



OPEN

FIN-PRINT a fully-automated multi-stage deep-learning-based framework for the individual recognition of killer whales

Christian Bergler¹✉, Alexander Gebhard¹, Jared R. Towers^{2,3}, Leonid Butyrev¹, Gary J. Sutton^{2,3}, Tasli J. H. Shaw^{2,3}, Andreas Maier¹ & Elmar Nöth¹

Biometric identification techniques such as photo-identification require an array of unique natural markings to identify individuals. From 1975 to present, Bigg's killer whales have been photo-identified along the west coast of North America, resulting in one of the largest and longest-running cetacean photo-identification datasets. However, data maintenance and analysis are extremely time and resource consuming. This study transfers the procedure of killer whale image identification into a fully automated, multi-stage, deep learning framework, entitled FIN-PRINT. It is composed of multiple sequentially ordered sub-components. FIN-PRINT is trained and evaluated on a dataset collected over an 8-year period (2011–2018) in the coastal waters off western North America, including 121,000 human-annotated identification images of Bigg's killer whales. At first, object detection is performed to identify unique killer whale markings, resulting in 94.4% recall, 94.1% precision, and 93.4% mean-average-precision (mAP). Second, all previously identified natural killer whale markings are extracted. The third step introduces a data enhancement mechanism by filtering between valid and invalid markings from previous processing levels, achieving 92.8% recall, 97.5% precision, and 95.2% accuracy. The fourth and final step involves multi-class individual recognition. When evaluated on the network test set, it achieved an accuracy of 92.5% with 97.2% top-3 unweighted accuracy (TUA) for the 100 most commonly photo-identified killer whales. Additionally, the method achieved an accuracy of 84.5% and a TUA of 92.9% when applied to the entire 2018 image collection of the 100 most common killer whales. The source code of FIN-PRINT can be adapted to other species and will be publicly available.

Biometric recognition typically relies on the visual differentiation of unique features on specific body parts of individuals. The best-known examples to distinguish identities of humans include analysis of individual fingerprint designs, retina features, and the composition of facial components^{1,2}. Individual recognition is also important in the field of wildlife biology, where images of specific body features are systematically used to differentiate between individuals of the same species. For example, repeated photo-identification of pigment patterns and appendage shape on individuals of various species of invertebrate³, aquatic⁴ and terrestrial mammals^{5,6}, birds⁷, fish^{8,9}, reptiles^{10,11}, and amphibians^{12,13} can be used to gain insights into the abundance, range, behaviour, ecology, and health of their populations.

The first systematic efforts to photo-identify free-ranging cetaceans began in the early 1970s⁴ and included studies on the population abundance of killer whales off the west coast of Canada¹⁴. It was found that individuals of this species could be recognized by the unique shapes of their dorsal fins as well as the shapes and pigment patterns on their saddle patches that were visible when the whales came to the surface. Thus, a combination of both attributes (dorsal fin and saddle patch) provides a distinct identification criterion¹⁵. Over time, several sympatric but genetically and behaviourally distinct populations of killer whales were discovered in the eastern North Pacific using photo-identification¹⁶. The “west coast transient” population of Bigg's killer whales is currently among the largest and most commonly photo-identified killer whale populations in this region. Individuals

¹Department of Computer Science - Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany. ²Bay Cetology, 257 Fir street, Alert Bay, BC V0N 1A0, Canada. ³Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada. ✉email: christian.bergler@fau.de

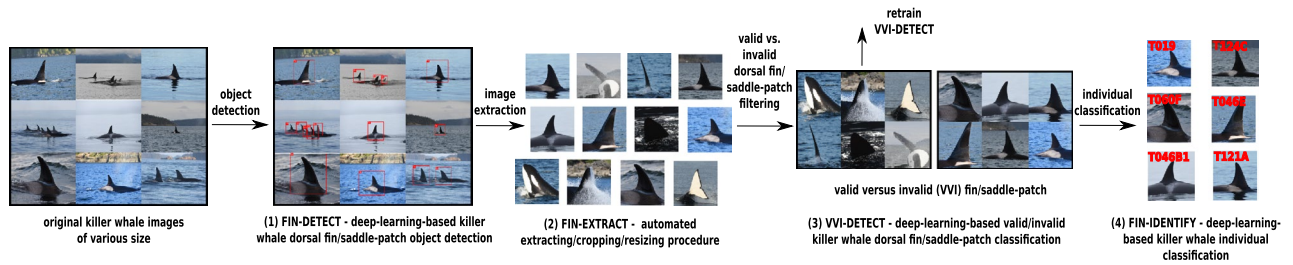


Figure 1. FIN-PRINT workflow including: (1) dorsal fin/saddle patch detection, (2) extraction of the detected killer whale markings, (3) valid versus invalid dorsal fin/saddle patch binary classification, and (4) multi-class killer whale individual identification.

have been systematically, but opportunistically photo-identified, from either or both, the left and right side, each year from 1975 to present resulting in one of the largest and longest-running cetacean photo-identification data archives in existence¹⁵.

The management and analysis of these photo-identification data currently require manual efforts which include labeling and sorting images, applying identification metadata to each photo¹⁷, entering resulting information into databases, and the periodic publication of reference material¹⁵. These tasks are typically best performed only by those who are intimately familiar with the unique physical features and social patterns of individuals in this population, as well as how they are likely to change over time. However, this requires an exceptional level of speciality and amount of time that may be expedited by taking advantage of developing technologies. Computers have assisted efforts to discern identities of individual cetaceans in identification images since the 1980s^{18,19} and over the following decades have been used increasingly to help manage workflow²⁰ and automate image analysis processes^{21–24}.

Most recently, machine (deep) learning algorithms have been setting new standards for image processing/analysis across various research areas and fields of application^{25–33}, due to increasing memory space and performance of central processing units (CPU) and graphics processing units (GPU)^{34–37}. Among many other image processing problems handled by deep learning, deep neural networks have recently also been applied to the detection and classification of individual animals of several species including amur tigers (*Panthera tigris altaica*)^{38–41}, elephants (*Proboscidea*)⁴², right whales (*Eubalaena*)^{43,44}, humpback whales (*Megaptera novaeangliae*)^{45–47}, brown bears (*Ursus arctos*)⁴⁸, giraffes (*Giraffa camelopardalis*)⁴⁹, pigs (*Sus scrofa domesticus*)⁵⁰, manta rays (*Mobula birostris*)⁵¹, common dolphins (*Delphinus delphis*)⁵², chimpanzees (*Pan troglodytes verus*)⁵³, red pandas (*Ailurus fulgens*)⁵⁴, giant pandas (*Ailuropoda melanoleuca*)⁵⁵, birds (e.g. sociable weaver (*Philetairus socius*), great tit (*Parus major*), zebra finch (*Taeniopygia guttata*)⁵⁶, gorillas (*Gorilla*)⁵⁷, primates (e.g. rhesus macaque (*Macaca mulatta*))⁵⁸, cattle (*Bos taurus*)⁵⁹, kiangs (*Equus kiang*)⁶⁰, zebras (*Equus quagga*)⁶⁰, and nyalas (*Tragelaphus angasi*)⁶¹, hawksbills (*Eretmochelys imbricata*)⁶², blue whales (*Balaenoptera musculus*)⁶³, and common bottlenose dolphins (*Tursiops truncatus*)⁶⁴. Besides deep learning-based detection and identification studies on single animal species, recent research also addresses cross-species recognition^{65–70}.

Despite some promising studies in the field of machine (deep) learning, it is difficult to transfer and apply existing approaches to model an end-to-end killer whale individual recognition pipeline, consisting of detection, extraction, enhancement, and classification (see Fig. 1). Several studies perform animal identification across different species^{65–70}, rather than recognition of individuals belonging to the same species. Others address only parts of an individual identification pipeline, such as only detection⁶⁰ or classification^{59,68,69}. Some approaches present a combination of modern deep learning techniques together with traditional machine learning algorithms^{39,42,48–51,57,69}. FIN-PRINT provides a modular, transferable, and state-of-the-art identification pipeline for killer whales, exclusively applying well-established deep learning concepts, to facilitate robust and task-specific feature learning at each stage. In comparison to traditional machine learning methods, all features were learned and derived in a data-driven fashion. Consequently, it was not necessary to perform any feature selection based on heuristic and/or analytical approaches. FIN-PRINT was trained and evaluated on a large, variable and complex dataset of approximately 121,000 human-annotated Bigg's killer whale identification images. In order to robustly handle the diversity in this dataset, FIN-PRINT integrates an automated, deep learning-based quality inspection, acting as a validation mechanism prior to the final classification. This guarantees that both the original image and the results obtained from upstream steps (e.g. detection), meet the standards for robust individual classification. A number of studies performed Deep Metric Learning along with the triplet loss^{71–73}, modifications of it, and/or combinations with other loss functions^{38,40,45,46,49,51,52,64}. However, specification of appropriate hard and semi-hard triplets⁷³ is extremely challenging, since: (1) killer whale individuals have been recorded from both body sides, resulting in different animal orientations besides potential deviating natural markings⁴⁵, (2) natural identifiers change over time (growth, acquisition of scars, etc.), (3) deviating saddle patch visibility, often obscured to some extent by water and/or other animals, as well as (4) variation of challenging image conditions (e.g. weather, exposure, etc.). Due to the mentioned difficulties, next to sufficiently large individual-specific data volumes, traditional supervised multi-class classification was applied to build an initial pilot system.

