



Multispecies facial detection for individual identification of wildlife: a case study across ursids

Melanie Clapham^{1,2} · Ed Miller² · Mary Nguyen² · Russell C. Van Horn³

Received: 9 March 2021 / Accepted: 6 August 2021 / Published online: 7 April 2022
© The Author(s) 2022

Abstract

To address biodiversity decline in the era of big data, replicable methods of data processing are needed. Automated methods of individual identification (ID) via computer vision are valuable in conservation research and wildlife management. Rapid and systematic methods of image processing and analysis are fundamental to an ever-growing need for effective conservation research and practice. Bears (ursids) are an interesting test system for examining computer vision techniques for wildlife, as they have variable facial morphology, variable presence of individual markings, and are challenging to research and monitor. We leveraged existing imagery of bears living under human care to develop a multispecies bear face detector, a critical part of individual ID pipelines. We compared its performance across species and on a pre-existing wild brown bear *Ursus arctos* dataset (*BearID*), to examine the robustness of convolutional neural networks trained on animals under human care. Using the multispecies bear face detector and retrained sub-applications of *BearID*, we prototyped an end-to-end individual ID pipeline for the declining Andean bear *Tremarctos ornatus*. Our multispecies face detector had an average precision of 0.91–1.00 across all eight bear species, was transferable to images of wild brown bears (AP=0.93), and correctly identified individual Andean bears in 86% of test images. These preliminary results indicate that a multispecies-trained network can detect faces of a single species sufficiently to achieve high-performance individual classification, which could speed-up the transferability and application of automated individual ID to a wider range of taxa.

Keywords Bears · Deep learning · Face recognition · Individual ID · Machine learning · Ursidae

Introduction

Conservation technologies can enhance the collection, analysis and sharing of wildlife-related data, and the implementation and evaluation of global conservation action (Lahoz-Monfort et al. 2019). Computer vision within ecology and

conservation increasingly facilitates the description of image features, counting within images, and identity classification (Weinstein 2018). This automated approach enables rapid processing and classification of large datasets in a standardized way that will be vital for global data sharing (Steenweg et al. 2017; Ahumada et al. 2020). Machine learning, and more specifically deep learning techniques, are now a focus of image classification, with emphasis on species and individual identification (ID) (Christin et al. 2019).

Computer vision enhances the integration of individual ID into ecological research, with the benefit of images being sourced from camera trap surveys and citizen science/ecotourists, allowing researchers to monitor species over broader scales (Berger-Wolf et al. 2017; Araujo et al. 2019; Schneider et al. 2019; Nipko et al. 2020). Individual photo ID has primarily focused on species with unique, stable body markings, such as the Grévy's zebra *Equus grevyi* (Crall et al. 2013) and the Northern giraffe *Giraffa camelopardalis* (Miele et al. 2020), other morphological traits and scars (Kelly and Holub 2008), or body parts such as fin shape

Handling editors: Leszek Karczmarski and Daniel I. Rubenstein.

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

✉ Melanie Clapham
melanie@bearid.org

¹ Department of Geography, University of Victoria, 3800 Finnerty Road, Victoria, BC V8P 5C2, Canada

² BearID Project, Sooke, BC, Canada

³ San Diego Zoo Wildlife Alliance, San Diego, CA, USA

(Hughes and Burghardt 2017). Face recognition is an alternative method of visual individual ID, especially when species lack distinguishing markings. Originally built for chimpanzees *Pan troglodytes* based on human facial recognition approaches (Loos and Ernst 2013), it has now been applied to other primates and selected large mammals (Deb et al. 2018; Körschens et al. 2018; Chen et al. 2020; Guo et al. 2020; Clapham et al. 2020), taking advantage of advances in deep learning techniques (Ravoor and T.S.B. 2020). Deep learning networks require large, labelled datasets that are difficult to acquire for wild animals, resulting in networks being trained primarily using images of individuals under human care (e.g., Freytag et al. 2016; Chen et al. 2020). Currently, a tradeoff exists between the need for reliable training data of individuals and the need for robust training data that reflects the contexts in which deep learning networks will be used (see Beery et al. 2018); the former favouring images taken ex situ and the latter favouring images of wild animals in situ (Schofield et al. 2019; Clapham et al. 2020). A possible solution is to train networks that generalise across species, sourcing larger datasets with diverse facial characteristics and background context, which may increase the robustness of ex situ-trained networks.

Bears (ursids) present an important focal taxonomic family for examining the application of computer vision to wildlife ecology, as they represent lineages that have diverged ecologically and morphologically over ~ 12.5 million years (Kutschera et al. 2014), inhabit habitats ranging from ice floes and deserts to forests, they are elusive, generally solitary and do not defend strict territories (Penteriani and Melletti 2020), which makes them challenging to research and monitor compared to other large carnivores. In addition, bear conservation and management would benefit from improved research tools. Six of the eight extant species of bear are considered vulnerable to extinction, with decreasing population trends for four of these six species (Scotson et al. 2017; Velez-Liendo and García-Rangel 2017; Dharaiya et al. 2020; Garshelis and Steinmetz 2020). Knowledge gaps exist across species, particularly for declining species across Asia (Asiatic black bear *Ursus thibetanus*, sun bear *Helarctos malayanus*, sloth bear *Melursus ursinus*) and South America (Andean bear *Tremarctos ornatus*). Tools in machine learning, coupled with monitoring techniques such as camera trapping, could help researchers to generate knowledge on population size, demography, and behaviour (Christin et al. 2019).

Within the *Ursidae*, the presence of distinguishing marks varies by species; Andean bears, Asiatic black bears, sun bears, and to some extent giant pandas *Ailuropoda melanoleuca*, generally possess varied markings on the face or chest that human observers have used for visual identification of individuals (Higashide et al. 2012; Ngoprasert et al. 2012; Zheng et al. 2016; Molina et al. 2017; Appleton et al.

