# Multispecies Animal Re-ID Using a Large Community-Curated Dataset

Lasha Otarashvili\*

Tamilselvan Subramanian

Jason Holmberg

Conservation X Labs

Conservation X Labs

Conservation X Labs

J.J. Levenson

Bureau of Ocean Energy Management: Washington, DC, US

Charles V. Stewart
Rensselaer Polytechnic Institute

#### **Abstract**

Recent work has established the ecological importance of developing algorithms for identifying animals individually from images. Typically, a separate algorithm is trained for each species, a natural step but one that creates significant barriers to wide-spread use: (1) each effort is expensive, requiring data collection, data curation, and model training, deployment, and maintenance, (2) there is little training data for many species, and (3) commonalities in appearance across species are not exploited. We propose an alternative approach focused on training multi-species individual identification (re-id) models. We construct a dataset that includes 49 species, 37K individual animals, and 225K images, using this data to train a single embedding network for all species. Our model employs an EfficientNetV2 backbone and a sub-center ArcFace loss function with dynamic margins. We evaluate the performance of this multispecies model in several ways. Most notably, we demonstrate that it consistently outperforms models trained separately on each species, achieving an average gain of 12.5% in top-1 accuracy. Furthermore, the model demonstrates strong zero-shot performance and fine-tuning capabilities for new species with limited training data, enabling effective curation of new species through both incremental addition of data to the training set and fine-tuning without the original data. Additionally, our model surpasses the recent MegaDescriptor on unseen species, averaging an 19.2% top-1 improvement per species and showing gains across all 33 species tested. The fully-featured code repository is publicly available on GitHub, and the feature extractor model can be accessed on HuggingFace for seamless integration with wildlife re-identification pipelines. The model is already in production use for 60+ species in a large-scale wildlife monitoring system. [2],

#### 1. Introduction

Individual animal identification has significant challenges that are not typically encountered in related problems such as human face recognition [7, 32] and human reid [38]:

- Ground truth training datasets tend to be small. This is because individual animal identity is difficult and timeconsuming to curate: it generally relies on humans searching for matching images, an effort that grows quadratically with the size of the dataset.
- Some species have large volumes of training data while others have relatively little. This is partially caused by difficult *in situ* imaging conditions that make it hard to obtain good quality photos. It is also a result of the aforementioned challenges of manual curation, where the resources necessary to build large training sets are available for only the most charismatic species.
- Standard practice is to develop and train animal re-id algorithms on a species-by-species basis, sometimes with per-species algorithmic features. This practice is reasonable in research settings and when just a few species are considered, but it is increasing impractical when supporting dozens of species, each with frequent, ongoing additions of new images and individuals

Our goal in this paper is to address these issues through training and testing a single model for identifying individual animals from many different species. We go beyond the ground breaking work of [39] and explore a dataset that is expert-curated from ongoing use of animal re-id in scientific and conservation practice. Our study focuses on two broad issues:

 How does the overall performance of models trained on multiple species compare to those trained on individual species? When performance is equivalent then individual species models are unnecessary, and when

<sup>\*</sup>lasha@conservationxlabs.org

performance is better on the multiple species model we also see a performance advantage to combined training on multiple species.

How well do multispecies models perform when applied to species unseen during training? This is crucial to addressing the shortage of annotated training data for many species. Ideally, a model would work as well as if training data were available, but practically if a multi-species model can significantly reduce the amount of training data needed obtain a given level of performance for a new species it is an important development.

As a result of our study, we make the following main contributions:

- 1. We train a multispecies animal re-id model MiewID on a community-curated dataset involving 49 species from 59 datasets, 37,138 individual animals and 225,374 total annotations. We publicly release the code as a GitHub repository and model weights on Huggingface. We show that our model consistently outperforms analogous single-species models, exhibiting an overall performance gain of 12.5% top-1 accuracy and per-species gains ranging from a minimum of -0.2% (negative for 1 species) to a maximum of 77.3%.
- We demonstrate the effectiveness of our model on previously species unseen during training, and show that it substantially outperforms MegaDescriptor [39] on all 33 tested species, with an average top-1 improvement of 19.2%.
- 3. We show the potential of our model for learning to effectively identify species that have only have a small set of data available during training. This is seen both when the new data are incorporated into the overall training set and when fine-tuning the published model.

# 2. Background

Historically, a wide range of methods have been used for animal re-identification, ranging from manual feature extraction to more recent advances in deep learning. In recent years, global descriptor-based methods leveraging deep learning [3, 8, 11, 21, 22, 39] have gained prominence over traditional manual feature-based approaches [24, 30, 33] and local feature algorithms [5, 18, 24, 28]. These deep learning models are becoming increasingly popular due to their ability to produce more robust, generalizable descriptors compared to hand-crafted or localized feature techniques. A major advantage of deep learning methods is their scalability and capability to streamline the re-identification pipeline. While they require larger datasets and greater computational resources to train, they minimize the need

for extensive preprocessing and can run end-to-end, making them highly efficient for large-scale animal ID tasks. In this work, we focus on end-to-end deep learning models due to their ability to handle complex identification tasks across multiple species with minimal manual intervention, thus improving scalability and applicability to realworld scenarios. The release of several large animal reidentification datasets has accelerated research in this field [1, 4, 8, 9, 12, 17, 23, 25, 31, 35–37]. These datasets cover a broad spectrum of species, providing the necessary data to train more effective and generalizable models. Recent studies have explored multi-species re-identification, which focuses on applying individual identification methods across diverse sets of species [4, 11, 39]. Our work reported here goes beyond these methods in several important ways, including the quality and diversity of species used in training and the extensive experiments in zero-shot and few-shot fine-tuning.

