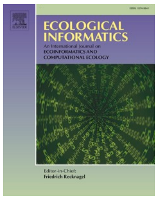
Ecological Informatics 72 (2022) 101876

Contents lists available at ScienceDirect

Ecological Informatics

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



Human vs. machine: Detecting wildlife in camera trap images

Scott Leorna [*], Todd Brinkman

University of Alaska Fairbanks, Institute of Arctic Biology, Fairbanks, AK, USA

ARTICLE INFO

Keywords:
Confusion matrix
Deep learning
Image analysis
Software
Trail camera

ABSTRACT

As the capacity to collect and store large amounts of data expands, identifying and evaluating strategies to efficiently convert raw data into meaningful information is increasingly necessary. Across disciplines, this data processing task has become a significant challenge, delaying progress and actionable insights. In ecology, the growing use of camera traps (i.e., remotely triggered cameras) to collect information on wildlife has led to an enormous volume of raw data (i.e., images) in need of review and annotation. To expedite camera trap image processing, many have turned to the field of artificial intelligence (AI) and use machine learning models to automate tasks such as detecting and classifying wildlife in images. To contribute understanding of the utility of AI tools for processing wildlife camera trap images, we evaluated the performance of a state-of-the-art computer vision model developed by Microsoft AI for Earth named MegaDetector using data from an ongoing camera trap study in Arctic Alaska, USA. Compared to image labels determined by manual human review, we found MegaDetector reliably determined the presence or absence of wildlife in images generated by motion detection camera settings ($\geq$94.6% accuracy), however, performance was substantially poorer for images collected with time-lapse camera settings ($\leq$61.6% accuracy). By examining time-lapse images where MegaDetector failed to detect wildlife, we gained practical insights into animal size and distance detection limits and discuss how those may impact the performance of MegaDetector in other systems. We anticipate our findings will stimulate critical thinking about the tradeoffs of using automated AI tools or manual human review to process camera trap images and help to inform effective implementation of study designs.

1. Introduction

Advancements of tools and technology for generating information have led to the collection and storage of massive expanses of data that can quickly become unwieldy to handle using traditional analytical methods. Often referred to as “Big Data”, these datasets are rapidly generated resulting in high volumes of stored data that needs to be processed to extract pertinent information (Chen et al., 2014; Fosso Wamba et al., 2015; Hariri et al., 2019). When the rate of data production exceeds analytical capacity, this causes a backlog or lag between data collection and use of those data to draw meaningful insights and conclusions. This imbalance between data collection and processing capacity is a common characteristic associated with Big Data and has become a significant challenge in the production of knowledge for many disciplines (Fosso Wamba et al., 2015; Philip Chen and Zhang, 2014). In ecology, a rapidly evolving technology and widely applied tool that has generated Big Data is the use of camera traps (i.e., remotely triggered cameras) to non-invasively survey wildlife (Burton et al., 2015; Farley et al., 2018; Glover-kapfer et al., 2019; Rowcliffe and Carbone, 2008; Trolliet et al., 2014). Today's camera traps are relatively simple to use and affordable and boast long battery life and immense data storage capacity. As the use of camera traps and their capacity to generate Big Data increases, so does the need for efficient and effective ways to process the data (Farley et al., 2018; Thomson et al., 2018; Young et al., 2018).

In the context of wildlife camera trapping, the two most basic data processing tasks are to determine whether (and in some cases where) an animal is present in an image (i.e., object detection) and to assign labels to detected animals (i.e., object classification) (Norouzzadeh et al., 2018; S. Schneider et al., 2018). These pieces of information are often used in assessments of species richness, diversity, distribution, abundance, behavior, and more (Burton et al., 2015; Caravaggi et al., 2017; Sollmann, 2018; Wearn and Glover-kapfer, 2017). Though seemingly simple and straightforward, these tasks can be daunting for individual researchers to undertake themselves as camera trap studies often deploy dozens of cameras and can quickly generate thousands to millions of

[*] Corresponding author at: University of Alaska Fairbanks, P.O. Box 757000, Fairbanks, AK 99775-7000, USA.
E-mail address: sleorna@alaska.edu (S. Leorna).