The FIN-PRINT pipeline (see Fig. 1) consists of (1) FIN-DETECT, a YOLOv3^{74–77}-based object detection network for recognizing killer whale dorsal fins and associated saddle patches in images with 1 to N individuals, (2) FIN-EXTRACT, an automatic extraction procedure cropping and equally resizing all detected dorsal fin/saddle patch markings within an image, (3) VVI-DETECT, a ResNet34⁷⁸-based convolutional neural network (CNN)

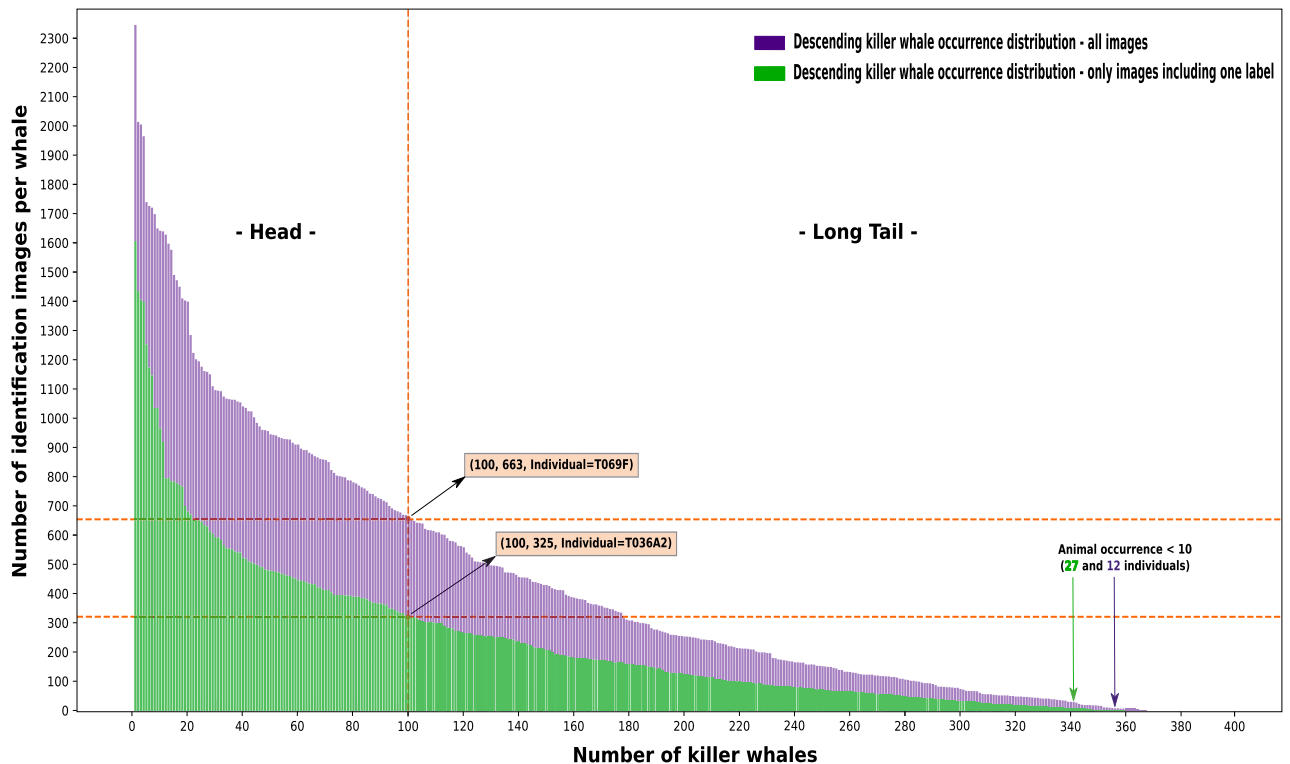


Figure 2. Bigg’s killer whale image long-tailed data distribution (2011–2018), summing up a total of 121,095 identification images, with 86,789 containing single labels, as well as 34,306 photos including multiple labels, resulting in 367 identified individuals (average number of images per individual ≈ 456 , standard deviation ≈ 442). The two colored graphs visualize the number of identification images per whale in descending order w.r.t. all images, including single and multiple labels (purple curve) and those only containing a single label (green curve). Furthermore, an exemplary data point is visualized for both curves, presenting the number of identification images in relation to a selected number of whales, here for the top-100, clearly describing the exponential decline. Moreover, the number of animals at which the total amount of identification images is < 10 were marked for both curves. In total, 367 individuals were encountered across 2011–2018. Among them, 128 and 125 were found at least once in each year when considering all images and only those with single labels, respectively.

performing data enhancement by classifying between previously detected/extracted valid versus invalid (VVI) killer whale identification sub-images (e.g. bad weather conditions, blurred, missing saddle patch, difficult angle, detection errors, etc.), and (4) FIN-IDENTIFY, a ResNet34⁷⁸-based CNN for multi-class killer whale individual classification modeling the 100 most commonly photo-identified killer whales. To the best of the authors’ knowledge, this is the first study transferring the analysis of killer whale image identification¹⁵ into a fully automated, multi-stage, sequentially ordered, deep-learning-based framework, in order to machine-identify individuals.

Materials and methods

Bigg’s killer whale photo-identification dataset. The dataset of this study includes photos of Bigg’s killer whale individuals accumulated over a period of 8 years (2011–2018), from the coastal waters of southeastern Alaska down to central California¹⁵. None of these animals were directly approached explicitly for this study. All photo-identification data was collected under federally authorized research licenses or from beyond mandated minimum viewing distances.

Supplementary Figure S1 visualizes a series of example images of this dataset. Each image contains one or more individuals. In addition to the identification name of the individual(s), further metadata such as photographer, GPS-coordinates, date, and time are provided. Every identification label is an alphanumeric sequence based on the animals’ ecotype (T—Transient), order of original documentation (e.g. T109), and order of birth (e.g. T109A2—the second offspring of the first offspring of T109)¹⁵.

A parsing procedure was designed to verify, analyze, and prepare the image data, guaranteeing adequate preparation for subsequent machine (deep) learning methods. Results of the entire data parsing procedure are presented in Fig. 2 and Supplementary Table S1. Figure 2 visualizes the number of identified individuals, together with the total amount of occurrences in descending order, considering (1) all images, and (2) only photos including a single label. General statistics with respect to the entire dataset are reported in the caption of Fig. 2. Supplementary Table S1 illustrates the 10 most commonly occurring individuals across all 8 years of data, considering all images including single and multiple labels, compared to photos only containing a single label.

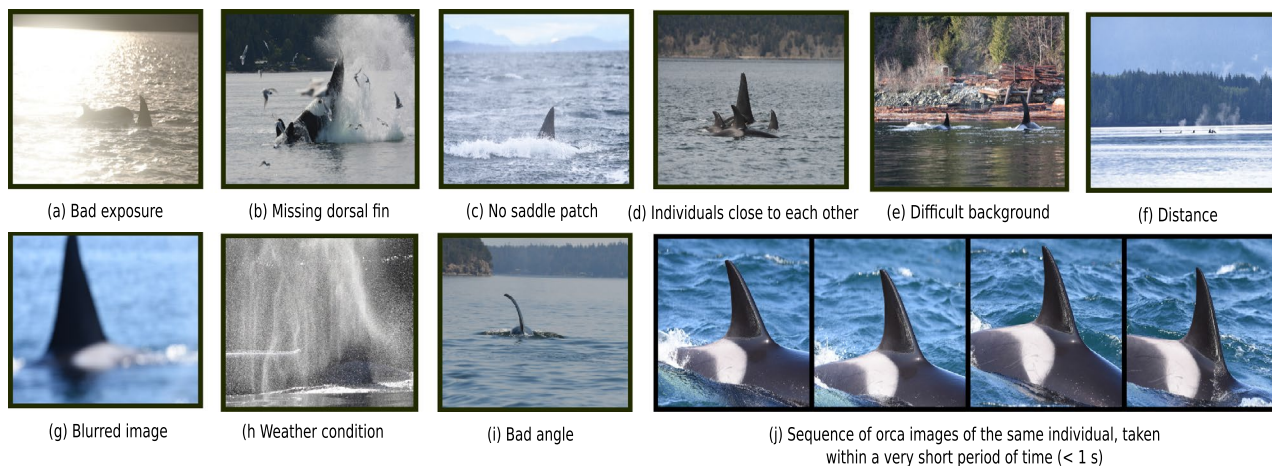


Figure 3. Examples of image content which either lead to completely unusable/invalid data samples, or which make a robust and correct detection/classification much more difficult.

The dataset exhibits a substantial class imbalance, as evidenced by the exponential decline in frequencies per killer whale individual (see Fig. 2). Especially for real-world datasets, such unbalanced data partitioning is a common and well-known phenomenon, also referred to as long-tailed data distribution⁷⁹. Such long-tailed data distributions are divided into two sections⁷⁹: (1) the *Head* region—representing the most commonly identified killer whale individuals, and (2) the *Long-Tail* region—visualizing a significantly larger number of killer whale individuals, however, with considerably less occurrences. For the purpose of this pilot study, the top-100 most commonly occurring killer whale individuals were selected for supervised classification and as boundary between the head and long-tail area (see Fig. 2). The defined boundary of the top-100 killer whales (head region) represents approximately 1/4 (100 out of 367) of the individuals, however, covering about 2/3 (55,305 out of 86,789) of the entire dataset of single-labeled images.

However, the number of usable and correctly labeled images which can actually be utilized for machine learning must be adjusted downward due to several circumstances. Figure 3a–i visualizes multiple examples of situations where images contain valid labels. However, the relevant biometric features are very difficult to recognize or not visible at all. These images cannot be labeled without contextual knowledge, for example by observing previous and/or subsequent images and/or knowing additional information about family-related structures. Therefore, such photos cannot be used for classification of individuals and have to be filtered out in advance.

Another scenario that impacts the final number of usable identification images is visualized in Fig. 3j. While conducting photo-identification in the field, several images are sometimes taken in very short intervals (< 1 s). However, this procedure leads to several very similar images. To avoid biasing the actual multi-class identification performance by including such images in validation and testing, only the first image of a photo series was machine-selected if the images were taken within a time interval $\delta \leq 5$ s, including the same date and photographer. Considering the photo series visualized in Fig. 3j, only the first image was utilized as a potential sample for network validation or testing. The training material for individual classification was unaffected by this time interval rule, since augmentation procedures change the images during training anyway.

Killer whale dorsal fin/saddle patch detection (FIN-DETECT). *Object detection.* In order to extract the regions of interest—killer whale dorsal fin(s) and saddle patch(es)—from the images, an automated and robust object detection has to be conducted. Object detection includes classification and localization of the corresponding object within the respective image³⁶. In this context, circumscribing rectangles, so-called bounding boxes, are utilized and drawn around the objects to be recognized. Between a ground truth bounding box and the predicted bounding box, a quality metric named *Intersection over Union (IoU)* ($= \frac{\text{Area of Overlap}}{\text{Area of Union}}$) is often used as a quality criterion⁸⁰.