#### 3. Data

We construct our multi-species dataset from a mixture of 1) publicly available datasets [1,4,8,9,12,23,25,31,37] and 2) ecological data and images contributed directly and privately to the Wild Me Lab's self-hosted Wildbook platforms [2], which support dynamic and secure data curation (to the individual animal level) by corresponding species experts. Inside Wildbook, users employ manual photo ID (i.e., "by eye") and earlier re-identification computer vision algorithms [5, 20, 34] to rank-order potential individual matches and then make curation decisions about individual animal identities across sightings directly in the software. Our work herein is in direct assistance to the Wildbook user community (2000+ users), which we transparently support via the Wild Me community platform (https://community.wildme.org/). Contributed data is used with permission to correspondingly improve individual matching capabilities in the Wildbook software platform. Where expressly permitted by data owners using Wildbook, we contribute these unique, species-specific re-ID datasets [13–16, 26] back for public access on LILA BC (https://lila.science/). Statistics on the individual species are summarized in Table 1.

Overall, there is an average of 6.25 annotations per individual and a median of 3. Clearly, this is a heavy-tailed distribution, with five species having more than 10,000 annotations, 11 having fewer than 1,000, 9 fewer than 500, and 6 fewer than 250.

#### 3.1. Preprocessing and Labeling

Images were preprocessed to detect animals of interest, locate them with an enclosing axis-aligned bounding box, and label their species. Low quality images were filtered out, although in many cases expert curators completed this

Table 1. Dataset summary, including name, source, number of annotations and number of individuals. Different image sets for a single species are combined when forming the training and test sets.

Dataset	Source	Annotations	Individuals
amur_tiger	contributed	1,015	103
beluga_whale	contributed	4,055	681
blue_whale	Happywhale	4,830	2348
bottlenose_dolphin	contributed	22,691	4,130
brydes_whale	Happywhale	154	44
capuchin	Lomas Capuchin	2,618	35
cheetah	contributed	3,005	403
chimpanzee	contributed	203	11
chimpanzee_ctai	C-Tai	4,662	71
chimpanzee_czoo	C-Zoo	2,109	24
commersons_dolphin	Happywhale	90	70
cuviers_beaked_whale	Happywhale	341	183
dog	DogFaceNet	8363	1,392
dusky_dolphin	Happywhale	3,139	2,731
eurasianlynx	contributed	4,716	1,642
fin_whale	Happywhale	1,324	466
frasiers_dolphin	Happywhale	14	13
giraffe	contributed	3,724	1,067
horse_wild+face	contributed	293	115
horse_wild_tunisian+face	contributed	1,403	47
hyena	contributed	2,948	498
hyperoodon_ampullatus	contributed	30,063	1,196
jaguar	contributed	1,876	257
japanese_monkey	PrimFace	297	18
killer_whale	Happywhale	2,455	472
leopard	contributed	6,052	848
lion	contributed	6,374	397
long_finned_pilot_whale	Happywhale	238	131
lynx_pardinus	contributed	2,573	752
macaque_face	MacaqueFaces	6,280	34
melon_headed_whale	Happywhale	1,689	1,323
nyala	contributed	1,942	237
pantropic_spotted_dolphin	Happywhale	145	44
pygmy_killer_whale	Happywhale	76	24
rhesus_monkey	PrimFace	763	38
rough_toothed_dolphin	Happywhale	60	46
seal	SealID	2,080	57
sei_whale	Happywhale	428	197
short_fin_pilot_whale+fin_dorsal	contributed	2,137	611
short_finned_pilot_whale	Happywhale	745	458
snow_leopard	contributed	4,715	241
spinner_dolphin	Happywhale	2,254	517
spotted_dolphin	Happywhale	490	280
turtle_green	contributed	5,640	109
turtle_green+head	contributed	6.282	109
turtle_hawksbill	contributed	5,906	109
turtle_hawksbill+head	contributed	7,625	112
turtle_loggerhead_ext	SeaTurtleID	7,405	438
turtle_loggerhead_ext+head	SeaTurtleID	7,166	438
whale_fin+fin_dorsal	contributed	2,998	942
whale_grey	contributed	8,294	2,343
whale_humpback+fin_dorsal	contributed	9,251	3,924
whale_humpback+fluke	contributed	67,06	4,767
whale_orca	contributed	4,045	1,208
		5,781	1,494
whole oreatfin dereal		J,/01	1,474
whale_orca+fin_dorsal	contributed	27.001	667
whale_sperm+fluke	contributed	27,001	667 530
whale_sperm+fluke whaleshark	contributed contributed	6,825	539
whale_sperm+fluke whaleshark white_shark+fin_dorsal	contributed contributed contributed	6,825 1,771	539 484
whale_sperm+fluke whaleshark white_shark+fin_dorsal white_sided_dolphin	contributed contributed contributed Happywhale	6,825 1,771 229	539 484 175
whale_sperm+fluke whaleshark white_shark+fin_dorsal	contributed contributed contributed	6,825 1,771	539 484

before contributing images. The result is a set of 0 or more "annnotations" per image, each a sighting of an individual animal. Each annotation also has its predominant view-

point assigned — left, right, front or back. As mentioned above, each annotation has been assigned an individual ID through a curation process that involved expert manual decision making, sometimes using the guidance of an earlier algorithm to rank-order and suggest potential matching annotations. These ID-ed annotations form the ground truth dataset from which the work in this paper starts.

