SC/68A/SH/07

Flukebook – Recent advances for cetacean photo identification and data archiving including automated fluke matching

Drew Blount, Jason Holmberg, Gianna Minton and Christin Khan

INTERNATIONAL
WHALING COMMISSION

Flukebook – Recent advances for cetacean photo identification and data archiving including automated fluke matching

Drew Blount[1], Jason Holmberg[1], Gianna Minton[2], Christin Khan[3]

1. Wild Me
2. Megaptera Marine Conservation/Arabian Sea Whale Network
3. National Oceanic and Atmospheric Administration

Abstract:

Flukebook (flukebook.org) is a non-profit, open source cetacean data archiving and photo-identification tool developed under the Wildbook platform (wildbook.org) that uses computer vision and machine learning to facilitate automated identification of individual animals in the wild. In 2016, the IWC approved funding for the development of a regional data platform for the Arabian Sea Whale Network (ASWN), to be implemented in collaboration with Wild Me, the developers of Flukebook. This collaboration resulted in expanded functionality of the Flukebook platform to allow storage and analysis of survey sightings data that do not include individual encounter or photo identification data, as well as other functions that will be of use to the global cetacean research community. Subsequent advances in machine learning research and engineering for cetacean photo-identification and broader adoption of Flukebook are now demonstrating the promise of this collaborative platform and opening new applied challenges in machine learning linked to cost and time savings in research data curation.

1. Background: Flukebook: Machine Learning, Open but Secure Science, and AI for Whales and Dolphins

Wild Me (wildme.org) actively develops the Wildbook (wildbook.org) open source platform to help scientists organize wildlife research, collect data from the public (e.g., photos and video), and integrate multi-stage, multi-modal machine learning to speed data curation. Through its web-based interface, Wildbook blends scientific collaboration, tourism, and the growing "citizen scientist" movement, bringing the concepts of broad sector inclusion to wildlife conservation while retaining a focus on researchers and conservation authorities as the primary end-user for collected and reconciled data. The application of machine learning in Wildbook can help researchers determine population sizes faster (reducing human labor) and with greater specificity (via increased data collected) and then subsequently adjust conservation action in shorter, iterative response cycles. Active Wildbook projects integrating photographic data and machine learning (especially computer vision) include whaleshark.org (*Rhincodon typus*), mantamatcher.org (*Manta birostris*), jaguar.wildbook.org (*Panthera onca*), giraffespotter.org (*Giraffa spp.*), and others. These projects have engaged over 500 researchers and over 10,000 members of the public in collaborative yet focused wildlife conservation across the globe. Figure 1a shows the broad ecosystem of the Wildbook platform and its users.

Figure 1b demonstrates the underlying “Theory of Change” for Flukebook.