Two additional evaluation attributes are of essential importance too³⁶: (1) *objectness score*—describes the probability that an object is present inside a given bounding box, and (2) *class confidences*—characterize the probability distribution over all distinct object classes. All objects which have to be localized inside an image can strongly vary not only in type and shape, but also in size. Hence, object detection algorithms usually predict a variety of potential bounding boxes. As a result, individual objects may be detected several times by circumscribing bounding boxes, locating at slightly different positions³⁶. To counteract this phenomenon, *non-maximum suppression*³⁶ (NMS) is executed to keep only the best fitting one. Since object detection requires both, correct classification and localization, the metrics per class are determined as follows⁸¹:

(1) *true positive (TP)*: the target object is within the predicted bounding box area, the bounding box objectness score is larger than a chosen threshold, the object classification and assignment are correct, and IoU between bounding box prediction versus ground truth is higher than a given threshold and all other IoUs of potential overlaying boxes (in case of overlaying boxes, only the box indicating the highest IoU is considered as TP, whereas all remaining boxes are false positives), (2) *false positive (FP)*: the bounding box objectness score is larger than a chosen threshold, but either the target object is not within the predicted circumscribing rectangle,

Dataset		Split					
		Training		Validation		Test	
		Samples		Samples		Samples	
		Photos	%	Photos	%	Photos	%
HADD ^a	2286	1686	73.8	300	13.1	300	13.1
EADD ^b	7511	5257	70.0	1127	15.0	1127	15.0

Table 1. Human-Annotated Detection Dataset (HADD), including human-labeled dorsal fin/saddle-patch bounding boxes, as well as Extended-Annotated Detection Dataset (EADD) containing human- and machine-labeled dorsal fin/saddle-patch bounding boxes. ^aHADD Human-Annotated Detection Dataset. ^bEADD Extended-Annotated Detection Dataset.

the classification hypothesis is wrong, and/or IoU is smaller compared to any other possible overlaying bounding boxes, (3) *false negative (FN)*: the target object is in the image, but no predicted bounding box hypothesis detected the corresponding object properly, (4) *true negatives (TN)*: object detection ignores TNs, since there are evidently an infinite number of empty boxes with an objectness score that is smaller than a chosen threshold. Based on these traditional binary classification scores, target metrics such as *precision*, *recall*, *F1-score*, *average precision (AP)*, and *mean average precision (mAP)* can be calculated³⁶. The average precision describes the *area-under-the-curve (AUC)* of a precision/recall graph, transformed into a monotonically decreasing curve beforehand, calculated on the basis of different IoU thresholds³⁶. The AP is calculated for each class, while the mAP refers to the average of all class-related AP scores³⁶. Consequently, AP and mAP are identical unless the number of classes is greater than one³⁶.

Detection data. The dataset which was utilized for training and evaluation of FIN-DETECT was generated via a two-step semi-automatic procedure. In a first step, 2,286 images, originating from various months in 2015, were manually annotated with bounding boxes resulting in the *Human-Annotated Detection Dataset (HADD)*—see Table 1. For this purpose, every dorsal fin and associated saddle patch, visible in each image, were individually circumscribed with a rectangle. FIN-DETECT was trained on HADD using the data distribution reported in Table 1.

The resulting and preliminary version of FIN-DETECT was utilized to automatically apply bounding boxes to randomly chosen unseen images from 2011, 2015, and 2018 in order to enlarge the HADD with machine-identified samples. These samples were not manually verified, but images with no bounding boxes, as well as those with more bounding boxes than labels, were discarded. After applying these rules, a joint dataset, named the *Extended-Annotated Detection Dataset (EADD)*, was created, consisting of the HADD and all valid machine-identified data samples. The resulting EADD (see Table 1) was utilized to retrain FIN-DETECT, which was ultimately applied to all future killer whale detections.

Network architecture, data preprocessing, training, and evaluation. FIN-DETECT, visualized in Supplementary Fig. S2, is based on an extended version of the original YOLOv3^{76,77}-based object detection architecture. YOLOv3^{74–76} (*You Only Look Once*) is a real-time, single-stage, multi-scale, and fully-convolutional object detection algorithm, which was first introduced as YOLOv1 by Redmon et al.⁷⁴ and continuous improvements have led to the most recent version known as YOLOv5⁸². At the development of FIN-PRINT, YOLOv3 was the most recent version. FIN-DETECT (see Supplementary Fig. S2) essentially consists of two major network parts^{74–76,83}: (1) *feature extraction network*, usually referred to as feature extractor and/or backbone network, learning compressed representations (feature maps) of a given input image, representing the foundation for subsequent detection, and (2) *feature pyramid network*, also named head-subnet and/or detector, responsible for detecting objects at three different scales. FIN-DETECT receives as network input a preprocessed, re-scaled, and square 416×416 px RGB-image (zero-padding in case of a none-square original image), resulting in an input shape of $3 \times 416 \times 416$. The network detects objects utilizing a 13×13 , 26×26 , and 52×52 grid to recognize large, medium, and small patterns^{76,83} (see Supplementary Fig. S2). FIN-DETECT predicts per cell a 1×21 detection vector, which contains $b = 3$ different bounding boxes and $c = 2$ classes (dorsal fin/saddle patch vs. no dorsal fin/saddle patch), combined with four 0/1-normalized bounding box coordinates (x, y, w, h) and one objectness score per box, resulting in $b * (5 + c) = 21$ elements per cell. Consequently, the scale-dependent detection outputs of FIN-DETECT comprised a final output shape of $13 \times 13 \times 21$, $26 \times 26 \times 21$, and $52 \times 52 \times 21$ (see Supplementary Fig. S2). More detailed information about YOLO in general, YOLOv3, and/or other YOLO versions can be found here^{74–76,82,84,85}.

The backbone network (Darknet-53⁷⁶) of FIN-DETECT was initialized with pre-trained weights on ImageNet⁸⁶. A detailed overview about all other network hyperparameters is given in Supplementary Table S2. Moreover, FIN-DETECT implements the following YOLOv3⁷⁶ detection parameters: objectness score threshold of 0.5 (training, validation) and 0.8 (testing), IoU threshold of 0.5, and NMS threshold equals to 0.5. FIN-DETECT reports precision, recall, F1-Score, and mean average precision as evaluation metrics. Based on a given input image, FIN-DETECT returns a text file containing 0/1-normalized bounding box information (x, y, w, h) of every detection hypothesis.

Dataset		Split											
		Training				Validation				Test			
		Samples				Samples				Samples			
		Valid	Invalid	Σ	%	Valid	Invalid	Σ	%	Valid	Invalid	Σ	%
VIKWID11-17 ^a	1590	700	509	1209	76.0	126	89	215	13.5	83	83	166	10.5

Table 2. Valid/Invalid Killer Whale Identification Dataset 2011–2017 (VIKWID11-17), a human-annotated dataset consisting of valid and invalid identification images (dorsal fin + saddle-patch), utilized to train, validate, and test VVI-DETECT, after applying the interval rule of 5 s with respect to the validation and test set. ^aVIKWID11-17 Valid/Invalid Killer Whale Identification Dataset 2011–2017.

Killer whale dorsal fin/saddle patch extraction (FIN-EXTRACT). FIN-EXTRACT facilitates automatic extraction and subsequent rescaling of previously detected and marked image sub-regions using the bounding box information derived by FIN-DETECT. For each identified bounding box, a square 512×512 px RGB-sub-image was cropped from the original photo. In a first step, the 0/1-normalized bounding box information (x, y, w, h) was multiplied by the original image shape to obtain the correct coordinates within the original image. In case a bounding box was not square, the larger of the two dimensions was utilized to reshape the original detection rectangle. Furthermore, it was verified whether a bounding box extended beyond the edge of an image and moved accordingly if necessary. In case the original image was smaller than 512×512 px, it was interpolated and resized respectively. Otherwise, a sub-image, based on the original bounding box size, was cropped and if applicable compressed and resized to 512×512 px. Depending on the resized bounding box, this may result in a bit more background content. However, any kind of zero-padding is avoided for subsequent individual classification. In addition, the image quality of the final extracted sub-image(s) depends on the original image resolution, along with the distance of the individual(s) within the captured photos.

Valid versus invalid (VVI) dorsal fin/saddle patch detection (VVI-DETECT). *VVI detection.* Considering potential detection errors (e.g. tail and/or pectoral fins, triangular formed head of the animal, etc.), besides all the different challenging situations visualized in Fig. 3a–i, additional data enhancement is indispensable (see also examples in Supplementary Fig. S3). All these scenarios either result in completely unusable/invalid (e.g. missing dorsal fin, no saddle patch, bad angle, distance, detection errors), or insufficient quality images (e.g. poor weather conditions, bad exposure, blurred image). Without sufficient domain knowledge and additional meta-information (e.g. images shortly taken before, other animals in the image, family-related structures, etc.), all the aforementioned situations lead to invalid identification images which are not able to be classified correctly by human or machine. Detected/extracted RGB-sub-images containing a single dorsal fin and saddle patch are considered as valid identification images. To filter the majority of such invalid samples originating from previous processing levels, a binary classification network was designed to distinguish between two classes—*Valid Versus Invalid* (VVI)—killer whale identification images prior to final multi-class individual recognition. Supplementary Fig. S3 visualizes some of the challenging pre-detected/-extracted sub-images, belonging to the *invalid* class.

Detection data. In order to train VVI-DETECT, a two-class dataset, named *Valid/Invalid Killer Whale Identification Dataset 2011–2017* (VIKWID11-17), was utilized. Table 2 describes VIKWID11-17 in combination with the respective data distribution. VIKWID11-17 is a manually labeled data archive based on randomly chosen, previously detected (FIN-DETECT), and extracted (FIN-EXTRACT) sub-images from 2011 to 2017. In addition to multiple valid pre-detected/-extracted identification images of different individuals, the dataset also includes examples of invalid sub-images covering the scenarios illustrated in Fig. 3a–i. Furthermore, the invalid class was extended by examples of images with potential detection errors (noise), such as water, boats, coastline, houses and/or other landscape backgrounds, to also filter such cases in advance. During data selection an interval of 5 s was applied to the validation and test set (see Fig. 3j) in order to not distort classification accuracy in any way.