### 3.2. Viewpoint

Viewpoint plays an important role in our experiments. For some species, such as belugas, there is only one (aerial / top) viewpoint. For many others, identifying information from different sides of the body — such as the spots on a cheetah or a giraffe — are so different that matching left and right views is impractical. For these, annotations from different viewpoints are considered different individuals in our dataset. For still other species, where the identifying information is predominantly on the outline, such as for the shape of the dorsal fin, opposite side views can match and therefore are not considered different animals for the purposes of our study.

# 3.3. Data Splitting

The dataset is divided into training and test sets without a separate holdout set. To assess model performance on both closed-set and open-set scenarios, we carefully designed the test set composition. Specifically, 50% of the individuals in the test set are also present in the training set, but with different annotations (i.e., different images). The remaining 50% of individuals in the test set were not part of the training set at all. This approach allows us to evaluate the model's ability to generalize to both known and unseen individuals.

To ensure the test set is balanced and to prevent the metrics from being skewed by individuals with a high number of annotations, we implemented several controls. The number of samples per individual was balanced across the test set, and the annotations in the test set were subsampled to include a maximum of 10 samples per individual.

The train/test split was performed with careful consideration of dataset size and individual representation. The selection process is a follows:

- For each individual with k annotations, we first determine whether the individual will be placed in the training or test set. A smaller training percentage is favored for smaller datasets to ensure a sufficient number of individuals in the test set.
- 2. If an individual is selected for the training set, a subset of their annotations is reserved for the test set.
- If an individual is not selected for the training set, all of their annotations are used for the test set.

#### Post-Split Filtering

- 1. Training Set: Individuals with fewer than 3 annotations are excluded. There is no upper limit on the number of annotations per individual.
- 2. Test Set: Individuals with fewer than 2 annotations in the test set are excluded. For individuals with more than 10 annotations, random 10 annotations are picked to remain in the test set. Finally, only one annotation per encounter is allowed in the test set to avoid inflating performance through matching of near identical images.

A separate validation set was not utilized in these experiments due to the limited data available for a significant number of species. The small sample sizes made it impractical to generate train, validation, and test splits of reasonable size, particularly for robust evaluation. Instead, model hyperparameters were optimized using data splits generated from random seeds that differed from those used in the final experiments, ensuring that the test set remained unseen during tuning.

#### 3.4. Evaluation

In evaluating the re-identification system, we employ a *one-vs-all* evaluation scheme, which avoids the pitfalls of traditional approaches that use separate query and reference sets. While typical methods reserve a reference set of labeled individuals and treat the query set as unknown, there is often insufficient data to create independent query and reference sets that are entirely separate from the training data. Previous works [1, 3, 4, 39] use the training set as the reference set and the test set as the query set. This introduces a *soft data leak*, as the model has been optimized on the reference set, even though it has not directly seen the query data. To mitigate this issue, we use the test set as both the query and reference set, while ensuring that annotations are never matched to themselves. The evaluation process is as follows:

For each annotation A in the test set:

- 1. Select annotation A as the query.
- 2. All other annotations in the test set, excluding *A*, are used as the reference set.
- 3. Compute accuracy at top-k ranks for the query A.
- 4. Repeat the process for all annotations in the test set.

Metrics are averaged across all annotations in the test set to generate the final evaluation result.

This approach ensures that the model is evaluated on unseen data without introducing biases from the training set. The final scores, averaged across all test annotations, provide robust performance metrics, yielding a more accurate evaluation of the model's ability to generalize to unseen data.

# 4. Model and Inference

Our model model starts from an EfficientNetV2-M [29] backbone pretrained on ImageNet1K [27]. It is trained using a sub-center (k=3) ArcFace loss function [6] with dynamic margins [10]. As will be discussed briefly in the experimental section below, unlike the MegaDescriptor [39] work, we find that using vision transformers did not improve the results. Each annotation's bounding box is used to crop the image and this crop is then scaled to a standard size prior to input. The production version size is 440x440 pixels, although for efficiency reasons the comparison experiments all use a smaller 256x256 resolution. Details on training and parameter settings are found in the appendix.

At inference time, the classification layer is discarded and an embedding vector of dimension 2,048, the default for EfficientNetV2 [29], is extracted when the trained model is applied to an annotation. A test database of the embedding vectors is formed for each species separately, and each test annotation's vector is matched against the other vectors in the database and the top-k nearest are returned in order of increasing cosine distance, regardless of identity label.

# 5. Experiments

We present a range of experiments showing the importance of multispecies training for individual animal id, demonstrating the power of using this large diverse dataset for single species id, for adding new species, and for finetuning.

#### 5.1. Multispecies vs. Individual Models

Our largest experiment compares the performance of the full model trained on N=49 species against that of N other models, each trained with the same architecture, loss function, and hyperparameters on each species separately. Top-1 ranking performances are shown in Figure 1 and Table 2, with species ordered by decreasing number of annotations. These results show the multispecies model consistently outperforming single-species models, by an average of 12.5% top-1 accuracy. The advantages of the multispecies model increase as the size of the population for any one species decreases (Figure 1). These results immediately give experimental justification for training and applying a single model across many species, an important practical consideration for managing multiple species at once. We are already seeing this in production use.

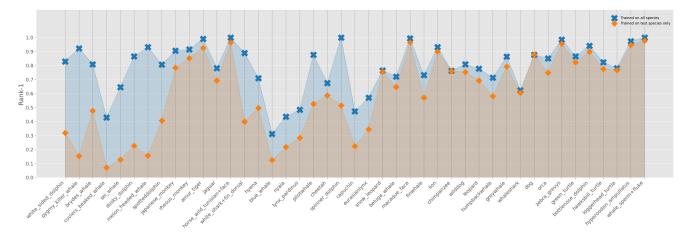


Figure 1. Comparison of top-1 performance of multi-species and single-species models. Species are ordered by increasing numbers of sightings.

