Mammalian Biology (2022) 102:915–929
https://doi.org/10.1007/s42991-021-00180-9

AUTOMATED INDIVIDUAL RECOGNITION

Advanced image recognition: a fully automated, high-accuracy photo-identification matching system for humpback whales

Ted Cheeseman[1,2] · Ken Southerland[1] · Jinmo Park[3] · Marilia Olio[1] · Kiirsten Flynn[4] · John Calambokidis[4] · Lindsey Jones[5] · Claire Garrigue[6,7] · Astrid Frisch Jordán[8] · Addison Howard[9] · Walter Reade[9] · Janet Neilson[10] · Christine Gabriele[10] · Phil Clapham[11]

Received: 3 August 2020 / Accepted: 30 August 2021 / Published online: 7 December 2021

Abstract

We describe the development and application of a new convolutional neural network-based photo-identification algorithm for individual humpback whales (*Megaptera novaeangliae*). The method uses a Densely Connected Convolutional Network (DenseNet) to extract special keypoints of an image of the ventral surface of the fluke and then a separate DenseNet trained to look for features within these keypoints. The extracted features are then compared against those of the reference set of previously known humpback whales for similarity. This offers the potential to successfully automate recognition of individuals in large photographic datasets such as in ocean basin-wide marine mammal studies. The algorithm requires minimal image pre-processing and is capable of accurate, rapid matching of fair to high-quality humpback fluke photographs. In real world testing compared to manual image matching, the algorithm reduces image management time by at least 98% and reduces error rates of missing potential matches from approximately 6–9% to 1–3%. The success of this new system permits automated comparisons to be made for the first time across photo-identification datasets with tens to hundreds of thousands of individually identified encounters, with profound implications for long-term and large population studies of the species.

Keywords Automated image recognition · Computer vision · Deep convolutional neural networks · Kaggle competition · Machine learning · Mark recapture · *Megaptera novaeangliae* · Photo-ID

Introduction

Reliable recognition of individual animals has proven to be a powerful tool in studies of behavior, ecology and population biology, and has been applied across many species

Handling editors: Leszek Karczmarski and Stephen C.Y. Chan.

This article is a contribution to the special issue on "Individual Identification and Photographic Techniques in Mammalian Ecological and Behavioural Research - Part 1: Methods and Concepts" – Editors: Leszek Karczmarski, Stephen C.Y. Chan, Daniel I. Rubenstein, Scott Y.S. Chui and Elissa Z. Cameron.

✉ Ted Cheeseman
ted@happywhale.com

1 Happywhale, Santa Cruz, CA 95060, USA

2 Marine Ecology Research Centre, Southern Cross University, Lismore, NSW 2480, Australia

3 Seoul 137-070, South Korea

4 Cascadia Research Collective, Olympia, WA 98501, USA

5 Allied Whale, College of the Atlantic, Bar Harbor, ME 04609, USA

6 UMR ENTROPIE, (IRD, Université de la Réunion, Université de la Nouvelle-Calédonie, CNRS, IFREMER, Laboratoire d'excellence-CORAIL), Nouméa Cedex, Nouvelle-Calédonie, France

7 Opération Cétacés, Nouméa, Nouvelle-Calédonie, France

8 Ecología y Conservación de Ballenas, A.C., 48325 Puerto Vallarta, Jalisco, Mexico

9 Google, Kaggle, Mountain View, CA 94046, USA

10 Glacier Bay National Park and Preserve, Gustavus, AK 99826, USA

11 Seastar Scientific, Vashon Island, WA 98070, USA

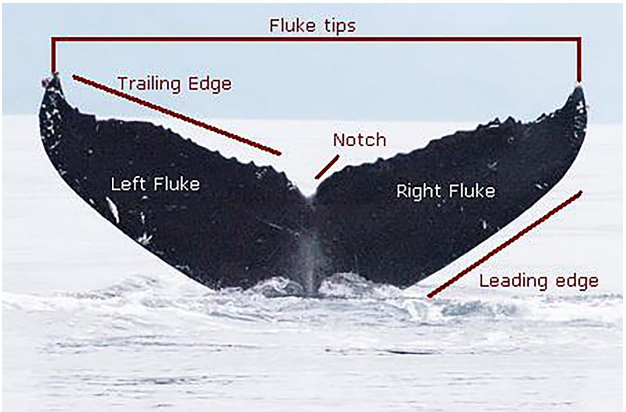

Fig. 1 Individually distinctive characteristics and landmarking features of a humpback whale fluke. Landmarking orients the image with fluke tips and the central V notch. Identification uses pigmentation, scarring and pattern on the surface of the fluke, and trailing edge shape. Used with permission from Jan Straley and Jennifer Cedarleaf

both marine and terrestrial (Hammond et al. 1990; Clutton-Brock and Sheldon 2010; Karczmarski et al. 2022a, b). Specific techniques of individual identification have improved considerably over the past decades, and in recent years, make extensive use of computer-assisted systems along with machine learning and artificial intelligence (Schneider et al. 2019; Clapham et al. 2022; Langley et al. 2022; Khan et al. 2022). As the digital technology is fast advancing, so are the improvements in the accuracy and effectiveness of automated systems for animal individual recognition.

Among great whales, humpback whales (*Megaptera novaeangliae*) were the first to be individually recognized through photo-identification methods (Schevill and Backus 1960). This was made possible by the individually distinctive and largely stable pigment pattern on the ventral surface of the flukes (Fig. 1), as well as by the unique shape of the flukes' trailing edge (Jurasz and Jurasz 1978; Katona et al. 1979). Often complemented by use of variations in the shape, size and scarring of the dorsal fin, photo-identification (photo-ID) offered a robust, economical and non-invasive alternative to lethal sampling (Hammond 1986; Franklin et al. 2020). In contrast to whaling-based studies, photo-ID allows researchers to resight individual whales over seasons, years, and decades, providing a wealth of life history data and facilitating ecological and behavioral studies, including abundance estimates, population structure and reproductive parameters. Photo-ID data used with increasingly sophisticated mark-recapture models have contributed enormously to understanding the post-whaling recovery of whale populations worldwide (Fleming and Jackson 2011). Humpback whales present a compelling subject as a species with complex migratory patterns and conservation status considered partially recovered from industrial whaling (Bettridge et al. 2015).

Photo-ID comparisons: the problem of scale

As image catalogs have grown, especially with the advent of digital photography, so has the cost of matching images across very large datasets. The landmark humpback whale studies to date, the 'Year of the North Atlantic Humpback' (YONAH, Smith et al. 1999; Stevick et al. 2003) and the 'Structure of Populations, Levels of Abundance and Status of Humpbacks' (SPLASH, Calambokidis et al. 2008) generated large datasets—4207 separate identified encounters of 2998 individual whales and 18,469 encounters of 7971 individual whales, respectively—requiring thousands of hours of manual image-matching effort. Recent studies have struggled to mount the resources necessary to gather and manage such datasets (Garrigue et al. 2011; Acevedo and Félix 2017; J. Calambokidis, unpublished). However, manual matching by pairwise image comparison has remained the primary method of dealing with photo-ID data, largely following the original methodology established by Katona and Whitehead (1981), updated with use of digital image management systems.

Pairwise image-matching effort increases multiplicatively with the growth of a photo-ID dataset; for example, a full reconciliation of the SPLASH study dataset's 18,469 images would have created over 170 million pairwise matches. An experienced photo-ID technician will recognize that the time-intensive element of the work lies in confirming if an unmatched whale is new to a dataset, rather than the false negative of a missed match. For example, for the North Atlantic Humpback Whale Catalog maintained at Allied Whale, College of the Atlantic, each new individual requires comparison by two experienced photo-ID technicians to a catalog of 10,314 known whales, requiring on average 51.5 min of effort per whale (L. Jones, unpublished). Therefore, to scale up photo-ID as a tool for the study of large populations, a clear need has arisen for efficiency improvements in image management and recognition.

The ultimate goal in this humpback whale photo-ID effort is to create a computer-based system that fully automates the matching process, including automatically scanning each digital image with no human input, and then rapidly comparing the result to animals in the existing database. The successful system should be able to rapidly and accurately match images, ideally including fair and even poor-quality images in order to maximize sample sizes from expensive fieldwork effort. Using manual matching, the SPLASH study achieved an estimated accuracy of 91% of potential matches found in a set of 18,469 good-to-excellent quality images (Barlow et al. 2011). The SPLASH study generated robust population estimates based on mark-recapture analyses while accounting for this error rate. Based on this precedent, for this current study we use the standard of at least 91% of potential matches found as a benchmark against which to judge automated photo-ID system accuracy.