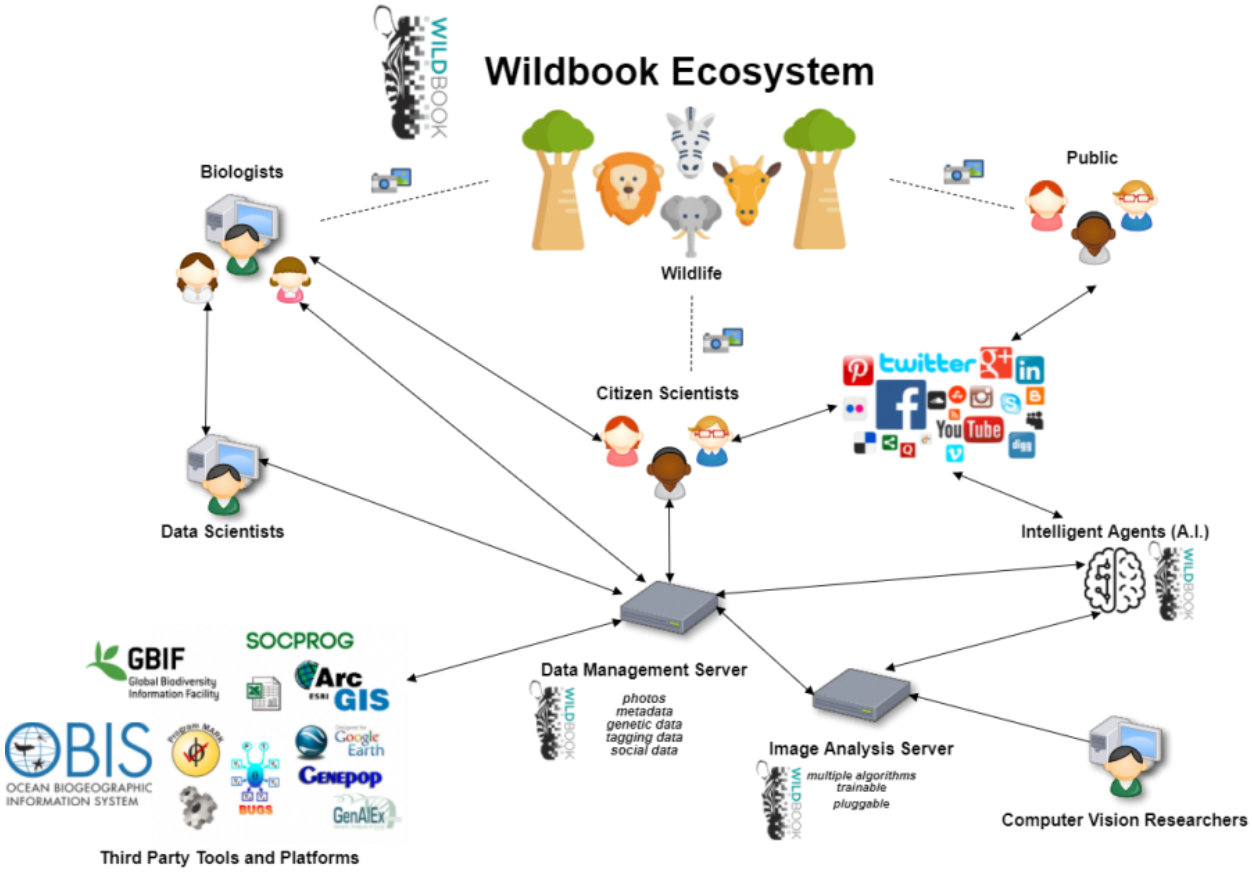

Figure 1a. The Wildbook Ecosystem of Scientists, Citizen Scientists, and A.I.