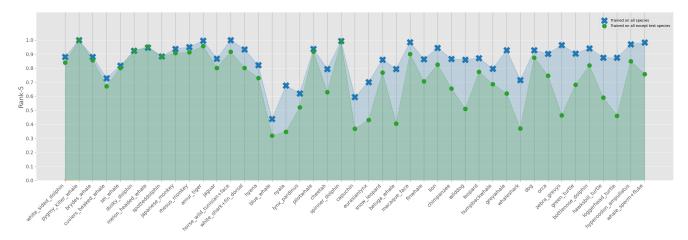


Figure 2. Top-5 performance comparison per species for model trained on the full dataset and model trained on all species except the test species

#### **5.2. Species Not Seen During Training**

One potential application of the multispecies model is to bootstrap identities for individuals from species with a paucity of labeled data. Our first study of this is the zero-shot case where no annotations from the species are seen during. To study this, we train N additional models, each with a single species left out, and measure the test performance on the left-out species. Figure 2 shows the results comparing the full model with the leave-one-out models. We show top-5 performance because that is more indicative of what's needed to guide human decision making when building up new species. (Top-1 results are provided in the supplementary material.) With species ordered by increasing numbers of annotations, the results show that for species with fewer annotations, the model uses aggregate information from training on many species to fill in missing

information for matching. This effect is less prominent for species with larger datasets.

#### 5.3. Our Model vs. MegaDescriptor

The significance of the results on unseen species can be better appreciated by comparing them with the results from MegaDescriptor [39]. Both MegaDescriptor-L-384 and our model are evaluated at their native resolutions—384px and 256px, respectively — and neither has been trained on the tested species. Table 3 presents the rank-1 test results per species between MegaDescriptor and our model, trained without including the species in the training set (as detailed in Section 5.2). The comparison is conducted on 33 species because these are the ones MegaDescriptor has not been trained on. In this zero-shot setting, our model outperforms MegaDescriptor on every species, with an average perfor-

Table 2. Comparison of metrics across various species for multispecies, leave-one-out and single species models

Source Multi-species Leave-one-out Single species 99.1 92.3 92.7 amur\_tiger 72.1 28.0 64.8 beluga\_whale 31.3 19.9 12.5 blue whale bottlenose\_dolphin 94.2 71.5 89.9 81.0 69.7 47.6 brydes\_whale 22.3 22.3 capuchin 47.3 cheetah 67.6 49.5 58.7 45.1 chimpanzee 76.4 76.1 cuviers\_beaked\_whale 42.9 37.3 7.1 877 75 1 88.0 dog dusky\_dolphin 86.5 84.8 22.8 57.1 25.2 34 6 eurasianlynx finwhale 73.2 49.2 57.1 giraffe 100.0 0.0 93.6 green\_turtle 86.5 48 3 82.4 greywhale 86.5 44.4 79.3 hawksbill\_turtle 82.5 38.1 77.6 100.0 96.6 horse wild tunisian+face 86.8 humpbackwhale 71.3 58.1 51.6 71.0 58.9 49.7 hvena 94.7 97.5 72.9 hyperoodon\_ampullatus jaguar 78.2 67.8 69.4 japanese\_monkey 90.7 84.6 78.5 leopard 77.8 61.3 69.4 93.2 75.2 90.0 lion loggerhead\_turtle 78.2 32.7 76.7 lynx\_pardinus 48.4 32.3 28.5 macaque\_face 99.5 71.8 96.5 melon\_headed\_whale 93.1 92.1 15.8 43.4 15.6 21.8 nyala orca 85.1 59 5 74.9 87.7 813 52.7 pilotwhale pygmy\_killer\_whale 92.3 87.6 15.4 rhesus\_monkey 91.6 74.5 85.3 63.9 seal 67.6 64.1 sei\_whale 64.7 62.5 12.9 snow\_leopard 76.6 65.2 75.5 100.0 spinner\_dolphin 95.2 51.6 74.8 40.8 spotteddolphin 80.7 whale\_sperm+fluke 100.0 62.0 97.7 whaleshark 62.4 24 2 60.5 white\_shark+fin\_dorsal 89.0 64.9 40.1 white\_sided\_dolphin 83.0 80.0 32.0 wilddog 80.9 40.2 75.3 zebra\_grevys 98.5 33.8 95.7

Table 3. Comparison of 0-shot rank-1 performace of our model to MegaDescriptor

Source	MegaDescriptor-L-384	Our model
blue_whale	19.2	19.9
bottlenose_dolphin	17.7	71.5
brydes_whale	66.7	81.0
capuchin	14.2	22.3
cheetah	38.9	49.5
cuviers_beaked_whale	18.6	42.9
dog	57.0	75.1
dusky_dolphin	28.5	84.8
eurasianlynx	15.7	25.2
finwhale	24.9	49.2
green_turtle	22.3	48.3
greywhale	32.0	44.4
hawksbill_turtle	25.0	38.1
horse_wild_tunisian+face	86.3	86.8
humpbackwhale	43.7	51.6
hyperoodon_ampullatus	35.3	72.9
jaguar	41.1	67.8
japanese_monkey	50.8	84.6
lion	69.6	75.2
lynx_pardinus	18.6	32.3
melon_headed_whale	38.2	92.1
orca	46.8	59.5
pilotwhale	51.8	81.3
pygmy_killer_whale	61.5	92.3
rhesus_monkey	47.3	74.5
sei_whale	39.7	64.7
snow_leopard	58.8	65.2
spinner_dolphin	19.1	95.2
spotteddolphin	65.0	74.8
whale_sperm+fluke	47.3	62.0
white_shark+fin_dorsal	11.2	64.9
white_sided_dolphin	72.0	80.0
wilddog	31.2	40.2

mance gain of 19.2%. We attribute this advantage in part to the size and diversity of our dataset, as well as the fact that it has been extensively curated through ongoing, productionlevel review by species experts.

