Global Ecology and Conservation 41 (2023) e02350

Contents lists available at ScienceDirect

Global Ecology and Conservation

journal homepage: www.elsevier.com/locate/gecco

Human expertise combined with artificial intelligence improves performance of snow leopard camera trap studies

Eve Bohnett[a,b,*], Sorosh Poya Faryabi[c], Rebecca Lewison[a], Li An[a], Xiaoxing Bian[d], Ali Madad Rajabi[c], Nasratullah Jahed[c], Hashim Rooyesh[c], Erica Mills[a], Saber Ramos[d], Nathan Mesnildrey[e], Carolina M. Santoro Perez[f], Janet Taylor[g], Vladimir Terentyev[h], Stephane Ostrowski[i]

[a] Department of Biology, San Diego State University, USA
[b] Department of Geography, San Diego State University, USA
[c] WIldlife Conservation Society, Afghanistan
[d] Department of Wildlife Ecology and Conservation, University of Florida, USA
[e] Department of Geography, University of Cambridge, USA
[f] Department of Life Sciences and Systems Biology, University of Turin, USA
[g] KwaZulu-Natal Department of Agriculture and Rural Development, South Africa
[h] Association for the Conservation of Biodiversity of Kazakhstan (ACBK), Nur-Sultan, Kazakhstan
[i] Wildlife Conservation Society, New York

A R T I C L E I N F O

Keywords:
Snow leopard
Artificial intelligence
Camera trap misclassification, individual ID, HotSpotter

A B S T R A C T

Camera trapping is the most widely used data collection method for estimating snow leopard (*Panthera uncia*) abundance; however, the accuracy of this method is limited by human observer errors from misclassifying individuals in camera trap images. We evaluated the extent Whiskerbook (www.whiskerbook.org), an artificial intelligence (AI) software, could reduce this error rate and enhance the accuracy of capture-recapture abundance estimates. Using 439 images of 34 captive snow leopard individuals, classification was performed by five observers with prior experience in individual snow leopard ID ("experts") and five observers with no such experience ("novices"). The "expert" observers classified 35 out of 34 snow leopard individuals, on average erroneously splitting one individual into two, thus resulting in a higher number than true individuals. The success rate of experts was 90 %, with less than a 3 % error in estimating the population size in capture-recapture modeling. However, the "novice" observers successfully matched 71 % of encounters, recognizing 25 out of 34 individuals, underestimating the population by 25 %. It was found that expert observers significantly outperformed novice observers, making statistically fewer errors (Mann Whitney U test P = 0.01) and finding the true number of individuals (P = 0.01). These differences were contrasted with a previous study by Johansson et al. 2020, using the same subset of 16 individuals from European zoos. With the help of AI and the Whiskerbook platform, "experts" were able to match 87 % of encounters and identify 15 out of 16 individuals, with modeled estimates of 16 $\pm$ 1 individuals. In contrast, "novices" were 63 % accurate in matching encounters and identified 12 out of 16 individuals, modeling 12 $\pm$ 1 individuals that underestimated the population size by 12 %. When comparing the performance of observers using AI and the Whiskerbook platform to observers performing the tasks manually, we found that observers using Whiskerbook made significantly fewer errors in splitting one

* Correspondence to: Department of Biology, San Diego State University, San Diego, CA 92104, USA.
E-mail address: evebohnett@yahoo.com (E. Bohnett).

https://doi.org/10.1016/j.gecco.2022.e02350
Received 12 October 2022; Received in revised form 1 December 2022; Accepted 2 December 2022
Available online 5 December 2022

individual into two (P = 0.04). However, there were also a significantly higher number of combination errors, where two individuals were combined into one (P = 0.01). Specifically, combination errors were found to be made by "novices" (P = 0.04). Although AI benefited both expert and novice observers, expert observers outperformed novices. Our results suggest that AI effectively reduced the misclassification of individual snow leopards in camera trap studies, improving abundance estimates. However, even with AI support, expert observers were needed to obtain the most accurate estimates.

1. Introduction

To get reliable wildlife population estimates, ecologists have long prioritized identifying individual animals (Borchers and Fewster, 2016). Camera traps have become increasingly popular in recent years, allowing wildlife biologists to collect non-invasive samples of elusive or secretive species in various settings (O'Connell et al., 2011). Using pattern recognition, individuals can be distinguished in camera trap imagery by their unique pattern of stripes or spots. Differentiating between individuals in camera trap images has traditionally been done manually, which is time-consuming and expensive and often requires multiple observers to cross-verify the classification. Even with multiple observers, misclassification has been shown to affect density estimates in camera trapping studies (Johansson et al., 2020). It is crucial to improve techniques for identifying individuals in camera trap images.

Ecologists are increasingly turning to artificial intelligence (AI) to help with the main tasks involved with image data preparation, such as identifying animals and classifying at the species level (Beery et al., 2019; Falzon et al., 2019; Nguyen et al., 2017; Norouzzadeh et al., 2019; Parham et al., 2018). While several studies have evaluated how well AI has assisted in automatically classifying images for individual identification, relatively few studies evaluate whether AI can improve the classification process for manual observers. Earlier research concluded that the software's algorithms alone could not automatically sort the imagery, requiring observers to verify classifications (Morrison et al., 2011, Nipko et al., 2020). AI image classifiers can only achieve a specific accuracy rate, which varies depending on the dataset. There is still a need for observers, but it is unclear whether software significantly improves their accuracy.

The primary approach used to estimate snow leopard population size, and density is camera trapping (Alexander et al., 2015), along with the less frequently used method of genetic analysis utilizing scat samples (Janečka et al., 2011; Laguardia et al., 2015). Analysis of camera trapping and genetic data to estimate population abundance generally follows two steps. First, the data are sorted to identify unique individuals. Following individual identification, these data are processed and employed in spatial capture-recapture models (SCR) to determine the density and population size (Royle and Young, 2008). Research has found that sampling bias in important habitats and incorrect individual classification in camera trap images have led to systematic overestimations of the size of the snow leopard population (Johansson et al., 2020; Suryawanshi et al., 2019). The manual observers must be able to differentiate between images of the same individual taken at different times, under different lighting conditions, and from different camera angles, mainly based on the pelage patterns on their midsection, flanks, tail, and face. The observer's ability is related to the magnitude of the overestimation error. The accurate estimation of population abundance and density using camera trap images remains a challenge for snow leopard conservation, and finding methodological answers would require correcting the significant inaccuracies that result from misclassifying individuals in camera trap imagery.

Accurate estimates of population size and density are essential for all species of conservation concern, particularly for the snow leopard. Snow leopards (*Panthera unica*) were recently delisted in 2017 by the IUCN from endangered to vulnerable, based on an estimated global population size of 2710 and 3386, with a decreasing tendency (McCarthy et al., 2017). Although the delisting was primarily based on increased knowledge of the snow leopard's distribution and population size, the decision was criticized because of the limited precision of the current population estimates (Mallon and Jackson, 2017). Snow leopards in Asia's high-altitude regions are under intense pressure due to the instability and fragmentation of their natural habitats brought on by various political, social, economic, and ecological challenges (Sultan et al., 2022). The twelve countries comprising the snow leopard's high-altitude home are Afghanistan, Bhutan, China, India, Kazakhstan, Kyrgyzstan, Mongolia, Nepal, Pakistan, Russia, Tajikistan, and Uzbekistan. Within the range countries, only 14–19 % of the snow leopard range overlaps with protected areas which are thought to be inadequate for protecting the species over the long term, with the species living in multi-use landscapes along a gradient of human pressure from pastoral communities (Sharma and Singh, 2021). Priority areas for snow leopard conservation have been identified (Li et al., 2020) within a large area of perceived suitable snow leopard habitat (Riordan et al., 2016), although density estimates largely remain unknown (Sharma and Singh, 2021). Less than 3 % of the snow leopard range has been sampled using acceptable population estimates (Sharma and Singh, 2021). In order to establish conservation programs, the "Population Assessment of the World's Snow Leopards" (PAWs) initiative and the Global Snow Leopard and Ecosystem Protection Program (GSLEP) are working to obtain a renewed global assessment of the snow leopard population (Sharma et al., 2020).