The evolution of computer-assisted matching

Manual, then computer-assisted, categorization systems evolved to narrow the search effort to a subset most likely to contain the test animal, if known within the reference set. In the case of humpback whales, this has often been achieved by simply dividing flukes into five ranked categories from all white to all dark, then matching only within the same and adjacent categories (Friday et al. 2000; Calambokidis et al. 2008). Enhanced efficiency was gained through finer scale categorization of flukes within the five pigment categories; this involved division of flukes into a more detailed grid, and manual assignment of pigment, scars and other markings on the whale to the appropriate grid section. The results of the assignment were stored in a database, and a computer program could then search the data for potential matches to any newly categorized image (Mizroch et al. 1990). This system was effective for an individual catalog but did not effectively scale to large datasets, and was limited by regionally specific variations in identification features, subjectivity of data entry among individual researchers, and/or a complexity that required substantial training for matchers.

Semi-automated image recognition has demonstrated robustness and success when applied to species that show clear consistent markings, such as a constellation-mapping algorithm applied to user-annotated spot patterns on whale sharks, *Rhincodon typus* (Arzoumanian et al. 2005). However, the combination of the complex patterns and shapes used to compare humpback whale flukes has challenged fully automating image recognition for matching. A powerful computer-aided fluke identification system using a planar transformation of the flukes, edge detection and multiple digitized key features achieved matching success of 91–96% of matches found within the top five ranked matches (Kniest et al. 2010); these results were particularly impressive for working well on feature-poor primarily white flukes from South Pacific humpback whales. However, the system did not scale well to large collections, because it required specific processing of subjective fluke criteria. This included a pre-processing time of approximately 4.5 min per image required to map control points and features on each fluke.

Development of accurate and efficient feature-matching-based image recognition requiring minimal to no pre-processing of images for humpback whale flukes made significant advances with programs such as 'Hotspotter' (Crall et al. 2013) and 'Curverank' (Weideman et al. 2017). Hotspotter was capable of managing datasets of thousands of images but exhibited limited success on indistinct, poorly featured flukes, with no published accuracy assessment for humpback whales. Curverank was based on trailing edge matching and had a stated accuracy of 80%.

The Kaggle platform

The Kaggle algorithm development platform allows data scientists and other researchers to have access to a collaborative competitive environment, a cost-effective way to bring current computer science into marine mammal biology. Applied to photo-ID of North Atlantic right whales (*Eubalaena glacialis*), a competition on the Kaggle platform achieved 87% accuracy in matching the individually distinctive callosity patterns, an individually distinctive feature on the head of right whales (Kaggle 2015[1] and Bogucki et al. 2019). The Kaggle platform casts a wide net, attracting competitors for the opportunity to work on problems which otherwise may not have easy solutions, and which are integral to a specific domain or area. In a standard competition—in the case of whale photo-ID—competitors can access the complete dataset and build models on it for automated scoring on the Kaggle platform against correct matches to known individuals. During the competition period, competitors are motivated by placement on a leaderboard and a discussion forum in which score-improving techniques are shared among many competitors. Hosting a competition on the Kaggle platform offered the possibility that an efficient and implementable algorithm could be developed and trained on a larger dataset to meet our goal of being able to find 91% of potential matches in any set of test fluke photo-ID images.

We gave the competitors five well-curated catalogs of humpback whale fluke images to develop, train, and test algorithms to achieve the desired automated matching of humpback whale individuals. The success of the resulting new system permits automated comparisons to be made for the first time across a photo-identification dataset exceeding 132,000 identified separate individual humpback whale encounters of over 56,000 individual whales, which carries profound implications for long-term and large-scale studies of this species across ocean basins and populations.

Methods

The algorithm described in this paper was the result of a Google-sponsored open competition on the Kaggle platform; we review the methods of developing the necessary training dataset for the competition and then describe the competition itself, algorithm testing, and accuracy criteria used on the selected winning entry.

[1] https://www.kaggle.com/c/noaa-right-whale-recognition.

Building a training dataset

Computer vision requires large training datasets, ideally with multiple repetitions of each subject or class of interest. What mammalian biologists might consider a very large dataset—thousands of images with one to several repetitions of each individual—is generally orders of magnitude smaller than the training dataset expectations of convolutional neural network (CNN) computer vision specialists. For example, the landmark DeepFace facial recognition algorithm (Taigman et al. 2014) was trained on 4 million images of 4000 faces, an average of 1000 images per individual. Our initial aim was to generate a large enough dataset to train development of a CNN algorithm as the most promising technique for automated image recognition. To gather a dataset of many thousands of identified humpback whale fluke images, we began with publicly shareable portions of four established well-curated humpback whale fluke photo-ID catalogs: those from Cascadia Research Collective and Ecologia y Conservacion de Ballenas for North American West Coast humpback whales, from Allied Whale at College of the Atlantic for North Atlantic and Antarctic Peninsula humpback whales, and from Opération Cétacés and Institut de Recherche pour le Développement, New Caledonia, for South Pacific humpback whales. These catalogs were built over decades, and manually matched through methodology established by Katona and Whitehead (1981). This base dataset represented the full diversity of humpback whale fluke colors and patterns, including poorly featured all-black flukes from the North Pacific, poorly featured all-white flukes from the South Pacific, and Antarctic flukes with pattern-obscuring stains of diatoms that grow seasonally with exposure to extremely cold water (Figs. 2, 3).

We established a fifth catalog adding to the collected dataset by developing Happywhale.com, a web application for citizen science and research collaboration. We invited public submission of images in the style of eBird (Sullivan et al. 2009) and iNaturalist,[2] requesting full-resolution jpeg format images with associated date, location and observation information. To manage large datasets of unsorted and unprocessed images, we built two projects on the Zooniverse platform where volunteer citizen scientists selected potentially identifiable humpback whale fluke images[3] and cropped and rotated fluke images.[4]

Progressive development of algorithms

To increase the size of the training set most efficiently, we used two generations of image recognition algorithms from 2014 through 2018, in combination with manual matching. From 2014 to 2016, we used a custom-developed modified scale-invariant feature transformation (SIFT) feature recognition algorithm to assist manual matching (Town, Van Oast, Southerland and Cheeseman, unpublished). The algorithm correctly found 36% of potential matches averaged across all pigment categories of flukes, but only 14% of potential matches on poorly featured all-black flukes. This reduced manual matching effort, but was biased by unequal detection probability and gave little confidence to determine if a whale with no match found was indeed new to the reference set. All matches were manually confirmed by data managers.

From 2016 through 2018, we used an open sourced pre-publication version of the Hotspotter algorithm (Crall et al. 2013), with feature matching capable of managing a dataset with over 20,000 reference images. We found this was successful in finding 70–90% of potential matches with flukes containing distinct features; however, accuracy dropped to 20–55% of potential matches with poorly featured all-black North Pacific flukes. For poorly featured all-white flukes from the eastern coast of Australia and Oceania, accuracy was below 20%. All matches found were manually confirmed by data managers. This algorithm contributed to our workflow of managing image submissions from citizen scientists and undertaking initial catalog comparisons between research collaborators. However, we recognized a bias towards distinctive flukes and found the results insufficiently accurate overall, not meeting our benchmark goal (established from the results of the SPLASH project) of 91% of potential matches.

Throughout this stage of dataset development, whales with no matches found by the algorithms were subsequently searched for manually. Matching was conducted among known whales within the same ocean basin; if no match was found for an individual, a new ID was assigned. By late 2018 we had gathered 33,321 globally distributed individually identified encounters of approximately 15,500 individuals as a reference set for algorithm training (Table 1).

The Kaggle competition

Competition dataset

From November 2018 to March 2019, the Kaggle platform hosted an open competition based on 33,321 identified fluke-ID images combined from our five catalogs[5]. Images

[2] https://www.inaturalist.org.

[3] https://www.zooniverse.org/projects/tedcheese/snapshots-at-sea.

[4] https://www.zooniverse.org/projects/tedcheese/whales-as-individuals/.

[5] https://www.kaggle.com/c/humpback-whale-identification.

