exploiting transfer learning to extract particular embeddings, instead of just freezing the series of convolutional and pooling layers, would boost the final model and be of great improvement for our problem of individuals identification if we had a considerable number of images too. However, we could rely on a high number of photographs and our focus was on building a model that is sustainable in the few-shot context.

Regarding the multi-class classification scheme, conventional approaches —such as a softmax top layer, where the number of predicted classes must match the number of known individuals— need to be upgraded for the task of photo-identification, as the appearance of new individuals (i.e., categories of the classification model) that are not known a priori is frequent and definitely an issue that needs to be addressed. This problem fits in the novelty detection domain, which can be defined as the task of recognising that test data differ somehow from the data that are available during training (Pimentel et al., 2014). As was disclosed, siamese networks provide a flexible framework in two steps that can be used to properly address this challenging issue and make up a system that could be applied not only to the re-identification of recaptures in closed-set (pre-defined and fixed support set of individuals) problems but also to open-set identification, which considers new incoming individuals. However, there is still need of further research inquiring into the probability of similarity so as the model not always chooses an individual from the support set.

As a last point in our methodology, the adaptation of ensemble learning encourages the prediction capability and fully exploits the reduced data set. Robust methods to assist in minimising errors and bias of animal identification need to be explored (Meek et al., 2013). They have been successfully applied before for handling complex problems in marine ecosystems (Kuhnert et al., 2012). However, to our knowledge, ensemble methods have not been previously included in few-shot problems nor in re-identification methodologies, and the theoretical reduction of the prediction error for regression problems (Hastie et al., 2009) is definitely a knack that should be considered. In most cases, a person at one time can take a few photographs of a single individual by for the same cost as taking just one, but if it is not the case, we still can benefit of averaging over the probabilities of being similar to each sample in the support set, just not voting over the prediction for each new image. However, in circumstances where very similar looking individuals coexist, in addition to changing environmental scenarios, a reliance on a single image may lead to mistaken identity.

5. Conclusion

Our findings suppose an advance with respect to the state of art, where the application of similarity comparison networks such as a siamese neural networks has recently emerged for the re-identification of marine individuals. We incorporated statistical results to existing deep learning techniques and we got to develop a novel methodology that proposes the most similar identities in a robust way without the need for filtering. Moreover, we stated that it is feasible to directly use feature extraction for a model to correctly identify the individuals. Our model indeed enables the re-identification of known/marked individuals, correctly identifying the 70% of the skates in the test set which included re-sightings of recaptures that happened about a year after the first shots. Hence, we overcame the few-shot learning problem and in a much more direct procedure.

This new tool for individual recognition is a critical step forward to implement a long-term monitoring program based on low cost and non-invasive techniques for a marine species with conservation and exploitation concerns, as it is the *R. undulata*. In addition, we kept it straightforward so as it can be easily extended to other species use cases within a quality and quantity limited scenario. We believe that this is what brings out the relevance of our study, as it is the standard for most research teams which are becoming familiar with deep learning approaches. We hope that our work will motivate other research studies and serve as a base to exploit deep learning for individual photo-identification.

Authors' contributions

LTA and AAF conceived the study; NGV developed the deep learning-based approach, setting up the computing architecture; RB, LTA and AAF supervised NGV; NGV and AAF developed the protocol for field image acquisition; AAF provided the photographs; NGV, RB, AAF and LTA performed the analysis of results. All authors contributed to the writing of the manuscript, contributed critically to the drafts and gave final approval for publication.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Alexandre Alonso Fernandez reports financial support was provided by Fundación Biodiversidad, Fondo Europeo Marítimo y de Pesca.

Data availability

Code available in https://github.com/nuriagomv/Application- of-Deep-Learning-techniques-for-the-photo-identification-of-fish-individuals

Acknowledgements

This work was developed with the collaboration of Fundación Biodiversidad of the Ministry for the Ecological Transition and the Demographic Challenge of the Spanish Government, through pleamar program, co-funded by the EMFF. The authors also want to acknowledge the invaluable collaboration of Martin and Sheilah Openshaw, coordinators the Undulate Ray project (http://undulateray.uk/) by providing a complete and valuable dataset of photos of individuals of the target species of this work undulate skate (*Raja undulata*) from South Coast of UK (West of Selsey Bill). Finally, Nuria Gómez-Vargas wants to thank the JAE Intro ICU grant awarded by the IIM-CSIC.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2023.102036.

References

Allken, V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiout, T., Malde, K., 2019. 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.

Alonso-Fernández, A., Otero, J., Bañón, R., Campelos, J.M., Quintero, F., Ribó, J., Filgueira, F., Juncal, L., Lamas, F., Gancedo, A., Molares, J., 2019. Inferring abundance trends of key species from a highly developed small-scale fishery off NE Atlantic. Fish. Res. 209, 101–116. https://doi.org/10.1016/J.FISHRES.2018.09.011.