Given the ongoing debate on the status of the snow leopard, accurate estimates of population parameters, such as density and abundance, are essential to address substantial uncertainties regarding their abundance across their range states. Using AI-supported technology has successfully automated the ability to distinguish snow leopards from photographs of other animals, such as blue sheep (Miguel et al., 2016), with convolutional neural networks showing 91 % accuracy for species differentiation (Tariq et al., 2018). A developing area of AI is also improving individual identification using the pelt patterns uniquely distinctive to each individual (Wäldchen and Mäder, 2018; Weinstein, 2018). It has been difficult for algorithms to differentiate snow leopard individuals due to the pelt patterns' similarity to the mountain background, which is primarily grey and white and easily misunderstood by deep learning algorithms. These errors have led to image segmentation research that sought to erase the background from images to help algorithms in focalizing regions of interest (Beery, 2016; Miguel et al., 2019). For snow leopards, deep learning algorithms were assessed independently through the rigorous algorithm training and evaluation phases, finding that the PIE and HotSpotter algorithms with background subtraction could find an individual ID match ~85 % of the time (Blount et al., 2022). These research developments have improved the potential to automate the identification of individual snow leopards in vast amounts of image data based on distinctive pelt patterns distinguished from the mountainous terrain.

A software application using image classification algorithms called Whiskerbook was developed based on advancements in AI technology to perform big cat species-specific individual identification tasks for camera trap images. The platform has been expanded to big cat species identification, including jaguar, cheetah, leopard, and snow leopard. However, research has not been conducted to determine if the platform can assist with specific identification tasks or if observers who use the platform identify individuals with greater accuracy.

This study aimed to determine how well observers could use the Whiskerbook AI tools to improve their ability to identify individual snow leopards. Two individual ID algorithms, namely HotSpotter (Crall et al., 2013) and pose invariant embeddings (PIE) (Moskvyak et al., 2019), along with a "visual matcher" allowing for a manual side-by-side comparison of images, were assessed for their ability to improve manual observer success in the individual classification. The primary goal of this project is to support the big cat community in deciding whether to integrate the platform into their protocols by 1) determining how well observers could use the technology; 2) whether it would improve their ability to classify the imagery more accurately in comparison to a previous study sorting the imagery

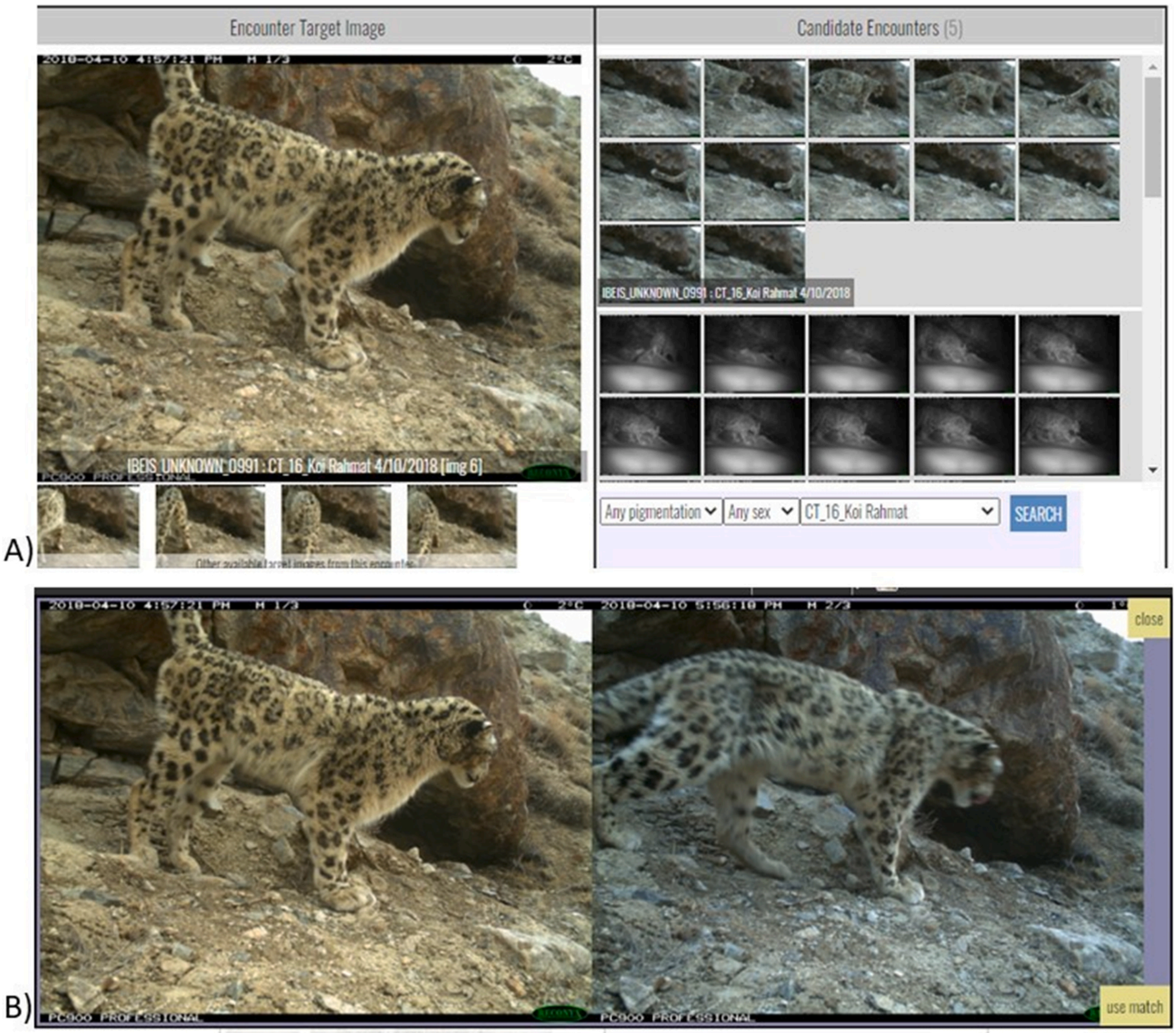

Fig. 1. Snapshot of visual matcher in Whiskerbook system. A) Users can compare images at camera trap stations. B) Side-by-side comparison of two images in the image matcher for manual comparison of images (without the HotSpotter algorithm). (Color Printing Not Necessary).

manually by Johansson et al. (2020); 3) consider manual observer perceptions of Whiskerbook for assisting in individual ID tasks in terms of usability and speed. We hypothesized that the online platform would help both groups of observers improve their classification accuracy. The results of this study can be used to improve population estimation efforts for future snow leopard assessments and clarify how the AI tools can be used in protocols of imagery classification for individual identification.

2. Methods

2.1. *Whiskerbook*

Whiskerbook (http://www.Whiskerbook.org) is a cloud-based data management system that batch-processes camera trap images for pattern recognition and conducting individual identification and offers a user-friendly interface for various user-oriented functions (Berger-Wolf et al., 2017). It is an open-source tool created by the non-profit group Wild Me (wildme.org). The program's "visual matcher" can be used to compare images manually, or algorithms can identify the animal's physical traits, patterns, and degree of pattern similarity before observers decide if two images show the same individual. The two algorithms, PIE and HotSpotter, were pre-trained and evaluated for individual snow leopard identification, incorporating background subtraction techniques (Blount et al., 2022).

Prior to use for snow leopard individual ID, the Whiskerbook was trained using detection annotations to recognize the cat and each of its body sides (front, rear, left, right, top, and bottom), as well as a complementary angle (for example, front-right, front-left, back-right, back-left).