Network architecture, data preprocessing, training, and evaluation. VVI-DETECT, visualized in Supplementary Fig. S3, is a ResNet34⁷⁸-based convolutional neural network (CNN), designed for binary classification between valid versus invalid (VVI) identification images. Residual networks⁷⁸ (ResNets) consist of a sequence of residual layers, which are built up from building blocks including concatenations of weight (e.g. convolutional/fully-connected), normalization (e.g. batch-norm⁸⁷), and activation layers (e.g. ReLU⁸⁸), together with residual-/skip-connections⁷⁸. These connections allow the network to optimize a residual mapping $F(x) = H(x) - x$ with respect to a given input x , rather than directly learning an underlying mapping $H(x)$ ⁷⁸. This type of learning, called residual learning, opens up the possibility to train deeper models⁷⁸. The use of different building block types, together with the number of blocks, results in various ResNet architectures, like ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152⁷⁸. For more detailed information about the concept of residual learning/networks, see He et al.⁷⁸. Compared to the original ResNet34 architecture, the size of the initial 7×7 convolution kernel was changed to 9×9 , in order to cover larger receptive fields at the initial stage. As network input, VVI-DETECT receives data of previously detected (FIN-DETECT) and extracted/reshaped (FIN-EXTRACT) $3 \times 512 \times 512$ -large RGB-pictures for both classes. The network output is a 1×2 probability vector, containing class-wise model prediction probabilities (see Supplementary Fig. S3). Based on preliminary investigations, ResNet34⁷⁸ proved to be the most efficient version for this entire study in terms of performance and computa-

Dataset	Split												
	Training				Validation				Test				
	Samples				Samples				Samples				
	Valid	Invalid	Σ	%	Valid	Invalid	Σ	%	Valid	Invalid	Σ	%	
KWID11-17 ^a	39,464	27,238	2227	29,465	74.7	4940	395	5,335	13.5	4279	385	4664	11.8
KWIDE11-17 ^b	65,713	48,200	2226	50,426	76.7	7729	392	8121	12.4	6811	355	7166	10.9

Table 3. Killer Whale Individual Dataset 2011–2017 (KWID11-17), including machine-annotated data of valid images (dorsal fin + saddle-patch) for the 100 most commonly photographed individuals satisfying the data constraints (one label per image + exactly one bounding box prediction), in combination with machine-annotated invalid data utilizing VVI-DETECT after applying the interval rule of 5 s. Killer Whale Individual Dataset Extended 2011–2017 (KWIDE11-17) extends the KWID11-17 data archive with images of the 100 most common individuals represented in images containing more than one label and classified via the first version of FIN-IDENTIFY, trained on KWID11-17. Notice that the distribution of the invalid photos differs slightly between KWID11-17 and KWIDE11-17 due to the different data splits and subsequent effect of the 5 s interval rule. Furthermore, additional statistics regarding the number of identification images (100 classes) are reported for both datasets. ^aKWID11-17 Killer Whale Individual Dataset 2011–2017—Statistics on the number of identification images (100 most common classes): mean = 364.57, stdv = 162.91, min = 135 (T073B), max = 916 (T019B) training stats (only valid images): mean = 272.38, stdv = 125.79, min = 107 (T073B), max = 695 (T019B) validation stats (only valid images): mean = 49.40, stdv = 20.81, min = 16 (T073B), max = 120 (T019B) testing stats (only valid images): mean = 42.79, stdv = 18.60, min = 8 (T121A), max = 101 (T019B). ^bKWIDE11-17 Killer Whale Individual Dataset Extended 2011–2017—Statistics on the number of identification images (100 most common classes): mean = 627.40, stdv = 245.06, min = 172 (T073B), max = 1442 (T019B) training stats (only valid images): mean = 482.00, stdv = 192.32, min = 139 (T073B), max = 1122 (T019B) validation stats (only valid images): mean = 77.29, stdv = 28.74, min = 17 (T073B), max = 174 (T019B) testing stats (only valid images): mean = 68.11, stdv = 26.98, min = 10 (T121A), max = 146 (T019B).

tion efficiency compared to other ResNet architectures. VVI-DETECT integrates an augmentation procedure consisting of eight different functions: (1) addition of random Gaussian noise to the image, (2) image rotation at maximum angle of ± 25 degree, (3) blurring the image by applying a gaussian blur, (4) mirroring the picture with respect to the y -axis, (5) edge enhancement within the image, (6) sharpening the input picture, (7) brighten/darken of the image, and (8) random color change by swapping the RGB channels. Out of this function pool, number, type, and arrangement of augmentation operations were randomly determined for each image within the training phase (no augmentation during validation and testing). The random number of augmentations per image was within an interval of $[1 : a_{max}]$ with $a_{max} \in [1 : 8]$ being constant across the entire training. In this study the maximum augmentation number per image was set to $a_{max} = 5$. VVI-DETECT reports accuracy, precision, recall, F1-Score, and false-positive-rate. A detailed description of all relevant network hyperparameters is illustrated in Supplementary Table S2.

Individual killer whale classification network (FIN-IDENTIFY). *Individual killer whale classification.* Robust multi-class killer whale individual classification requires representative and high-quality animal-specific image data in sufficient quantity. However, significant variations can be observed in the total number of animal-specific images (see Fig. 2). In addition, multiple and essential data constraints have been introduced which strongly affect the actual amount of usable identification images per individual, such as (1) only single-labeled images together with exactly one predicted bounding box hypothesis, (2) data enhancement by pre-filtering invalid identification images to avoid situations visualized in Fig. 3a–i, and (3) time interval rule of 5 s during network validation and testing to counteract the effect of classifying very similar photos, visualized in Fig. 3j. Moreover, all photos from 2018 were completely ignored for additional network evaluation purposes. Additionally, all images including more than a single label (in total 34,306 pictures, 2011–2018, see Fig. 2) could not be used for training an initial multi-class identification network due to the *label assignment problem*. The *label assignment problem* describes the situation where an image contains multiple individuals and labels, however, it is unknown which label belongs to which individual. All these data restrictions and constraints led to a significant, qualitative improvement of the material, but also considerably reduced the amount of usable data. In summary, these data limitations led to a final representation of the 100 (out of 367) most commonly single-labeled Bigg's individuals (see Fig. 2), present across all years (2011–2018), representing about 64% (55,305 photos) of the entire single annotated and original data from 2011 to 2018 (86,789 images). Based on the top-100 killer whales, the smallest individual-specific number of remaining data samples comprised 135 images (see Table 3), to still provide sufficient variation and data diversity combined with various image augmentation techniques during model training. Despite previous filtering by VVI-DETECT and to avoid potential errors caused by previous processing levels, the proposed *invalid* class was also included at this stage resulting in a final 101-class (100 individuals, 1 rejection class) procedure.

Identification data. FIN-IDENTIFY was trained on two different datasets, both illustrated in Table 3. The first dataset, named *Killer Whale Individual Dataset 2011–2017 (KWID11-17)*, consisted of 39,464 excerpts including

Metric	Dataset			
	HADD ^a		EADD ^b	
	Validation [%]	Test [%]	Validation [%]	Test [%]
Recall	95.0	89.0	95.1	94.4
Precision	80.0	85.0	94.2	94.1
F1-Score	86.9	87.0	94.7	94.2
mAP ^c	92.0	82.0	94.2	93.4

Table 4. Detection results while training two versions of FIN-DETECT with respect to HADD and EADD. ^aHADD Human-Annotated Detection Dataset. ^bEADD Extended-Annotated Detection Dataset. ^cmAP mean Average Precision.

only a single label, distributed across 101 classes, and recorded between 2011 and 2017 (see Table 3). All excerpts were machine-annotated, applying FIN-DETECT, FIN-EXTRACT, and VVI-DETECT in a sequential order, following the previously mentioned data constraints and restrictions. VVI-DETECT considered an image to be invalid if the network confidence was $p_{invalid} > 0.85$. The VIKWID11-17 dataset (see Table 2), on which VVI-DETECT was trained on, is completely independent from the entire data listed in Table 3. KWID11-17 consists of 36,457 images being assigned to the valid class, whereas 3007 photos were added to the invalid class, representing a small portion of the overall amount of detected invalid images across 2011 to 2017 in order to not bias class distributions. Table 3 presents the final data distribution of KWID11-17 as well as dataset-specific statistics.

To add additional data and simultaneously counteract the *label assignment problem*, the first version of FIN-IDENTIFY, trained on KWID11-17, was applied to all images from 2011 until 2017, including those with multiple labels and either one or more of the trained 100 individuals. FIN-IDENTIFY classified all potential detected (FIN-DETECT) and extracted (FIN-EXTRACT) labels for each image containing more than one animal. If the best classification hypothesis (class with the highest probability) per sub-image matches one of the original labels applied to that image, it was considered as correctly classified and added to the respective class. The resulting extended dataset, entitled *Killer Whale Individual Dataset Extended 2011–2017 (KWIDE11-17)*, together with the corresponding data distribution, was utilized to train an updated and more robust version of FIN-IDENTIFY (see Table 3). KWIDE11-17 consists of KWID11-17, extended by the additional machine-identified multi-label material, leading to a total number of 65,713 excerpts, distributed across 101 classes. The total number of valid identification images is 62,740, whereas the invalid class comprises 2,973 images. KWID11-17 and KWIDE11-17 use the same portion of machine-annotated invalid data excerpts, however, the overall number of samples slightly differs (KWID11-17—3007 versus KWIDE11-17—2973) due to a different split, in combination with the applied interval rule of 5 s during validation and testing.

Network architecture, data preprocessing, training, and evaluation. FIN-IDENTIFY, visualized in Supplementary Fig. S4, is a ResNet34⁷⁸-based convolutional neural network (CNN), created for multi-class individual classification. The network architecture is identical to VVI-DETECT (see Supplementary Fig. S3) except for the final 101-class output layer (1×101 probability vector). FIN-IDENTIFY was trained on the $3 \times 512 \times 512$ sub-images, generated by FIN-EXTRACT and if necessary filtered by VVI-DETECT (see Fig. 1 and Supplementary Fig. S4). Besides the same network architecture, identical interval rule conditions (5 s) were applied during training. Data augmentation and preprocessing was also identical to VVI-DETECT and all other required network hyperparameters are listed in Supplementary Table S2. Next to the overall accuracy, FIN-IDENTIFY reports a top-3 weighted (TWA) and unweighted accuracy (TUA). TWA describes whether the target class probability is within the top-3 and if so, a rank-dependent weight is assigned ($\omega_1 = 1$, $\omega_2 = 0.5$, and $\omega_3 = 0.25$). TUA illustrates, if the target individual is within the top-3, it is counted as correct, independent of the respective rank. For both metrics, either the sum of all weighted, or correct predictions is divided by the total number of classifications.