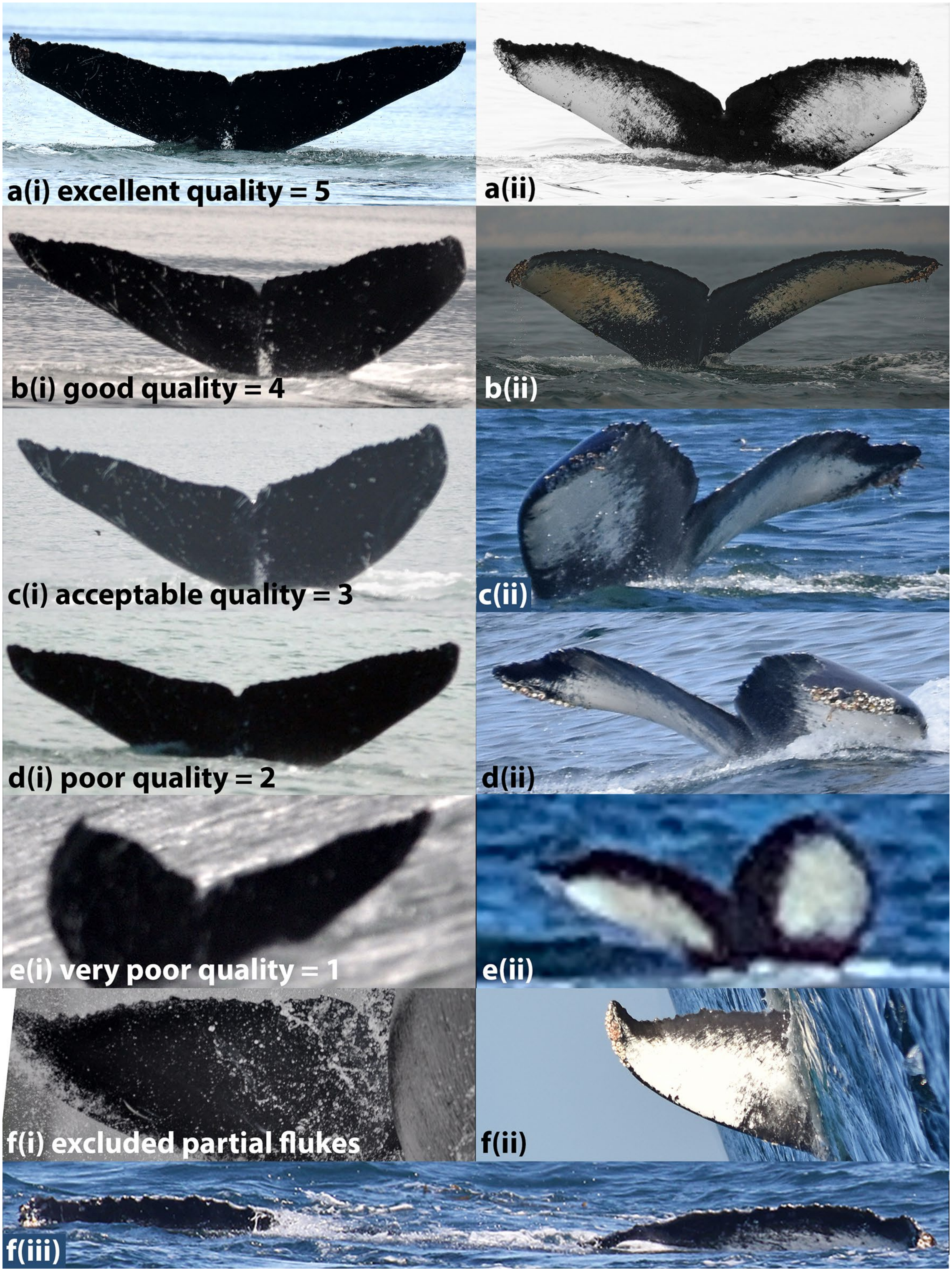

Fig. 2 Image quality scoring in descending order from excellent (5) to very poor quality (1). Images compare (i) the minimally featured all black fluke of SEAK-2476 with (ii) the well-featured fluke "Fran" CRC-12049. (f) Partial flukes were mostly excluded from training and testing data, with the exception of Allied Whale effort and accuracy test. Credits: **a–f(i)** Glacier Bay National Park Humpback Whale Monitoring Program, **a(ii)** Kate Cummings, **b(ii)** Slater Moore, **c**, **d**, **f(ii)** Marilia Olio, **e(ii)** Sean Aucoin, **f(iii)** Jodi Frediani

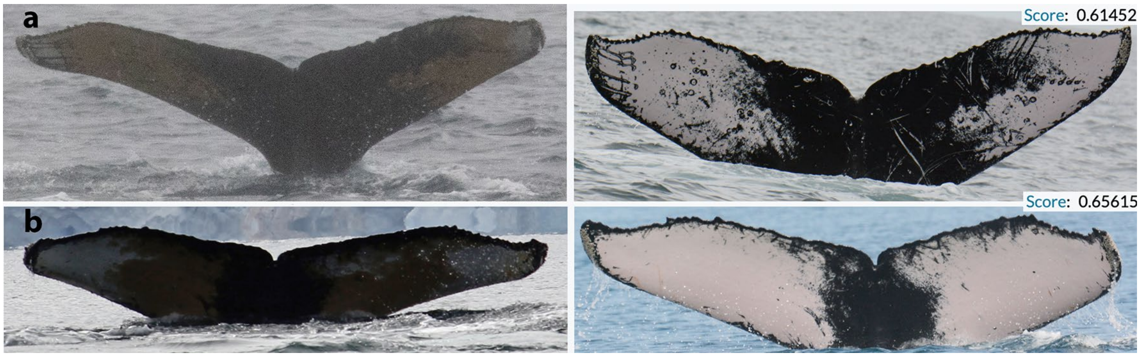

Fig. 3 A growth of diatoms obscuring coloration pattern on the ventral side of the fluke. Acceptable quality test images from Antarctica with substantial accumulation of pattern-obscuring diatom growth, matched to clean flukes. Credits: **a** left: Steph and Oli Prince, **b** left: Madoc Troup. **a** and **b** right: Kristin Rasmussen, Panacetacea

Table 1 Geographic distribution of Kaggle competition dataset, by Distinct Population Segment (after Bettridge et al. 2015)

Distinct population segment	Individuals	Encounters
1. West Indies	2843	4078
2. Cape Verde Islands/Northwest Africa	411	476
3. Hawaii	2910	7696
4. Central America	121	1129
5. Mexico	4108	14,029
6. Okinawa/Philippines	3	3
7. Second West Pacific	4	4
8. West Australia	5	5
9. East Australia	1372	1492
10. Oceania	2061	2432
11. Southeastern Pacific	1465	1784
12. Brazil	75	75
13. Gabon/Southwest Africa	54	54
14. Southeast Africa/Madagascar	62	64
15. Arabian Sea	0	0
Total	15,494	33,321

ranged in quality from very poor to excellent, with all images including a full or nearly full view of ventral surfaces of humpback whale flukes. Images were loosely cropped around target flukes. Images with partial flukes were minimally represented in the training set, so that most images displayed localization points of the trailing edge, notch, and—in the great majority of cases—both tail tips (Figs. 1, 2, 3). To mimic a realistic matching scenario where a whale might be seen frequently or might be previously unknown, individuals in the reference set were represented by 1–17 unique images. This forced algorithms to include "new class" as a possible answer for whales not found in the reference set, and presented a challenge of limited training data (compared to, for example, an average of 1000 images per individual in the DeepFace dataset (Taigman et al. 2014)). At the time of the competition, it was unclear how far the algorithms could be pushed given the task of recognizing thousands of individual whales with a median of only 2–3 training images per individual.

Competition process

Competitors matched 7960 test images of humpback flukes with no identities provided against 25,361 training images with identities provided. Competitors submitted sets of proposed IDs for the 7960 test images for automated scoring against actual correct match identities, with their scores displayed on a public leaderboard. Competitors were not required to submit actual algorithms, only proposed answers, with the agreement that at the competition closure, the five top-scoring teams would submit the full working code of their solutions for validation and publication in an open-sourced forum. An active discussion forum facilitated peer-to-peer exchange of development techniques; with input from a subset of the authors as competition hosts (AH, WR, KS and TC), which helped computer-science-focused competitors understand the relevant characteristics of humpback whales, such as accumulation and persistence of scars and seasonal growth in pattern-obscuring diatoms.