# 5.4. Bootstrapping New Species

As noted in the introduction, a major challenge in individual animal identification is bootstrapping new species from a limited number of annotated examples. In this subsection, we address this by selecting a representative set of species and retaining only a small fraction of the annotations in the training set. For each species separately, we then recombine this with the training set from all other species and train the model. By conducting this experiment at various data scales for the target species, we analyze the impact of more and more limited datasets. The results from Table 4 clearly show the advantages of including species, even in small amounts, in multi-species training. Moreover, joint training consistently achieves higher accuracy for the target species than the equivalent single-species model. This suggests that the shared representation learned from multi-species data enhances the model's generalization to new species, even when annotations are scarce.

Table 4. Comparison of training the model on single species vs. joint training of target species with the full multi-species dataset.

Source	# of Individuals	# of added annots	Species-only	Joint Training
	0	0	_	27.9
Beluga	25	557	32.5	46.7
	50	990	42.4	56.0
	100	1,541	57.3	64.3
	200	2,210	63.9	67.0
	0	0	_	58.8
	25	254	32.4	62.8
	50	558	40.6	63.3
Hyena	100	713	46.3	66.6
	200	1,154	48.0	70.4
	0		_	31.2
	25	3,632	35.7	36.7
I d Toud	50	5,552	45.0	47.2
Loggerhead Turtle	100	8,136	53.0	62.3
	200	11,084	60.9	73.8
	0	0	_	40.4
	25	552	25.7	54.8
C Wh-1-	50	995	40.4	61.6
Gray Whale	100	1,773	56.9	70.8
	200	2,992	67.1	75.3
White Shark	0	0	_	64.9
	25	331	21.1	77.6
	50	526	25.1	81.3
	100	864	31.0	85.4
	200	1,067	38.8	86.3
	0	0	_	74.5
	25	1,364	67.7	81.7
T :	50	2,180	74.4	84.7
Lion	100	3,094	85.2	87.2
	200	3,846	89.2	90.1

This ability to "bootstrap" new species can be further studied by showing the model's effectiveness in adding new individuals to a population. Here we are not interested in the ability to recognize that an annotation shows a new animal — while an important problem, it is not our focus but rather how easily annotations are identified when there are  $1, 2, 3, \ldots$  annotations in the database to correctly match a query annotation. We explore this for the same subset of species by showing average top-1 and top-5 rates for different numbers of potentially matching annotations, as shown in Figure 3. Matching performance is lower when there are less (k=1, k=2) annotations to match against in the database. However, the drop is not too significant, and performance quickly asymptotes, indicating strong potential for using the model to bootstrap new species since it does not require many sightings of an individual to reliably identify it.

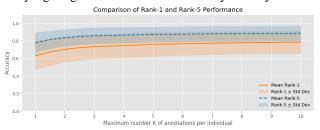


Figure 3. Mean top-1 accuracy and standard deviation across species as a function of the maximum number of annotations per individual in the database.

## 5.5. Fine-Tuning on New Species

Table 5. Comparison of training the model from scratch vs. finetuning the multi-species model on varying amounts of individuals in the training set

Source	# of Individuals	# total of annots	Species-only	Fine-tuning
	0	0	_	27.9
Beluga	25	557	32.5	44.5
	50	990	42.4	49.9
	100	1,541	57.3	54.2
	200	2,210	63.9	56.5
	0	0	_	58.8
	25	254	32.4	56.3
11	50	558	40.6	61.9
Hyena	100	713	46.3	62.5
	200	1,154	48.0	65.4
	0		_	31.2
	25	3,632	35.7	41.3
I dr d Tdl -	50	5,552	45.0	48.6
Loggerhead Turtle	100	8,136	53.0	60.5
	200	11,084	60.9	70.6
	0	0	_	40.4
	25	552	25.7	53.5
Gray Whale	50	995	40.4	56.6
	100	1,773	56.9	62.7
	200	2,992	67.1	66.6
White Shark	0	0		64.9
	25	331	21.1	66.7
	50	526	25.1	72.3
	100	864	31.0	74.8
	200	1,067	38.8	75.9
Lion	0	0		74.5
	25	1,364	67.7	79.1
	50	2,180	74.4	83.5
	100	3,094	85.2	85.8
	200	3,846	89.2	87.3

Next we explore the fine-tuning capabilities of our model. The goal is to use our model without needing access to the full training set to quickly apply to new species. To explore this in Table 5 we compare single species training to fine-tuning of our model — starting from the version trained without the species — for various numbers of annotations. This shows that fine-tuning is effective, especially for small datasets, with single-species training only showing better performance for larger numbers of individuals and annotations. Finally, comparing the right columns of Tables 4 and 5 we see that incorporating the data directly into training the multispecies model produces consistently better results.

#### 5.6. Vision Transformers

In our experiments, we evaluated the impact of using a vision transformer backbone, specifically SwinV2-Base [19], and compared its performance to the standard EfficientNetV2-M backbone. While vision transformers have shown strong results in various vision tasks, EfficientNetV2 consistently outperformed SwinV2 in our setting, achieving an average performance improvement of 4.5%. This suggests that, for our task, EfficientNetV2 provides better feature extraction and generalization capabilities compared to the transformer-based approach.