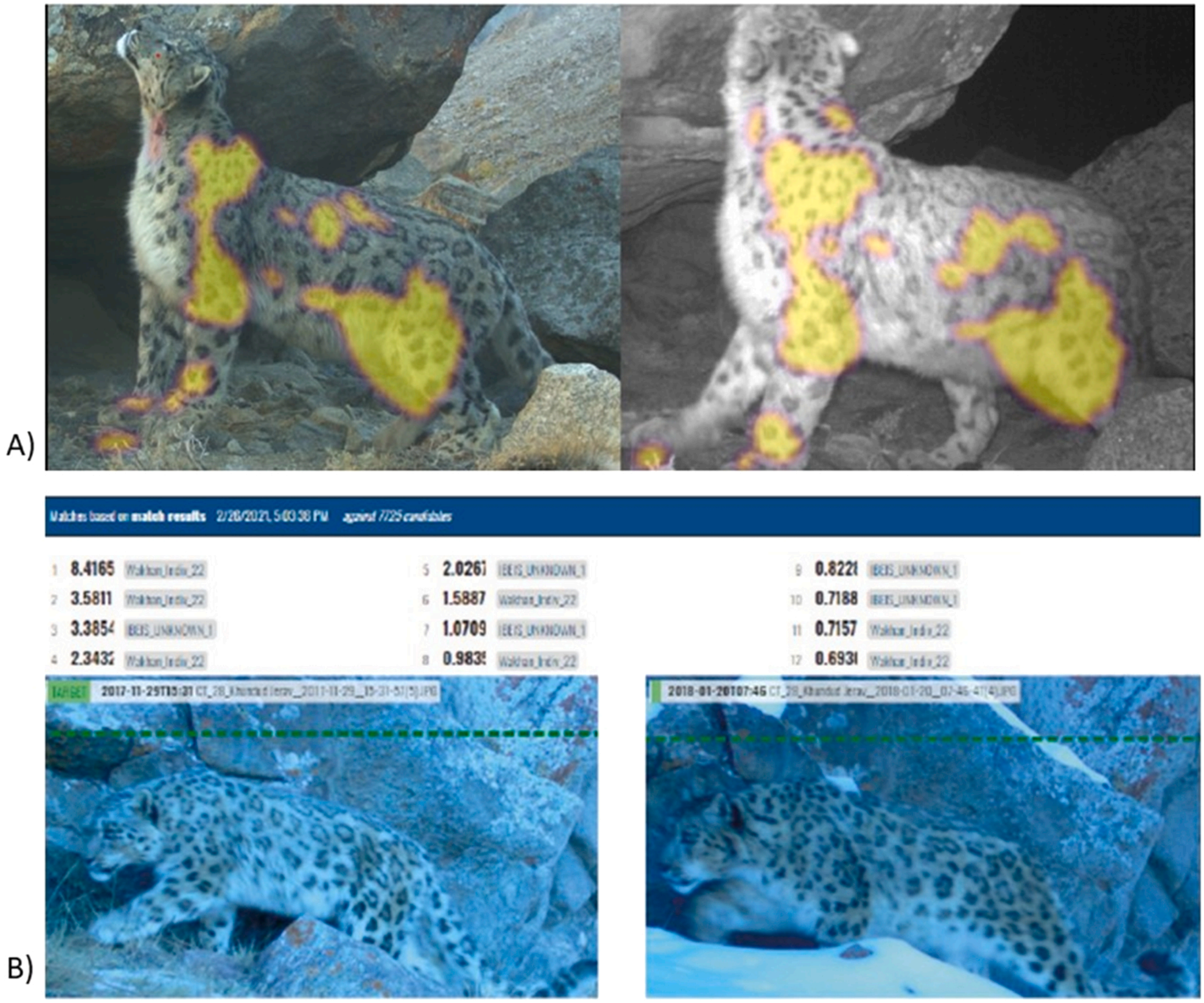

Fig. 2. HotSpotter algorithm on two individual images showing algorithm detected matching hotspots, which the analyst would classify as A) matching or B) not matching. Matching was conducted based on the HotSpotter classification score (Color Printing Not Necessary).

2.2. Visual matcher

The software's "visual matcher" lets analysts manually compare a focal image against images grouped by camera trap station (Fig. 1). There is a dropdown menu to switch camera stations at the bottom of the righthand pane. On the left, one encounter of several images with one focal image can be compared to the camera station's multiple encounters on the right. The manual observer can switch between camera stations, scroll through many encounters checking similar sides and angles of the focal image, and click on any image on the right for a side-by-side comparison with limited zoom. The observer can "use match" to assign a new identity or merge the individual with a previously assigned identity after finding matching individuals.

2.3. HotSpotter

The HotSpotter algorithm, developed by Crall et al. (2013), created methods for either matching two images against each other or one image against a database. The algorithm is a scale-invariant feature transform (SIFT) based comparison of significant visual texture areas. Where numerous points are cast onto the images, scoring is used to find the most distinct key points and descriptors using k nearest neighbors of any descriptor. This k nearest neighbor cluster is then highlighted within the image. The algorithm ranks the image matches using a numerical score, and the analyst can go through each image pair and check them individually, view the clusters that match within the images, and decide whether the individuals in the two images are the same.

The algorithm allows for multiple images of an individual previously identified to match against the image of interest and separately perform a database-wide search for the best matching images or similar images that depict the same individual (Blount et al., 2018). The capabilities are integrated into the Wildbook platform and can sequentially compare images to one another, using image similarity metrics in a one-vs.-many scoring mechanism (Fig. 2). Previously published research by Crall et al. (2013) identifies limitations in how the algorithm can deal with cases with overlapping animals, matching against the background, and matching failure in poor-quality images.

2.4. Pose Invariant Embeddings (PIE)

The Pose Invariant Embeddings (PIE) algorithm (Moskvyak et al., 2019) is a convolutional neural network (CNN) that uses a series of labeled and classified images, using a specific network architecture, in this case, ResNeXt (Xie et al., 2017). CNN's reduce images to significant features and patterns, which the algorithm extracts separately in a feature-reduction process to isolate specific patterns. In a series of iterations, the algorithm "learns" from these extracted features, thereby constructing a model for classification tasks that can detect similarly shaped patterns. In order to find underlying patterns like those found on the snow leopard's fur, the algorithm mines the image for shapes and features to create an "embedding" for a database of labeled images. After the embedding is created, the k-nearest neighbor classifier can return the closest matching images to the focal image. The observers then run the algorithm on an image in Whiskerbook, like HotSpotter, which returns ranked images for manual confirmation.

2.5. Camera-trap data

Nine zoos in the US and Europe collected a total of 439 images of 34 snow leopard individuals. 212 images of 18 snow leopards came from two zoos in the United States (WCS-managed Bronx and Central Park zoos, New York City). Researchers then assigned these images to 51 different "encounter" events. An additional 227 images came from 16 snow leopards from seven European zoos (Helsinki and Ätheri Zoos in Finland, Kolmården Zoo, Nordens Ark and Orsa Bear Park in Sweden, and Köln and Wuppertal Zoos in Germany) that were previously used in Johansson et al. (2020). These data were subset into 36 encounter events. Five encounters were of a single flank, with no chance of being matched. The encounters were divided into eight locations to imitate a camera-trapping study. The data were given random names, locations, and dates for each manual observer on the platform, making it impossible for them to share their data or information.

An "encounter" consisted of between one to eleven images of each snow leopard taken at a specific capture event when the animal crossed in front of the camera - the images comprised of left, right, front, or back angled imagery. The markings on the left and right flanks of snow leopards' pelts are not identical, a condition known as bilateral asymmetry, such that the photographs of the two sides cannot be matched (Augustine et al., 2018). As a result, encounters were examined for flank side consistency to ensure they could be matched with other encounters of the same flank.

2.6. Observers

In November 2021, we invited 16 observers to participate in the study, solicited from a broad diversity of educational institutions in Europe and the US, and researchers directly involved with snow leopard conservation at NGOs in Central Asia. Ten of the observers completed the task following the requirements, and observers were required to name and classify all encounters. They represented numerous nations, including Afghanistan (3), the United States (3), France (1), South Africa (1), China (1), and Italy (1). The spectrum of skill was extremely wide, ranging from having no experience with camera trapping studies to having experience with snow leopard individual ID. Five of the ten observers who finished the study were regarded as "experts" and had prior individual ID experience, although only two had done so for snow leopards. The remaining five observers were regarded as "novices" because they lacked any background in individual identification or camera trap data. All ten participants who finished the study are pursuing jobs in environmental sciences or ecology. All observers held bachelor's degrees, seven had master's degrees, and one had a Ph.D. Additionally, the work experience of these observers ranged from one to fifteen years, with the average being seven years.