Algorithm selection and implementation

Upon competition close, we reviewed the top five winning algorithms for validity and ease of implementation, choosing for clearly documented coding with well-separated processes. The successfully implemented algorithm required:

1. a process to find target humpback whale flukes within the image (localization),
2. a process to individually identify humpback whale flukes against a set of pre-processed flukes of known individuals within a reference dataset, and,

3. efficient use of computer processing to manage and match a test set of images, with delivery of proposed match results and a match ranking confidence score.

The selected algorithm was built upon a modification of the ArcFace algorithm (Deng et al. 2018) and uses a Densely Connected Convolutional Network (DenseNet, Huang et al. 2016), itself a type of Convolutional Neural Network (CNN). It was written in python using PyTorch, which is a package that provides high-level functions such as tensor computation, neural network creation and training modules. We start with a DenseNet model pretrained on ImageNet (Deng et al. 2009), a publicly available database of 3.2 million images; its use simply avoids many hours of initial training, for which any images will do, and gives us a head start on training for our specific purposes. This pretrained model is then further trained on images of flukes which have been manually reviewed to record the locations of the right tip, left tip, notch and base of the fluke (Fig. 1). After training, the resulting model is able to auto-extract these special keypoints of an image of the ventral surface of the fluke.

Separately, the original pretrained model is then trained on a large set of known whales to look for features (embeddings) common to these given individuals and within the bounding box provided by the results of the first model. After successful training the model is then used to generate feature maps for all flukes in the set of known whales. These feature maps are keyed (labeled) with the known whale's ID.

When identifying new images, the extracted features (embeddings) from the model of these images are then compared against those of the reference set of previously known humpback whales via a simple matrix product of the two feature arrays. The output is a score for each whale ID it is compared against, with a higher score indicating a higher degree of similarity much like a facial recognition system. The algorithm recognition process uses fluke shape, edge pattern and surface markings to locate images in a hypersphere space where proximity becomes a measure of similarity.

We scored results for each test image on a unitless 0.0–1.0 match confidence scale, 0.0 being no correlation, 1.0 being a perfect match (the same image in test and reference sets). Through repeated trials of match results against the Kaggle platform's automated competition scoring process, we calculated a match rank score threshold that balanced errors of omission (missing potential matches) and errors of commission (proposing incorrect matches), to efficiently separate proposed correct and incorrect matches. We configured algorithm results to display either proposed correct matches only, or the five highest scoring match results regardless of match rank score. The algorithm then required further refactoring from its roots in the Kaggle competition to fit into the information architecture underlying Happywhale.com for accessibility and functionality compatible with a workflow of automated curation of image reference sets and processes for sending and receiving match attempts.

For a user-friendly interface, we developed a Node server written in Typescript that, upon uploading images to it, runs the algorithm python code against these newly uploaded images and returns the results to the user allowing visual confirmation or rejection of the findings of the algorithm.

Algorithm accuracy test

To facilitate testing, we trained the chosen algorithm to separately match against two reference datasets, (1) the complete dataset of 33,321 identified humpback whale fluke images used for the Kaggle competition for accuracy and resolution sensitivity testing, and (2) a reference set of flukes of all known North Atlantic humpback whale individuals curated by Allied Whale for manual versus automated matching comparison.

Accuracy testing

To test accuracy, we matched Glacier Bay National Park and Preserve (GBNP) novel fluke images ($n = 4670$) with known IDs from manual matching of 227 individuals to separate images within the complete Kaggle reference set, correlated by individual IDs in the Southeast Alaska humpback whale catalog.[6] 70% of flukes in this catalog are all dark, often without distinguishing pattern. Matching accuracy could be expected to vary by image quality across multiple variables, such as clarity, contrast, completeness of presentation (that is, how much of the flukes are visible), and angle (Friday et al. 2000). We combined quality variables into a single overall quality score of 1–5, representing very poor-to-excellent photo-ID image quality (Fig. 2). We measured accuracy per quality score category by the percentage of possible matches found as first ranked match, and among the top five ranked matches.

Resolution sensitivity test

Digital images from cameras currently used by field biologists and whale watching citizen scientists typically measure 6000 pixels wide or greater. However, successful image recognition with low-resolution images increases the potential sample size of usable images; this is especially valuable for sampling whales in the open ocean where closer approach is unrealistic, for sampling whales without needing to engage in disruptive approaches, and for using images from older or lower quality camera gear. To understand the impact of

[6] https://alaskahumpbacks.org/Catalog/SEAK_2012.pdf.

image resolution on match success, we selected excellent quality fluke images of 50 individuals, resampled the images to dimensions of 1600, 800, 400, 200, 100, 50 and 25 pixels wide, and tested match accuracy for all images at each resolution (Supplementary Fig. 1).

Manual versus automated matching: effort and accuracy test

To test the relative efficiency and accuracy of manual versus automated matching, we tasked experienced photo-ID technicians, and the algorithm, to match a set of 601 images of 371 individual whales against the Allied Whale reference set, recording the time required and match accuracy.

Internal accuracy test

Our five reference catalogs have relied on many thousands of hours of manual matching over decades, undertaken by a wide range of researchers. This effort has been subject to potential errors of omission and commission both from the matching effort and from clerical mistakes. We expected internal error rates in the range of 9% as found by Barlow et al. (2011) in the landmark SPLASH study. After initial accuracy assessment work, we trained the algorithm to treat each image as a separate class (as if each image was a unique whale) and thus to detect missed matches within our known set of whales, thereby reducing the error rate within reference catalogs.

Results

Kaggle competition

The Kaggle competition attracted submissions from 2129 participating teams. The top five Kaggle competition winning algorithms achieved accuracy scores of 0.9678–0.9731, implying that any of these algorithms, once implemented, should be able to find approximately 97% of potential humpback whale fluke-ID matches in a set of good to excellent quality fluke-ID images. The top 135 teams posted scores meeting or exceeding a threshold of 0.91, presenting many options for solutions that we could expect to meet or exceed our goal of 91% accuracy. Four of the top five algorithms used CNN approaches to image recognition. The fourth-place scoring algorithm was the one exception, utilizing a classic feature recognition approach applied with computing power estimated at 10–100 times that required by the CNN approaches, in order to handle extremely intensive feature recognition calculations.

We implemented the third-place winning algorithm (posted score: 0.9711) by co-author Jinmo Park. Reference set feature building with 28,902 individual humpback whales represented in 59,295 images required 25 h of processing time on a NVIDIA Tesla K80 33 MHz GPU, an average of 1.5 s per image. Once the reference set features were built, each test image was matched in an average of five seconds, based on 60 images ranging from 0.1 to 5.5 mb uploaded and batch identified in 5.25 min (1 min to upload, 4.25 min to batch identify). We sought to define a similarity measure threshold above which a proposed match would be expected to be correct. We found a threshold of 0.385 optimally minimized false positives (incorrect matches proposed) and false negatives (matches missed, therefore representing an incorrect proposal of a 'new whale' class). We used this threshold for operational efficiency and convenience to separate likely correct matches from individuals likely unrepresented in the reference dataset.

Algorithm accuracy: Glacier Bay National Park and Preserve dataset

Testing 4670 new images of known whales identified by Glacier Bay National Park and Preserve biologists, we found a correct match as the first ranked match in 99.2% of results for excellent quality images and in 84.6% for very poor-quality images (Table 2, Fig. 4). Poor-quality images showed much wider variation in match confidence scores than high-quality photos (Fig. 5).

Resolution sensitivity test

Images retained near 100% match success with reduced resolution until just 50 pixels wide (Table 3). This equates to a cropped image of under 0.75% of the frame from a Canon 5D mark IV camera. Even at 50 pixels wide, the correct match was found in 86% of test images within the top five ranked matches. At this level of resolution, manual verification of a correct match becomes very difficult to impossible (Supplementary Fig. 1).