Alonso-Fernández, A., Otero, J., Bañón, R., 2021. Indicators of body size variability in a highly developed small-scale fishery: ecological and management implications. Ecol. Indic. 121, 107141 https://doi.org/10.1016/J.ECOLIND.2020.107141.

Álvarez-Ellacuría, A., Palmer, M., Catalán, I.A., Lisani, J.L., 2020. 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.

Angelov, P., Soares, E., 2020. Towards explainable deep neural networks (xDNN). Neural Netw. 130, 185–194. https://doi.org/10.1016/J.NEUNET.2020.07.010.

Antelo, L., Ovalle, J., Vilas Fernández, C., Abad, E., Valeiras, J., Velasco, E., Pennino, M. G., Landeira, F., 2019. Nuevas herramientas de monitorización electrónica para una gestión sostenible y en línea de los recursos pesqueros: el proyecto SICAPTOR. XXII Foro Dos Recursos Mariños e Da Acuicultura Das Rías Galegas.

Benjamins, S., Dodd, J., Thorburn, J., Milway, V.A., Campbell, R., Bailey, D.M., 2018. Evaluating the potential of photo-identification as a monitoring tool for flapper skate

(Dipturus intermedius). Aquat. Conserv. Mar. Freshwat. Ecosyst. 28 (6), 1360–1373. https://doi.org/10.1002/AQC.2937.
Bouma, S., Pawley, M.D.M., Hupman, K., Gilman, A., 2018. Individual common dolphin identification via metric embedding learning. Int. Conf. Image Vision Comp. New Zealand 1–6. https://doi.org/10.1109/IVCNZ.2018.8634778.
Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R., 1993. Signature verification using a "Siamese" time delay neural network. Adv. Neural Inf. Proces. Syst. 6.
Chicco, D., 2021. Siamese Neural Networks: An Overview. Artificial Neural Networks 73–94. https://doi.org/10.1007/978-1-0716-0826-5_3.
Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. Methods Ecol. Evol. 10 (10), 1632–1644. https://doi.org/10.1111/2041-210X.13256.
Coelho, R., Bertozzi, M., Ungaro, N., Ellis, J., 2009. Raja undulata. In The IUCN Red List of Threatened Species.
Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., Lester, S.E., 2012. Status and solutions for the world's unassessed fisheries. Science 338 (6106), 517–520. https://doi.org/10.1126/SCIENCE.1223389/SUPPL_FILE/COSTELLO.SM.V2.PDF.
Dala-Corte, R.B., Moschetta, J.B., Becker, F.G., 2016. Photo-identification as a technique for recognition of individual fish: a test with the freshwater armored catfish Rineloricaria aequalicuspis Reis & Cardoso, 2001 (Siluriformes: Loricariidae). Neotrop. Ichthyol. 14 (1), 150074 https://doi.org/10.1590/1982-0224-20150074.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, Kai, Fei-Fei, Li, 2010. ImageNet: a large-scale hierarchical image database. IEEE Conf. Comp. Vision Pattern Recognition 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
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.
Hassaballah, M., Hosny, K.M., 2019. Recent advances in computer vision. Stud. Comput. Intell. 804.
Hastie, T., Friedman, J., Tibshirani, R., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. https://doi.org/10.1007/978-0-387-21606-5.
He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.
He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks. Lect. Notes Comput. Sci 9908 LNCS, 630–645. https://doi.org/10.1007/978-3-319-46493-0_38.
He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969.
Hilborn, R., Amoroso, R.O., Anderson, C.M., Baum, J.K., Branch, T.A., Costello, C., De Moor, C.L., Faraj, A., Hively, D., Jensen, O.P., Kurota, H., Little, L.R., Mace, P., McClanahan, T., Melnychuk, M.C., Minto, C., Osio, G.C., Parma, A.M., Pons, M., Ye, Y., 2020. Effective fisheries management instrumental in improving fish stock status. Proc. Natl. Acad. Sci. U. S. A. 117 (4), 2218–2224. https://doi.org/10.1073/PNAS.1909726116.
Hirsch, P.E., Eckmann, R., 2015. Individual identification of Eurasian perch Perca fluviatilis by means of their stripe patterns. Limnologica 54, 1–4. https://doi.org/10.1016/J.LIMNO.2015.07.003.
James, K., Bradshaw, K., 2020. Detecting plant species in the field with deep learning and drone technology. Methods Ecol. Evol. 11 (11), 1509–1519. https://doi.org/10.1111/2041-210X.13473.
Jessop, B.M., 2000. Estimates of population size and instream mortality rate of American Eel Elvers in a Nova Scotia River. Trans. Am. Fish. Soc. 129, 514–526.