https://doi.org/10.1016/j.ecoinf.2022.101876
Received 21 July 2022; Received in revised form 17 October 2022; Accepted 17 October 2022
Available online 27 October 2022

images that need to be reviewed and analyzed (Norouzzadeh et al., 2018; Swanson et al., 2015). These data processing tasks often represent a significant bottleneck in the research process from the collection of images to having useful information to answer research questions. In response to this challenge, many have turned to the field of artificial intelligence (AI) to build computer vision models that leverage machine learning to help automate these data processing tasks (Christin et al., 2019; Miao et al., 2019; Norouzzadeh et al., 2018; S. Schneider et al., 2018; Tabak et al., 2019; Thomson et al., 2018; Tuia et al., 2022; Vélez et al., 2022). With rapid adoption and advancement of automated methods, the time-consuming burden of manually reviewing and labeling images has been significantly reduced, ultimately reducing the lag between data collection and application and alleviating some of the Big Data challenges associated with wildlife camera trap data (Farley et al., 2018). However, as new computer vision models are developed and become more accessible, having a thorough understanding of their strengths and limitations is imperative to inform wise decisions for their use in project workflows (Greenberg, 2020a; Vélez et al., 2022; Young et al., 2018). Frequent and critical evaluations of model performance can help contribute to this understanding, however, outcomes are largely dependent on the availability and variety of well-labeled (e.g., accurate, precise, representative) image sets to verify the accuracy of model output and the level of scrutiny manual human reviewers put into identifying wildlife in images (Greenberg, 2020a).

To create a computer vision model for automatically detecting wildlife in camera trap images, typically a subset of data is processed by manual human review to generate a “training” dataset which is then used to train computer algorithms and build a model. Once a model has been developed, it is then applied to unseen raw data (i.e., “test” dataset) and its performance is evaluated by comparing its predictions with the intended outcomes. This process of training and testing a model is often iterative and can be used to refine a model's performance as additional training data become available and new insights are gained from model evaluations (Christin et al., 2019; Norouzzadeh et al., 2018). How well a model performs is dependent on the characteristics of the data the model is confronted with and the associated challenges they present (Greenberg, 2020a; S. Schneider et al., 2020). For example, a model would be expected to perform better with an image set where wildlife were consistently captured near the camera and clearly visible in images compared to if they were generally further from the camera and difficult to see in images. As the vast majority of camera trap studies use a camera's passive infrared (PIR) motion detection sensor to trigger the camera to record an image, the resulting data used to train and evaluate automated models largely consist of wildlife restricted to the camera's motion detection sensor range (typically ≤30 m for large animals under ideal conditions) (Apps and Mcnutt, 2018; Beery et al., 2018; Driessen et al., 2017; Norouzzadeh et al., 2018, 2021; S. Schneider et al., 2018, 2020; Swanson et al., 2015; Trolliet et al., 2014). This leaves some uncertainty about how well automated models may perform at detecting wildlife beyond the camera motion detection sensor range (e.g., in images captured using time-lapse camera settings which record images at preset time intervals where animals can be captured anywhere in the camera's viewshed).

When image sets generated by motion detection are used to assess the performance of an automated object detection model, the task being evaluated is essentially how well the model can detect wildlife that were detected by the camera's motion detection sensor. However, the question of a model's performance is often intended to relate to how well an automated model can detect wildlife in an image that could have been detected by manual human review. The nuances between this actual and intended evaluation are particularly important for camera trap studies conducted in landscapes where targeted animals can be seen beyond the camera's motion detection range or in cases where targeted animals are unlikely to trigger the camera's PIR motion detection sensor (e.g., very small or well-insulated animals) (Meek et al., 2014; Rovero et al., 2013). In such cases, time-lapse settings may expand the area sampled at each camera site and/or capture animals that would not have otherwise been detected (Hamel et al., 2013). The extent to which time-lapse triggered images may be beneficial is dependent on how well the method used to process images can reliably detect wildlife beyond the motion detection range of the camera. Therefore, critically evaluating a computer vision object detection model using a rigorously curated image set generated by both motion detection and time-lapse camera settings (i.e., where animals are assumed to be randomly distributed in the viewshed rather than biased toward the camera) may lead to novel insights into how well automated models perform compared to manual human review and inform practical decisions about the most efficient and effective way to process camera trap images for specific study conditions.