2018; Penteriani et al. 2020; Rodríguez et al. 2020; Morrell et al. 2021), whereas American black bears *U. americanus*, brown bears *U. arctos*, polar bears *U. maritimus*, and sloth bears generally do not (although see Shimozuru et al. 2017). However, there is concern regarding the standardization of manual individual ID approaches across different researchers, for wildlife with and without distinguishing markings (Choo et al. 2020; Johansson et al. 2020). In addition, markings may not be entirely stable over time for some species (Yoshizaki et al. 2009; Van Horn et al. 2015) or easily distinguished under different environmental conditions from photographs (Reyes et al. 2017), and visual individual ID is labor-intensive, especially as multiple trained observers may be required (Ramsey et al. 2019; Morrell et al. 2021). Automated individual ID was recently developed for brown bears (Clapham et al. 2020) and giant pandas (Chen et al. 2020) using deep-learning approaches of facial recognition. Deep convolutional neural networks (CNN) could make use of distinguishing facial markings depending on their stability, but also benefit from other biometric features of an animal's face. Automated facial recognition may provide a standardized and reproducible approach of individual ID across multiple taxa that could be more broadly accessible to researchers.

We used deep-learning techniques to develop a multispecies facial detector across all eight members of the *Ursidae* family. We used this multispecies detector, trained on images of bears at zoos and rescue centers, to: (1) evaluate the variation in its performance for each ursid species, (2) compare its performance to a single-species trained detector on a pre-existing test set of wild brown bears, and (3) develop an example end-to-end pipeline in combination with a retrained face encoder and SVM (classifier) for automated individual ID of a novel species, the Andean bear. We compared results to those presented in Clapham et al. (2020), where networks were trained and tested on wild brown bears. A multispecies facial detector could reduce development time and advance the application of automated individual ID across a broader range of species. In practice, face detectors could be linked to pre-existing animal object detectors and species classifiers, to pull images of focal species for individual ID.

Methods

Data collection

Images were sourced from bears housed in zoos and sanctuaries across North America, Europe, and Asia. Images were initially collected for separate research (Van Horn et al. 2014, 2015), and the dataset was supplemented for this study (Table 1). The only criteria for image inclusion was that both eyes of the bear needed to be visible for the object detector

Table 1 Summary of multispecies image dataset

Species	Number of face images (face chips)	Number of individuals	Mean face chips per individual
American black bear <i>Ursus americanus</i>	84	22	3.8
Andean bear <i>Tremarctos ornatus</i>	609	32	19.0
Asiatic black bear <i>U. thibetanus</i>	54	25	2.2
Brown bear <i>U. arctos</i>	588	49	12.0
Giant panda <i>Ailuropoda melanoleuca</i>	185	73	2.5
Polar bear <i>U. maritimus</i>	481	175	2.8
Sloth bear <i>Melursus ursinus</i>	92	34	2.7
Sun bear <i>Helarctos malayanus</i>	99	35	2.8
Total	2192	445	4.9

to find the face (see below) of a bear of known identity. Identification of individual bears were provided by the image source, with bear birthdate and sex sometimes sourced from individual zoos or, most commonly, from studbooks. The habitat of bear enclosures varied across the dataset, creating variation in the background of images (Fig. 1). The dates on which images were taken were sourced from the image metadata if present. The exact cameras (brand and model) used to take the photographs were unknown due to the post hoc nature of image collection, however, image resolution ranged from 0.06 to 3.84 megapixels. Images were JPEG, PNG and TIFF format. TIFF images were converted to JPEG prior to use.

Andean bear subset

We used the Andean bear subset (Table 1) for inclusion in our end-to-end example application. To minimize the impact of ontogenetic changes in morphology on overall success, only images of Andean bears over 2 years of age were included in the dataset, and only individuals with seven or more images were included to allow sufficient number of images for testing and training. This allowed a minimum number of two images per individual available for testing, a requirement of the testing methodology for evaluating similarity comparison networks (see “Testing methodology”). Images of Andean bears were taken from 2004 to 2021, with a mean span of 7.4 (\pm SD 4.1) years per individual. Eight

percent of total images were missing information for the year the image was taken. Of the individual bears included in the dataset, 13 were female and 19 were male. We used a randomly selected 80/20% split of the 609 images for training ($n = 488$) and testing ($n = 121$), respectively.

Golden dataset

The golden dataset is the manually annotated ‘gold standard’ dataset ($n = 2192$) and is used for training networks (with the training dataset) and evaluating their performance (with the testing dataset). We ran all images through the application *bearface* (Clapham et al. 2020) to speed-up creating a golden dataset, and manually adjusted erroneous or missing bounding boxes and facial landmarks using *imglab* from the Dlib toolkit (King 2009). The golden dataset also included labels of individual identification for the Andean bear subset.

Multispecies face detector development

We trained a multispecies face detector network (*face_allbears.dat*), using the sub-application (*bearface*) from *BearID* developed by Clapham et al. (2020), to evaluate its performance across species and on an existing wild brown bear dataset. We split the multispecies dataset (Table 1) into 80% for training and 20% for testing within species, and then combined these together across species, resulting in 1754 images for training and 438 for testing. The *BearID* pipeline is based on the FaceNet approach (Schroff et al. 2015) and consists of: (1) face detection (*bearface* sub-application), (2) face reorientation and cropping (*bearchip* sub-application), (3) face encoding (*bearembd* sub-application) and (4) face classification (*bearsvm* sub-application) (see Clapham et al. 2020 for a full description of the *BearID* application pipeline and hardware used). For this study, we used Microsoft Azure NC6s_v3 cloud computing instances to train and evaluate the multispecies face detector.

Bearface with the new network *face_allbears.dat*, hereafter *bearface*, finds faces and facial landmarks (i.e., the outer corners of the eyes) of multiple bear species in images. As in Clapham et al. (2020), it consists of an object detector (sliding window and CNN (Dalal and Triggs 2005; King 2015)) and a shape predictor [face alignment with an ensemble of regression trees (King 2009; Kazemi and Sullivan 2014)]. The object detector and shape predictor were trained using the bounding boxes and facial landmark labels from the golden dataset. See Clapham et al. (2020) for full equivalent training procedures. *Bearface* accepts JPEG and PNG file types as input images (other formats would need pre-conversion), and outputs an XML file that includes a list of the images with predicted face and landmark data for each.