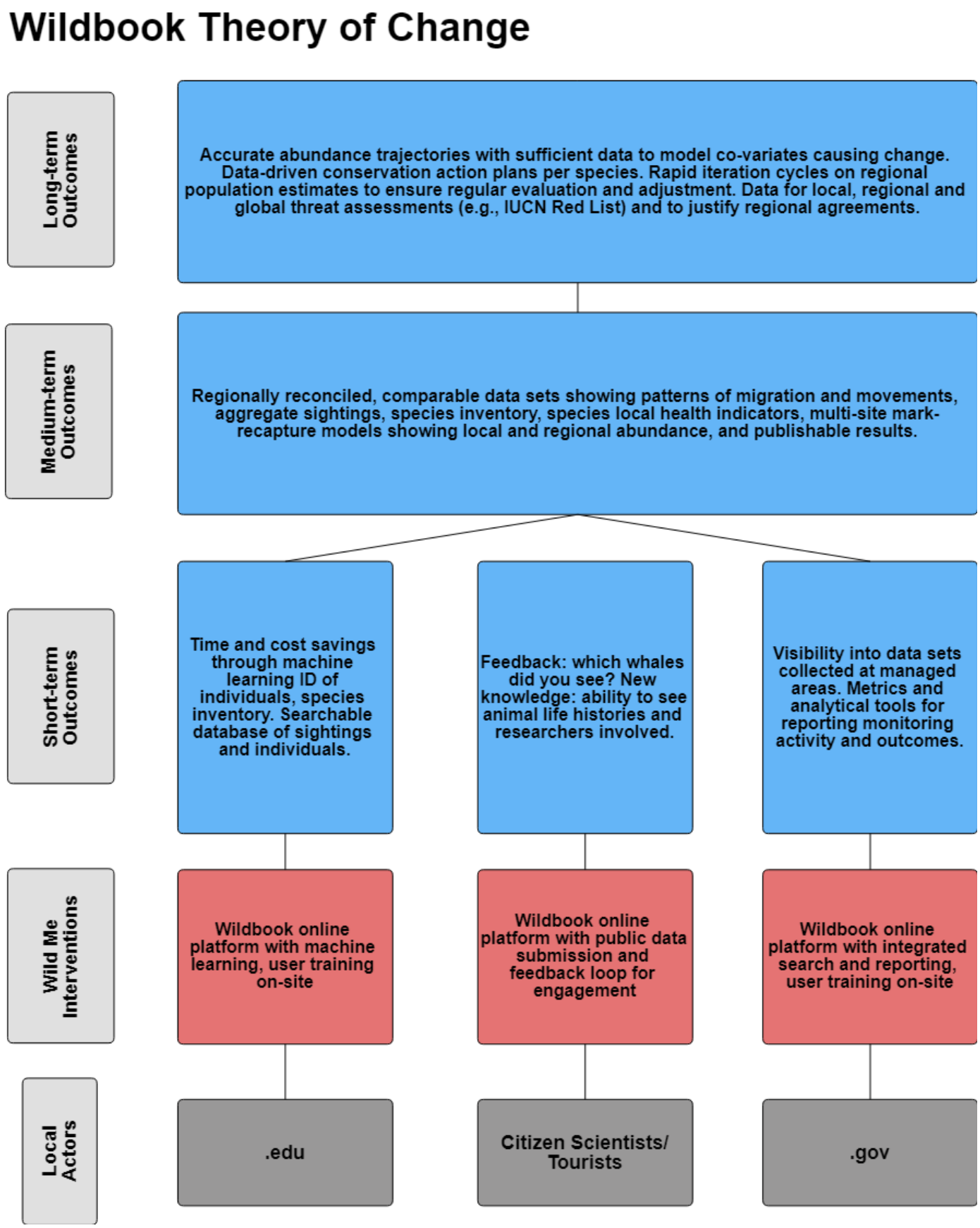

Figure 1b. The Wildbook Theory of Change

Flukebook (https://www.flukebook.org) is an instance of Wildbook tailored for regional- and global-scale research and collaboration on whales and dolphins. Project collaborators include NOAA, Cascadia Research Collective, Dominica Sperm Whale Project, Arabian Sea Whale Network, Indian OceanNetwork for Cetacean Research (IndoCet), and Sarasota Dolphin Research Program, and others.

Figure 2. Flukebook.org landing page provides web-based data submission, curation, and execution of integrated machine learning models in multiple languages.

Flukebook currently provides automated computer vision (algorithm- and deep learning-based) for multiple species of cetaceans [1,2,3,4], including:

- *Megaptera novaeangliae*, three algorithms: one white/black texture of fluke [3], two trailing edge of fluke [2][4]
- *Tursiops truncatus*, one algorithm using trailing edge of dorsal fin [2] and a second being integrated [5]. While each algorithm uses the trailing edge, their approaches with machine learning to individual identification differ and thus succeed and fail under different combinations. Their combined application has a strong potential for future meta-scoring (i.e. a single score and ranking based on both results) with ensemble machine learning techniques.
- *Physeter macrocephalus*, two algorithms using the trailing edge of the fluke [2][4]
- *Eubalaena australis*/*Eubalaena glacialis*, one algorithm using callosity patterns on the head [3] and a second being integrated [6].
- ...and more species planned...

Computer Vision in Flukebook is automated and follows two stages: Detection and Identification:

Detection

Our detection pipeline is a cascade of deep convolutional neural networks (DCNNs) that applies a fully-connected classifier on extracted features. Three separate networks produce: (1) whole-scene classifications looking for specific species of animals in the photograph, (2) object annotation bounding box localizations, and (3) the viewpoint, quality, and final species classifications for the candidate bounding boxes.

