Supplementary Information

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

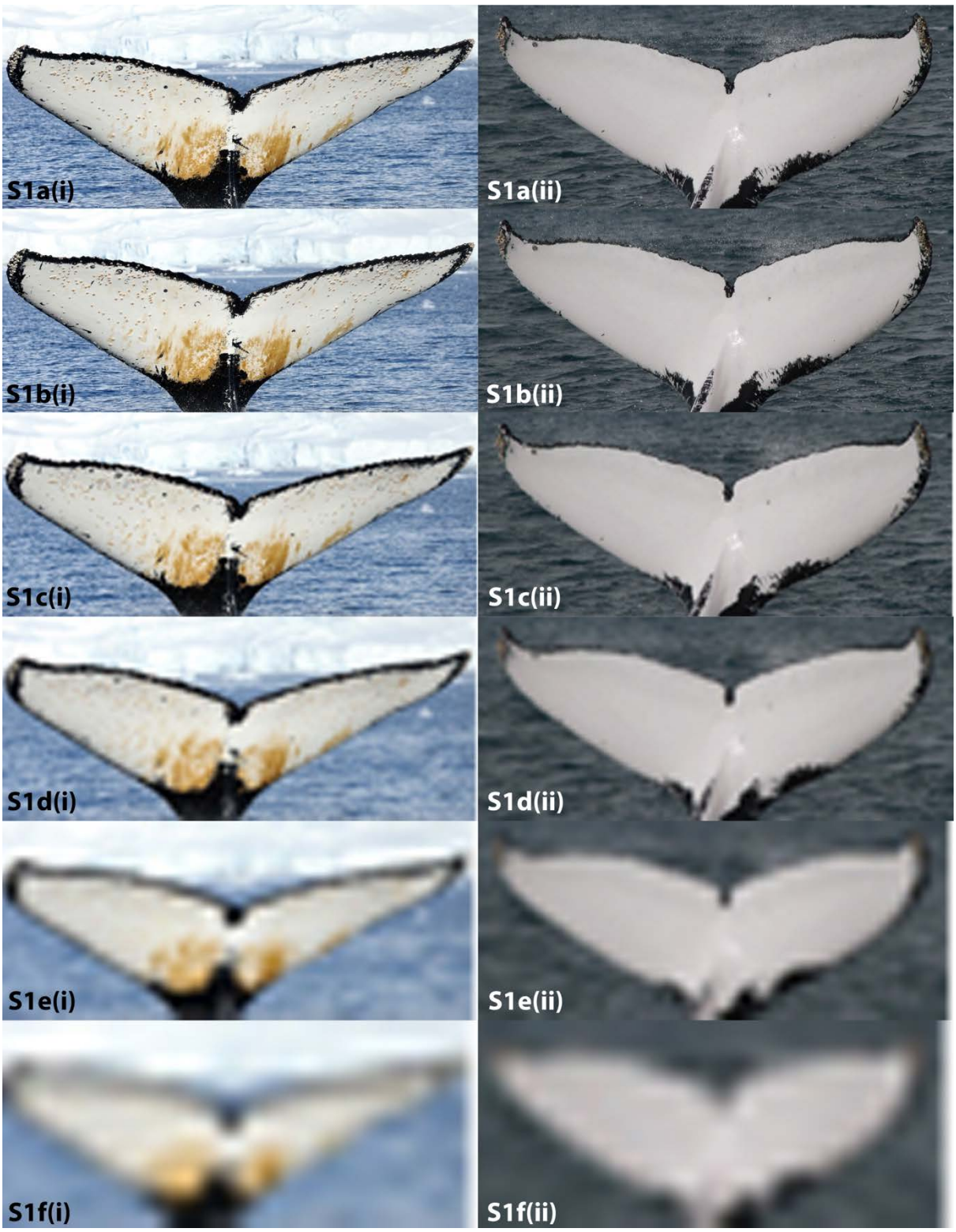
Ted Cheeseman[1,2*], Ken Southerland[1], Jinmo Park[3], Marilia Olio[1], Kiirsten Flynn[4], John Calambokidis[4], Lindsey Jones[5], Claire Garrigue[6, 7], Astrid Frisch[8], Addison Howard[9], Walter Reade[9], Janet Neilson[10], Christine Gabriele[10], Phil Clapham[11]

[1]Happywhale, Santa Cruz, California 95060, USA
[2]Marine Ecology Research Centre, Southern Cross University, Lismore, NSW, 2480, Australia
[3]Seoul, South Korea, 137-070
[4]Cascadia Research Collective, Olympia, WA 98501, USA
[5]Allied Whale, College of the Atlantic, Bar Harbor, Maine 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]Ecologia y Conservacion de Ballenas, AC., Puerto Vallarta, Jalisco 48325, Mexico
[9]Google, Kaggle, Mountain View, CA 94046, USA
[10]Glacier Bay National Park & Preserve, Gustavus, AK 99826, USA
[11]Smithsonian Institution, National Museum of Natural History, Washington, DC 20560, USA

* Corresponding author email: ted@happywhale.com

Supplementary figures

Supplementary Fig 1. Image resolution sensitivity test.
Supplementary Fig 2. Examples of difficult matches.
Supplementary Fig 3. A successful match with all white flukes.
Supplementary Fig 4. Highly rotated flukes.