Table 2 Accuracy assessment with Glacier Bay National Park fluke photo-ID dataset ($n = 2670$)

ID quality	n	% of matches found in first proposed result	% of matches found in top five proposed results
5-Excellent	377	99.20	99.47
4-Good	1557	98.72	99.36
3-Acceptable	1929	96.68	97.77
2-Poor	781	93.21	95.90
1-Very poor	26	84.62	88.46

Images of known whales were ranked by quality, demonstrating increased match success with image quality. The algorithm was successful in finding more than our benchmark of 91% of potential matches found in the first proposed match in all quality categories better than "very poor"

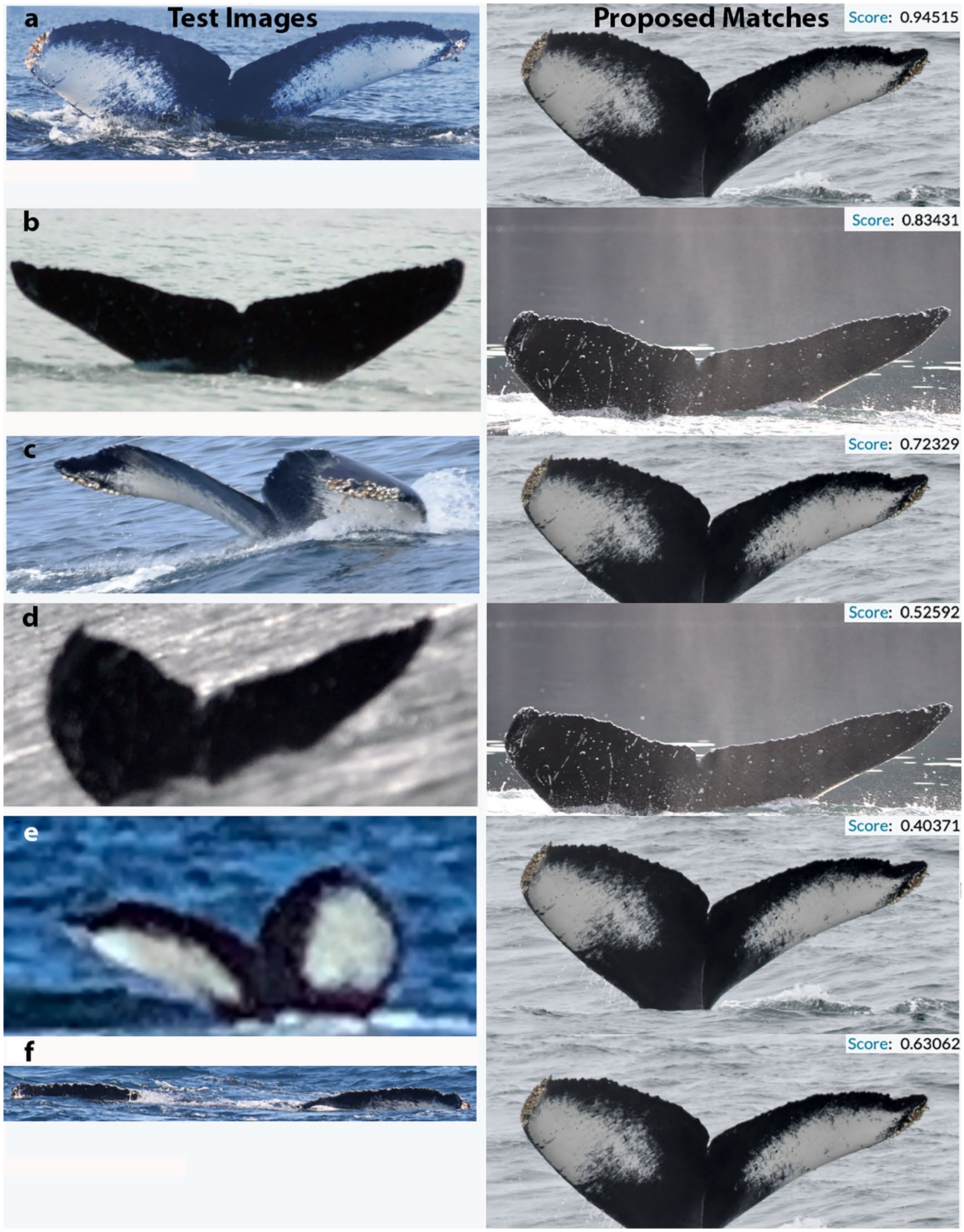

Fig. 4 Example match results by image quality. Images compare match results with a minimally featured all black fluke (SEAK-2476) and a well-featured fluke ("Fran" CRC-12049). Test image (left) has been successfully matched in all cases to reference image (right). Score displays variation in match confidence on a scale of 0 (no similarity) to 1 (exact match). **a** Excellent quality test fluke image successfully matched with high match rank score. Credit: (left) Nico Ransome and (right) Kate Cummings. **b** Poor-quality fluke with limited visibility of surface features and soft focus, successfully matched with high match rank score. Credit: Glacier Bay National Park Humpback Whale Monitoring Program. **c** Poor-quality highly-rotated fluke, successfully matched with high match rank score. Credit: (left) Marilia Olio and (right) Kate Cummings. **d**, **e** Very poor-quality low-resolution image of poorly featured fluke with soft focus, successfully matched with moderate match rank score. Manual visual confirmation is difficult to impossible at this resolution; for this image, the correct identification was confirmed by Glacier Bay National Park biologists, however without supporting confirmation such as dorsal fin matching, the match may be deemed unusable. Credit: **d** Glacier Bay National Park Humpback Whale Monitoring Program, **e** (left) Sean Aucoin and (right) Kate Cummings. **f** Very poor-quality image with fluke mostly obscured. Sufficient trailing edge is visible, successfully matched with a high match rank score. Credit: (left) Jodi Frediani and (right) Kate Cummings

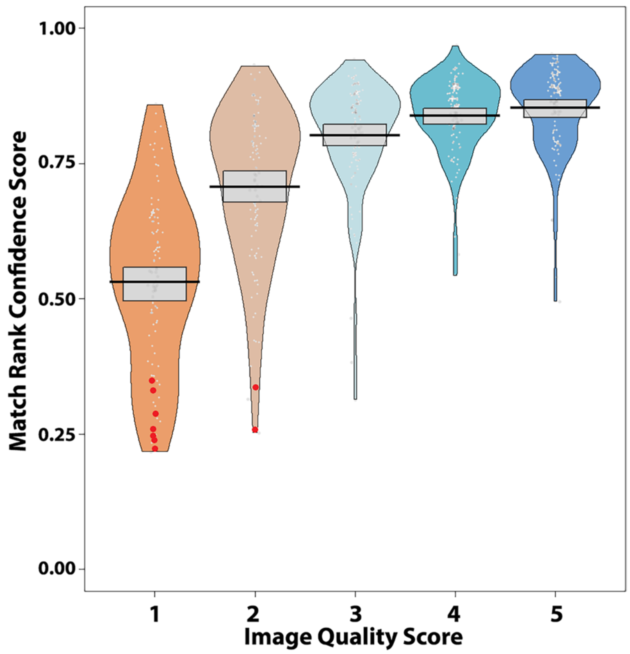

Fig. 5 Match rank confidence score as a function of image quality categories. This violin plot displays match rank confidence scores by image quality, with 100 randomly chosen Glacier Bay National Park and Preserve humpback whale fluke-ID images of each quality category, from very poor (1) to excellent (5). Poor quality images show much wider variation in match confidence scores than high quality photos. Boxes display one standard deviation from the mean confidence score of each quality category. Red dots denote match attempts that did not succeed in finding the correct match in the top five proposed results, including seven missed matches within quality score 1 images, two missed matches within quality score 2 images, and zero missed matches images of quality 3 and higher

Manual versus algorithm matching: effort and accuracy test

We compared time required and accuracy of manual versus automated matching within the North Atlantic Humpback Whale Catalog of Allied Whale at College of the Atlantic, using a test set of 594 images of very poor to excellent quality, including partial flukes, of 371 humpback whales (Table 4). Manual matching of the complete set required 318.4 total hours of experienced photo-ID technician effort, including match verification time. Automated image recognition required a total of 4.29 h, including 61.1 min of upload and system run time (6 s per image) and an average of 32.7 min of match verification time per batch of 99 images. Manual matching, including match verification time, required an average of 51.5 min of effort per whale with an average of 67.6 min per whale for new individuals and an average of 22.6 min when the whale was successfully matched. The minimum time expended for manually matching was 27 s. On average, the automated matching required 1.35% of the time of manual matching, excluding image management time from camera to match-ready image. Both manual and automated matching committed errors of missing matches; error rates were 6.7% for manual matching compared to 1.7% for automated matching methods. The algorithm did not successfully find the correct match for four whales (1.08% of whales in test set) which had been successfully matched using manual methods. Images of these four whales were of poorer quality or partial flukes.