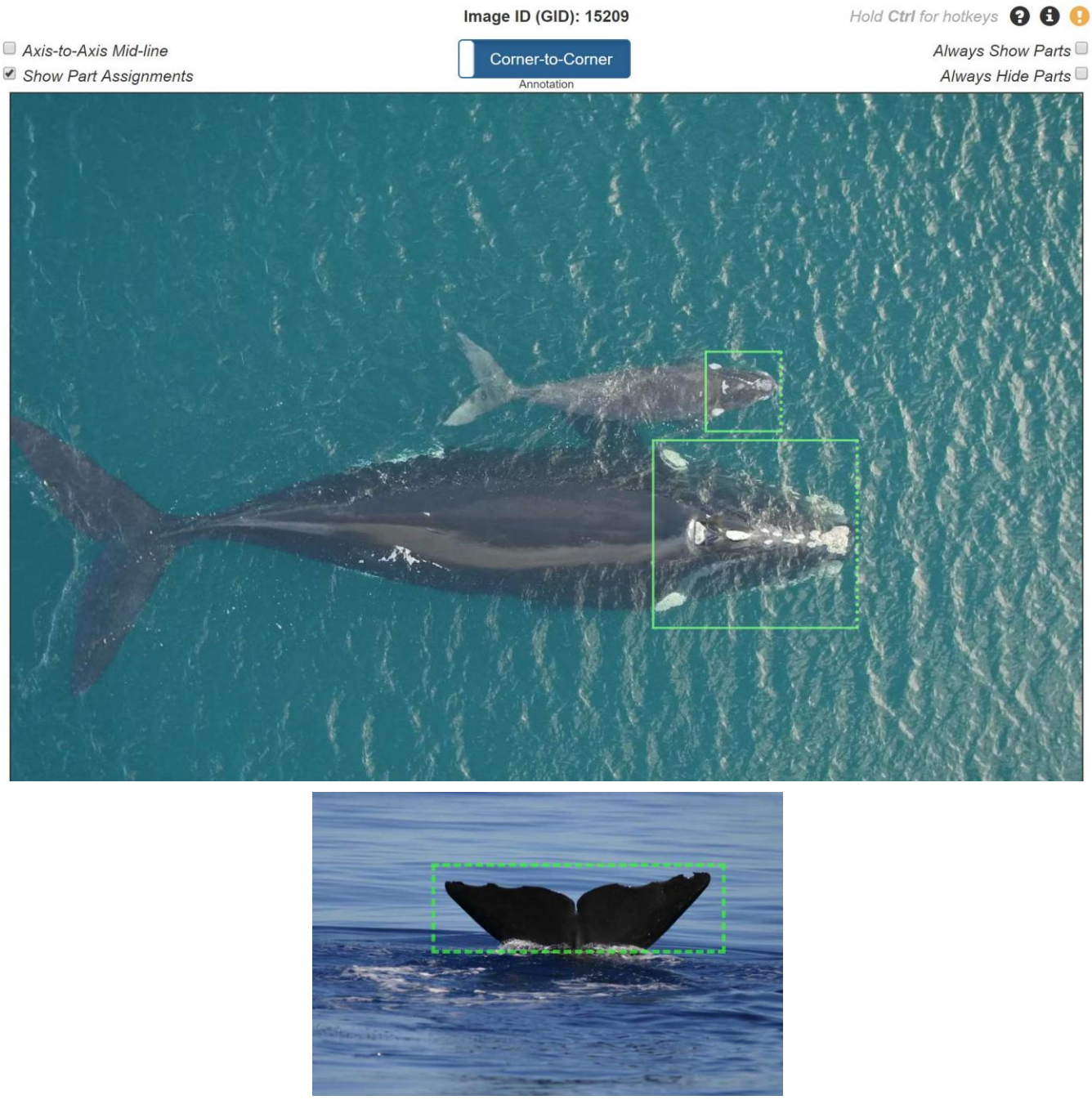

Figure 3a. A right whale image is manually processed as training data input for machine learning models that can then repeat the task automatically without human review. Figure 3b. A sperm whale (Physeter macrocephalus) fluke is detected automatically with machine learning in Flukebook, allowing downstream machine learning identification tasks to focus on only areas of interest in a photo.

Identification

The second major computer vision step is identification, which seeks to assign a name or ID to each fin, fluke, or head from detection using one or more algorithms capable of matching individualized features from a species (e.g., the individually distinct flukes of humpback whales). The matching animals of the same species in the database are then ranked by their accumulated scores and displayed to the user in a list.