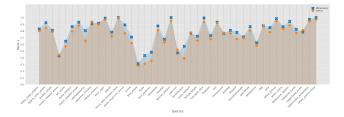


Figure 4. Comparison of Swin-V2 and Efficientnet-V2 backbones for MiewID training

#### 6. Discussion and Conclusion

We have presented and experimentally evaluated a multispecies individual animal identification model called MiewID. Trained on 49 different species from 59 datasets, we showed that it consistently outperforms the same model trained on each species individually, with the greatest gains for species with small datasets. Our model also demonstrated strong test performance on unseen (during training) species, uniformly and significantly better than the recent MegaDescriptor. Next, we showed that our model can be used as the starting point for fine-tuning, even with small dataset sizes. We also showed that incorporating these same datasets into multispecies training produced better results. This argues for frequent retraining as more and more species are added, which seems expensive but perhaps no more than training on each species individually, with a much simpler training and deployment process since just a single model is needed.

Our model is already in open-source production use on our Wildbook system. More species are being added frequently and new models are periodically released on our Huggingface page. The architecture, code, training procedure, and weights are freely available, and some of the datasets are as well. More should be distributed openly as the advantages of multispecies foundation models become clear. Overall, we have achieved a significant advance in multispecies animal identification and intend to continue development as more and more well-curated datasets and advanced architectures and training methods become available.

# 7. Acknowledgements

The Gordon and Betty Moore Foundation directly supported our experiments. Precursor funding for cetacean-specific experiments was also received from the Bureau of Ocean Energy Management (BOEM) under awards 140M0121P0030 and 140M0121D0004. We would like to thank Dr. Ted Cheeseman for providing inspiration and community-sourced open cetacean data [4] that benefited our work. We thank Paul Kalil and Maureen Reilly of T4C for coordinating the acquisition of several terrestrial carni-

vore datasets in Wildbook from a broad research community. Data used in this work was either publicly available or directly contributed to Wild Me via our hosted Wildbook platforms. Data contributors include: NOAA, Sarasota Dolphin Research Project, Botswana Predator Conservation Trust, Dr. Kasim Rafiq, Lion Guardians, Dr. Brad Brad Norman of ECOCEAN, Giraffe Conservation Foundation, WWF Spain, WWF Finland, Norwegian Orca Survey, Cascadia Research Collective, MarEcoTel, the Dominica Sperm Whale Project, African Parks, Ol Jogi Wildlife Conservancy, Ongava Research Centre, Endangered Wildlife Trust, Dalhousie University, Dr. Salvador Jorgensen, Olive Ridley Project, Dr. Eve Bohnett, Office français de la biodiversité, San Diego Zoo, and Cheetah Conservation Foundation. C. Stewart's contributions were also partially supported by the US National Science Foundation under Award 2118240.

# References

- [1] Lukáš Adam, Vojtěch Čermák, Kostas Papafitsoros, and Lukas Picek. Seaturtleid2022: A long-span dataset for reliable sea turtle re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 7146–7156, January 2024. 2, 4
- [2] Tanya Y. Berger-Wolf, Daniel I. Rubenstein, Charles V. Stewart, Jason A. Holmberg, Jason Parham, Shyamala Menon, Josh Crall, John Van Oast, Emre Kiciman, and Lucas Joppa. Wildbook: Crowdsourcing, computer vision, and data science for conservation. arXiv preprint arXiv:1710.08880, 2017. 1, 2
- [3] Vojtěch Cermak, Lukas Picek, Lukáš Adam, Lukáš Neumann, and Jiří Matas. Wildfusion: Individual animal identification with calibrated similarity fusion, 2024. 2, 4
- [4] Ted Cheeseman, Ken Southerland, Jinmo Park, Marilia Olio, Kiirsten Flynn, John Calambokidis, Lindsey Jones, Claire Garrigue, Astrid Frisch Jordán, Addison Howard, Walter Reade, Janet Neilson, Christine Gabriele, and Phil Clapham. Advanced image recognition: a fully automated, highaccuracy photo-identification matching system for humpback whales. *Mamm. Biol.*, 102(3):915–929, June 2022. 2, 4, 8
- [5] James P. Crall, Charles V. Stewart, Tanya Y. Berger-Wolf, Daniel I. Rubenstein, and Siva R. Sundaresan. Hotspotter patterned species instance recognition. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision* (WACV), 2013. 2
- [6] Jiankang Deng, Jia Guo, Tongliang Liu, Mingming Gong, and Stefanos Zafeiriou. Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In *Proceedings* of the European Conference on Computer Vision (ECCV), pages 741–757. Springer, 2020. 4
- [7] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5962–5979, Oct. 2022. 1