Table 3 Resolution sensitivity assessment

Resolution in pixels ($n=50$)	% of original	ID quality score	1st proposed correct match (%)	Top 5 proposed correct match (%)
Full	100	5	100	100
1600	24	5	100	100
800	12	4	100	100
400	6	3	100	100
200	3	2	100	100
100	1.5	2	100	100
50	0.75	1	78	86
25	0.37	1	0	0

Resolution of 50 excellent quality images were successively reduced, resampled to dimensions of 1600, 800, 400, 200, 100, 50 and 25 pixels wide, and matched at each resolution. Images successfully matched down to 100 pixels wide, under 2% of the frame of most current digital cameras

Table 4 Manual vs. automated matching accuracy

	# Images	# Whales
Manual errors	40	30
Manual error rate	**6.66%**	**8.09%**
Automated errors 1st match	5	5
Automated error rate	**0.83%**	**1.35%**
Automated errors top five matches	4	4
Automated error rate	**0.67%**	**1.08%**
Totals	601	371

Manual matching of 601 images by experienced photo-ID technicians at Allied Whale, College of the Atlantic resulted in a rate of 6.66% of potential matches missed, compared to 0.83% of potential matches missed with automated matching

Internal accuracy test

Upon training the algorithm to treat each fluke-ID image as a separate individual, we found an 8% overall internal duplicate rate in our set of known whales across the entire reference set in the Happywhale dataset, approximately equivalent to the error rate found in the SPLASH study (Barlow et al. 2011). This dataset integrated catalogs from Cascadia Research Collective, Allied Whale and

Opération Cétacés—Institut de Recherche pour le Développement, New Caledonia where many of the images had been reviewed for decades without discovery of the internal missed matches. Consequently, the internal matches found by the automated algorithm were disproportionately cases of difficult matches. Sources of the missed match were highly changed flukes, such as calf-to-non-calf matches (Fig. 6 and Supplementary Fig. 2), flukes with few distinct markings (Supplementary Fig. 3), flukes with patterns obscured by diatom growth (Fig. 3), and/or with poor-quality photographs (Figs. 2, 4; Supplementary Fig. 4).

Discussion

Since its inception almost five decades ago, photo identification of humpback whales has depended in large part upon manual comparisons of images. Here, we have developed and implemented CNN-based image recognition that is much faster and more accurate than traditional manual matching, reducing image identification time by over 98% and reducing error rates from approximately 6–9% to 1–3%. With this development, the time-intensive aspects of manual photo-ID matching can be considered a secondary method for most applications, allowing efficient and inexpensive scaling up of mark-recapture studies to enable better understanding of humpback whale populations across ocean basins.

Quality considerations and sources of error

As with manual matching, image quality remains an important factor with an impact on the reliability of automated matching. Friday et al. (2000) identified that quality is affected by variables of clarity, contrast, angle (i.e., rotation), and presentation. The algorithm developed and implemented here proved to be highly resilient against limitations of clarity and contrast, and consistently exceeds the human eye's ability to confirm apparent proposed matches. Extreme rotation angles do cause an increase in missed matches, though far less so than if fluke images were flattened with a scaling transformation to plot flukes onto a two-dimensional surface, as some algorithms have sought to do (Kniest et al. 2010).

Presentation issues: partial flukes

The algorithm was not trained on partial flukes, thus it often fails to correctly understand the context of what portion of flukes are presented. This is a failure of object detection and landmarking, when part of the fluke trailing edge is not visible, especially if presented without a view of the central V notch (Fig. 1). This demonstrates an important lesson about the capability of computer vision: computers are excellent at detecting patterns if trained to understand context. When object detection and landmarking fails, however, the context that is obvious to the trained human matcher is not recognized by this artificial intelligence, with resulting high error rates. We eliminated most partial flukes from the training dataset, not anticipating how successful the Kaggle competition would prove to be. Retraining the algorithm to include partial flukes with either an automated or manually augmented object detection, and landmarking that recognizes context, would allow automated recognition of what are often high-value flukes, such as a partial fluke displayed by a whale entangled in fishing gear.

Use of low-resolution images

A test in which image resolution was successively reduced (Table 3 and Supplementary Fig. 1) demonstrated that many images previously considered unusable may in fact be of value to photo-ID studies. When an image of a fluke that takes only a small portion of the frame of a modern digital camera retains sufficient resolution to find 99% of potential matches, such images can successfully contribute to marine

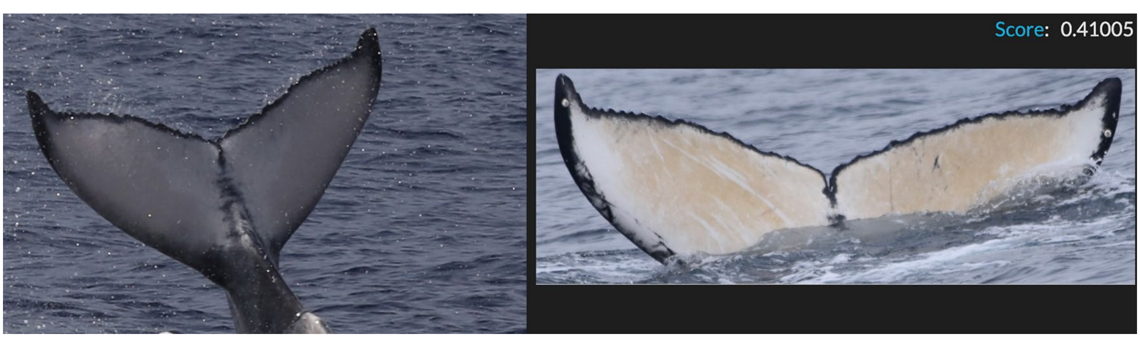

Fig. 6 Example successful match of a newborn calf to a first summer juvenile, displaying substantial change in shape, appearance of natural pigmentation, and appearance of acquired scars. The correct match is difficult to confirm visually despite good quality photos, and is found with a relatively low match confidence score. Credit: (left) Pacific Whale Foundation—Larry Hauser and (right) Sami Dean, Salish Sea Citizen Scientists

mammal science; their utility will depend upon the scientific question being addressed and the biases inherent in the inclusion of such images. Such images include those captured on the high seas at a distance, images sourced from citizen science, older datasets, and images from non-invasive platforms not approaching the subject. We have conclusively demonstrated that lower-resolution images can be used with error rates still notably lower than comparable manual matching; this is based upon a documented 9% error rate in the SPLASH study (Barlow et al. 2011) with very high image quality standards, and a 6%—8% error rate shown here in manual matching of images curated at Allied Whale. Use of low-resolution and poor-quality images may be limited by the user's ability to conclusively confirm match accuracy (e.g. Fig. 4d, e; Supplementary Fig. 1), or may require match confirmation by alternative means such as dorsal fin or multi-mark based identification (Franklin et al. 2020). Photo-ID based study design should always address quality standards with consistent methodology to define how individual ID is confirmed, estimated error rates and, as appropriate, accounting for estimated error with model correction factors as in the case of Barlow et al. (2011).

Sources of error: reference catalog quality control and change over time

In most cases when testing excellent- or good-quality images of flukes where the potential match was not found, the missed match could be attributed to a poor-quality reference image and/or a fluke pattern or shape that had changed substantially over time. Our reference catalogs were not consistently controlled for image quality apart from limited use of partial flukes. If a near-perfect match rate is desired, quality controlling the reference image set to include only good- to excellent-quality images could maintain an expected match rate that correctly identified more than 99% of potential matches.

Humpback whale flukes change over time, most substantially from calf to juvenile to adulthood (Fig. 6 and Supplementary Fig. 2a, 2b and 2c), and as well through physically disfiguring events such as ship strikes or entanglements (Blackmer et al. 2000). Nonetheless, elements of pattern and trailing edge features persist over time, even if superficially changing in appearance, such as barnacles on flukes of juvenile whales becoming barnacle scars in adults. During our internal accuracy test, we discovered multiple previously undetected matches of first summer calves (aged three to eight months) and juveniles to well-known adults where the pigment category changed enough during growth to cause the match to be missed despite repeated catalog reviews (Fig. 6 and Supplementary Fig. 2). We have noted lower match rank scores and higher error rates for calf to non-calf matches, but we do not currently have the dataset to quantify accuracy over time in this context. Newborn calves photographed on breeding grounds may appear to be very rarely matchable before at minimum approximately 3 months of age, while once on the feeding grounds during the first summer of a calf's life, some permanence of pattern and shape appears to allow automated matching. Two dynamics are confounded in detecting match success rate from calf to adults fluke-ID images, both expected different survivorship rates and change in fluke shape and pattern. In the future, we see the need to assess accuracy specifically on known series of calf to adult flukes, especially to enable more accurate estimates of calf to adult survivorship.