Fig. 1 Example images showing variation of captive bears images across all eight species. Red boxes indicate faces detected using the multispecies face detector. Species from top left to bottom right:

American black bear, Asiatic black bear, Andean bear, brown bear, giant panda, polar bear, sloth bear, and sun bear

End-to-end Andean bear example pipeline

We used the Andean bear subset of the golden dataset with the multispecies face detector and the pre-existing bear recognition pipeline (*BearID*: Clapham et al. 2020), maintaining the same 80/20% split of the data for training and testing, respectively. *Andean BearID* (*BearID* with new networks), consists of four sub-applications: (1) *Bearface* finds bear faces, (2) *Bearchip* creates face chips from found faces (Fig. 2), (3) *Bearembd* creates embeddings for the face chips, and (4) *Bearsvm* determines individual ID from embeddings (Fig. 3). *Bearface* uses *face_allbears.dat* network, which required no additional training from that described in the previous section. Both the face encoder (*bearembd*) and face classifier (*bearsvm*) required individual-specific retraining with Andean bear images. *Bearchip* required no retraining. Face chips produced by *bearchip* (Fig. 2) using the Andean bear subset of the golden dataset were used to retrain the similarity comparison network *bearembd*, resulting in a new face embedding network, *embed_andeanbear.dat* (Fig. 3). The embeddings produced by that network were used to retrain *bearsvm*, resulting in a new face classifier network, *svm_andeanbear.dat* (see Code Availability for all code required for retraining networks). Hereafter, *bearembd* and *bearsvm* refer to their use collectively with the Andean bear-trained networks.

Testing methodology

Multispecies face detector

We followed the testing methodology of Clapham et al. (2020) and tested the object detector and shape predictor separately. We focus on interpolated average precision (area under a precision-recall curve) as a focal performance metric for the face detector overall, but also present precision ($\frac{\text{truepositive}}{\text{truepositive}+\text{falsepositive}}$) and recall ($\frac{\text{truepositive}}{\text{truepositive}+\text{falsenegative}}$). We tested performance of the facial detector across the trained species (Table 1), as well as on the test set of wild brown bears from Clapham et al. (2020) ($n=934$), to examine the performance of a captive-trained detector on wild-tested images.

Andean bear end-to-end pipeline

We tested the full end-to-end Andean bear application from input file to ID classification, but also considered the performance of sub-applications *bearembd* and *bearsvm* separately, to assess performance without cumulative error by comparing results to labels from the golden dataset. The Andean bear face encoder (*bearembd*) was tested using pairs of images generated from the test split of the golden

dataset ($n=121$). The paired test set represented 230 matching pairs of images (same individual) and 230 non-matching pairs (different individual). There were even numbers of matching and non-matching pairs, no face chip was compared to itself, and each pair within the test set was unique. We further evaluated the predictive capability of the embedding network for both known individuals and unknown (or new) individuals by applying fivefold validation across two test regimes:

1. Folds across all face chips, whereby different chips for the same individual appear in every fold. Paired tests represent 214 matching and 214 non-matching pairs.
2. Folds by ID label, whereby all chips from an individual appear in only one fold. Paired tests represent 400 matching and 400 non-matching pairs.

We evaluated performance using accuracy: $(\text{true positive rate} \times \text{positive ratio}) + (\text{true negative rate} \times \text{negative ratio})$, positive and negative ratio = 0.5, and F1 score: $2 \times \left(\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right)$, which is the harmonic mean of precision and recall. Precision here refers to correctly identified matching pairs from all predicted matching pairs. Recall refers to correctly identified matching pairs out of all the actual matching pairs. F1 score may be a preferred metric of performance when classes are imbalanced. In all cases, we used a closed-set approach (Deb et al. 2018). *Bearsvm* was evaluated by comparing the test set accuracy (number of correct ID predictions/total number of ID predictions) of predicted ID labels to those in the golden dataset.

Results

Multispecies face detector

The average precision (intersection over union = 0.5) of the multispecies object detector varied among bear species, with an overall performance across species of 0.959 (Table 2). The overall mean normalised distance between the facial landmarks of the golden dataset and those predicted by the shape predictor was 0.083 ± 0.115 (Table 2); in other words, ~8% of the distance between the outer corners of the eyes.

The average precision of the multispecies object detector (trained on captive bears) when tested on a wild brown bear dataset was 0.929, which is similar to the performance of a detector trained on a wild brown bear dataset (Table 3; Fig. 4). For the wild brown bear dataset tested with the multispecies detector, the mean normalised distance between the facial landmarks of the golden dataset and those predicted by the shape predictor was 0.161 ± 0.155 (Table 3); ~16% of the distance between the outer corners of the eyes.



Fig. 2 Example face chips produced by the multispecies detector (*bearface*) and the reorientation/cropping application (*bearchip*); **a–d** different individual Andean bears displaying variation in facial appearance within and among individuals, **a** and **b** are males, while **c** and **d** are females

Fig. 3 Schematic of *BearID* pipeline including programming languages and file formats, adapted from Clapham et al. (2020). Blue boxes indicate the new networks trained in this study. The process begins with an input image, the face of the bear is detected using the multispecies network of *bearface*, the face is cropped and rotated using *bearchip*, an embedding is created using the Andean bear network of *bearembded*, the embedding is then matched using Andean bear *bearsvm* to produce a classification and output individual ID, e.g., Andean bear 'I843')