Experiments

The following major experiments were conducted: (1) training/evaluating FIN-DETECT on the dataset listed in Table 1 (HADD, EADD), to derive a robust dorsal fin/saddle patch detection network, (2) training/evaluating VVI-DETECT on the data (VIKWID11-17) presented in Table 2, (3) training/evaluating FIN-IDENTIFY with respect to the datasets (KWID11-17, KWIDE11-17) reported in Table 3, and (4) applying the entire FIN-PRINT pipeline (see Fig. 1), while utilizing the best previously trained networks, to all original, unseen, and single-labeled images from 2018, containing individuals which are modeled and represented within the 100 classes of FIN-IDENTIFY (see Supplementary Table S1).

Results

FIN-DETECT and FIN-EXTRACT. Table 4 reports validation and test results (recall, precision, F1-score, mAP) of FIN-DETECT evaluated on both detection datasets—HADD and EADD (see Table 1). Despite the fact that both data archives are not directly comparable, because of different data volumes and distributions, the automated and machine-driven data enlargement shows significant improvements with respect to the validation and test metrics. The version of FIN-DETECT trained on the EADD data material was utilized within all subse-



Figure 4. Dorsal fin/saddle patch detection and extraction results based on randomly chosen identification images from various years (2011–2017), applying FIN-DETECT, trained on the machine-extended EADD data archive (see Table 1), and FIN-EXTRACT.

Metric	Dataset	
	VIKWID11-17 ^a	
	Validation [%]	Test [%]
Recall	97.8	92.8
Precision	92.6	97.5
FPR	5.6	2.4
F1-Score	95.1	95.1
Accuracy	95.8	95.2

Table 5. Detection results of VVI-DETECT to filter between valid versus invalid identification images (data enhancement), while training VVI-DETECT on VIKWID11-17. ^aVIKWID11-17 Valid/Invalid Killer Whale Identification Dataset 2011–2017.

quent machine detection tasks. In addition to the traditional object recognition metrics listed in Table 4, various detection and extraction examples have been visualized in Fig. 4. All detection results, visualized in Fig. 4, were computed by applying FIN-DETECT, trained on the machine-extended EADD, to some random and unseen images from different years. Next to the detected and valid identification sub-images, represented by the red circumscribing bounding boxes (see Fig. 4), the associated extractions were created applying FIN-EXTRACT, together with the corresponding bounding box information. The image pairs, visualized in Fig. 4, consist of detection results and corresponding extractions. Besides valid fin/saddle patch detection results, example images of invalid, but correctly detected identification images, are displayed as well (see Fig. 4, last row). In all these cases the dorsal fin was detected correctly, however, due to lack of information and/or very challenging scenarios, the extracted sub-images are unusable/invalid for future individual identification (bad angle, no saddle patch, individuals close to each other, bad exposure, difficult background—see also Fig. 3a–i).

VVI-DETECT. Table 5 reports validation and test results of VVI-DETECT evaluated on VIKWID11-17 (see Table 2). This model was utilized for all required valid versus invalid image predictions. Besides validation and test metrics, example images of various, correctly predicted and filtered invalid identification photos from the unseen 2018 dataset are visualized in Fig. 5. The sub-images, presented in Fig. 5, reflect the previously mentioned variety of challenging scenarios shown in Fig. 3a–i.

The photos from 2018 that are shown in Fig. 5 visualize examples of invalid identification images due to poor image quality (lighting, exposure, etc) or poor subject representation (bad angle, too distant, dorsal fin and saddle patch not shown, etc.) (see also Fig. 3a–i). The problem regarding such detection errors is that at least one appendage (tail, pectoral, and/or dorsal fin) is present in most of these images (see Fig. 5, detection errors—last row). Furthermore, there are also cases where the shape of the recognized object is very close to the triangular structure of the fin (e.g. a spyhop where the killer whale lifts its head out of the water, see last row in Fig. 5). All these invalid data samples were successfully pre-filtered utilizing VVI-DETECT as an additional data enhancement step, to avoid subsequent misclassifications during final individual recognition (see FIN-IDENTIFY).

FIN-IDENTIFY. The last step of the entire FIN-PRINT pipeline, visualized in Fig. 1, describes final individual multi-class identification. Due to reasons of comparison, the results for both models—the preliminary

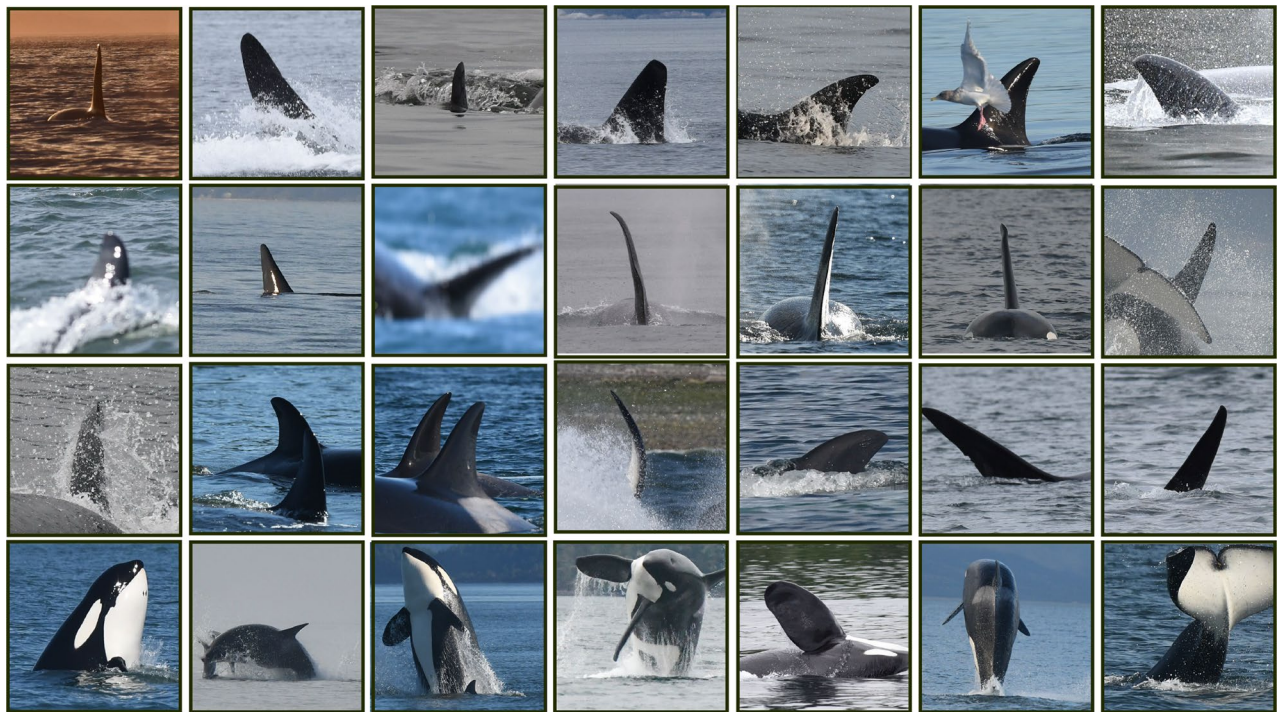


Figure 5. Detected (FIN-DETECT) and extracted (FIN-EXTRACT) unseen identification images from 2018, which were successfully categorized and filtered as invalid identification images by VVI-DETECT, trained on VIKWID11-17, reported in Table 2.

Metric	Dataset			
	KWID11-17 ^a		KWIDE11-17 ^b	
	Validation [%]	Test [%]	Validation [%]	Test [%]
Accuracy	85.8	86.7	91.1	92.5
TWA ^c	89.0	89.9	93.4	94.6
TUA ^d	93.2	94.3	96.3	97.2

Table 6. Individual killer whale classification results (101-classes), while training two versions of FIN-IDENTIFY, using the initial KWID11-17 or KWIDE11-17 datasets. ^aKWID11-17 Killer Whale Individual Dataset 2011–2017. ^bKWIDE11-17 Killer Whale Individual Dataset Extended 2011–2017. ^cTWA Top-3 Weighted Accuracy. ^dTUA Top-3 Unweighted Accuracy.

version and the final FIN-IDENTIFY network—are reported. In both cases the overall 101-class accuracy, the top-3 weighted (TWA) and unweighted (TUA) accuracy, is presented on the validation and test set, all together visualized in Table 6. Both FIN-IDENTIFY models show similar validation and test metrics, which thus provide no evidence of overfitting. Although both datasets (KWID11-17 and KWIDE11-17) are not comparable due to different splits and distributions, the additional machine-annotated images of the 100 most common individuals result in a significant improvement in model performance, generalization, and transferability. For all pending unseen classification events, FIN-IDENTIFY trained/evaluated on KWIDE11-17, was applied. Moreover, such consistently promising multi-class classification results prove feasibility and quality of the entire FIN-PRINT pipeline (see Fig. 1).

FIN-PRINT—Unseen Year 2018. To further verify performance and generalization, the entire FIN-PRINT pipeline (see Fig. 1) was applied to unseen data from 2018. The best FIN-DETECT, VVI-DETECT, and FIN-IDENTIFY model was applied in a sequential order (see FIN-PRINT workflow in Fig. 1) to predict identification labels for the 100 most commonly photographed individuals, being covered by FIN-IDENTIFY. All single-labeled images in the 2018 dataset, which include one of these 100 individuals, were automatically processed by FIN-PRINT (detection, extraction, filtering, and classification—see Fig. 1). A total of 5,768 single-labeled sub-images, each of them belonging to one of the 100 most commonly photographed animals, were detected and extracted applying FIN-DETECT/-EXTRACT, while considering the previous data constraint of a single label together with exactly one bounding box. Afterwards, VVI-DETECT was applied to pre-filter the 5,768 identification images, which machine-identified 1057, either challenging, and/or unusable/invalid excerpts (see Fig. 3a–i) resulting in 4711 valid identification sub-images of the 100 most commonly photographed individuals.