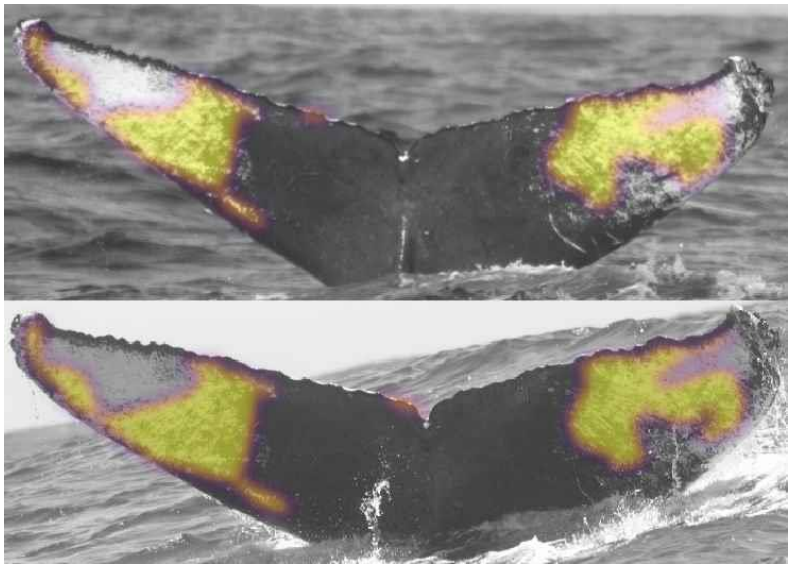

Figure 4. A humpback whale (Megaptera novaeangliae) fluke is matched using a comparison of white and black contrast of its fluke in Flukebook. Flukebook provides multiple techniques for matching several species using an automated computer vision pipeline.

While humpback whales were the first species for which computer vision matching algorithms were developed, the model has been applied to several additional species, with multiple algorithms for some species, and a growing number of algorithms that can be adapted/re-trained for an even wider range of species in the near future. Figure 5 presents a species-specific view of Flukebook's machine learning pipeline.