Many automated image processing models have been developed to address specific computer vision challenges under a specific set of circumstances or conditions, however, these models often do not perform well in broader applications (Thomson et al., 2018). For example, a model developed specifically to detect birds would likely not perform well at identifying large mammals (Beery et al., 2019; S. Schneider et al., 2020). With this in mind, Microsoft AI for Earth set out to develop a universal image processing model focused on object detection that would perform well with camera trap images of wildlife from around the world in a variety of ecosystems; the result of this effort was the creation of a state-of-the-art computer vision model named MegaDetector V4.1 (Beery et al., 2019; Microsoft and for Earth., 2020). MegaDetector uses a two-stage process known as a Faster Region-based Convolutional Neural Network (Faster R-CNN) which first seeks to identify all regions in an image that contain an object and then examines attributes of those specific regions to assign objects to a particular class (Ren et al., 2015). In other words, the model first searches the image to determine areas that contain an object not part of the background scene and draws bounding boxes around those objects. Then, each object is reviewed and assigned an object classification and confidence value indicating the model's confidence in the selected object class. MegaDetector has been trained using hundreds of thousands of images labeled by manual human review from a wide variety of ecosystems and classifies objects as either an animal (any non-human animal), human (any person), or truck (any vehicle) (Beery et al., 2019; Microsoft and for Earth., 2020). This two-step process of Faster R-CNNs leads to slower processing speeds compared to other single-step methods, however, they have been shown to result in greater accuracy in detecting objects in complex images (Huang et al., 2017; S. Schneider et al., 2018; Vecvanags et al., 2022). To facilitate visualization of MegaDetector's output (i.e., bounding-boxes around detected objects with class labels and confidence values), it has been integrated with software developed for manually processing camera trap images such as TimeLapse2 (Greenberg et al., 2019) which allows users to visually inspect MegaDetector's results for each image and organize the dataset according to various attributes from MegaDetector's output. For example, within TimeLapse2, users can organize images based on MegaDetector's confidence thresholds for labeled objects or their classes (e.g., select images where MegaDetector detected an animal with a confidence value > 75%) which can then be easily reviewed and further manually annotated with additional relevant information (e.g., species, behavior, count, etc.) (Greenberg, 2020a; Vélez et al., 2022). The generalizability of MegaDetector and its integration with existing software to manually review model results familiar to many camera trap researchers provides a streamlined and adaptable workflow that facilitates the transition between automated and manual image processing and allows manual human reviewers to make corrections to image labels and/or add supplemental annotations to images to address specific study objectives. Additionally, there are few “off-the-shelf” computer vision models that are readily available that can be applied to any camera trap project by ecologists with limited computer science or coding skills. These qualities of MegaDetector make it an appealing option for a wide diversity of camera trap projects and its use and popularity has been growing around the world (Microsoft and for Earth., 2020).

To gain a greater understanding of how automated AI tools can contribute to Big Data challenges associated with wildlife camera trap projects, we evaluated performance of Microsoft AI for Earth's MegaDetector (V4.1) using camera trap images generated by both motion detection and time-lapse camera settings from an ongoing wildlife study in Arctic Alaska, USA. We decided to focus our study on MegaDetector because it was specifically developed with wildlife camera trapping in mind, is freely available and accompanied with excellent documentation and supporting information facilitating its ease of use and accessibility, and because developers were eager to help with running the model, interpreting results, and interested in understanding how the model was being used by wildlife camera trapping projects and how it could be improved for end-users (Microsoft and for Earth., 2020). Our main objectives were to 1) compare the performance of manual human review and MegaDetector for detecting wildlife in camera trap images by evaluating the proportion of motion detection and time-lapse images correctly labeled with animal presence or absence by each image processing method and 2) identify the minimum detection size (i.e., smallest pixel area occupied by an animal in the image) and maximum detection distance (i.e., furthest distance between the camera and animal) at which each camera trigger type (i.e., motion detection and time-lapse) and image processing method failed to reliably detect wildlife (i.e., detection limits). We discuss the implications of our findings in terms of the maximum area sampled in the camera's viewshed (i.e., area where animals can be detected). We anticipated wildlife captured in motion detection triggered images would generally be larger and/or closer to the camera (i.e., easier to detect and identify in images) compared to those in time-lapse triggered images and that manual human review would detect and identify smaller objects and/or objects further from the camera (i.e., represented by animals occupying a smaller pixel area in the image) compared to MegaDetector. Consequently, we expected fewer discrepancies in images labeled with/without animals (i.e., comparable performance) between manual human review and MegaDetector for motion detection triggered images compared to time-lapse triggered images. Based on our findings, we provide considerations for selecting an image processing method to balance efficiency and accuracy for various camera trapping project needs. Our findings and manual human-reviewed dataset may also be used to inform and improve future development of computer vision models for automated image analysis.