Supplementary Fig 1. Image resolution sensitivity test

Supplementary Fig 1. Image resolution sensitivity test. Resolution of fifty excellent quality images were successively reduced, resampled to dimensions of 1600, 800, 400, 200, 100, 50 and 25 pixels wide, with matching attempted at each resolution. Images retained near 100% match success with reduced resolution through 100 pixels wide, with match success declining to 86% within the top five ranked matches at 50 pixels wide, and no match success at 25 pixels wide. Credit (i) Ted Cheeseman, (ii) Vicki Neville

Fig S1a(i) and (ii) Full resolution, 5052 pixels wide after cropping

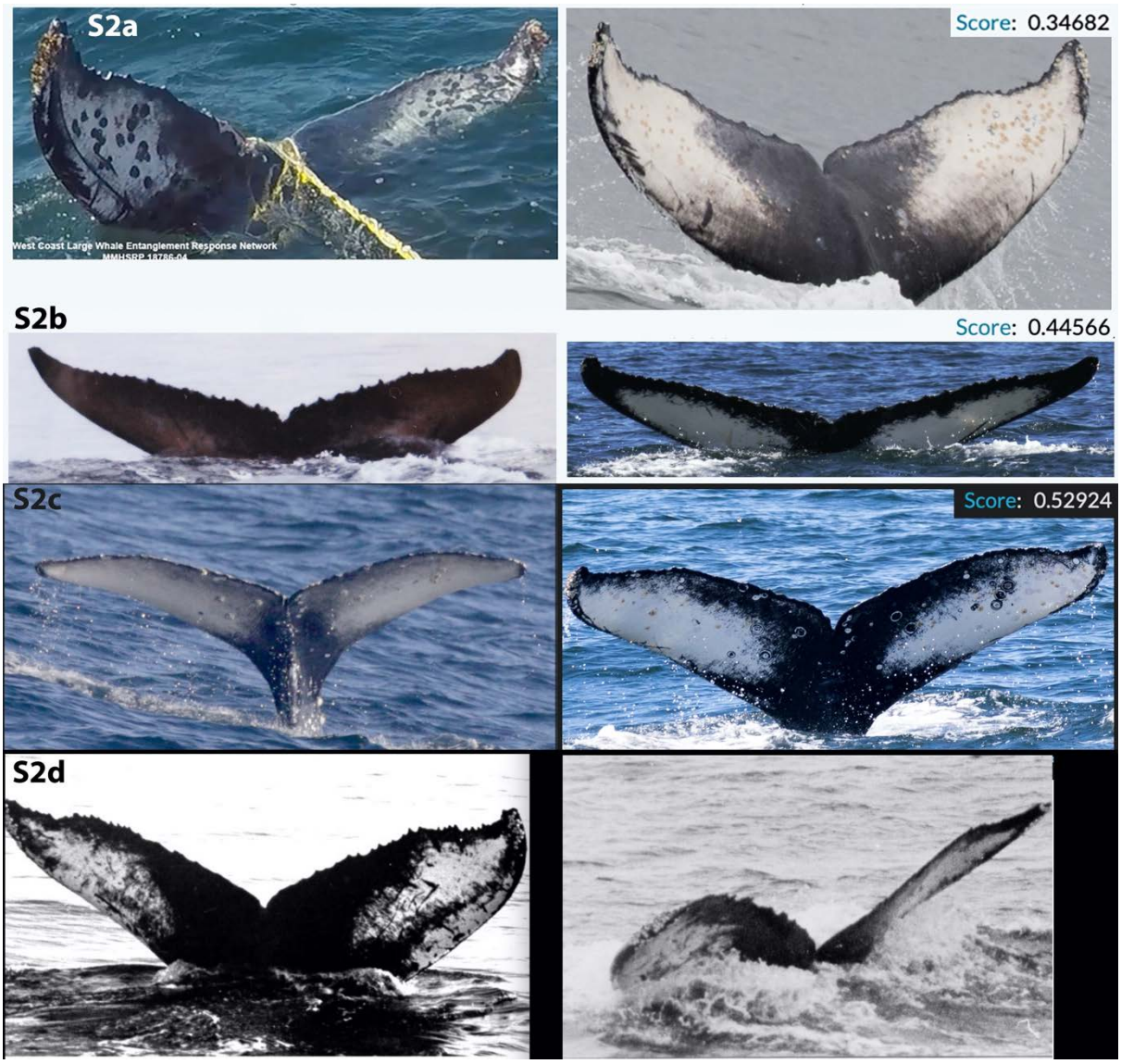
Fig S1b(i) and (ii) 1600 pixels wide, showing no visible loss of resolution