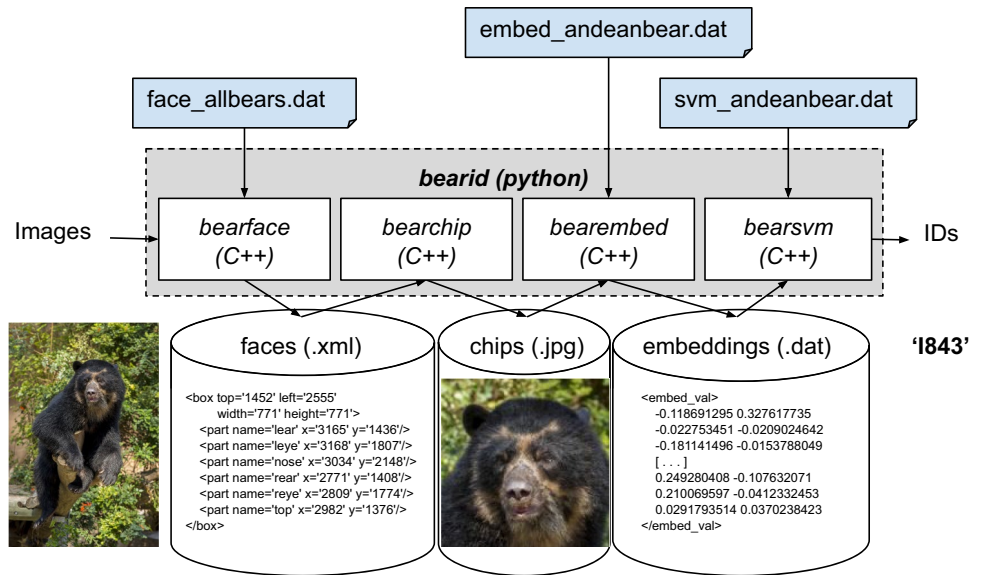


Table 2 Testing performance of the multispecies-trained facial detector

Test species	Test <i>n</i>	Precision (OD)	Recall (OD)	Average precision, IoU=0.5 (OD)	Mean (\pm SD) normalised distance (SP)
American black bear	17	1.000	1.000	1.000	0.144 \pm 0.188
Andean bear	121	0.975	0.983	0.978	0.061 \pm 0.057
Asiatic black bear	11	1.000	0.909	0.909	0.150 \pm 0.148
Brown bear	118	0.983	0.966	0.966	0.085 \pm 0.103
Giant panda	37	1.000	0.973	0.973	0.103 \pm 0.229
Polar bear	96	0.989	0.979	0.969	0.072 \pm 0.090
Sloth bear	18	0.947	1.000	1.000	0.127 \pm 0.122
Sun bear	20	0.905	0.950	0.920	0.091 \pm 0.080
All species	438	0.975	0.970	0.959	0.083 \pm 0.115

OD object detector, SP shape predictor, IoU intersection over union

Table 3 Comparing the testing performance of the multispecies detector (trained on images of captive bears) on images of wild brown bears and to results from a species-specific detector

Trained detector	Test set	Test <i>n</i>	Precision (OD)	Recall (OD)	Average precision, IoU=0.5 (OD)	Mean (\pm SD) normalised distance (SP)
Multispecies (captive-trained)	Wild brown bear ^a	934	0.980	0.939	0.929	0.161 \pm 0.155
Brown bear (wild-trained) ^b	Wild brown bear	934	0.986	0.983	0.977	0.111 \pm 0.122

OD object detector, SP shape predictor, IoU intersection over union

^aRepresents a wild bear dataset (from Clapham et al. 2020) tested on a captive bear-trained detector

^bBaseline comparative results from Clapham et al. (2020)

End-to-end Andean bear pipeline

Using the multispecies detector network with *bearface* (Table 3), the original *bearchip* and the newly trained networks for *bearembded* and *bearsvm*, the end-to-end Andean bear pipeline correctly predicted the ID for 104 out of 121

images (closed-set accuracy = 86.0%). *Bearface* detected 120 of 121 Andean bear faces, plus 2 erroneous faces were detected in one image for a total of 122 faces detected. We manually removed these erroneous detections from the pipeline at this stage to maintain an accurate evaluation of



Fig. 4 Automated face detection (red boxes) using the multispecies face detector on images of wild brown bears

Table 4 Comparing the performance of the similarity comparison network across three test methods: golden test set, folds by face chips, and folds by ID label

Test method	Test $n \pm \text{SD}$	Pairs n	Accuracy (%) $\pm \text{SD}$	TPR (%) $\pm \text{SD}$	TNR (%) $\pm \text{SD}$	F1-score (%) $\pm \text{SD}$
Golden test split	121	230	90.9	85.2	96.5	90.3
Folds by chip	121.8 \pm 3.2	214	90.3 \pm 3.0	83.4 \pm 5.7	97.2 \pm 0.9	89.5 \pm 3.6
Folds by ID	121.8 \pm 31	400	78.9 \pm 6.3	75.2 \pm 5.5	82.5 \pm 9.5	78.1 \pm 6.2

Pairs n = number of equal matching and non-matching pairs of face chips per test regime, *TPR* true positive rate, *TNR* true negative rate

bearembd and *bearsvm*. Of the 120 correctly detected faces, 104 were correctly identified.

For sub-application analysis using the test split of the Andean bear subset of the golden dataset ($n = 121$), the face encoder (*bearembd*) is effective at predicting matching and non-matching pairs with an accuracy of 90.9% (Table 4). A receiver operating characteristic curve (ROC) displays the performance of *bearembd* at different thresholds of true positive rate [TPR (recall/sensitivity)] and false positive rate [FPR (specificity); Fig. 5]. Further evaluation of *bearembd* using the two fivefold test regimes previously described, folds across all face chips and folds by ID label,

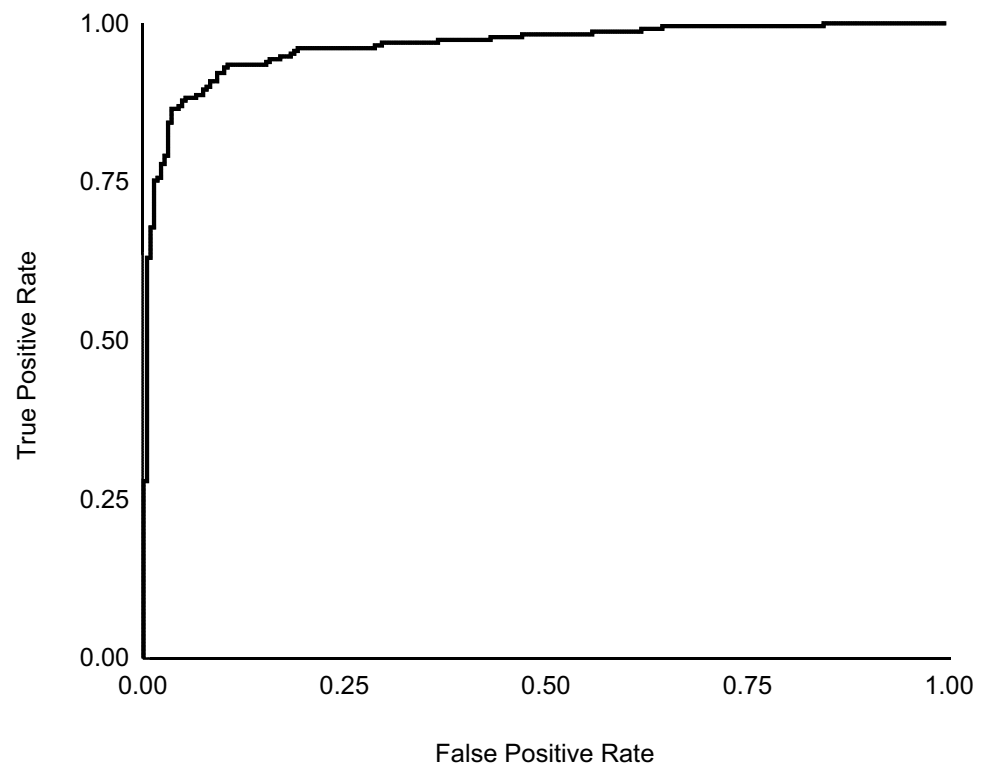
shows mean accuracies of $90.3 \pm 3.0\%$ and $78.9 \pm 5.5\%$, respectively (Table 4).