2. Methods

2.1. Study area

We used images from an ongoing camera trap project (2019–2023) in Arctic Alaska, USA with a particular emphasis on barren-ground caribou (*Rangifer tarandus*). Images from this ecosystem provide useful characteristics to evaluate the performance of MegaDetector under ideal conditions for several reasons. First, the open landscape of Alaska's Arctic tundra has limited tall vegetation (e.g., treeless) and relatively flat topography. Therefore, the camera's viewshed and the distance at which animals can be seen is well beyond the motion detection range of the camera, maximizing the potential for time-lapse images to record additional detections and provide novel insights (e.g., identifying minimum detection size, maximum detection distance). Second, the background scene is generally uniform which is ideal for detecting objects using both automated models and manual human review. Third, there is a relatively low diversity of wildlife species, most of which have physical characteristics similar to those which are well represented in MegaDetector's training data (e.g., large mammals). Fourth, species most likely to be captured in images have distinctive morphological and behavioral characteristics which can be used by manual human reviewers to help determine the correct identity of objects. For example, it is very uncommon in this region to observe large mammals form groups larger than ~10 individuals other than caribou. These characteristics are useful because they enhance our ability and confidence in manually detecting and correctly labeling wildlife, providing a rigorous dataset for ground-truthing which may allow us to identify MegaDetector's detection limits for various wildlife under optimal conditions and help establish realistic expectations for automated image processing of wildlife camera trap images.

2.2. Image collection

We used Reconyx HyperFire 2 HF2X camera traps which were mounted ~1 m above the ground and recorded images with a 2048 × 1440 pixel resolution. Cameras recorded 5 min interval time-lapse images during a 24 h period and a 3-image sequence when motion-triggered with 10s between images and a 30s quiet period. Motion sensitivity was set to very high. We selected images from 20 sampling sites during a 24 h period when each site experienced relatively high animal activity (≥288 images/site/day) to represent variation in camera placement and background and capture a wide diversity of animal characteristics while maintaining a sufficient and manageable dataset for comparison. The time period from which data were used spanned from May 2019, when the landscape was largely snow-covered and melting, to the beginning of September 2019, when the landscape was snow-free and temperatures were returning to freezing.

2.3. Image processing

To facilitate direct comparison between image processing methods, the same image set was used for both manual human review and automated processing, and all images were processed independently before results were reviewed (i.e., each processing method was blind to the results of the other). We restricted our analysis and discussion to MegaDetector's performance in only the animal class (i.e., objects labeled by MegaDetector as either “person” or “truck” were not considered). Also, our evaluation focused on the most common first step for camera trap image processing which is to separate images containing animals of interest (detections) from images without (non-detections; commonly referred to as “empty”). Therefore, with each processing method, each image received a binary label of either a 1 (indicating ≥1 animal detected) or 0 (indicating no animals detected).