Fig S1c(i) and (ii) 200 pixels wide, showing substantial loss of resolution, with reduced ability to distinguish fine features and edge shapes

Fig S1d(i) and (ii) 100 pixels wide, showing major loss of resolution, with limited ability to distinguish all but the most distinct features and shapes

Fig S1e(i) and (ii) 50 pixels wide, showing severe loss of resolution, with very limited ability to distinguish all but the most distinct features and shapes

Fig S1f(i) and (ii) 25 pixels wide, showing total loss of resolution, where the whale tail is difficult to distinguish from the image background



Supplementary Fig 2 Examples of difficult matches. Test image (left) has been successfully matched to reference image (right).

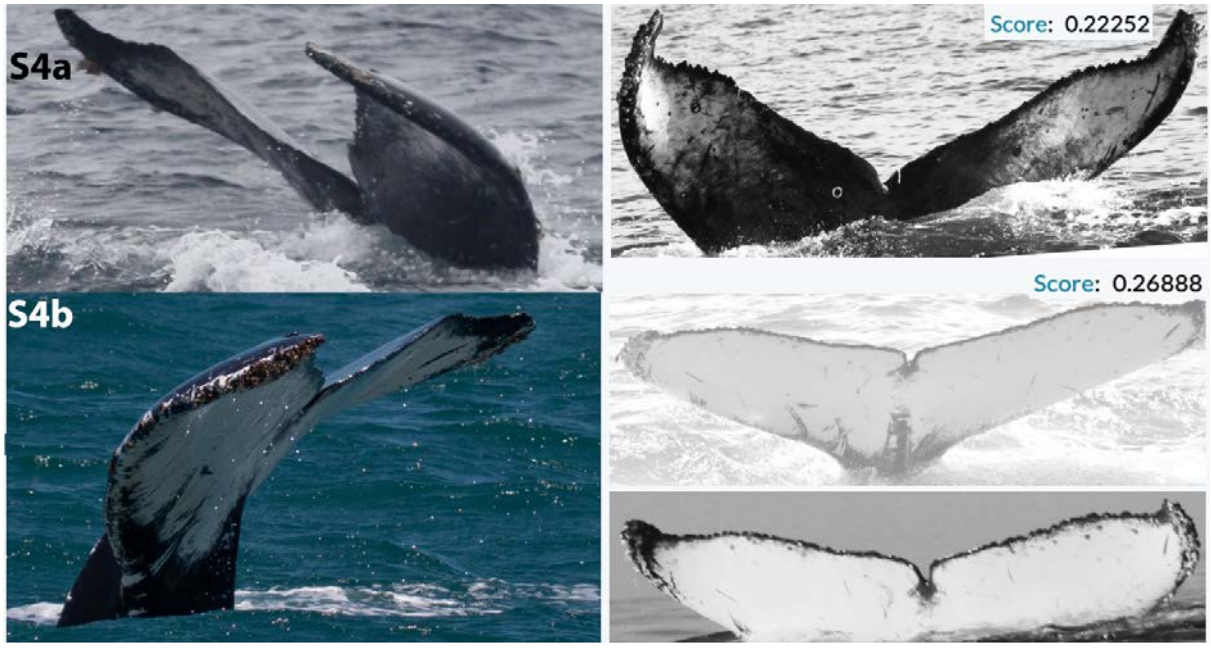
Fig S2a A successfully matched substantially changed fluke. Reference image (right) is a first summer calf in 2018, matched to the same whale as an entangled juvenile (left) in the summer of 2019. Credit: (left) West Coast Large Whale Entanglement Response Network and (right) Tim Huntington

Fig S2b A successfully matched first summer calf matched to adult. Pigmentation lightened dramatically, causing manual matching to fail, resulting in the same individual with two separate IDs. Credit: Mark Malleson

Fig S2c A successfully matched first summer calf fluke matched to juvenile. Barnacles on the calf fluke have become barnacle scars on the juvenile. Credit: Alisa Schulman-Janiger

Fig S2d A successfully matched fluke despite rotation away from the plane of the camera. This match remained undetected with two IDs in the North Atlantic Humpback Whale Catalog for over 30 years. Credit: Allied Whale, College of the Atlantic

Supplementary Fig 3. A successful match with all white flukes. Matching flukes with minimal to no pattern relies on shape, particularly using the trailing edge. In this case, 1995 (left) to 2015 (right) the trailing edge shape has become less distinct, which caused manual matching to fail where image recognition succeeded. Credit: Claire Garrigue