To ensure the homogeneity of this study, the algorithms were evenly distributed to the observers; five evaluated the HotSpotter method, while another group of five evaluated the PIE algorithm. The "expert" category had three individuals use the HotSpotter, and two use the PIE. The "novice" category had two individuals use the HotSpotter and three use the PIE.

The observers had instructions on operating the platform's software to do individual IDs by watching four 15-minute introductory videos on how to use the Whiskerbook platform to perform individual IDs of snow leopards. The video tutorials could be accessed via web link at any time convenient to the observers and described the processes they needed to follow. To ensure we answered any questions and the observers were comfortable using the platform tools, we scheduled two conference calls, one for each group (Hotspotter or PIE).

Before starting to classify the data, observers followed specific methodological steps for cleaning the data to ensure the images were accurately tagged by the detection annotations with green dotted line boxes surrounding the snow leopards in the images and that the system did not mistakenly identify two animals. Then the observers employed one of the two algorithms and the visual matcher to perform further classification tasks. Except for a few interactions that would have no match, observers were told to name and match every encounter with another snow leopard. The instructions were comparable to the guidelines provided in field-based research, where it is assumed that almost all the snow leopard images will classify to an individual. They were then requested to perform one more final round of checking, in which they went back and examined the individuals they named and checked that the encounters were indeed the same.

2.7. Validation assessment

We estimated the frequency of identification errors following methods presented by Johansson et al. (2020) to evaluate the capacity of observers to perform individual ID. Encounters were labeled as being incorrect if they were 1) Split- an individual's encounters were split into a separate individual (thus adding a new individual), 2) Combine-all encounters from the individual were incorrectly merged with another individual (and an entire individual was lost), 3) Shift- the encounter was combined with encounters from another individual. 4) Exclude- there was an NA value instead of a name.

We employed the Mann Whitney U- test to compare our study's results between observer groups since our data did not follow a normal distribution. We compared the outcomes —HotSpotter vs. PIE, Expert vs. Novice, and Whiskerbook vs. Manual Observers – for the entire dataset and the subset of the European data from the Johansson et al. (2020) study. The Mann Whitney U- test was performed using program R (R Core Team, 2022) and the stats package to determine the one-way significance ($\alpha = 0.05$).

2.8. Capture-recapture

We employed statistical modeling to determine whether the estimates of the number of unique individuals from the observer groups produced accurate population estimates. To determine this, we fit closed population capture-recapture models to estimate population size. In a closed population, the total number of individuals does not change due to births, deaths, immigration, or emigration, whereas in an open population, the number of individuals can fluctuate during the study (Baillargeon and Rivest, 2007). The data were organized to represent the capture histories of each individual, which are the encounter events assigned to the individual thought to occur in continuous time. All models were run in the Rcapture package (Baillargeon and Rivest, 2007; Rivest and Baillargeon, 2022) in program R (R Core Team, 2022). The models operate so that a maximum likelihood estimate is obtained by fitting a Poisson generalized linear model with log-linear parameters. By maximizing the loglikelihood, optimization was carried out iteratively using least squares to determine the size of the population. We chose the models for continuous-time captures, which allowed estimators M0, Mh Chao (LB), Mh Poisson2, Mh Darroch, and Mh Gamma3), which the best model was chosen using AIC (Rivest and Daigle, 2004).

The closed population capture-recapture models illustrated how the population size estimates changed with the errors made by the observers. The Johansson et al. (2020) study also performed simulation experiments that qualify these techniques. For example, even though the manual observer may have identified the correct number of true individuals, the density estimates might differ if the individual images have been misclassified. The capture-recapture models also illustrate how modeled density estimates change in a scenario when the observer mistakenly splits one individual into two, attributes multiple encounters to the wrong individual, or combines unique individuals into a single individual.

In the study by Johansson et al. (2020), the shifting and splitting errors resulted in density estimates that were increasing the number of individuals, with the manual observers initially classifying 18 individuals (based on splitting one individual into two) instead of the actual number of 15 individuals and also incorrectly classifying many of the encounters (many of which involved shifting errors), finding a final capture-recapture estimate of 22 ± 3. Understanding how the errors generated by observers using the Whiskerbook may affect the final population size estimate was our aim in employing this approach.

2.9. Perception of user experience questions

After the classification was completed, manual observers were asked several questions in a follow-up survey. Questions included 1) how many hours did it take to match the data? (Observers were told prior to the study that it would take them ~20 h, and although they were not asked to time themselves, the responses are an estimate) 2) How easy was it to clean the data? 3) Were there any major challenges to cleaning data in the program? 4) Was it easier to use the visual matcher or the algorithm? 5) Approximately how many times did you comb the data? 5) Would you recommend the program for individual ID? 6) What are some of the shortcomings of the program? 7) Would you have preferred to do it manually?

3. Results

3.1. USA and European dataset results combined

The Whiskerbook provided the results of observers classifying 87 snow leopard encounters to 34 individuals, using a hybrid of algorithms and their side-by-side comparisons with the visual matcher's output.

The five "expert" observers that had previously performed individual ID had an overall success rate of 90 % in classifying every encounter correctly (Table 1). On average, these five observers identified 35 ± 1 out of 34 snow leopard individuals, splitting one individual into two, thus resulting in a higher number. The expert observers using the Whiskerbook had less than a 3 % error for overinflating the population size in capture-recapture modeling.

The five "novice" observers who had never performed individual IDs had a success rate of 71 % in matching encounters and correctly identified 25 out of 34 individuals. Two or more individuals being incorrectly combined into one individual was the leading cause of novice user error. By merging the data from individual snow leopards with other individuals, untrained observers using Whiskerbook were more likely to underestimate the number of snow leopards by 25 %.

For the data from USA and Europe, the ten observers showed an average success rate of 81 % for classifying the encounters, with the ability to classify 30 out of 34 individuals correctly, with similar abundance estimates reported from capture-recapture data. The results from our study indicate that, on average, an observer who uses the Whiskerbook will tend to underinflate the density estimates by 12 %.

The overall findings of the study show a clear distinction between observers with individual identification experience and those who have never performed individual ID for any species, where expert observers demonstrated a significantly higher accuracy (Mann Whitney U test $P = 0.01$) for successfully determining the correct number of individuals (Table 2). Expert observers were also demonstrated to have fewer errors from combining encounters with the wrong individual ($P = 0.01$) and significantly fewer individuals being lost by incorrectly merging individuals ($P = 0.01$) than novice observers.

The results show that there was no significant difference between the groups of users for HotSpotter and PIE.

Table 1

Snow leopard classification results from American and European datasets from Whiskerbook observers. Results indicated the number of individuals recognized out of the 34 true individuals, the estimated population size from capture-recapture, the proportion of correct encounters, and the number of encounters that were split, combined, shifted, or classed as NA for each observer. How many individuals were gained due to a split, how many individuals were lost due to combination with other individuals, and how many individuals were eliminated due to NA. **Highlighted in bold are the averages for the experts, novices, and the total for all observers.** The ± value represent the standard errors for the capture recapture models.

Observer	Algorithm	Expert	Individuals Identified (true = 34)	Capture Recapture	% Correct Encounters	Split	Combine	Shift	NA	Indv. Add from Split	Indv. Lost in Combine	Indv.Lost from NA
Obs 1	HotSpotter	Yes	34	36 ± 2	88 %	3	5	2	20	3	3	4
Obs 2	HotSpotter	Yes	34	35 ± 1	93 %	2	0	2	1	1	0	0
Obs 3	HotSpotter	Yes	36	36 ± 0	94 %	3	2	0	0	2	1	0
Obs 6	PIE	Yes	35	37 ± 2	84 %	7	1	5	4	5	1	2
Obs 10	PIE	Yes	31	33 ± 2	93 %	1	6	1	0	1	4	0
Expert	**Average**		**34**	**35 ± 1**	**90 %**	**3**	**3**	**2**	**5**	**2**	**2**	**1**
Obs 4	HotSpotter	No	25	25 ± 1	82 %	0	11	4	2	0	8	0
Obs 5	HotSpotter	No	27	27 ± 1	81 %	2	9	4	1	1	5	1
Obs 7	PIE	No	22	21 ± 1	48 %	6	24	15	0	3	9	0
Obs 8	PIE	No	21	20 ± 1	62 %	0	26	7	0	0	14	0
Obs 9	PIE	No	31	30 ± 0	84 %	5	10	1	0	3	5	0
Novice	**Average**		**25**	**25 ± 1**	**71 %**	**3**	**16**	**6**	**1**	**1**	**8**	**0**
Total	**Average**		**30**	**30 ± 1**	**81 %**	**3**	**9**	**4**	**3**	**2**	**5**	**1**

Table 2

The Mann Whitney U test results for the entire dataset of 34 individuals from European and American zoos are presented as p-values between HotSpotter vs. PIE and Expert vs. Novices. Categories include the total number of individuals, the number of split errors, the number of combine errors, the number of shift errors, the number of NA values, the number of individuals added from splitting errors, the number of individuals lost in combination errors, and the number of individuals lost from NA values. Significant p-values ($\alpha < 0.05$) are highlighted in bold.