Kristensen, E., Sand-Jensen, K., Martinsen, K.T., Madsen-Østerbye, M., Kragh, T., 2020. Fingerprinting pike: the use of image recognition to identify individual pikes. Fish. Res. 229, 105622 https://doi.org/10.1016/J.FISHRES.2020.105622.
Kuhnert, P.M., Duffy, L.M., Young, J.W., Olson, R.J., 2012. Predicting fish diet composition using a bagged classification tree approach: a case study using yellowfin tuna (Thunnus albacares). Mar. Biol. 159 (1), 87–100. https://doi.org/10.1007/S00227-011-1792-6/FIGURES/7.
Leeb, K., Villegas-Ríos, D., Mucientes, G., Garci, M., Gilcoto, M., Alonso-Fernández, A., 2021. Drivers of spatial behaviour of the endangered undulate skate, Raja undulata. Aquat. Conserv. Mar. Freshwat. Ecosyst. 31 (12), 3466–3479. https://doi.org/10.1002/AQC.3714.
Marshall, A.D., Pierce, S.J., 2012. The use and abuse of photographic identification in sharks and rays. J. Fish Biol. 80 (5), 1361–1379. https://doi.org/10.1111/J.1095-8649.2012.03244.X.
McCully, S., Walls, R.H.L., Ellis, J.R., 2015. Raja undulata. In: In *The IUCN Red List of Threatened Species*.
Meek, P.D., Vernes, K., Falzon, G., 2013. On the reliability of expert identification of small-medium sized mammals from camera trap photos. Wildl. Biol. Pract. 2, 1–19. https://doi.org/10.2461/wbp.2013.9.4.
Moskvyak, O., Maire, F., Dayoub, F., Armstrong, A.O., Baktashmotlagh, M., 2021. Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. In: DICTA 2021–2021 International Conference on Digital Image Computing: Techniques and Applications. https://doi.org/10.1109/DICTA52665.2021.9647359.
Mucientes, G., Irisarri, J., Villegas-Ríos, D., 2019. Interannual fine-scale site fidelity of male ballan wrasse Labrus bergylta revealed by photo-identification and tagging. J. Fish Biol. 95 (4), 1151–1155. https://doi.org/10.1111/JFB.14111.
Nepovinnykh, E., Eerola, T., Kalviainen, H., 2020. Siamese network based pelage pattern matching for ringed seal re-identification. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV) workshops, pp. 25–34.
Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J., 2021. A deep active learning system for species identification and counting in camera trap images. Methods Ecol. Evol. 12 (1), 150–161. https://doi.org/10.1111/2041-210X.13504.
Pimentel, M.A.F., Clifton, D.A., Clifton, L., Tarassenko, L., 2014. A review of novelty detection. Signal Process. 99, 215–249. https://doi.org/10.1016/J.SIGPRO.2013.12.026.
Sackett, D.K., Catalano, M., 2017. Spatial heterogeneity, variable rewards, tag loss, and tagging mortality affect the performance of mark–recapture designs to estimate exploitation: an example using red snapper in the Northern Gulf of Mexico. N. Am. J. Fish Manag. 37 (3), 558–573. https://doi.org/10.1080/02755947.2017.1303007.
Schneider, S., Taylor, G.W., Linquist, S., Kremer, S.C., 2019. Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods Ecol. Evol. 10 (4), 461–470. https://doi.org/10.1111/2041-210X.13133.
Schneider, S., Taylor, G.W., Kremer, S.C., 2020. Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops, pp. 44–52.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48. https://doi.org/10.1186/S40537-019-0197-0/FIGURES/33.
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.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. Adv. Neural Inf. Proces. Syst. 25.
Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4. In: Inception-ResNet and the Impact of Residual Connections on Learning. Thirty-First AAAI Conference on Artificial Intelligence.
Tamaya, I., Shimdate, M., Kubuta, N., Nomura, Y., 1982. Application for optical sensor to fish sorting. *Reito (Tokyo)*. Refrigeration 57, 1146–1150.
Torrey, L., Shavlik, J., 2010. Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques, 242–264. https://doi.org/10.4018/978-1-60566-766-9.CH011.
Tseng, C.-H., Kuo, Y.-F., Tseng, C.-H., Kuo, Y.-F., 2020. Detecting and counting harvested fish and identifying fish types in electronic monitoring system videos using deep convolutional neural networks. ICES J. Mar. Sci. 77 (4), 1367–1378. https://doi.org/10.1093/ICESJMS/FSAA076.
Xian, Y., 2020. Learning from Limited Labeled Data - Zero-Shot and Few-Shot Learning.
Zhang, C., Ma, Y., 2012. Ensemble Machine Learning: Methods and Applications. Springer.