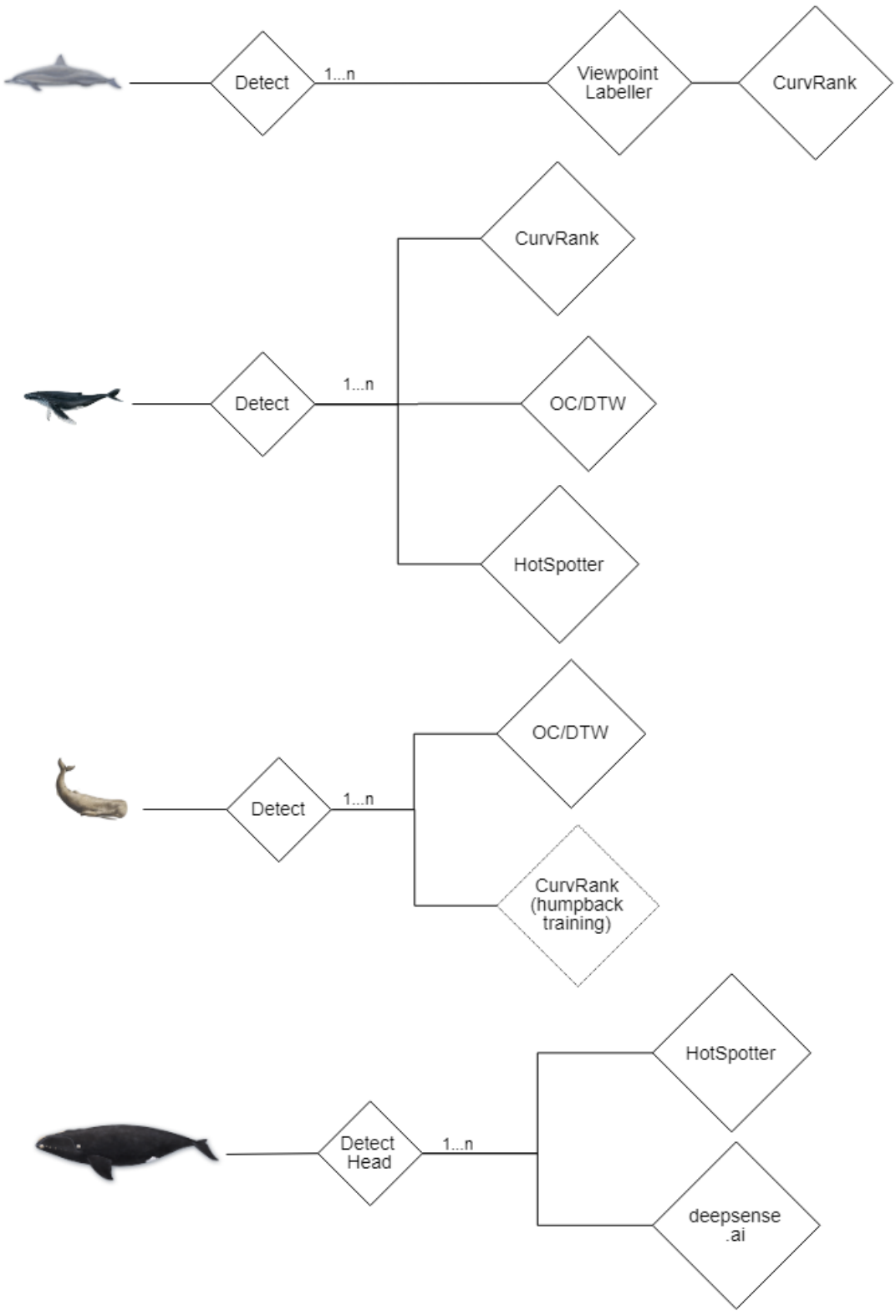

Figure 5. The Flukebook machine learning pipeline for identification of individuals from multiple species demonstrates the application of multiple stages and combinations of machine learning models and algorithms for individual ID of bottlenose dolphins, humpback whales, sperm whales, and right whales. Importantly, additional techniques can be integrated as plugins or retrained and re-purposed using exemplar data for other species.

Flukebook Videos

The following videos of Flukebook demonstrate in very practical terms how two aspects of data querying and interface operate:

- Computer Vision Matching: https://www.youtube.com/watch?v=3_QQrJ5lCxg
- Search and Exporting Data: https://www.youtube.com/watch?v=EYylAOAzX50

2. Recent Technical Advances in Flukebook

A number of technical advances have been made on Flukebook since the last report to the IWC [11]. These include, but are not limited to the following:

Multiple IDs per individual

Flukebook is a collaborative platform designed especially to allow multiple historical and ongoing catalogs to be rapidly reconciled using applied machine learning. To allow differing projects to have a shared concept of individual "identity" when sighting the same whales and dolphins, Flukebook recently switched to a UUID-based identity system that allows for truly unique and non-duplicating IDs. However, to make this unique system (common in software development for large datasets) more usable and human-readable to varying projects with their own legacy IDs, nicknames, alternate IDs, field IDs, etc., Flukebook can now associate multiple IDs with each individual animal, allowing for a shared concept of identity across contributors but direct reference and traceability back to multiple contextual IDs recorded locally in differing studies. This new feature supports broader collaboration without the loss of local autonomy.

Addition of Computer Vision for Individual ID of Sperm Whales, Bottlenose Dolphins, and Right Whales

Flukebook recently added computer vision for individual ID of sperm whales (by fluke trailing edge), bottlenose dolphins (by fin shape), and right whales (by callosity patterns) in addition to its extant ID algorithms for humpback whales. An important function of Flukebook (and Wildbook more broadly) is to serve as an open source repository and applied platform for novel computer vision research originating out of academia [1,2,3,4], investment [5], or competitions [6], and we anticipate the addition of new algorithms and machine learning approaches as well as their cross-application to additional species. For example, the CurvRank algorithm [2] was developed with Waitt Foundation funding for dolphin application, subsequently re-purposed for humpback whale fluke ID, and then re-trained for sperm whale flukes. In this capacity, Flukebook serves as a co-investment model for new photo ID techniques and as a "maximum impact" focal point for machine learning techniques developed with cetaceans in mind (e.g., "make this work in Flukebook"), ensuring that the broadest number of users can quickly access and reliably utilize new developments.

Further Investment in Internationalization

Flukebook.org has been internationalized and can be used in English, French (see Figure 2), Italian, and Spanish. Because of its global userbase and the ease of contributing to open source projects, some users of Flukebook and other Wildbooks have eagerly contributed to these translations by both manually providing translations, and proof-reading machine-generated translations. In particular, the CARI'MAM consortium[7] has helped with French translations for Flukebook, and WWF Spain provided a large amount of effort on Spanish translations for Wildbook via lynx.wildbook.org.