Supplementary Fig 4. Highly rotated flukes. Images were successfully matched as the first proposed result, however with a low match rank score. This figure illustrates an extreme where matches often fail. Credits: (a) (left) Nico Ransome, (b) (left) Dane McDermott, (a) and (b) (right) Cascadia Research Collective

Supplementary Material no. 5

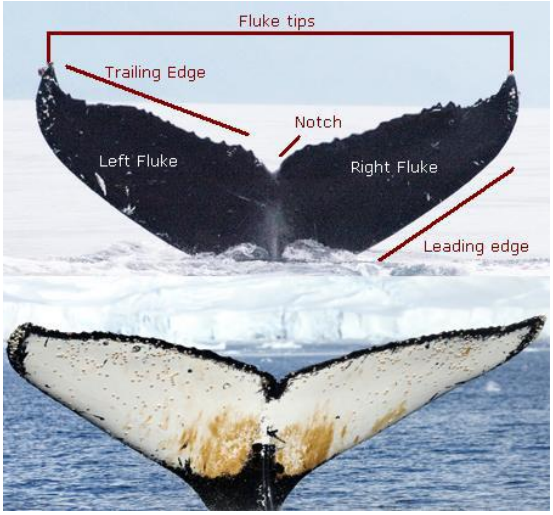
Advanced image recognition: a fully automated, high-accuracy photo-identification matching system for humpback whales

Ted Cheeseman*, Ken Southerland, Jinmo Park, Marilia Olio, Kiirsten Flynn, John Calambokidis, Lindsey Jones, Claire Garrigue, Astrid Frisch, Addison Howard, Walter Reade, Janet Neilson, Christine Gabriele, Phil Clapham

* ted@happywhale.com

Highlights:

- We developed and applied a new convolutional neural network-based photo-identification algorithm for individual humpback whales (*Megaptera novaeangliae*).
- The algorithm requires minimal image pre-processing to deliver accurate, rapid ID matching for humpback whale fluke photographs.
- Compared to manual image matching, the algorithm reduces image management time by at least 98% and reduces error rates from approximately 6-9% to 1-3%.
- The success of this new system has enabled building a single global dataset of over 56,000 individual humpback whales.



This article is part of a thematic collection of articles (Special Issue) of *Mammalian Biology* and covers the following topics and taxa (marked with ☑) addressed in the Special Issue:

Article Type

☐ Original Research ☑ Techniques ☐ Review ☐ Short Communication ☐ Concept Note

Taxon

Terrestrial

☐ Bats *(Order Chiroptera)*
☐ Carnivores : Bears *(Family Ursidae)*
☐ Carnivores : Canids *(Family Canidae)*
☐ Carnivores : Felids *(Family Felidae)*
☐ Carnivores : Hyenas *(Family Hyaenidae)*
☐ Carnivores : Mustelids *(Family Mustelidae)*
☐ Elephants *(Family Elephantidae)*
☐ Primates : Great Apes *(Family Hominidae)*
☐ Primates : Old World monkeys *(Family Cercopithecidae)*
☐ Ungulates : Bovids *(Family Bovidae)*
☐ Ungulates : Deers *(Family Cervidae)*
☐ Ungulates : Giraffes *(Family Giraffidae)*
☐ Ungulates : Horses *(Family Equidae)*
☐ Multiple taxa *(3 or more Families/Orders)*

Marine

☐ Baleen whales : Right whales *(Family Balaenidae)*
☑ Baleen whales : Rorquals *(Family Balaenopteridae)*
☐ Carnivores : Bears *(Family Ursidae)*
☐ Carnivores : Mustelids *(Family Mustelidae)*
☐ Dolphins *(Family Delphinidae)*
☐ Large toothed whales *(Families Delphinidae & Hyperoodontidae)*
☐ Pinnipeds : True seals *(Family Phocidae)*
☐ Porpoises *(Family Phocoenidae)*
☐ Sirenians : Manatees *(Family Trichechidae)*
☐ Multiple taxa *(3 or more Families/Orders)*

Topic

☐ Acoustic ID
☐ Aerial surveys
☐ Analytical innovations
☑ Automated pattern recognition
☐ Behavioural ecology
☐ Camera-trapping
☐ Conservation management
☐ Data management
☐ Demographic parameters
☐ Field methodology
☐ Genetic ID
☐ Health conditions
☐ Other: *(please specify)*
☑ Identification techniques
☐ Life-history
☐ Machine learning
☐ Mark-recapture analysis
☐ Morphometrics
☐ Network analysis
☐ Photogrammetry
☐ Population ecology
☐ Site fidelity & Movement
☐ Social ecology
☐ Software/Package development
☐ Thermal imagery

References

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. *Mammalian Biology* (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. *Mammalian Biology* (Special Issue), 102 (4) https://link.springer.com/journal/42991/volumes-and-issues/102-4