Figure 6. Examples of detected (FIN-DETECT), extracted (FIN-EXTRACT), and pre-filtered (VVI-DETECT) unseen and valid killer whale identification images from 2018 which were successfully classified by FIN-IDENTIFY, trained on KWIDE11-17.

On average, each animal occurred 47.1 times, with a standard deviation of 30.2. In the 2018 dataset, *T109* was the least photographed individual with only 2 images, whereas *T100C* was the most frequently photographed with 132 identification images. Finally, FIN-IDENTIFY, trained on KWID11-17 and KWIDE11-17, was applied to predict the respective identification labels. Within a real-world scenario, one would not need to continue looking at the previously machine-filtered 1057 invalid material in case of individual classification and directly process the remaining 4711 valid samples. However, to demonstrate and prove the necessity of introducing a rejection class also at the final stage of individual classification (FIN-IDENTIFY), all 5768 unseen images were used for prediction. Furthermore, in practice, only the final version of FIN-IDENTIFY, trained on KWIDE11-17, would be applied to unseen data.

FIN-IDENTIFY, trained on KWID11-17, achieved an accuracy of 82.8%, next to a top-3 weighted and unweighted accuracy of 86.6%, as well as 91.7%, based on the 101-class task. Training FIN-IDENTIFY on KWIDE11-17, resulted in an accuracy of 84.5%, next to a top-3 weighted and unweighted accuracy of 88.1%, as well as 92.9%. Figure 6 visualizes correct classification examples of the extended classifier version for 9 individual killer whales from the unseen data from 2018.

Discussion

The current study presents a fully machine-based, multi-stage, deep-learning pipeline, named FIN-PRINT (see Fig. 1), with the aim to automatize and support the analysis of killer whale photo-identification data. Dorsal fin and saddle patch detection, the first step of FIN-PRINT, was performed via a two-stage training procedure. The initial version of FIN-DETECT achieved promising results (see Table 4), hence additional machine-annotated data was generated by applying the model to unseen data from 2011, 2015, and 2018 (see Table 1). Whereas validation and test results on the smaller HADD dataset slightly diverge, they both significantly and consistently improved while training/evaluating FIN-DETECT on the machine-extended EADD (see Table 4). However, a direct comparison between both models is difficult because the volume and distribution of data were different (see Table 1). Based on the detected bounding box coordinates, equally-sized 512×512 RGB-sub-images were extracted and if necessary interpolated or compressed (no zero-padding), using FIN-EXTRACT, the second step of FIN-PRINT. However, the quality of detected and extracted sub-images is not solely dependent on the performance of FIN-DETECT, but also on the original image content and quality (see Figs. 3 and 5).

Most of these images contain dorsal fins, leading to correct identifications by FIN-DETECT, however they are useless for downstream individual classification. Besides these cases, images of other body parts, such as tail flukes, pectoral flippers, or other triangular structures (e.g. head of a killer whale), often exist. Such false detections do have strong similarities, hence making them difficult to avoid. Consequently, it is imperative to conduct a data enhancement procedure to filter such invalid identification images beforehand. For this reason, VVI-DETECT, the third step of FIN-PRINT, was trained and evaluated on the manually labeled VIKWID11-17 (see Table 2). Binary classification metrics of VVI-DETECT on the unseen test set (see Table 5) provide no indication of overfitting. In addition, several examples of invalid pre-detected/-extracted identification images, correctly identified by VVI-DETECT, are visualized in Fig. 5, representing all the challenging situations previously described in Fig. 3a–i and clearly proving the enormous importance of such a preliminary data enhancement procedure. The final step of FIN-PRINT—killer whale individual classification—was conducted in a two-step process, similar to FIN-DETECT. First, a preliminary version was trained and evaluated on KWID11-17 (see Table 3), showing no evidence of overfitting. The top-3 classification hypothesis (TWA/TUA) greatly improves the chance of observing the correct prediction, while simultaneously reducing the dimensionality of potentially eligible individuals by an order of magnitude (101 versus 3 classes).

The final version of FIN-IDENTIFY was trained and verified on KWIDE11-17, whereby the overall classification performance was significantly improved by the data expansion (86.7% versus 92.5%) and no sign of overfitting was observed. A 5.8% increase in accuracy results in an error reduction rate of 43.6%. Considering the difference of 2.9% regarding the top-3 unweighted accuracy (94.3% versus 97.2%) an error reduction rate of 50.9% was achieved. Due to different data volumes and distributions, results of the preliminary and final model (see Table 6) cannot be directly compared. However, the consistent improvements on validation and test are a good indication for a working FIN-PRINT pipeline.

Despite all the promising dataset-specific results, an additional real-world evaluation scenario was simulated. Identification image data are typically labeled at the end of an annual fieldwork period. While considering such a procedure, the year 2018 was disregarded, to provide FIN-PRINT with new and unseen data. Due to evaluation purposes, the number of images in 2018 was limited to only those containing the 100 most common individuals. Moreover, only single-labeled identification images, together with exactly one bounding box hypothesis, were analyzed. Contrary to the previous changing datasets, a direct comparison of the classification models is now possible. Within this real-world evaluation scenario the performance of both 101-class classifiers clearly shows a working FIN-PRINT pipeline. Furthermore, a significant performance improvement is shown in the analysis of the 2018 dataset, with respect to the dataset the classifier was trained on. An accuracy difference of 1.7% (82.8% versus 84.5%) led to an error reduction of 9.9%, whereas a TUA difference of 1.2% (91.7% versus 92.9%) resulted in an error reduction rate of 14.5%. Considering how fine details in the appearance of individuals change naturally over time, in combination with completely different environmental conditions (weather, water, background, and/or changing cameras), the results are very promising.

A one-to-one comparison with results from other machine-learning studies identifying individuals proved to be very difficult due to: (1) different species and use-cases, (2) variability in datasets (amount of data, type of annotations, labeling granularity, data distribution, etc.), (3) completely different or slightly deviating approaches, and (4) varying evaluation scales and metrics. However, to emphasize and clearly demonstrate the value of this work, FIN-PRINT was compared to the most recent studies and state-of-the-art concepts addressing detection and classification of individuals represented in image data.

Animal localization and classification (object detection) are often modeled within a single network (e.g. YOLO⁷⁴⁻⁷⁷) at the same time⁶⁷. Such an approach is not recommended for the identification of individuals belonging to a certain species, as it can cause significant reduction in the system's robustness. On the one hand, there is no possibility to filter out potential object localization errors by subsequent algorithms. On the other hand, the joint feature representations, learned for localization and classification, generally prove not to be ideal especially when distinguishing very similar objects, as is the case when recognizing individuals within a species, rather than cross-species recognition.

Recent studies also apply approaches such as posture identification^{38,40} to incorporate additional information. Moreover, alignment points (landmarks) are frequently used^{43,45,48,54,55} to adjust, orientate, and standardize images regarding their final alignment to receive homogeneous data samples and consequently counteract the scale and rotation invariance of CNNs. In case of killer whale individual identification, such concepts are not relevant. Images are taken from either the left and/or right side of the animal's body as soon as they surface to fully identify both, fin and saddle patch. These body features are often the only ones visible as well as the only ones necessary for identification (see Fig. 6). Images where the fin and saddle patch are hidden and/or not sufficiently visible because of a poor angle (see Fig. 3i and examples in Fig. 5) can not be used, even after rotation, making an alignment procedure superfluous.

Several recent methods designed for automated image identification were evaluated on considerably smaller and less complex datasets^{38,39,42,43,50,51,54,59,61,64,68,69}, shorter time series datasets^{50,59}, and data collected from geographically limited locations^{50,54,55,59}. FIN-PRINT, however, was analyzed on a large-scale dataset (roughly 121,000 images of 367 individuals), collected over 8 years within a huge territory. This introduced complexity to the dataset, which was intensified through changing killer whale markings over time.

The work of Thompson et al.⁶⁴ is to some extent a similar study, which includes several sequentially-ordered steps to automate and expedite the individual recognition of common bottlenose dolphins (*Tursiops truncatus*). It must be considered that for common bottlenose dolphins only the fin is used as identification criterion, whereas killer whales also have the saddle patch. However, the system achieved a top-ranked accuracy of 88.1%, top-10 of 93.6% and top-50% of 97.0%, evaluated on 672 images and 420 unique animals. FIN-PRINT, by comparison, achieved 97.2% top-3 accuracy on the unseen test data (7166 images, 100 animals), as well as 92.9% top-3 accuracy on the entire and unseen year 2018 (5768 images, 100 animals).

Data distribution is also very important next to the mentioned data complexity. Most of the research approaches did not have uniformly distributed image data for each individual^{42,48,55,61}, which means that some animals are observed significantly more often than others, leading to the aforementioned long-tailed distribution. Exactly the same long-tailed phenomenon can be observed in our case (see Fig. 2), which strongly affects the number of killer whales being represented within the final classification model due to a limited number of training samples. In order to address these problems, most studies either use traditional classifiers^{42,48} (e.g. SVM), which do not require such large data volumes compared to deep learning methods, but usually also provide worse classification results, or apply Deep Metric Learning^{38,40,45,46,49,51,52,61,64}, especially in combination with the triplet loss⁷¹⁻⁷³. Considering the aforementioned difficulties regarding the initial usage of the triplet loss and identification of appropriate triplets, traditional supervised classification was performed as an initial step. However, together with FIN-IDENTIFY, it is now possible to automatically generate appropriate hard and semi-hard triplets⁷³ for 100 individuals, based on the top-N classification hypothesis. Thus, robust and efficient Deep Metric Learning will be possible in the future, allowing an extension to all 367 individuals, regardless of the number of images per killer whale, which consequently also solves the previously mentioned problem regarding

the long-tailed data distribution. In addition, it is not necessary to retrain the classification system in case new animals have to be added.