The cumulative error of the face detection and embedding resulted in a drop in classification (*bearsvm*) accuracy from 89.3% (108 out of 121 correct IDs: golden dataset test) to 86.7% (104 out of 120 correct IDs).

Discussion

Our multispecies facial detector network, trained on images of bears under human care, performed well, resulting in an average precision of 0.9–1.0 for every bear species used

Fig. 5 A receiver operating characteristic (ROC) curve showing the probability of the face encoder (*bearembed*) predicting matching (same individual) or non-matching pairs (different individual) of face chips under different thresholds. The steep curve towards the top left of the graph indicates high predictive utility at separating matching/non-matching pairs



to train the network. Despite the relatively low number of training images ($n = 1754$ total) compared to large datasets typical of deep learning approaches, our results are comparable to single-species trained facial detection networks [African forest elephant *Loxodonta cyclotis*: 0.98, $n = 1573$ training images, Körschens et al. (2018); Western gorilla *Gorilla gorilla*: 0.91, $n = 2000$ training images, Brust et al. (2017); Giant panda: 1.0, $n = 5854$ training images, Chen et al. (2020)], and other detectors that focus on whole-body shape or focal body areas [Northern giraffe: 0.89 accuracy, Buehler et al. (2019); luderick *Girella tricuspidata*: 0.93, Ditria et al. (2020)]. Multispecies-trained detectors have more commonly been used in whole-body detection for species recognition, with performance in the range of 0.55 (AP_n) to 0.97 (accuracy) (Loos et al. 2018; Norouzzadeh et al. 2018). MegaDetector (Beery et al. 2019b) is a multispecies (generalised) whole-body detector that can be used to remove empty frames and train project-specific classifiers from camera trap images. Our multispecies facial detector is intended to perform a similar function for the detection of faces for use in individual ID of bears, which could be replicated for other taxonomic families using our open source code (see Code Availability). Guo et al. (2020) recently developed a multispecies facial detector trained on 41 primate species and 4 carnivores resulting in detection accuracies of 0.91, 0.98, and 0.98 for golden snub-nosed monkeys *Rhinopithecus roxellana*, Tibetan macaques *Macaca thibetana*, and tigers *Panthera tigris*, respectively. In addition,

Khan et al. (2020) present AnimalWeb, an annotated dataset of animal faces for 334 species across 21 orders, which achieves a class-wise face detection mean average precision of 0.64.

We only used images of bears where an individual identification was known, to avoid unintentionally training and testing a detector on images of a very low number of individuals, which could influence performance. This resulted in sample sizes being skewed by species, which could have biased the facial detector network in favour of those species with a larger dataset. However, two of the three species with the lowest sample sizes (American black bear, sloth bear), had the highest possible average precision (1.000), whereas the third species (Asiatic black bear) had the lowest average precision of all species (0.909), indicating the influence of additional variance beyond sample size. Further testing, on a larger sample size per species, should provide more information on the network's performance. In addition, due to the post hoc nature of image collection, we could not account for the date the images were taken across the whole dataset, leading to potential data leakage if images recorded on the same day were mixed between the training and testing datasets. Future research in this area should attempt to maintain a 24-h window (image independence) between image inclusion across the training or test datasets.

We have demonstrated how a detector trained across all eight bear species under human care, can be effective at detecting the faces of wild bears. This finding suggests

that images of wildlife under human care may be useful in training deep-learning networks for use with images in field settings, whose collection can be challenging, and whose manual processing can be labor-intensive. We postulate that the relatively high average precision of our detector on wild brown bears could be due to variation in housing environments of the different species included in the dataset, as well as variation in facial biometrics from the inclusion of multiple species for training. Using a wild brown bear test dataset, when comparing the performance of a wild brown bear-trained detector to the performance of our new multispecies detector, precision was consistent between detectors (0.99 and 0.98, respectively), but recall was reduced (0.98 and 0.94, respectively). This suggests that the multispecies detector missed some faces of bears taken in situ, which may be due to additional background complexity (see Beery et al. 2018), pose variation, or size of the face within the image. The shape predictor improved at detecting the facial landmarks of bears under human care (all species combined: 0.083 ± 0.115 ; brown bears: 0.085 ± 0.103), compared to the shape predictor trained and tested on wild brown bears (0.111 ± 0.122 ; Clapham et al. 2020). However, when the multispecies shape predictor was tested on wild brown bears, its performance declined (0.161 ± 0.155), which could indicate a slight dissimilarity between the dataset of bears under human care compared to wild bears. Further analyses of performance should test images of wild bears for the other seven members of the *Ursidae*, to better evaluate its multispecies application in situ. Combining images of wildlife taken under human care with those taken in situ, or augmenting backgrounds of ex situ images (Beery et al. 2019a), could enhance the robustness of these datasets.

Using the multispecies face detector, we developed a pipeline for its use in automated individual identification of an example species, the Andean bear. The retrained face encoder (*bearembbed*) obtained an accuracy of 90.9% for predicting matching and non-matching pairs of images of Andean bears in the golden test set. Five-fold analysis by chip and by label obtained accuracies of 90.3% and 78.9%, respectively, which outperforms the face encoder developed for wild brown bears for both methods (84.2% and 71.3%; Clapham et al. 2020). Other studies evaluating the performance of wildlife-focused deep learning networks for variants of face verification found similar or slightly reduced accuracies [lemurs *Lemuroidea* spp.: 83.1% (Deb et al. 2018); golden monkeys *Cercopithecus mitis kandti*: 78.7% (Deb et al. 2018); chimpanzees: 59.9% (Deb et al. 2018), 0.811 (mAP@1) (Schneider et al. 2020)], although our dataset has fewer images in comparison to these examples [this study: 609 images; Deb et al. (2018): 3000, 1450, and 5599 images, respectively; Schneider et al. (2020): 5599]. We suspect the facial markings of Andean bears may have contributed to the relatively high performance of our face