To manually label images, we used a two-stage human review process using the free software TimeLapse2 v2.2.3.6 (Greenberg, 2020b; Greenberg et al., 2019). First, all images were reviewed and manually labeled independently by three trained human reviewers. Images were played at rapid frame rates (up to 15 frames/s) to detect subtle changes between consecutive images. When objects were detected, they were carefully reviewed and assigned to one of the following classes: caribou, bird (Aves spp.), microtine (Arvicolinae spp.), ground squirrel (*Urocitellus parryii*), or fox (*Vulpes vulpes*). Second, labeled datasets from all reviewers were combined and discrepancies in image labels were given a secondary detailed review to make a final determination. The zoom function in TimeLapse2 was often used when carefully inspecting images and when toggling between consecutive images. The image trigger type (motion detection or time-lapse) was extracted from image file metadata using TimeLapse2 and assigned to each image in the dataset (Greenberg et al., 2019).

For automated image labeling, we provided our image set to the Microsoft AI for Earth Team who ran the images through MegaDetector and returned to us an image recognition file that specified MegaDetector's output for each image. Through the integration of MegaDetector with TimeLapse2, we uploaded the image recognition file and visually reviewed MegaDetector's results for each image which contained bounding boxes around detected objects with class labels and confidence values (Greenberg, 2022). We did not make any corrections to MegaDetector's labels and no specific optimization or customization was made specifically for this image set. The resulting labeled dataset was therefore the initial raw output from MegaDetector. The intent of this was to provide an honest and unbiased perspective on how

MegaDetector performed compared to manual human review and describe possible advantages and disadvantages of using readily available automated image processing tools.

2.4. Data analysis

With labels from manual human review as a reference, we determined true positives (TP; ≥1 of MegaDetector's bounding boxes was confirmed to contain an animal), true negatives (TN; neither MegaDetector nor manual human review detected any animals), false positives (FP; MegaDetector labeled ≥1 object in an image which was determined to have no animals present by manual human review), and false negatives (FN; MegaDetector labeled an image as empty which was confirmed to contain ≥1 animal by manual human review) (Fig. 1). In addition, we determined false-true positives (FTP) which were cases when MegaDetector's bounding-box(s) around a detected animal(s) did not contain an animal(s), however, an animal(s) was found elsewhere in the image from manual human review (Fig. 1). FTP were combined with FP for analysis.

As objects labeled by MegaDetector are influenced by the object detection threshold used (i.e., value indicating the model's confidence in the assigned label), we included thresholds of > 0%, ≥ 25%, ≥ 50%, and ≥ 75% to provide a broader perspective on how this can influence performance results. To ensure labels determined from manual human review served as a quality “truth” reference, we visually reviewed all images (i.e., TP, TN, FP, FTP, and FN) from MegaDetector's > 0% confidence threshold (i.e., greatest number of objects labeled) with a particular focus on FP and FN since we wanted to ensure manual human reviewers did not miss any relevant objects in images or incorrectly label objects.

To examine how image sets generated by motion detection or time-lapse camera trigger settings influenced the performance of MegaDetector compared to manual human review at detecting wildlife in camera trap images (Obj. 1), we separated motion-triggered images from time-lapse images and report performance metrics on these separately. To describe the detection limits of wildlife for each trigger type and image processing method (Obj. 2), we chose a subset of images where animals were at distances corresponding to the maximum camera motion detection range, maximum MegaDetector detection range, and maximum manual human reviewer detection range and measured their pixel area (i.e., minimum detection size) using the free software ImageJ (C. A. Schneider et al., 2012). We felt using the pixel size of objects would be the most broadly applicable and generalizable metric for describing detection size limits because we assumed large objects far away and small objects closer to the camera would be comparably difficult to identify when they occupied the same number of pixels in an