Robust representation learning is essential for final classification. Hu et al.⁸⁹ introduced an impressive representation learning approach for multi-label images applying a Graph Attention Network (RRL-GAT). Results on two well-known image datasets have shown significant performance improvements compared to all current state-of-the-art methods⁸⁹. This promising approach could benefit even further from the strong limitations of potential objects/labels present in killer whale identification images, which in turn could improve the focus on interesting image regions, all of which will be the task of future research activities.

Due to the promising accuracy, together with a high performance during inference, FIN-PRINT will be the key element of an interactive web-based server/client labeling system in the future, supporting biologists during their daily work (data maintenance and analysis). In addition, it will also be possible for anyone to access and upload killer whale images worldwide via a web interface. Consequently, FIN-PRINT must be able to process images of widely varying quality (different cameras, locations, photographers, environmental conditions, etc.) as accurately as possible, making a deep learning-based quality inspection (VVI-DETECT) indispensable. Thus, FIN-PRINT facilitates efficient and robust processing of large volumes of killer whale photo-identification data. The overall classification accuracy as well as efficient response time during network inference allow FIN-PRINT to be used in conjunction with video recordings for real-time detection and classification, as well as offline evaluation of the recorded video footage.

Future work will also include artificial data enlargement to counteract the mentioned long-tailed data distribution phenomenon and accompanying data sparsity for most of the individuals in the population (see Fig. 2). For this purpose, deep learning-based algorithms in connection with 3D-modeling approaches will be examined. Besides data augmentation techniques, additional investigations will be conducted to counteract current data limitations visualized in Fig. 3. In the context of this study, photos with bad weather conditions, next to originally blurred images (see Fig. 3g,h), and/or vague examples caused through the magnification of detected and extracted distant dorsal fins (see Fig. 3f), were machine-filtered via VVI-DETECT beforehand. In future work super-resolution techniques will be investigated to recover high-resolution images based on given low-resolution photos to allow the use of such material. Zhu et al.⁹⁰ introduced an auspicious end-to-end CNN-based super-resolution network, entitled *Cross View Capture network* (CVCnet), outperforming state-of-the-art super-resolution methods. Furthermore, other data enhancement approaches, such as binary mask segmentation⁵⁵ and/or contour detection^{63,64} of incoming images will be also of essential interest in the near future. Finally, the use of contextual knowledge is also a powerful and very promising avenue for improving FIN-PRINT, since killer whales have very distinctive and well documented social patterns and structures¹⁵. Such data can be used to actively adapt posterior probabilities, which in turn reduces the dimensionality of a potential classification hypothesis.

Data availability

Data to replicate the analyses are available from Bay Cetology and Fisheries and Oceans Canada upon reasonable request. Contact details can be obtained from the corresponding author. Upon acceptance, the code for FIN-PRINT will be made publicly available at <https://github.com/ChristianBergler91>, listing all single modules with a detailed description.

Received: 2 August 2021; Accepted: 17 November 2021

Published online: 06 December 2021

References

- Jain, A. K., Ross, A. & Prabhakar, S. An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Technol. Special Issue on Image- Video-Based Biom.* **14** (2004).
- Tripathi, K. P. A comparative study of biometric technologies with reference to human interface. *Int. J. Comput. Appl.* **14**, 10–15 (2011).
- Frisch, A. J. & Hobbs, J. A. Photographic identification based on unique, polymorphic colour patterns: A novel method for tracking a marine crustacean. *J. Exp. Mar. Biol. Ecol.* **351**, 294–299 (2007).
- Hammond, P. S., Mizroch, S. A., Donovan, G. P. & Commission, I. W. *Individual Recognition of Cetaceans: Use of Photo-identification and Other Techniques to Estimate Population Parameters: Incorporating the Proceedings of the Symposium and Workshop on Individual Recognition and the Estimation of Cetacean Population Parameters*. Reports of the International Whaling Commission: Special issue (International Whaling Commission, 1990). <https://books.google.de/books?id=xMccAQAIAAJ>.
- Patton, F. J. & Campbell, P. E. Using eye and profile wrinkles to identify individual white rhinos. *Pachyderm* **50**, 84–86 (2011).
- Möcklinghoff, L., Schuchmann, K.-L. & Marques, M. I. New non-invasive photo-identification technique for free-renaging giant anteaters (*Myrmecophaga tridactyla*) facilitates urgently needed field studies. *J. Nat. Hist.* **52**, 2397–2411 (2018).
- Williams, E. R. & Thomson, B. Improving population estimates of glossy black-cockatoos (*Calyptorhynchus lathami*) using photo-identification. *Emu - Austral Ornithol.* **115**, 360–367. <https://doi.org/10.1071/MU15041> (2015).
- Marshall, A. D. & Pierce, S. J. The use and abuse of photographic identification in sharks and rays. *J. Fish Biol.* **80**, 1361–1379 (2012).
- Gore, M. A., Frey, P. H., Ormond, R. F., Allan, H. & Gilkes, G. Use of photo-identification and mark-recapture methodology to assess basking shark (*Cetorhinus maximus*) populations. *PLoS ONE* **11**, 1–22 <https://doi.org/10.1371/journal.pone.0150160> (2016).
- Perera, A. & Perez-Mellado, V. Photographic identification as a non invasive marking technique for Lacertid lizard. *Herpetol. Rev.* **35**, 349–350 (2004).
- Schofield, G., Katselidis, K. A., Dimopoulos, P. & Pantis, J. D. Investigating the viability of photo-identification as an objective tool to study endangered sea turtle populations. *J. Exp. Mar. Biol. Ecol.* **360**, 103–108 (2008).
- Gamble, L., Ravela, S. & McGarigal, K. Multi-scale features for identifying individuals in large biological databases: An application of pattern recognition technology to the marbled salamander *Ambystoma opacum*. *J. Appl. Ecol.* **45**, 170–180 (2008).
- Zaffaroni Caorsi, V., Santos, R. & Grant, T. Clip or Snap? An evaluation of toe-clipping and photo-identification methods for identifying individual southern red-bellied toads, *Melanophryniscus cambaraensis*. *South Am. J. Herpetol.* **7**, 79–84 (2012).
- Bigg, M. A. An assessment of killer whale (*Orcinus orca*) stocks off Vancouver Island, British Columbia. *Rep. Int. Whal. Comm.* **32**, 655–666 (1982).

15. Towers, J. R. *et al.* Photo-identification catalogue, population status, and distribution of bigg's killer whales known from coastal waters of British Columbia, Canada. *Can. Tech. Rep. Fish. Aquat. Sci.* **3311**, vi + 299 p (2019).
16. Ford, J. K. B. & Ellis, G. M. You are what you eat: Foraging specializations and their influence on the social organization and behaviour of killer whales. In Yamagiwa, J. & Karczmarski, L. (eds.) *Primates and Cetaceans: Field Research and Conservation of Complex Mammalian Societies*, 75–98 (Springer, 2014).
17. Towers, J. R., Ford, J. K. B. & Ellis, G. M. Digital photo-identification dataset management and analysis: Testing protocols using a commercially available application. *Can. Tech. Rep. Fish. Aquat. Sci.* **2978**, iv + 16 p (2012).
18. Mizroch, S. A., Beard, J. A. & M., L. Computer assisted photo-identification of humpback whales (*Megaptera novaeangliae*). *Rep. Int. Whal. Commn. Special Issue* **12**, 63–70 (1990).
19. Whitehead, H. Computer assisted individual identification of sperm whale flukes. *Rep. Int. Whal. Commn. Special Issue* **12** (1990).
20. Adams, J. D., Speakman, T., Zolman, E. & Schwacke, L. H. Automating image matching, cataloging, and analysis for photo-identification research. *Aquat. Mamm.* **32**, 374–384 (2006).
21. Araabi, B. N., Kehtarnavaz, N., McKinney, T., Hillman, G. & Würsig, B. A string matching computer-assisted system for dolphin photo-identification. *Ann. Biomed. Eng.* **28**, 1269–1279 (2000).
22. Hiby, L. & Lovell, P. A note on an automated system for matching the callosity patterns on aerial photographs of southern right whales. *J. Cetacean. Res. Manage. (special issue)* **2**, 291–295 (2001).
23. Gope, C., Kehtarnavaz, N., Hillman, G. & Würsig, B. An affine invariant curve matching method for photo-identification of marine mammals. *Pattern Recogn.* **38**, 125–132 (2005).
24. Kniest, E., Burns, D. & Harrison, P. Fluke Matcher: A computer-aided matching system for humpback whale (*Megaptera novaeangliae*) flukes. *Mar. Mam. Sci.* **26**, 744–756 (2010).
25. Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N. & Ramirez, G. Optimal deep learning model for classification of lung cancer on ct images. *Futur. Gener. Comput. Syst.* **92**, 374–382 (2019).
26. Frid-Adar, M. *et al.* GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321**, 321–331 (2018).
27. Falcini, F., Lami, G. & Costanza, A. M. Deep learning in automotive software. *IEEE Softw.* **34**, 56–63 (2017).
28. Chen, X., Ma, H., Wan, J., Li, B. & Xia, T. Multi-view 3d object detection network for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1907–1915 (2017).
29. Patel, K. *et al.* Deep learning-based object classification on automotive radar spectra. In *2019 IEEE Radar Conference (RadarConf)*, 1–6 (IEEE, 2019).
30. Zhu, Z. *et al.* Traffic-sign detection and classification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2110–2118 (2016).
31. Covington, P., Adams, J. & Sargin, E. Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198 (2016).
32. Silver, D. *et al.* A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science* **362**, 1140–1144 (2018).
33. Mahmud, T., Rahman, M. A. & Fattah, S. A. CovXNet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization. *Comput. Biol. Med.* **122**, 103869 (2020).
34. Hilbert, M. & López, P. The world's technological capacity to store, communicate, and compute information. *Science* **332**, 60–65 (2011).
35. Sood, D., Kour, H. & Kumar, S. Survey of computing technologies: Distributed, utility, cluster, grid and cloud computing. *JNCET* **6**, 99–102 (2016).
36. Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (O'Reilly Media, 2019), 2 edn.
37. Owens, J. D. *et al.* GPU computing. *Proc. IEEE* **96**, 879–899 (2008).
38. Liu, C., Zhang, R. & Guo, L. Part-pose guided amur tiger re-identification. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, (2019).
39. Shukla, A. *et al.* A hybrid approach to tiger re-identification. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, (2019).
40. Li, S., Li, J., Tang, H., Qian, R. & Lin, W. ATRW: A benchmark for amur tiger re-identification in the wild. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, 2590–2598 (Association for Computing Machinery, New York, NY, USA, 2020). <https://doi.org/10.1145/3394171.3413569>.
41. CVWC2019. The 2019 Computer Vision for Wild life Conservation Challenge (CVWC2019). <https://cvwc2019.github.io/challenge.html> (January 2021).
42. Körschens, M., Barz, B. & Denzler, J. Towards automatic identification of elephants in the wild. arXiv preprint [arXiv:1812.04418](https://arxiv.org/abs/1812.04418) (2018).
43. Bogucki, R. *et al.* Applying deep learning to right whale photo identification. *Conserv. Biol.* **33**, 676–684 (2019).
44. Kaggle2016. Right Whale Recognition—Identify endangered right whales in aerial photographs. <https://www.kaggle.com/c/noaa-right-whale-recognition> (January 2021).
45. Simões, H. & Meidanis, J. Humpback whale identification challenge: A comparative analysis of the top solutions. (2020). <https://www.ic.unicamp.br/~meidanis/PUB/IC/2019-Simoes/HWIC.pdf>. Accessed Nov 2021.
46. Wang, W., Solovyev, R. A., Stempkovsky, A. L., Telpukhov, D. V. & Volkov, A. A. Method for whale re-identification based on siamese nets and adversarial training. *Opt. Memory Neural Netw.* **29**, 118–132 (2020).
47. Kaggle2019. Humpback whale identification—Can you identify a whale by its tail? <https://www.kaggle.com/c/humpback-whale-identification/overview> (January 2021).
48. Clapham, M., Miller, E., Nguyen, M. & Darimont, C. T. Automated facial recognition for wildlife that lack unique markings: A deep learning approach for brown bears. *Ecol. Evol.* **10**, 12883–12892 (2020).
49. Miele, V. *et al.* Revisiting animal photo-identification using deep metric learning and network analysis. *Methods Ecol. Evol.* (submitted) (2021).
50. Hansen, M. *et al.* Towards on-farm pig face recognition using convolutional neural networks. *Comput. Ind.* **98**, 145–152 (2018).
51. Moskvyak, O., Maire, F., Armstrong, A., Dayoub, F. & Baktash, M. Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. [arXiv:1902.10847](https://arxiv.org/abs/1902.10847) (2019).
52. Bouman, S., Pawley, M., Hupman, K. & Gilman, A. Individual common dolphin identification via metric embedding learning. 1–6 (2018).
53. Schofield, D. *et al.* Chimpanzee face recognition from videos in the wild using deep learning. *Sci. Adv.* **5**, eaaw0736 <https://doi.org/10.1126/sciadv.aaw0736> (2019).
54. He, Q. *et al.* *Distinguishing Individual Red Pandas from Their Faces*, 714–724 (2019).
55. Chen, P. *et al.* A study on giant panda recognition based on images of a large proportion of captive pandas. *Ecol. Evol.* **10**, 3561–3573. <https://doi.org/10.1002/ece3.6152> (2020).
56. Ferreira, A. *et al.* Deep learning-based methods for individual recognition in small birds. *Methods Ecol. Evol.* **11**, 1072–1085. <https://doi.org/10.1111/2041-210x.13436> (2020).