encoder, but more test images are required to fully interpret its performance. Although our dataset is modest and we present this software as proof of concept rather than immediately applicable, the images used to train *bearembbed* represent a long-term dataset of Andean bears under human care taken over at least 17 years. This undoubtedly added variation to the dataset beyond images taken on the same day, for example; especially as the facial appearance of Andean bears, and other bear species, can change over time (Yoshizaki et al. 2009; Van Horn et al. 2015; Clapham et al. 2020). The habitat of bear enclosures varied across the dataset, creating variation in the background of images, however, enclosure similarity within images of the same individuals may be a confounding variable positively influencing performance. While accuracy of *bearembbed* was reduced when testing folds by ID label, it still shows good predictive utility (78.9%) for use in matching of new individuals. Face encoders that use similarity comparison networks, such as *bearembbed*, are a promising tool for individual ID of wildlife due to their ability to generalise, allowing for the identification of new individuals that are not contained in the training dataset (Schneider et al. 2020). This is vital for wildlife studies looking to use automated approaches of individual ID for population assessments, such as spatial capture–recapture (SCR).

We developed a multispecies bear face detector using images of bears under human care that achieved a high level of performance across all eight bear species. This performance was transferable to a wild brown bear dataset, although further analysis using images of wild counterparts across all bear species is required to fully determine its application. Our automated end-to-end Andean bear example application correctly identified the individual in 85.9% of images inputted. These initial results indicate that a multispecies-trained face detection network can detect faces of a single species sufficiently to achieve high performance for individual classification. This could be important as within the bounds of human capacity, funding availability for wildlife research, and timeliness of conservation action required, it may not be possible to develop separate detectors and classifiers for every species at risk that could benefit from automated methods of individual identification. Automated methods are increasingly needed due to an ever-growing need for effective conservation (Dirzo et al. 2014) and expanding awareness that objective and replicable methods to identify individuals are needed to avoid undesired conservation outcomes (Choo et al. 2020; Johansson et al. 2020). Developing robust pipelines of automated individual ID will enable rapid and systematic data collection and processing on taxa of conservation concern, boosting the existing power of research to better inform conservation planning and management.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42991-021-00168-5>.

Acknowledgements We are grateful to several people who volunteered their time to the San Diego Zoo Wildlife Alliance and collected images of individual bears living under human care. Numerous organizations and individual photographers generated the images from which *bearface*, *bearembed*, and *bearsym* extracted information. The images of zoo-housed bears used in Figures 1, 2 and 3 were compiled by Lisa Bissi and are used courtesy of the San Diego Zoo Wildlife Alliance. Two images in Fig. 1 were provided by STIFTUNG für BÄREN who, in addition to Bears in Mind, provided additional image datasets that were used for training and testing *bearface*. Images of wild brown bears used in analysis were taken in the traditional territory of the Da'naxda'xw Awaetlala First Nation and Katmai National Park and Preserve, Alaska.

Author contributions MC, EM, MN and RCVH conceived the ideas and designed the methodology; MC and RCVH collected the data; EM and MN developed the models and analysed the results; MC, EM, MN and RCVH led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Funding Cloud computing (Microsoft Azure) credits for this project were granted by Microsoft AI for Earth program. MC is supported by the Natural Sciences and Engineering Research Council of Canada (CRDPJ 523329-18; ALLRP 559534-20), in combination with industry partners (Knight Inlet Lodge and Wild Bear Lodge) and Nanwakolas Council. RCVH is supported by the San Diego Zoo Wildlife Alliance. Google for Nonprofits funded Google Workspace which helped to facilitate this collaborative project.

Availability of data and materials (data transparency) Not applicable.

Code availability *BearID* is an open-source application available on GitHub at <https://github.com/hypraptive/bearid> (version 21.03). The pretrained models are available at <https://github.com/hypraptive/bearid-models> (version 21.03).

Declarations

Conflict of interest We declare no competing interests.

Ethical approval Images of wild brown bears used in this study were collected in accordance with the University of Victoria Animal Care Committee (2014-031(1-3)).

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

References

- Ahumada JA, Fegraus E, Birch T, Flores N, Kays R, O'Brien TG, Palmer J, Schuttler S, Zhao JY, Jetz W, Kinnaird M, Kulkarni S, Lyet A, Thau D, Duong M, Oliver R, Dancer A (2020) Wildlife Insights: a platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. *Environ Conserv* 47:1–6. <https://doi.org/10.1017/S0376892919000298>
- Appleton RD, Van Horn RC, Noyce KV, Spady TJ, Swaisgood RR, Arcese P (2018) Phenotypic plasticity in the timing of reproduction in Andean bears. *J Zool* 305:196–202. <https://doi.org/10.1111/jzo.12553>
- Araujo G, Agustines A, Tracey B, Snow S, Labaja J, Ponzo A (2019) Photo-ID and telemetry highlight a global whale shark hotspot in Palawan, Philippines. *Sci Rep* 9:1–12. <https://doi.org/10.1038/s41598-019-53718-w>
- Beery S, Van Horn G, Perona P (2018) Recognition in Terra Incognita. *Lect Notes Comput Sci* 11220 LNCS:472–489. https://doi.org/10.1007/978-3-030-01270-0_28
- Beery S, Liu Y, Morris D, Pivavsky J, Kapoory A, Meister M, Perona P (2019a) Synthetic examples improve generalization for rare classes. *arXiv*: 1904.05916
- Beery S, Morris D, Yang S (2019b) Efficient pipeline for camera trap image review. *arXiv*: 1907.06772
- Berger-Wolf TY, Rubenstein DI, Stewart C V., Holmberg JA, Parham J, Menon S (2017) Wildbook: Crowdsourcing, computer vision, and data science for conservation. *arXiv*: 1710.08880
- Brust CA, Burghardt T, Groenenberg M, Käding C, Kühl HS, Mangue ML, Denzler J (2017) Towards automated visual monitoring of individual gorillas in the wild. In: *Proceedings—2017 IEEE international conference on computer vision workshops, ICCVW 2017*. pp 2820–2830. <https://doi.org/10.1109/ICCVW.2017.333>
- Buehler P, Carroll B, Bhatia A, Gupta V, Lee DE (2019) An automated program to find animals and crop photographs for individual recognition. *Ecol Inform* 50:191–196. <https://doi.org/10.1016/j.ecoinf.2019.02.003>
- Chen P, Swarup P, Matkowski WM, Kong AWK, Han S, Zhang Z, Rong H (2020) A study on giant panda recognition based on images of a large proportion of captive pandas. *Ecol Evol* 10:3561–3573. <https://doi.org/10.1002/ece3.6152>
- Choo YR, Kudavidanage EP, Amarasinghe TR, Nimalrathna T, Chua MAH, Webb EL (2020) Best practices for reporting individual identification using camera trap photographs. *Glob Ecol Conserv* 24:e01294. <https://doi.org/10.1016/j.gecco.2020.e01294>
- Christin S, Hervet E, Lecomte N (2019) Applications for deep learning in ecology. *Methods Ecol Evol* 10:1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Clapham M, Miller E, Nguyen M, Darimont CT (2020) Automated facial recognition for wildlife that lack unique markings: a deep learning approach for brown bears. *Ecol Evol* 10:12883–12892. <https://doi.org/10.1002/ece3.6840>
- Crall JP, Stewart CV, Berger-Wolf TY, Rubenstein DI, Sundaresan SR (2013) HotSpotter-patterned species instance recognition. In: *2013 IEEE workshop on applications of computer vision (WACV)*. IEEE, Clearwater Beach, FL, pp 230–237. <https://doi.org/10.1109/WACV.2013.6475023>
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp 886–893. <https://doi.org/10.1109/CVPR.2005.177>
- Deb D, Wiper S, Gong S, Shi Y, Tymoszek C, Fletcher A, Jain AK (2018) Face recognition: primates in the wild. In: *2018 IEEE 9th international conference on biometrics theory, applications and systems, BTAS 2018*, pp 1–10. <https://doi.org/10.1109/BTAS.2018.8698538>