USA + European Dataset	Individuals	Split	Combine	Shift	NA	Indv. Add from Split	Indv. Lost in Combine	Indv.Lost from NA
HotSpotter vs. PIE	0.34	0.40	0.25	0.34	0.15	0.39	0.25	0.52
Expert vs. Novice	**0.01**	0.53	**0.01**	0.14	0.37	0.33	**0.01**	0.37

3.2. European zoo data: whiskerbook vs. manual observers

The 16 individuals from European zoos were previously utilized in a study on manual observers by Johansson et al. (2020), which makes our study directly comparable. It was reported that the manual observers in Johansson study's had the propensity to split individuals or shift encounters to the wrong individuals, which led to an overestimation of density in the capture-recapture modeling. The manual observers from Johansson et al. (2020) reported 88 % success in matching encounters and classifying 18 out of 15 individuals (with one being removed on average from denoting NA). There was no statistically significant difference in performance between experts and novices across error categories in the manual study without AI. By mistakenly splitting one individual into two or more, splitting errors added several individuals; on average, three individuals were added to both the "expert" and "novice" categories. In addition, shifting errors occurred when encounters were added to the wrong individual (without losing or gaining an entire individual), with both "expert" and "novice" observers reporting an average of three shifting errors. On average, no individuals were lost due to combination errors. The Johansson study showed that shifting and splitting errors were sufficient to boost the capture-recapture density estimates by 37 %, reporting 22 ± 3 individuals instead of the 16 known true individuals.

In comparison, the ten observers that used the Whiskerbook had an average success of 75 % matching the encounters, which classified an average of 14 out of 16 individuals correctly. The modeled estimates reported a 12 % underestimate in population density overall.

In our study, the five "expert" observers had better classification accuracy and estimations of capture-recapture, indicating a significant reduction in error. Expert observers classified 87 % of all contacts in the European data and, on average, classified most known snow leopards as 15 out of 16 individuals (Table 3). Using the Whiskerbook platform led to accurate-modeled capture-recapture estimates, finding 16 out of 16 true individuals on average. The five expert observers using the Whiskerbook were exact in their estimate of population size.

However, our study's five "novice" observers were 63 % accurate in matching encounters and, on average, matched 12 out of 16 individuals. The modeled estimates showed that the population size was underestimated by 25 %.

The results of the Mann-Whitney U-test for the subset of 16 individuals from the European data demonstrate substantial differences between observers using Whiskerbook and Johansson et al. (2020) (Table 4). The results reveal statistically significant differences where Whiskerbook observers made fewer errors, including the number of individuals identified ($P = 0.01$), the number of individuals added from splitting the individuals into two ($P = 0.03$), the number of shifting encounters to the wrong individuals (which did not

Table 3

Observer results for European zoo dataset for the Whiskerbook observers. Results indicated the number of individuals recognized out of the 16 true individuals, the estimated population size from capture-recapture, the proportion of correct encounters, and the number of encounters that were split, combined, shifted, or classed as NA. How many individuals were gained due to a split, how many individuals were lost due to a combination with other individuals, and how many individuals were eliminated due to NA. Highlighted in bold are the averages for the experts, novices and the total for all observers. The ± value represents the standard errors for the capture-recapture models.

Observer	Algorithm	Expert	Individuals Identified (True = 16)	Capture Recapture	% Correct Encounters	Split	Combine	Shift	NA	Indv. Add from Split	Indv. Lost in Combine	Indv. Lost from NA
Obs 1	HotSpotter	Yes	16	17 ± 2	89 %	1	1	1	8	1	1	3
Obs 2	HotSpotter	Yes	15	14 ± 1	97 %	0	0	0	1	0	0	0
Obs 3	HotSpotter	Yes	16	17 ± 1	94 %	0	2	0	0	0	1	0
Obs 6	PIE	Yes	18	18 ± 1	75 %	4	1	3	3	3	1	2
Obs 10	PIE	Yes	12	13 ± 1	81 %	0	5	1	0	0	4	0
Expert	**Total Average**		**15**	**16 ± 1**	**87 %**	**1**	**2**	**1**	**2**	**1**	**1**	**1**
Obs 4	HotSpotter	No	12	11 ± 1	83 %	0	6	0	0	0	5	0
Obs 5	HotSpotter	No	13	12 ± 1	94 %	0	1	0	1	0	1	1
Obs 7	PIE	No	11	10 ± 1	14 %	2	13	8	0	1	6	0
Obs 8	PIE	No	11	15 ± 4	38 %	0	21	1	0	0	11	0
Obs 9	PIE	No	12	11 ± 0	84 %	0	9	0	0	0	4	0
Novice	**Total Average**		**12**	**12 ± 1**	**63 %**	**0**	**10**	**2**	**0**	**0**	**5**	**0**
Total	**Average**		**14**	**14 ± 1**	**75 %**	**1**	**6**	**1**	**1**	**1**	**3**	**1**

Table 4

The Mann Whitney U test results for the European Zoo data are presented as p-values between HotSpotter vs. PIE, Expert vs. Novices, and Whiskerbook vs. Johansson. The categories included the number of individuals, split errors, combine errors, shift errors, NA values, number of individuals added from splitting errors, number of individuals lost in combination errors, and number of individuals lost from NA values. Significant p-values ($\alpha < 0.05$) are highlighted in bold.

European Subset	Individuals	Split	Combine	Shift	NA	Indv. Add from Split	Indv. Lost in Combine	Indv.Lost from NA
HotSpotter VS PIE	0.17	0.37	0.07	0.06	0.29	0.44	0.08	0.52
Expert vs. Novice	**0.03**	0.52	**0.04**	0.73	0.16	0.44	**0.04**	0.37
Whiskerbook vs. Johansson	**0.01**	**0.03**	**0.01**	**0.04**	0.69	**0.004**	**0.01**	0.54

lead to a loss or gain of the individuals) ($P = 0.04$). However, there were also significant differences where the Whiskerbook observers made considerably more errors in combining two individuals into one ($P = 0.01$).

In the results from only observers using the Whiskerbook, there was a difference in the number of individuals detected between experts and novices ($P = 0.03$). This discrepancy in the number of individuals identified arose because novices made considerably more errors by combining individuals into one individual than experts ($P = 0.04$).

3.3. Perception of user experience results

Observers were asked to complete a short survey after completing the Whiskerbook tasks (Table 5). The observers sorted the data for 40 h on average (min 8, max 90), with experts averaging 32 h and novices 42 h. There was a considerable range in the time that novices spent on the data, where surprisingly, the novice that spent 8 h had an above-average level of accuracy (showing 81 % accuracy for the whole dataset (USA + Europe), which was the average of all observers, and 94 % accuracy on the European data subset which was much higher than average). Participants were informed that the task would likely take 20 h during study recruitment. Since we did not ask participants to time themselves and only asked how long the task took in the follow-up survey, our results were biased and represented only a rough estimate.

Most expert users combed through the data on average 3–4 times (min 2, max 10), whereas one expert user was an outlier combing the data 10 times. Compared to inexperienced observers, who spent 2.5 times searching through the datasets, expert observers spent, on average, 5 times combing the data (4 times if we remove the outlier). Although it does not necessarily mean that expert observers spend more time than novices, it seems that they scanned the data more in fewer hours, perhaps because experts spent less time on the individual images than the novices.