- [8] Nkosikhona Dlamini and Terence L van Zyl. Automated identification of individuals in wildlife population using siamese neural networks. In 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI), pages 224–228. IEEE, 2020. 2
- [9] Alexander Freytag, Erik Rodner, Marcel Simon, Alexander Loos, Hjalmar Kühl, and Joachim Denzler. Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates. In *German Conference on Pattern Recognition (GCPR)*, 2016. 2
- [10] Qishen Ha, Bo Liu, Fuxu Liu, and Peiyuan Liao. Google landmark recognition 2020 competition third place solution, 2020. 4
- [11] Bingliang Jiao, Lingqiao Liu, Liying Gao, Ruiqi Wu, Guosheng Lin, PENG WANG, and Yanning Zhang. Toward re-identifying any animal. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 2
- [12] Labeled Information Library of Alexandria (LiLA). Atrw dataset: Ai city challenge track 5 - animal track re-id in the wild, 2022. Accessed: 2024-07-02.
- [13] Labeled Information Library of Alexandria (LiLA). Beluga id 2022 dataset, 2022. Accessed: 2024-07-02.
- [14] Labeled Information Library of Alexandria (LiLA). Hyena id 2022 dataset, 2022. Accessed: 2024-07-02. 2
- [15] Labeled Information Library of Alexandria (LiLA). Leopard id 2022 dataset, 2022. 2024-07-02. 2
- [16] Labeled Information Library of Alexandria (LiLA). Whale shark id dataset, 2022. Accessed: 2024-07-02.
- [17] LILA BC. Labeled information library of alexandria: Biology and conservation. 2
- [18] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed, 2023.
- [19] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin transformer v2: Scaling up capacity and resolution. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 7
- [20] Olga Moskvyak, Frédéric Maire, Asia O. Armstrong, Feras Dayoub, and Mahsa Baktash. Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. 2021 Digital Image Computing: Techniques and Applications (DICTA), pages 1–8, 2019. 2
- [21] Olga Moskvyak, Frederic Maire, Asia O. Armstrong, Feras Dayoub, and Mahsa Baktashmotlagh. Robust reidentification of manta rays from natural markings by learning pose invariant embeddings, 2019. 2
- [22] Guillaume Mougeot, Dewei Li, and Shuai Jia. A deep learning approach for dog face verification and recognition. In Abhaya C. Nayak and Alok Sharma, editors, *PRI-CAI 2019: Trends in Artificial Intelligence*, pages 418–430, Cham, 2019. Springer International Publishing. 2
- [23] Guillaume Mougeot, Dewei Li, and Shuai Jia. A deep learning approach for dog face verification and recognition. In Abhaya C. Nayak and Alok Sharma, editors, *PRI-CAI 2019: Trends in Artificial Intelligence*, pages 418–430, Cham, 2019. Springer International Publishing. 2

- [24] Ekaterina Nepovinnykh, Ilia Chelak, Tuomas Eerola, and Heikki Kälviäinen. Norppa: Novel ringed seal reidentification by pelage pattern aggregation, 2022. 2
- [25] Ekaterina Nepovinnykh, Tuomas Eerola, Vincent Biard, Piia Mutka, Marja Niemi, Mervi Kunnasranta, and Heikki Kälviäinen. Sealid: Saimaa ringed seal re-identification dataset. Sensors, 22(19), 2022. 2
- [26] Jason Parham, Jonathan Crall, Charles Stewart, Tanya Berger-Wolf, and Daniel I. Rubenstein. Animal population censusing at scale with citizen science and photographic identification. In AAAI Spring Symposium-Technical Report, Palo Alto, CA, January 1 2017. AAAI Press. 2
- [27] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 4
- [28] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks, 2020. 2
- [29] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021. 4
- [30] Jaime Thompson, Victoria Zero, Lori Schwacke, Todd Speakman, Brian Quigley, Jeanine Morey, and Trent Mcdonald. finfindr: Automated recognition and identification of marine mammal dorsal fins using residual convolutional neural networks. *Marine Mammal Science*, 38, 07 2021. 2
- [31] Visiome Platform. Primface dataset, 2024. Accessed: 2024-07-02. The face images used in this study are provided by PrimFace database: http://visiome.neuroinf.jp/primface, funded by Grant-in-Aid for Scientific research on Innovative Areas, "Face Perception and Recognition" from Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. 2
- [32] Xinyi Wang, Jianteng Peng, Sufang Zhang, Bihui Chen, Yi Wang, and Yandong Guo. A survey of face recognition, 2022. 1
- [33] Hendrik Weideman, Charles Stewart, Jason Parham, Jason Holmberg, Kiirsten Flynn, John Calambokidis, D. Paul, Anka Bedetti, Michelle Henley, Jerenimo Lepirei, and Frank Pope. Extracting identifying contours for african elephants and humpback whales using a learned appearance model. pages 1265–1274, 03 2020. 2
- [34] Hendrik J. Weideman, Zachary M. Jablons, Jason A. Holmberg, Kiirsten Flynn, John Calambokidis, Reny B. Tyson, Jason B. Allen, Randall S. Wells, Krista Hupman, Kim Urian, and Charles V. Stewart. Integral curvature representation and matching algorithms for identification of dolphins and whales. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pages 2831–2839, 2017. 2
- [35] WildMe. Beluga ID 2022, 2022. 2
- [36] WildMe. Leopard ID 2022, 2022. 2
- [37] Claire Witham and Emily Bethell. Macaque faces. figshare. Dataset, 2019. 2

- [38] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person reidentification: A survey and outlook, 2021.
- [39] Vojtěch Čermák, Lukas Picek, Lukáš Adam, and Kostas Papafitsoros. Wildlifedatasets: An open-source toolkit for animal re-identification, 2023. 1, 2, 4, 5

# A. Appendix

# A.1. Training

Optimized with optuna.

Augmentation: rand color sharpening, clahe, shift (25%), scale (20%) and rotation (15 degree), color jitter

Learning rate schedule: linear warmup up in first 15 epoch (1.5e-5, 1.5e-3), with exponential decay 0.8 per epoch

Batch size: 112 Image size: 256

Subcenter ArcFace with dynamic margins: k = 3, scale of 51.5, and margin is dynamic (started at 0.5), but calculated by class frequency.

Backbone is EfficientNetV2-M pre-trained on ImageNet-1K.