- Dharaiya N, Bargali HS, Sharp T (2020) *Melursus ursinus* (amended version of 2016 assessment). IUCN Red List Threat Species 2020:e.T13143A166519315. <https://doi.org/10.2305/IUCN.UK.2020-1.RLTS.T13143A166519315.en>
- Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJB, Collen B (2014) Defaunation in the Anthropocene. *Science* (80-) 345:401–406. <https://doi.org/10.1126/science.1251817>
- Ditria EM, Lopez-Marciano S, Sievers M, Jinks EL, Brown CJ, Connolly RM (2020) Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning. *Front Mar Sci* 7:1–9. <https://doi.org/10.3389/fmars.2020.00429>
- Freytag A, Rodner E, Simon M, Loos A, Kühl HS, Denzler J (2016) Chimpanzee faces in the wild: Log-euclidean CNNs for predicting identities and attributes of primates. In: Rosenhahn B, Andres B (eds) *Pattern recognition*. GCPR 2016. Lecture notes in computer science, vol 9796. Springer, Cham, pp 51–63. https://doi.org/10.1007/978-3-319-45886-1_5
- Garshelis D, Steinmetz R (2020) *Ursus thibetanus* (amended version of 2016 assessment). IUCN Red List Threat Species 2020:e.T22824A166528664. <https://doi.org/10.2305/IUCN.UK.2020-3.RLTS.T22824A166528664.en>
- Guo S, Xu P, Miao Q, Shao G, Chapman CA, Chen X, He G, Fang D, Zhang H, Sun Y, Shi Z, Li B (2020) Automatic identification of individual primates with deep learning techniques. *iScience* 23:101412. <https://doi.org/10.1016/j.isci.2020.101412>
- Higashide D, Miura S, Miguchi H (2012) Are chest marks unique to Asiatic black bear individuals? *J Zool* 288:199–206. <https://doi.org/10.1111/j.1469-7998.2012.00942.x>
- Hughes B, Burghardt T (2017) Automated visual fin identification of individual Great White Sharks. *Int J Comput vis* 122:542–557. <https://doi.org/10.1007/s11263-016-0961-y>
- Johansson Ö, Samelius G, Wikberg E, Chapron G, Mishra C, Low M (2020) Identification errors in camera-trap studies result in systematic population overestimation. *Sci Rep* 10:1–10. <https://doi.org/10.1038/s41598-020-63367-z>
- Kazemi V, Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. In: 2014 IEEE conference on computer vision and pattern recognition. pp 1867–1874. <https://doi.org/10.1109/CVPR.2014.241>
- Kelly MJ, Holub EL (2008) Camera trapping of carnivores: trap success among camera types and across species, and habitat selection by species, on Salt Pond Mountain, Giles County, Virginia. *North-east Nat* 15:249–262. [https://doi.org/10.1656/1092-6194\(2008\)15\[249:CTOCTS\]2.0.CO;2](https://doi.org/10.1656/1092-6194(2008)15[249:CTOCTS]2.0.CO;2)
- Khan MH, McDonagh J, Khan S, Shahabuddin M, Arora A, Khan FS, Shao L, Tzimiropoulos G (2020) AnimalWeb: a large-scale hierarchical dataset of annotated animal faces. In: *Proceedings of IEEE computer society conference on computer vision pattern recognition*, pp 6937–6946. <https://doi.org/10.1109/CVPR42600.2020.00697>
- King DE (2009) Dlib-ml: a machine learning toolkit. *J Mach Learn Res* 10:1755–1758
- King DE (2015) Max-margin object detection. *arXiv*: 1502.00046
- Körschens M, Barz B, Denzler J (2018) Towards automatic identification of elephants in the wild. *arXiv*: 1812.04418
- Kutschera VE, Bidon T, Hailer F, Rodi JL, Fain SR, Janke A (2014) Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Mol Biol Evol* 31:2004–2017. <https://doi.org/10.1093/molbev/msu186>
- Lahoz-Monfort JJ, Chadès I, Davies A, Fegraus E, Game E, Guiller-Arroita G, Harcourt R, Indraswari K, McGowan J, Oliver JL, Refisch J, Rhodes J, Roe P, Rogers A, Ward A, Watson DM, Watson JEM, Wintle BA, Joppa L (2019) A call for international leadership and coordination to realize the potential of conservation technology. *Bioscience* 69:823–832. <https://doi.org/10.1093/biosci/biz090>
- Loos A, Ernst A (2013) An automated chimpanzee identification system using face detection and recognition. *EURASIP J Image Video Process*. <https://doi.org/10.1186/1687-5281-2013-49>
- Loos A, Weigel C, Koehler M (2018) Towards automatic detection of animals in camera-trap images. In: *European signal processing conference 2018-September*, pp 1805–1809. <https://doi.org/10.23919/EUSIPCO.2018.8553439>
- Miele V, Dussert G, Spataro B, Chamailé-Jammes S, Allainé D, Bonenfant C (2020) Revisiting giraffe photo-identification using deep learning and network analysis. *bioRxiv* 2020.03.25.007377. <https://doi.org/10.1101/2020.03.25.007377>
- Molina S, Fuller AK, Morin DJ, Royle JA (2017) Use of spatial capture–recapture to estimate density of Andean bears in northern Ecuador. *Ursus* 28:117. <https://doi.org/10.2192/URSU-D-16-00030.1>
- Morrell N, Appleton RD, Arcese P (2021) Roads, forest cover, and topography as factors affecting the occurrence of large carnivores: the case of the Andean bear (*Tremarctos ornatus*). *Glob Ecol Conserv* 26:e01473. <https://doi.org/10.1016/j.gecco.2021.e01473>
- Ngoprasert D, Reed DH, Steinmetz R, Gale GA (2012) Density estimation of Asian bears using photographic capture–recapture sampling based on chest marks. *Ursus* 23:117–133. <https://doi.org/10.2192/URSUS-D-11-00009.1>
- Nipko RB, Holcombe BE, Kelly MJ (2020) Identifying individual jaguars and ocelots via pattern-recognition software: Comparing HotSpotter and Wild-ID. *Wildl Soc Bull* 44:424–433. <https://doi.org/10.1002/wsb.1086>
- Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, Clune J (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc Natl Acad Sci USA* 115:E5716–E5725. <https://doi.org/10.1073/pnas.1719367115>
- Penteriani V, Melletti M (2020) *Bears of the world*. Cambridge University Press, Cambridge
- Penteriani V, Te WS, May CL, Wah SY, Crudge B, Broadis N, Bombieri G, Valderrábano E, Russo LF, Delgado MM (2020) Characteristics of sun bear chest marks and their patterns of individual variation. *Ursus* 2020:1–8. <https://doi.org/10.2192/URSUS-D-19-00027.1>
- Ramsey AB, Sawaya MA, Bullington LS, Ramsey PW (2019) Individual identification via remote video verified by DNA analysis: a case study of the American black bear. *Wildl Res* 46:326–333. <https://doi.org/10.1071/WR18049>
- Ravoor PC, T.S.B. S (2020) Deep learning methods for multi-species animal re-identification and tracking – a survey. *Comput Sci Rev* 38:100289. <https://doi.org/10.1016/j.cosrev.2020.100289>
- Reyes A, Rodríguez D, Reyes-Amaya N, Rodríguez-Castro D, Restrepo H, Urquijo M (2017) Comparative efficiency of photographs and videos for individual identification of the Andean bear (*Tremarctos ornatus*) in camera trapping. *Therya* 8:83–87. <https://doi.org/10.12933/therya-17-453>
- Rodríguez D, Reyes A, Quiñones-Guerrero A, Poveda-Gómez FE, Castillo-Navarro Y, Duque R, Reyes-Amaya NR (2020) Andean bear (*Tremarctos ornatus*) population density and relative abundance at the buffer zone of the Chingaza National Natural Park, cordillera oriental of the Colombian Andes. *Pap Avulsos Zool* 60:1–7. <https://doi.org/10.11606/1807-0205/2020.60.30>
- 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>
- Schneider S, Taylor GW, Kremer SC (2020) Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. In: *Proceedings of 2020 IEEE winter conference applied computational vision work WACVW 2020*, pp 44–52. <https://doi.org/10.1109/WACVW50321.2020.9096925>