Equal detection probability

Mark-recapture models can be biased by unequal capture probabilities in the sampled population (see e.g. Link 2003), a problem for photo-ID studies using natural markings where some animals are more distinctly marked than others. We found no bias against poorly featured flukes, both with minimally featured all-white flukes in the South Pacific (Supplementary Fig. 3) and minimally featured all black flukes in the North Pacific (Figs. 2, 4). Furthermore, the algorithm successfully ignored seasonally changing diatom growth concentrations in cold Antarctic waters (Fig. 3).

Image quality and error detection in an evolving photo-ID workflow

Photo-ID studies with manual image matching through the techniques of Katona and Whitehead (1981) typically follow a procedure of primary and secondary review, each positive or negative identification reviewed by two experienced technicians. In this early stage of automation, we have effectively replaced the entire primary review as well as a portion of the time-intensive secondary review. We rely on the algorithm to confirm the absence of a match with an error rate of only 1–3% of potential matches missed, as compared to 6–9% missed by manual matching. In practice, this allows assignment of a new individual ID on any good- to high-quality image where no match is found.

Looking forward, a consistent methodology of quality scoring designed around algorithm strengths (such as detection despite poor resolution and/or soft focus) and weaknesses (such as partial or oddly presented flukes) applied against match confidence scores could define thresholds for automated acceptance or rejection of matches with no manual review. This could further speed large efforts at cross-catalog matching as well as reduce human error; implicit in matching with human review is an assumption of consistent methodology but standards do vary for how clear a match must be to confirm a match. At this stage of system development, however, sample sizes are not so large as to make manual review unwieldy, and there is substantial value added in maintaining a level of human observation and review of images for the sake of knowledge of the individuals of our study.

Scaling up and recent developments

Until now, manual matching has been the highest quality standard available, limiting the scale and extent of many photo-ID studies. We have developed a fast and accurate automated image recognition system surpassing the accuracy of manual matching. Manual matching can now be eliminated except in the case of partial flukes and the important quality assurance step of manually confirming all matches found. We have successfully removed the time-intensive bottleneck of manual searching to confirm if a whale is new to the collection when no match is found.

In the time since we initially developed this algorithm, we have implemented further improvements to increase the efficiency of the match verification step, reducing match time to approximately 0.1 s and functional processing time for an experienced operator to under 30 s per image to upload, match and confirm or reject match results for most images. Furthermore, we have improved reference catalog updates (the process of adding extracted features of flukes for newly identified individuals to the reference set) to be performed in approximately 0.1 s per image and to be incremental, reducing the time needed for catalog updates from greater than 24 h to several minutes. This accurate and efficient automated image recognition now allows scaling of photo-ID to long time series and large-area collaborative studies at very low cost. We have seen no saturation of algorithm accuracy despite growth of a single global-scale reference catalog.

Implications for the marine mammal research community

The intent of this study is to create an image recognition tool for humpback whale photo-ID that is broadly accessible to the marine mammal research community (see Code Availability). Our primary immediate application has been to extend the SPLASH project (Calambokidis et al. 2008) from its original 3-year synoptic study to an ongoing capacity for continuously updated knowledge of the status of the species in the North Pacific. Achieving the full benefit of this automation requires a curated, integrated ocean basin-wide reference catalog, necessitating long-term and broad research collaboration. Such collaborations can be logistically and politically difficult to develop and maintain, and can present potentially challenging questions of data accessibility, use, ownership, as well as authorship of resulting publications.

We have built a single integrated global reference catalog that as of August 2021 contained over 56,000 known individual humpback whales. This dataset is global, with particular concentrations in the North and South Pacific, Antarctic Peninsula, North and Southwest Atlantic. Interested researchers are encouraged to contact the authors to discuss system use. With future development we intend to create installable code for use in offline settings matching against regional subset image catalogs.

Technological development brings change; this scale-shifting development, allowing a 98% reduction in photo-ID matching effort, has substantially improved workflows in our associated studies. However, for research organizations that use photo-ID as an accessible but time-intensive task undertaken by entry-level personnel, automation may, by eliminating such tasks, remove an educational tool that has often been used for teaching observation skills.

Next steps

The fluke of humpback whales contains a large amount of individual-specific distinctive features, making them a strong candidate for an early case of automated image recognition. However, the Kaggle competition results suggest that we could have created a more difficult image recognition task; competing algorithm top scores were bunched against 100% accuracy (after accounting for errors later found in the reference set), implying that competitors could have developed further if the recognition task was harder. By providing mostly cropped images of whole whale flukes, for example, we minimized the need to develop around object detection and landmarking, instead choosing to focus the competition primarily on recognition. This solved the largest single problem limiting photo-ID based studies, but at the same time suggests substantial further potential in deep learning neural network computer vision. Based upon our current experience, we can envision neural network algorithms able to identify individuals in many other taxa when provided a consistent context, such as lateral views of dorsal fins and flank patterns of multiple marine mammal species.

Conclusions

The utility of photo-ID as a research technique is often limited by the time and effort required to process the photographic data obtained in the field; an effort that increases greatly as the datasets grow larger. This limitation is now eliminated in studies of humpback whales with the fully automated matching system presented here. When compared against the workflow of manual matching, the automated workflow including manual match confirmation reduced the time required for matching by at least 98%, and reduced error rates from approximately 6–9% to 1–3% depending on image quality thresholds, and to near-zero for high-quality images. Access to this algorithm and supporting information architecture is available for use at no cost via the web platform www.happywhale.com. Collaborative research groups are now able to conduct low-cost ocean-basin-wide or even global studies of humpback whales. We believe this delivers great potential to better understand short- and long-term population demographics, survival and behavior of individuals that are the basis of informing conservation on ocean-basin scales.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42991-021-00180-9.

Acknowledgements We extend special thanks and acknowledgement to all data contributors; a full list of 970 image contributors other than the authors, is given as Supplementary Information. This includes all contributors for whom we have complete attribution information whose images were included in the algorithm development competition dataset. The competition dataset would not have been possible to create without efforts from Deana Glenz, Kate Spencer, Denny Zwiefelhofer, Jenny Grayson, Lucy Payne and Tory Johnson. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. The Zooniverse data processing would not have been successful without help from Peter Mason. We also thank Jan Straley, Dan Burns, Tom Hart, Leszek Karczmarski and anonymous peer reviewers for substantial manuscript contributions.

Author contributions TC conceived the study, secured funding, collected and managed data, oversaw algorithm development, analyzed results and led manuscript preparation, KS directed and executed information architecture coding and contributed to manuscript preparation. JP authored the implemented Kaggle competition algorithm, MO managed data, KF and JC contributed to study conception and contributed image and identification data, LJ contributed image and identification data, directed testing manual versus automated matching and contributed to manuscript preparation, CG, JN and CG contributed image and identification data and contributed to manuscript preparation, AF contributed image and identification data, AH and WR directed the Kaggle competition, and PC contributed to study conception and contributed to manuscript preparation.

Funding Funding for this research was provided by a grant from Cheesemans' Ecology Safaris, San Jose, California (www.cheesemans.com).

Declarations

Conflict of interest Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and materials (data transparency) Most of the photo-ID data utilized in this research is visible via www.happywhale.com, including individual sighting histories, available for use within contributor-set licensing as specified with each image. Images used for the Kaggle competition are available via https://www.kaggle.com/c/humpback-whale-identification.

Code availability (software application or custom code) Open source repositories for the top five scoring algorithms, as posted at the completion of the Kaggle competition, are available via https://www.kaggle.com/c/humpback-whale-identification/discussion. Access to the implemented algorithm and supporting information architecture is available for use at no cost via the web platform www.happywhale.com and via the authors.