3. New Organizational Usage

In addition to ASWN and other collaborators mentioned above, the following organizations have recently contributed data and/or funding to Flukebook development:

- **National Oceanic and Atmospheric Administration (NOAA):** NOAA has contracted Wild Me to further develop Flukebook.org for the photo-identification of North Atlantic and Southern right whales, applying HotSpotter [3] and deepsense.ai [6] algorithms to aerial identification of individuals using callosity patterns. This project is underway and will complete in August 2019. Contact: christin.khan@noaa.gov
- **IndoCet:** The Indian Ocean Network for Cetacean Research--IndoCet[8]--is currently in its second contract with Wild Me to develop Flukebook features specific to consortium-level needs. Features developed for IndoCet include a custom visual style to present IndoCet branding when its users are on Flukebook, and restricting drop-down menus (e.g. observed animal behavior categories) to consortium-defined options**.** The most significant new feature facilitated by IndoCet is an upgrade to the way Flukebook stores names ("Multiple IDs per individual," above) such that one individual whale can now have different names used by different users and research groups. Previously, Wildbooks used only an authoritative name, a nickname, and an "alternate ID" for each individual animal. Because IndoCet consists of three research groups in the same region (Globice[9], Megaptera, and CetaMada[10]), different groups might have different catalog names for the same individual whale. Flukebook can now store an indefinite number of labeled names on any individual, with those labels (e.g. "IndoCet Regional Catalog Number") providing context for each name.
- **Cascadia Research Collective (CRC):** Cascadia Research Collective is contributing its eastern Pacific humpback whale catalog to Flukebook, allowing the platform to grow into a photo-identification and collaboration hub for the region. Wild Me and CRC are also partnering to add computer vision for other species using existing CRC catalogs as baseline training for machine learning and for broader matching across catalogs.
- **BOEM:** The Bureau of Ocean Energy Management (BOEM) of the U.S. Department of the Interior is currently investing in the Flukebook platform for citizen science and sanctuary cetacean observations. The project is called MAPS and is running 2018-2021. Contact: jacob.levenson@coem.gov

4. Conclusions

Flukebook provides a dedicated platform for collaborative photo-identification of cetaceans with the application of machine learning to generate significant time and cost savings in data curation and reconciliation across catalogs. Investment by the governments of France and the United States as well as NGOs has advanced its technical capabilities and furthered adoption in the research community. As an open source platform, each additional collaboration contributes to its further development and improvement. Investments made by one research project or consortium benefit the wider cetacean research community, as the newly developed algorithms, data archiving or analyses functions become available to all users. The WildMe development team welcome new collaborations that will enable the addition of new species and dimensions to the Flukebook Platform, and hope that existing functions can continue to be refined and applied to cetacean research and conservation efforts around the globe.

References

[1] Parham, Jason & Stewart, Charles & Crall, J.P. & Rubenstein, Daniel & Holmberg, Jason & Berger-Wolf, Tanya. (2018). An Animal Detection Pipeline for Identification. 1075-1083. 10.1109/WACV.2018.00123.

[2] Weideman, H.J, Z.M. Jablons, J. Holmberg, K. Flynn, J. Calambokidis, R. B. Tyson, J.B. Allen, R.S. Wells, K. Hupman, K. Urian, C.V. Stewart. 2017. Integral Curvature Representation and Matching Algorithms for Identification of Dolphins and Whales. doi: 10.1002/ar.23650

[3] Crall JP, Stewart CV, Berger-Wolf T, Rubenstein DI and Sundaresan SR, "HotSpotter — Patterned species instance recognition," 2013 IEEE Workshop on Applications of Computer Vision (WACV 2013), pp. 230-237.

[4] Z. Jablons. Identifying humpback whale flukes by sequence matching of trailing edge curvature. Master's thesis, Rensselaer Polytechnic Institute, 2016.

[5] finFindR: https://www.eurekalert.org/pub_releases/2018-10/nmmf-nft103018.php

[6] Bogucki, R. , Cygan, M. , Khan, C. B., Klimek, M. , Milczek, J. K. and Mucha, M. (2018), Applying deep learning to right whale photo identification. Conservation Biology. doi:10.1111/cobi.13226

[7] Garnier S., Marras P. (2016) Cari'Mam: Caribbean marine mammals preservation network; Strengthen regional cooperation for the conservation of marine mammals within the Caribbean region and beyond: https://www.cbd.int/doc/meetings/mar/soiws-2017-01/other/soiws-2017-01-aamp-en.pdf

[8] http://indocet.org

[9] https://www.globice.org

[10] https://www.cetamada.org

[11] Blount D, Holmberg J, and Minton G. (2018) Flukebook – A tool for cetacean photo identification, data archiving and automated fluke matching. International Whaling Commission report SC/67B/PH/03.