- Schofield D, Nagrani A, Zisserman A, Hayashi M, Matsuzawa T, Biro D, Carvalho S (2019) Chimpanzee face recognition from videos in the wild using deep learning. *Sci Adv* 5:1–10. <https://doi.org/10.1126/sciadv.aaw0736>
- Schroff F, Kalenichenko D, Philbin J (2015) FaceNet: a unified embedding for face recognition and clustering. In: 2015 Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 815–823. <https://doi.org/10.1109/CVPR.2015.7298682>
- Scotson L, Fredriksson G, Augeri D, Cheah C, Ngoprasert D, Wai-Ming W (2017) *Helarctos malayanus* (errata version published in 2018). IUCN Red List Threat Species 2017:e.T9760A123798233. <https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T9760A45033547.en>
- Shimozuru M, Yamanaka M, Nakanishi M, Moriwaki J, Mori F, Tsujino M, Shirane Y, Ishinazaka T, Kasai S, Nose T, Masuda Y, Tsubota T (2017) Reproductive parameters and cub survival of brown bears in the Rusha area of the Shiretoko Peninsula, Hokkaido, Japan. *PLoS ONE* 12:1–17. <https://doi.org/10.1371/journal.pone.0176251>
- Steenweg R, Hebblewhite M, Kays R, Ahumada J, Fisher JT, Burton C, Townsend SE, Carbone C, Rowcliffe JM, Whittington J, Brodie J, Royle JA, Switalski A, Clevenger AP, Heim N, Rich LN (2017) Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Front Ecol Environ* 15:26–34. <https://doi.org/10.1002/fee.1448>
- Van Horn RC, Zug B, Lacombe C, Velez-Liendo X, Paisley S (2014) Human visual identification of individual Andean bears *Tremarctos ornatus*. *Wildl Biol* 20:291–299. <https://doi.org/10.2981/wlb.00023>
- Van Horn RC, Zug B, Appleton RD, Velez-Liendo X, Paisley S, LaCombe C (2015) Photos provide information on age, but not kinship, of Andean bear. *PeerJ* 3:e1042. <https://doi.org/10.7717/peerj.1042>
- Velez-Liendo X, García-Rangel S (2017) *Tremarctos ornatus* (errata version published in 2018). IUCN Red List Threat Species 2017:e.T22066A123792952. <https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T22066A45034047.en>
- Weinstein BG (2018) A computer vision for animal ecology. *J Anim Ecol* 87:533–545. <https://doi.org/10.1111/1365-2656.12780>
- Yoshizaki J, Pollock KH, Brownie C, Webster RA (2009) Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. *Ecology* 90:3–9. <https://doi.org/10.1890/08-0304.1>
- Zheng X, Owen MA, Nie Y, Hu Y, Swaisgood RR, Yan L, Wei F (2016) Individual identification of wild giant pandas from camera trap photos—a systematic and hierarchical approach. *J Zool* 300:247–256. <https://doi.org/10.1111/jzo.12377>

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