References

Acevedo J, Félix F (2017) Report of the Workshop on Southern Hemisphere blue, fin and humpback whale photo-identification catalogues from the Central and Eastern South Pacific and the Antarctic Peninsula. Rep Int Whal Commun

Arzoumanian Z, Holmberg J, Norman B (2005) An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. J Appl Ecol 42:999–1011. https://doi.org/10.1111/j.1365-2664.2005.01117.x

Barlow J, Calambokidis J, Falcone EA, Baker CS, Burdin AM, Clapham PJ, Ford JKB, Gabriele CM, LeDuc R, Mattila DK, Quinn TJ, Rojas-Bracho L, Straley JM, Taylor BL, Urbán RJ, Wade P, Weller D, Witteveen BH, Yamaguchi M (2011) Humpback whale abundance in the North Pacific estimated by photographic capture-recapture with bias correction from simulation studies. Mar Mamm Sci 27:793–818. https://doi.org/10.1111/j.1748-7692.2010.00444.x

Bettridge S, Baker CS, Barlow J, Clapham PJ, Ford M, Gouveia D, Mattila DK, Pace RMI, Rosel PE, Silber GK, Wade PR (2015) Status review of the humpback whale (Megaptera novaeangliae) under the endangered species act. NOAA-TM-NMFS-SWFSC-540. National Marine Fisheries Service, Silver Spring, MD. p 240

Blackmer AL, Anderson SK, Weinrich MT (2000) Temporal variability in features used to photo-identify humpback whales (Megaptera novaeangliae). Mar Mamm Sci 16:338–354. https://doi.org/10.1111/j.1748-7692.2000.tb00929.x

Bogucki R, Cygan M, Khan CB, Klimek M, Milczek JK, Mucha M (2019) Applying deep learning to right whale photo identification. Conserv Biol 33:676–684. https://doi.org/10.1111/cobi.13226

Calambokidis J, Falcone EA, Quinn TJ, Burdin AM, Clapham PJ, Ford JKB, Gabriele CM, LeDuc R, Mattila D, Bracho LR-, Janice M. Straley, Taylor BL, R. JU, Weller D, Witteveen BH, Yamaguchi M, Bendlin A, Camacho D, Flynn K, Havron A, Huggins J, Maloney N (2008) SPLASH: Structure of populations, levels of abundance and status of humpback whales in the North Pacific. Final report for Contract AB133F-03-RP-00078, available from Cascadia Research, Olympia, WA

Clapham M, Miller E, Nguyen M, Van Horn RC (2022) Multispecies facial detection for individual identification of wildlife: a case study across ursids. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00168-5

Clutton-Brock TH, Sheldon BC (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends Ecol Evol 25:562–573. https://doi.org/10.1016/j.tree.2010.08.002

Crall JP, Stewart C V, Berger-Wolf TY, Rubenstein DI, Sundaresan SR (2013) HotSpotter—patterned species instance recognition. In: 2013 IEEE workshop on applications of computer vision (WACV), pp 230–237

Deng J, Deng J, Dong W, Socher R, Li L, Li K, Fei-fei L (2009) Imagenet: a large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition
Deng J, Guo J, Xue N, Zafeiriou S (2018) ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: Proc IEEE comput soc conf comput vis pattern recognit 2019-June, pp 4685–4694
Fleming A, Jackson J (2011) Global review of humpback whales (*Megaptera novaeangliae*). NOAA technical memorandum. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service
Franklin T, Franklin W, Brooks L, Harrison P, Burns D, Holmberg J, Calambokidis J (2020) Photo-identification of individual humpback whales (*Megaptera novaeangliae*) using all available natural marks: Implications for misidentification and automated algorithm matching technology. J Cetacean Res Manag 21:71–83. https://doi.org/10.47536/JCRM.V21I1.186
Friday N, Smith TD, Stevick PT, Allen J (2000) Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales, Megaptera novaeangliae. Mar Mamm Sci 16:355–374. https://doi.org/10.1111/j.1748-7692.2000.tb00930.x
Garrigue C, Franklin T, Constatine R, Russell K, Burns D, Poole M, Paton D, Hauser N, Oremus M, Mattila D, Gibbs N, Franklin W, Robbins J, Clapham P, Baker C (2011) First assessment of interchange of humpback whales between Oceania and the East coast of Australia. J Cetacean Res Manag Special Issue 3:269–274
Hammond PS (1986) Estimating the size of naturally marked whale populations using capture-recapture techniques. Rep Int Whal Commun 8:253–282
Hammond PS, Mizroch SA, Donovan GP (1990) 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 parame. International Whaling Commission. Special Issue 12:448
Huang G, Liu Z, van der Maaten L, Weinberger KQ (2016) Densely connected convolutional networks. In: Proc—30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017-January, pp 2261–2269
Jurasz C, Jurasz V (1978) Results of 1977 Studies on Humpback Whales in Glacier Bay National Monument. Final Rept.
Karczmarski L, Chan SCY, Rubenstein DI, Chui SYS, Cameron EZ (2022a) Individual identification and photographic techniques in mammalian ecological and behavioural research – Part 1: Methods and concepts. Mamm Biol (Special Issue) 102(3). https://link.springer.com/journal/42991/volumes-and-issues/102-3
Karczmarski L, Chan SCY, Chui SYS, Cameron EZ (2022b) Individual identification and photographic techniques in mammalian ecological and behavioural research – Part 2: Field studies and applications. Mamm Biol (Special Issue) 102(4).https://link.springer.com/journal/42991/volumes-and-issues/102-4
Katona SK, Whitehead HP (1981) Identifying Humpback Whales using their natural markings. Polar Rec (gr Brit) 20:439–444. https://doi.org/10.1017/S003224740000365X
Katona S, Baxter B, Brazier O, Kraus S, Perkins J, Whitehead H (1979) Identification of Humpback Whales by Fluke Photographs. Behavior of marine animals. Springer, New York, pp 33–44
Khan C, Blount D, Parham J, Holmberg J, Hamilton P, Charlton C, Christiansen F, Johnston D, Rayment W, Dawson S, Vermeulen E, Rowntree V, Groch K, Levenson JJ, Bogucki R (2022) Artificial intelligence for right whale photo identification: from data science competition to worldwide collaboration. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-022-00253-3
Kniest E, Burns D, Harrison P (2010) Fluke Matcher: a computer-aided matching system for humpback whale (Megaptera novaeangliae) flukes. Mar Mamm Sci 26:744–756. https://doi.org/10.1111/j.1748-7692.2009.00368.x
Langley I, Hague E, Arso Civil M (2022) Assessing the performance of open-source, semi-automated pattern recognition software for harbour seal (*P. v. vitulina*) photo ID. Mamm Biol (Special Issue) 102(3). https://doi.org/10.1007/s42991-021-00165-8
Link WA (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics 59:1123–1130
Mizroch SA, Beard JA, Lynde M (1990) Computer assisted photo-identification of humpback whales. Rep Int Whal Commun 12:63–70
Schevill WE, Backus RH (1960) Daily Patrol of a Megaptera. J Mamm 41:279. https://doi.org/10.2307/1376380
Schneider S, Taylor GW, Linquist S, Kremer SC (2019) Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods Ecol Evol 10:461–470. https://doi.org/10.1111/2041-210X.13133
Smith TD, Allen J, Clapham PJ, Hammond PS, Katona S, Larsen F, Lien J, Mattila D, Palsboll PJ, Sigurjonsson J, Stevick PT, Oien N (1999) An ocean-basin-wide Mark-Recapture Study of the North Atlantic Humpback Whale (*Megaptera novaeangliae*). Mar Mamm Sci 15:1–32. https://doi.org/10.1111/j.1748-7692.1999.tb00779.x
Stevick P, Allen J, Clapham P, Friday N, Katona S, Larsen F, Lien J, Mattila D, Palsbøll P, Sigurjónsson J, Smith T, Øien N, Hammond P (2003) North Atlantic humpback whale abundance and rate of increase four decades after protection from whaling. Mar Ecol Prog Ser 258:263–273. https://doi.org/10.3354/meps258263
Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S (2009) eBird: a citizen-based bird observation network in the biological sciences. Biol Conserv 142:2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006
Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp 1701–1708
Weideman HJ, Jablons ZM, Holmberg J, Flynn K, Calambokidis J, Tyson RB, Allen JB, Wells RS, Hupman K, Urian K, Stewart CV (2017) Integral Curvature Representation and Matching Algorithms for Identification of Dolphins and Whales. ICCV 2017 Workshop on Visual Wildlife Monitoring. arXiv:1708.07785v1

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