There were mixed results on whether the users preferred the algorithm or the visual matcher, with three observers favoring the algorithm and three observers favoring the visual matcher, two preferring a mix of both, and two observers not responding to the question.

Observer preference dictated which tool is most suitable to use. However, nine out of ten users would suggest/ recommend the program. One manual observer would not suggest the program. Three of the ten observers would have opted to carry out the jobs manually. These decisions are due to several issues raised with the program during the trials. These include issues with scrolling images, inability to zoom into images to examine finer details, the inability of novice observers to identify snow leopards using either the visual matcher or the algorithm, poor performance of the PIE algorithm, lengthy algorithm matching times, poorly functioning algorithms, and program defects appearing as buggy or incorrect error messages that would confuse the observers. Any program-related issues were reported to the Wild ME community board and investigated by the program personnel. The application was updated during our study to speed up the algorithms and improve the user interface.

Observer	What are some of the shortcomings of the program?	Was it easier to use the visual matcher or the algorithm?	Were there any major challenges in getting the data cleaned in the program?
Obs 1	When I run the HotSpotter in few photos bring matches, which was wrong, because the spot were less visible	Algorithm was easier that visual matcher, and visual matcher taking time and boring	I did not have any specific challenges in HotSpotter
Obs 2	Na	mix of both	Photos where it didn't match the algorithm and the leopard didn't give me a good angle to see patterns
Obs 3	It is not very convenient when I want to check some detailed patterns of an encounter.		I still have several encounters unmatched but I could not make sure of the results from both auto-matching or visual matching.
Obs 4	Hottspoter is very strong but there are some problems with weak poorly? taken pictures which if you match by visual matcher or algorithm, you face the problem.	The algorithm Hottspotter is very easy and fast and that is completely useful.	Just for removing additional annotation which I removed but it removed the whole picture.
Obs 5	Slow to run, some bugs	Visual matcher didn't work well for me	No
Obs 6	1) The fact that when the cursor is moved the image changes, in this way you can't use more than 30 pictures (approximately) performing the algorithm, and 2) in visual matcher you can't zoom in the pictures	Yes, but I didn't feel well with the program	
Obs 7		In PIE, the first visual matcher was the only way to match the images, but later on the algorithm worked better than the visual matcher.	Yes! At the beginning the matching option were not functioning well in the PIE algorithm and even the visual matcher was so difficult to match an image.
Obs 9	The PIE algorithm didn't always choose pictures with the same side of the animal. Not sure if it is operator (me) error or not: when comparing photos visually I wasn't able to see the full portrait orientated pics when comparing them to landscape pictures	Visual	No
Obs 10	The PIE algorithm didn't seem to work very well, very time consuming to use, and the matches were often not correct.	Visual Matcher	No

Table 5 continued. Perceptions of user experience with the Whiskerbook platform

Table 5
Perceptions of the Whiskerbook platform as experienced by both expert and novice users.

Observer	Expert?	How many hours to finish the task?	How easy was it to clean the data?	How many times did you comb the data?	Recommend?	Prefer to perform tasks manually?
Obs 1	Yes	40	Mostly Easy	2	yes	no
Obs 2	Yes	40	Easy	5	yes	no
Obs 3	Yes	30	Mostly easy	5	yes	no
Obs 6	Yes		Very easy	10	no	yes
Obs 10	Yes	20	Easy	3	Yes	no
Obs 4	No	40	Easy	2	yes	yes
Obs 5	No	8	Easy	4	yes	no
Obs 7	No	90	Mostly easy	2	yes	no
Obs 9	No	30	Very easy	2	yes	yes

4. Discussion

Camera traps are the most accessible method for estimating large felid species' population size and density. To model with a high degree of precision, it is necessary to have methods that can extract individual identities from camera-trapped wildlife. While snow leopard markings exhibit substantial variation in pelt pattern rosettes, they have traditionally exhibited high levels of misclassification by manual observers (Johansson et al., 2020) and proved challenging for machine learning-based image classification (Blount et al., 2022). This study provides the first observer-based evaluation of the Whiskerbook platform, using image classification algorithms HotSpotter, PIE, and a visual matcher for side-by-side classification for the individual ID of snow leopards. The results show that the Whiskerbook program significantly improves population estimate accuracy when used by observers with experience, as compared to manual identification. Observers can benefit from AI-assisted technologies when performing individual ID tasks, with experts making fewer errors and thus enabling reliable estimates of population size.

Our study also sought to understand which errors were more prevalent in the Whiskerbook system and how those errors impacted the population estimates. Compared to fully manual observers in Johansson et al. (2020), the Whiskerbook observers were significantly less likely to split a single individual into two or more individuals incorrectly ($P = 0.004$) or to assign an encounter to the wrong individual ($P = 0.003$). However, Whiskerbook users were likelier to combine multiple individuals into one ($P = 0.01$). For the cohort using Whiskerbook software, experts made fewer combination errors that merged individuals into one than novices ($P = 0.04$), and the errors experts made were distributed so that they did not compound to affect the model estimate results. The novices were more likely to combine multiple individuals into one entity, thereby underestimating population densities. The software-assisted methods in this study enhanced accuracy, and subsequent capture-recapture estimations overall since the modeled estimates from the Whiskerbook for all observers were 12 % underestimated, and the Johansson study was 37 % overestimated. The five observers who were considered "experts" showed a middle path and estimated the true population size more precisely as 16 ± 1 in the raw classification results and after modeling estimates via capture-recapture modeling.

Few studies conducted in recent years have also confirmed that experts outperform novices when it comes to identifying individuals. One study on the Interactive Individual Identification System (I3S Pattern) software discovered a discrepancy when attempting to match the Australian skink (*Egernia group*) and Slater's skinks (*Liopholis slateri*), where experienced observers matched more images correctly than those without experience (Treilibs et al., 2016). However, the Johansson et al. (2020) study found no significant difference for manual observers between experts and novices for any classification categories, indicating the need for an online training tool for any observers to assess their ability which is available at camtraining.globalsnowleopard.org. Some studies have used the camtraining online tool to gauge the observer's performance based on the accuracy from 30 side by side image matches (Pal et al., 2022). Based on the findings from our research, we advise inexperienced observers first to use the camtraining tool, take the same zoo data challenge, or gauge their aptitude using a subset of field data to familiarize themselves with snow leopard individual ID before attempting to match on a new field dataset. Thorough training and this zoo test evaluation can help those unsure of their performance understand their mistakes and improve. Allowing novice observers to match field-based data, followed by a rigorous double-check by someone with greater experience, could also ensure consistency.

We discovered some evidence that the less experienced observers were also combing through the data more slowly. From the perception questions, we can understand that the experts used their time to comb through the data around 4–5 times on average, compared with 2.5 times for the novice observers. The expert observers used their time to continually run over the data numerous times while spending about 10 h less on average (~ 30 compared to ~40 h). Before the study, we informed the participants that it would take ~20 h when we were recruiting them for the study so that the observers would have an idea of how long it would take. Although we only relied on self-reported data on time spent processing data, our results suggest that experienced observers can process camera image data more efficiently.

In recent years, AI and deep learning methods have demonstrated significant improvements in identifying individual animals from photographs and capture-recapture studies. Researchers have evaluated software programs with varying algorithmic performance, such as Wild-ID (Bolger et al., 2011), I3S Pattern+ (Hartog and Reijns, 2013), APHIS (Moya et al., 2015), HotSpotter, SIFT-based algorithms, and AmphIdent (Matthé, 2015), to identify amphibians, harbor seals (*Phoca vitulina vitulina*), Masai giraffe (*Giraffa camelopardalis tippelskirchi*), and wildebeest (*Connochaetes taurinus*) (Bardier et al., 2020; Matthé et al., 2017; Dawson, 2021). Different systems have yielded varying degrees of precision for various taxa. Our study did not make attempts to compare the Whiskerbook to other software available, mainly due to the increased capabilities of the Whiskerbook to incorporate image segmentation to remove the background from the images, which is an improvement over other software specifically for snow leopard individual ID.