Final layer: GeM pooling + BatchNorm

Feature dimension: 2,152

# A.2. Top-K Evaluation Results

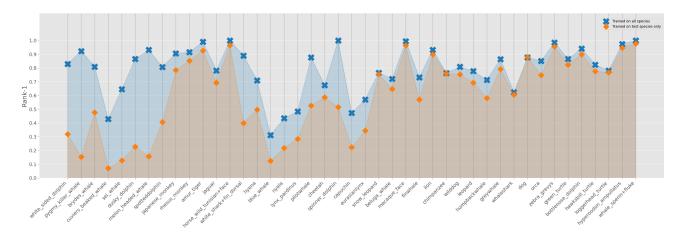


Figure 5. Comparison of top-1 performance of multi-species and single-species models. Species are ordered by increasing numbers of sightings.

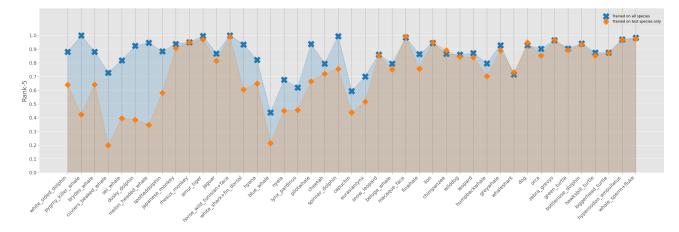


Figure 6. Comparison of top-5 performance of multi-species and single-species models. Species are ordered by increasing numbers of sightings.

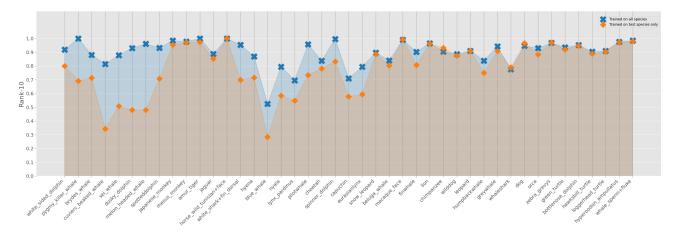


Figure 7. Comparison of top-10 performance of multi-species and single-species models. Species are ordered by increasing numbers of sightings.

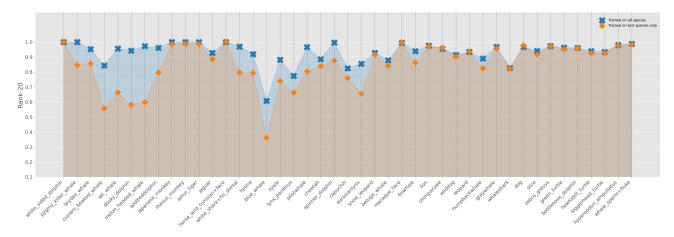


Figure 8. Comparison of top-20 performance of multi-species and single-species models. Species are ordered by increasing numbers of sightings.

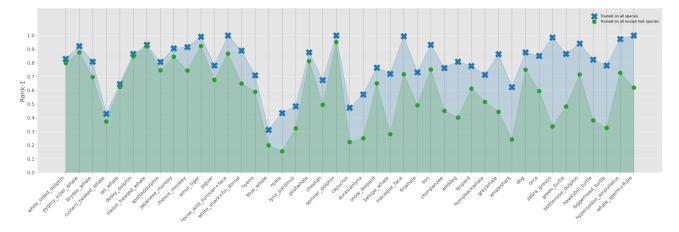


Figure 9. Top-1 performance comparison per species for model trained on the full dataset and model trained on all species **except the test species**.

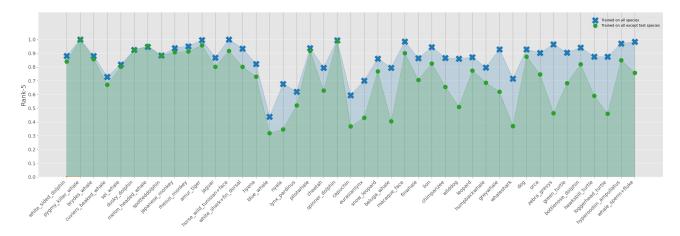


Figure 10. Top-5 performance comparison per species for model trained on the full dataset and model trained on all species **except the test species**.

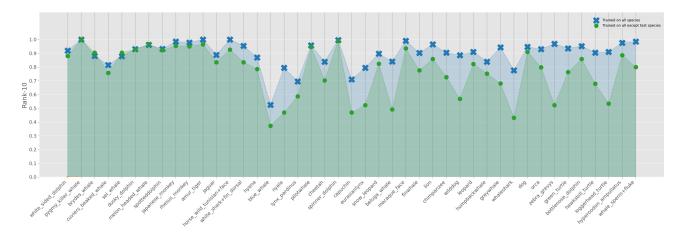


Figure 11. Top-10 performance comparison per species for model trained on the full dataset and model trained on all species **except the test species**.

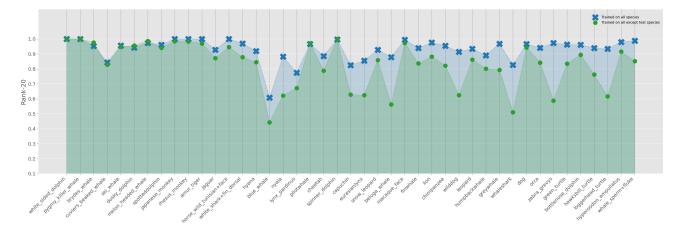


Figure 12. Top-20 performance comparison per species for model trained on the full dataset and model trained on all species **except the test species**.