57. Brust, C.-A. *et al.* Towards automated visual monitoring of individual gorillas in the wild. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2820–2830 (2017).
58. Shukla, A., Cheema, G. S., Anand, S., Qureshi, Q. & Jhala, Y. Primate face identification in the wild. In Nayak, A. C. & Sharma, A. (eds.) *PRICAI 2019: Trends in Artificial Intelligence*, 387–401 (Springer International Publishing, Cham, 2019).
59. Li, Z., Shen, S., Ge, C. & Li, X. Cow individual identification based on convolutional neural network. 1–5 (2018).
60. Peng, J. *et al.* Wild animal survey using UAS imagery and deep learning: Modified faster r-cnn for kiang detection in tibetan plateau. *ISPRS J. Photogramm. Remote Sens.* 364–376 (2020).
61. Van Zyl, T. L., Woolway, M. & Engelbrecht, B. Unique animal identification using deep transfer learning for data fusion in siamese networks. In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 1–6 (2020).
62. Dunbar, S. G. *et al.* HotSpotter: Using a computer-driven photo-id application to identify sea turtles. *J. Exp. Mar. Biol. Ecol.* 535, 151490 (2021).
63. Ramos-Arredondo, R. *et al.* PhotoId-Whale: Blue whale dorsal fin classification for mobile devices. *PLoS ONE* 15, 1–19. <https://doi.org/10.1371/journal.pone.0237570> (2020).
64. Thompson, J. W. *et al.* finFindR: Computer-assisted recognition and identification of bottlenose dolphin photos in r. *bioRxiv* (2019).
65. Lopez-Vazquez, V. *et al.* Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. *Sensors* 20 (2020). <https://www.mdpi.com/1424-8220/20/3/726>.
66. Parham, J. *et al.* An animal detection pipeline for identification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1075–1083 (2018).
67. Rosli, M. S. A. B., Isa, I. S., Maruzuki, M. I. F., Sulaiman, S. N. & Ahmad, I. Underwater animal detection using YOLOV4. In *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 158–163 (2021).
68. Al-Araj, R. S. A., Abed, S. K., Al-Ghoul, A. N. & Abu-Naser, S. S. Classification of animal species using neural network. *Int. J. Acad. Eng. Res. (IJAER)* 4, 23–31 (2020).
69. Jamil, S. *et al.* Deep learning and computer vision-based a novel framework for himalayan bear, marco polo sheep and snow leopard detection. In *2020 International Conference on Information Science and Communication Technology (ICISCT)*, 1–6 (2020).
70. Moallem, G., Pathirage, D., Reznick, J., Gallagher, J. & Sari-Sarraf, H. An explainable deep vision system for animal classification and detection in trail-camera images with automatic post-deployment retraining. *Knowl.-Based Syst.* 216, 106815 (2021).
71. Hoffer, E. & Ailon, N. Deep metric learning using triplet network (2014).
72. Dong, X. & Shen, J. *Triplet Loss in Siamese Network for Object Tracking: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part XIII*, 472–488 (2018).
73. Schroff, F., Kalenichenko, D. & Philbin, J. FaceNet: A unified embedding for face recognition and clustering. 815–823 (2015).
74. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788 (2016).
75. Redmon, J. & Farhadi, A. YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7263–7271 (2017).
76. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018).
77. Linder-Norèn, E. PyTorch-YOLOv3 – A minimal PyTorch implementation of YOLOv3, with support for training, inference and evaluation. <https://github.com/eriklindernoren/PyTorch-YOLOv3> (January 2021).
78. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (2016).
79. Wang, Y.-X., Ramanan, D. & Hebert, M. Learning to model the tail. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017). <https://proceedings.neurips.cc/paper/2017/file/147ebe637038ca50a1265abac8dea181-Paper.pdf>.
80. Tan, R. J. Breaking down mean average precision (mAP). <https://towardsdatascience.com/breaking-down-mean-average-precision-map-ae462f623a52> (2019). Last visited: 6 October, 2020.
81. Aidouni, M. E. Evaluating object detection models: Guide to performance metrics. <https://manalelaidouni.github.io/manalelaidouni.github.io/Evaluating-Object-Detection-Models-Guide-to-Performance-Metrics.html> (2019). Last visited: 6 October, (2021).
82. Jocher, G. *et al.* ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations (2021). <https://doi.org/10.5281/zenodo.4679653>.
83. Zhang, C., Chang, C. & Jamshidi, M. Bridge damage detection using single-stage detector and field inspection images. *CoRRarXiv:1812.10590* (2018).
84. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. YOLOv4: Optimal speed and accuracy of object detection (2020). [arXiv:2004.10934](https://arxiv.org/abs/2004.10934).
85. Long, X. *et al.* PP-YOLO: An effective and efficient implementation of object detector (2020). [arXiv:2007.12099](https://arxiv.org/abs/2007.12099).
86. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255 (IEEE, 2009).
87. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, 448–456 (2015).
88. Nair, V. & Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 807–814 (2010).
89. Hu, B., Guo, K., Wang, X., Zhang, J. & Zhou, D. RRL-GAT: Graph attention network-driven multi-label image robust representation learning. *IEEE Internet Things J.* 1–1. <https://doi.org/10.1109/JIOT.2021.3089180> (2021).
90. Zhu, X. *et al.* Cross view capture for stereo image super-resolution. *IEEE Trans. Multimed.* 1–1. <https://doi.org/10.1109/TMM.2021.3092571> (2021).
91. Bergler, C. GitHub Repository. <https://github.com/ChristianBergler>.

Acknowledgements

The authors would like to acknowledge funding by the German Research Council (DFG) in the project “Deep Animal Linguistic Analysis (DALA) - Decoding animal communication using a hybrid approach between bio-acoustics and machine learning” (project number MA-4898/18-1), the Fisheries and Oceans Canada Species At Risk Program and Eagle Wing Tours. The authors would also like to thank “The Paul G. Allen Frontier’s Group” for their initial grant for the pilot research, Elysanne Durand for logistical support and the many data contributors for making this analysis possible.

Author contributions

C.B., E.N., and J.R.T. devised the FIN-PRINT project, the main conceptual ideas and proof outline. C.B., A.G., and E.N. worked out all the technical details, and performed the numerical calculations, implementations and verifications for the proposed experiments. L.B. initially performed a pilot study on killer whale dorsal fin/saddle patch detection. C.B. wrote the entire manuscript with support and feedback from A.G., E.N., J.R.T., L.B., G.J.S., T.J.H.S. and A.M. C.B., E.N., and J.R.T. helped supervise the project as well as the master thesis of A.G. Labeling

and formatting of the 2011–2018 image data was conducted by J.R.T., G.J.S and T.J.H.S. All authors discussed the results and commented on the manuscript, as well as agreed on the final version.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-02506-6>.

Correspondence and requests for materials should be addressed to C.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021