Our study attempted to determine whether observers were more successful or preferred using one of the two Whiskerbook algorithms (PIE vs. HotSpotter). There were no discernable differences between the observers using HotSpotter and PIE to perform the tasks, reporting the same number of snow leopard individuals, and all categories of error were considered non-significant between groups by Mann Whitney U-test. We had split our cohort for the two algorithms so that our five expert observers (3 HotSpotter and 2 PIE) and five novice observers (2 HotSpotter and 3 PIE) were distributed evenly throughout the algorithms. Both datasets (34 individuals from Europe and USA and 16 individuals from Europe only) showed no significant differences between findings in the cohorts split between the algorithms. According to a prior study comparing the algorithms separately, to achieve the best results, both algorithms should be used simultaneously (Blount et al., 2022). The perception survey also indicated that the users used a range of tools, with three preferring algorithms, three preferring the visual matcher, three choosing to use a mix of both, and two not responding to the question. The researchers utilized different platform aspects throughout the study, each according to their respective preferences.

5. Conclusions

This study indicated that the Whiskerbook tools provided a promising alternative to traditional approaches for identifying individual animals in images. Expert observers were helped by the online Whiskerbook individual identification technology to identify individual snow leopards with a high degree of accuracy, finding the true number of individuals $\pm$ 1. Novices who used the application more commonly combined individuals incorrectly, underestimating the population size by 12–25 %. These findings were directly compared to a prior study by Johansson et al. (2002) that used manual classification without software, finding that observers had the propensity to separate one individual into two or shift encounters to the wrong individual, overestimating abundance by 37 %. The results of our study show that, primarily for the expert observers, using Whiskerbook software is more accurate than manually sorting photos. Before starting camera image classification research, inexperienced observers should receive rigorous training in individual identification and the software, and their work may benefit from oversight and review of their classification decisions. The Whiskerbook program has a visual matcher and algorithm tools, both preferred equally depending on the observer. Our results showed no significant difference in the performance of observers using the PIE or Hotspotter algorithms, although it is advised that Whiskerbook observers combine the PIE and HotSpotter algorithms simultaneously when using the program (Blount et al., 2022). The AI technology has the potential to enhance individual ID greatly, but according to our tests, the capabilities do not yet include entirely automated detection and still call for skilled observers to complete the tasks with a high degree of accuracy. Individual identification via Whiskerbook with AI has the potential to increase accuracy and efficiency to support future population abundance monitoring for snow leopards and may be applicable to support population monitoring of other big cat species.

Funding

UNDP GEF grant AA/Pj/PIMS: 00076820/0088001/5038. and European Union (Contract ACA/2018/399–742). The National Science Foundation funded this research under the Coupled Natural and Human Systems program [BCS-1826839]".

CRediT authorship contribution statement

Eve Bohnett co-designed the study, conducted the analysis, developed training materials, and wrote the manuscript. Sorosh Poya Frayabi co-designed the study, brought together the observers, methods, discussion and reviewed the manuscript, Xiaoxing Bian, Ali Madad Rajabi, Nasratullah Jahed, Hashim Rooyesh, Erica Mills, Saber Ramos, Nathan Mesnildrey, Carolina M. Santoro Perez, Janet Taylor, Vladimir Terentyev, Rebecca Lewison, Li An, reviewed manuscript, the PIE and Hotspoter programs. Stephane Ostrowski contributed to the concept, methods, and discussion and reviewed the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgments

We thank the Government of Afghanistan, GEF/UNDP, and the European Union for granting the Wildlife Conservation Society to execute the projects. The study contents are the authors' sole responsibility and do not necessarily reflect the views of the European Union. Special thanks go to the Bronx and Central Park zoos, New York City, USA, Helsinki and Ätheri Zoos in Finland, Kolmården Zoo, Nordens Ark and Orsa Bear Park in Sweden; and Köln and Wuppertal Zoos in Germany for creators of their camera trap imagery freely available. We also thank Wild ME for adapting their Whiskerbook platform for snow leopard identification and providing technical support. Finally, we thank the anonymous reviewers for their time and effort in reviewing the manuscript, which significantly improved the draft contents. We also would like to acknowledge the reviewers in the Lewison lab for their comments and edits that assisted in the production of this work, especially Greta Schmidt.

References

Alexander, J.S., Gopalaswamy, A.M., Shi, K., Riordan, P., 2015. Face value: towards robust estimates of snow leopard densities. PLoS ONE 10, e0134815. https://doi.org/10.1371/journal.pone.0134815.

Augustine, B.C., Royle, J.A., Kelly, M.J., Satter, C.B., Alonso, R.S., Boydston, E.E., Crooks, K.R., 2018. Spatial capture–recapture with partial identity: an application to camera traps. Ann. Appl. Stat. 12. https://doi.org/10.1214/17-AOAS1091.

Baillargeon, S., Rivest, L.-P., 2007. Rcapture: loglinear models for capture-recapture in R. J. Stat. Softw. 31.

Bardier, C., Székely, D., Augusto-Alves, G., Matínez-Latorraca, N., Schmidt, B.R., Cruickshank, S.S., 2020. Performance of visual vs. software-assisted photo-identification in mark-recapture studies: a case study examining different life stages of the Pacific Horned Frog (Ceratophrys stolzmanni). Amphib. -Reptil. 1–12. https://doi.org/10.1163/15685381-bja10025.

Beery, S., 2016. Orientation Invariant Autonomous Recognition of Individual Snow Leopards 9.

Beery, S., Morris, D., Yang, S., Simon, M., Norouzzadeh, A., Joshi, N., 2019. Efficient pipeline for automating species ID in new camera trap projects. BISS 3, e37222. https://doi.org/10.3897/biss.3.37222.

Berger-Wolf, T.Y., Rubenstein, D.I., Stewart, C.V., Holmberg, J.A., Parham, J., Menon, S., Crall, J., Van Oast, J., Kiciman, E., Joppa, L., 2017. Wildbook: crowdsourcing, computer vision, and data science for conservation. arXiv 1710, 08880 ([cs]).

Blount, D., Minton, G., Khan, C., Levenson, J., Dulau, V., Gero, S., Parham, J., Holmberg, J., 2018. Flukebook – continuing growth and technical advancement for cetacean photo identification and data archiving. Incl. Autom. Fin., Fluke, Body Matching 13.

Blount, D., Bohnett, E., Holmberg, J., Parham, J., Faryabi, S.P., Johansson, Ö., An, L., Ahmad, B., Khan, W., Ostrowski, S., 2022. Comparison of two individual identification algorithms for snow leopards after automated detection. (Prepr.). Bioinforma. https://doi.org/10.1101/2022.01.20.477059.

Bolger, D., Vance, B., Morrison, T., Farid, H., 2011. WildID User Guide: WildID Camera Trap Image Processing.

Borchers, D., Fewster, R., 2016. Spatial capture–recapture models. Stat. Sci. 31, 219–232. https://doi.org/10.1214/16-STS557.

R. Core Team, 2022. R: A language and environment for statistical computing.

Crall, J.P., Stewart, C.V., Berger-Wolf, T.Y., Rubenstein, D.I., Sundaresan, S.R., 2013. HotSpotter; Patterned species instance recognition, in: 2013 IEEE Workshop on Applications of Computer Vision (WACV). Presented at the 2013 IEEE Workshop on Applications of Computer Vision (WACV), IEEE, Clearwater Beach, FL, USA, pp. 230–237. https://doi.org/10.1109/WACV.2013.6475023.

Dawson, J., 2021. Comparisons of image-matching software when identifying pool frog (Pelophylax lessonae) individuals from a reintroduced population. HJ 55–59. https://doi.org/10.33256/31.1.5559.

Falzon, G., Lawson, C., Cheung, K.-W., Vernes, K., Ballard, G.A., Fleming, P.J.S., Glen, A.S., Milne, H., Mather-Zardain, A., Meek, P.D., 2019. ClassifyMe: a field-scouting software for the identification of wildlife in camera trap images. Animals 10, 58. https://doi.org/10.3390/ani10010058.

Hartog, J., Reijns, R., 2013. About I3S.

Janečka, J.E., Munkhtsog, B., Jackson, R.M., Naranbaatar, G., Mallon, D.P., Murphy, W.J., 2011. Comparison of noninvasive genetic and camera-trapping techniques for surveying snow leopards. J. Mammal. 92, 771–783. https://doi.org/10.1644/10-MAMM-A-036.1.

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, 6393. https://doi.org/10.1038/s41598-020-63367-z.

Laguardia, A., Wang, J., Shi, F.-L., Shi, K., Riordan, P., 2015. Species identification refined by molecular scatology in a community of sympatric carnivores in Xinjiang, China. Zoological. Research 11.

Li, J., Weckworth, B.V., McCarthy, T.M., Liang, X., Liu, Y., Xing, R., Li, D., Zhang, Y., Xue, Y., Jackson, R., Xiao, L., Cheng, C., Li, S., Xu, F., Ma, M., Yang, X., Diao, K., Gao, Y., Song, D., Nowell, K., He, B., Li, Y., McCarthy, K., Paltsyn, M.Y., Sharma, K., Mishra, C., Schaller, G.B., Lu, Z., Beissinger, S.R., 2020. Defining priorities for global snow leopard conservation landscapes. Biol. Conserv. 241, 108387 https://doi.org/10.1016/j.biocon.2019.108387.

Mallon, D.P., Jackson, R.M., 2017. A downlist is not a demotion: red list status and reality. Oryx 51, 605–609. https://doi.org/10.1017/S0030605317000606.

Matthé, M., 2015. About AmphibIdent.

Matthé, M., Sannolo, M., Winiarski, K., Spitzen, -, van der Sluijs, A., Goedbloed, D., Steinfartz, S., Stachow, U., 2017. Comparison of photo-matching algorithms commonly used for photographic capture-recapture studies. Ecol. Evol. 7, 5861–5872. https://doi.org/10.1002/ece3.3140.

McCarthy, T., Mallon, D., Jackson, R., Zahler, P., McCarthy, K., 2017. Snow Leopard (Panthera uncia) [WWW Document]. IUCN red list of threatened species. URL. https://doi.org/10.2305/IUCN.UK.2017-2.RLTS.T22732A50664030.en accessed 6.24.21.

Miguel, A., Beery, S., Flores, E., Klemesrud, L., Bayrakcismith, R., 2016. Finding areas of motion in camera trap images, in: 2016 IEEE International Conference on Image Processing (ICIP). Presented at the 2016 IEEE International Conference on Image Processing (ICIP), IEEE, Phoenix, AZ, USA, pp. 1334–1338. https://doi.org/10.1109/ICIP.2016.7532575.

Miguel, A.C., Bayrakçismith, R., Ferre, E., Bales-Heisterkamp, C., Beard, J., Dioso, M., Grob, D., Hartley, R., Nguyen, T., Weller, N., 2019. Identifying individual snow leopards from camera trap images, in: Mao, K., Jiang, X. (Eds.), Tenth International Conference on Signal Processing Systems. Presented at the Tenth International Conference on Signal Processing Systems, SPIE, Singapore, Singapore, p. 36. https://doi.org/10.1117/12.2521856.

Morrison, T.A., Yoshizaki, J., Nichols, J.D., Bolger, D.T., 2011. Estimating survival in photographic capture-recapture studies: overcoming misidentification error: *Unbiased survival estimation in photograph-ID*. Methods Ecol. Evol. 2, 454–463. https://doi.org/10.1111/j.2041-210X.2011.00106.x.

Moskvyak, O., Maire, F., Armstrong, A.O., Dayoub, F., Baktashmotlagh, M., 2019. Robust Re-identification of manta rays from natural markings by learning pose invariant embeddings. arXiv 1902, 10847.

Moya, Ó., Mansilla, P.-L., Madrazo, S., Igual, J.-M., Rotger, A., Romano, A., Tavecchia, G., 2015. APHIS: A new software for photo-matching in ecological studies. Ecol. Inform. 27, 64–70. https://doi.org/10.1016/j.ecoinf.2015.03.003.

Nguyen, H., Maclagan, S.J., Nguyen, T.D., Nguyen, T., Flemons, P., Andrews, K., Ritchie, E.G., Phung, D., 2017. Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring, in: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Presented at the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, Tokyo, pp. 40–49. https://doi.org/10.1109/DSAA.2017.31.

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J., 2019. A deep active learning system for species identification and counting in camera trap images. arXiv:1910.09716 [cs, eess, stat].

O'Connell, A.F., Nichols, J.D., Karanth, K.U. (Eds.), 2011. Camera Traps in Animal Ecology. Springer Japan, Tokyo. https://doi.org/10.1007/978–4-431–99495-4.

Pal, R., Sutherland, C., Qureshi, Q., Sathyakumar, S., 2022. Landscape connectivity and population density of snow leopards across a multi-use landscape in Western Himalaya. Anim. Conserv. 25, 414–426. https://doi.org/10.1111/acv.12754.

Parham, J., Stewart, C., Crall, J., Rubenstein, D., Holmberg, J., Berger-Wolf, T., 2018. An Animal Detection Pipeline for Identification, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Presented at the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, Lake Tahoe, NV, pp. 1075–1083. https://doi.org/10.1109/WACV.2018.00123.

Riordan, P., Cushman, S.A., Mallon, D., Shi, K., Hughes, J., 2016. Predicting global population connectivity and targeting conservation action for snow leopard across its range. Ecography 39, 419–426. https://doi.org/10.1111/ecog.01691.

Rivest, L., Daigle, G., 2004. Loglinear models for the robust design in mark-recapture experiments. Biometrics 60, 100–107 https://doi.org/10.1111/j.0006-341X.2004.00157.x. PMID: 15032779.

Rivest, L.-P., Baillargeon, S., 2022. Package 'Rcapture.'

Royle, J.A., Young, K.V., 2008. A hierarchical model for spatial capture–recapture data. Ecology 89, 2281–2289. https://doi.org/10.1890/07-0601.1.

Sharma, K., Borchers, D., MacKenzie, D.I., Durbach, I., Sutherland, C., Nichols, J.D., Lovari, S., Zhi, L., Ale, S., Khan, A., Modaqiq, W., McCarthy, T., Alexander, J.S., Mishra, C., 2020. PAWS Guidelines. Population Assessment of the Worlds Snow Leopards.

Sharma, R.K., Singh, R., 2021. Over 100 Years of Snow Leopard Research: A Spatially Explicit Review of the State of Knowledge in the Snow Leopard Range. https://doi.org/10.13140/RG.2.2.25514.54721.

Sultan, H., Rashid, W., Shi, J., Rahim, I. ur, Nafees, M., Bohnett, E., Rashid, S., Khan, M.T., Shah, I.A., Han, H., Ariza-Montes, A., 2022. Horizon scan of transboundary concerns impacting snow leopard landscapes in Asia. Land 11, 248. https://doi.org/10.3390/land11020248.

Suryawanshi, K.R., Khanyari, M., Sharma, K., Lkhagvajav, P., Mishra, C., 2019. Sampling bias in snow leopard population estimation studies. Popul. Ecol. 61, 268–276. https://doi.org/10.1002/1438-390X.1027.

Tariq, N., Saleem, K., Mushtaq, M., Nawaz, M.A., 2018. Snow Leopard Recognition Using Deep Convolution Neural Network, in: Proceedings of the 2nd International Conference on Information System and Data Mining - ICISDM '18. Presented at the the 2nd International Conference, ACM Press, Lakeland, FL, USA, pp. 29–33. https://doi.org/10.1145/3206098.3206114.

Treilibs, C.E., Pavey, C.R., Hutchinson, M.N., Bull, C.M., 2016. Photographic identification of individuals of a free-ranging, small terrestrial vertebrate. Ecol. Evol. 6, 800–809. https://doi.org/10.1002/ece3.1883.

Wäldchen, J., Mäder, P., 2018. Machine learning for image based species identification. Methods Ecol. Evol. 9, 2216–2225. https://doi.org/10.1111/2041-210X.13075.

Weinstein, B.G., 2018. A computer vision for animal ecology. J. Anim. Ecol. 87, 533–545. https://doi.org/10.1111/1365-2656.12780.

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated Residual Transformations for Deep Neural Networks.