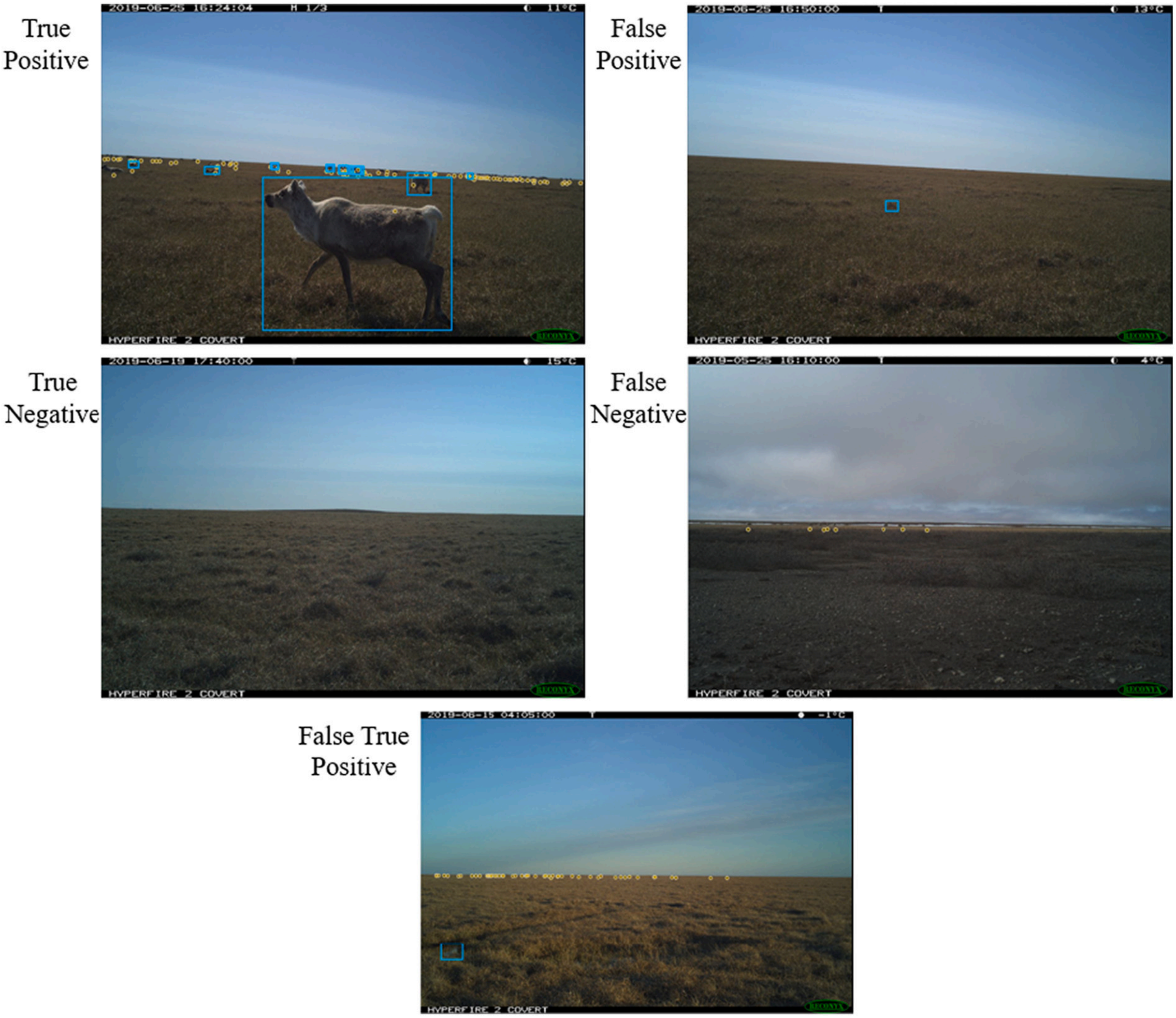


Fig. 1. Example of final image results determined by comparing animal labels from Microsoft AI for Earth's MegaDetector (i.e., blue bounding boxes) with animal labels from manual human reviewers (i.e., yellow dots) as a truth reference. Animals shown here are caribou from Arctic Alaska. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

image. For example, we thought identifying a 50-pixel small bird relatively nearby would likely be as challenging as identifying a 50-pixel large mammal relatively further away, provided the model was adequately trained on both objects. To estimate the maximum detection distance, we used information available on the captured animals' morphometrics (e.g., nose-to-tail length) and the corresponding apparent size of selected morphometrics in the image (i.e., pixel length measured using ImageJ) to estimate the distance between the camera and the animal using a photogrammetric technique described by Leorna et al. (2022). Then, using the maximum detection distance and the angular field of view of the camera (i.e., approximately 37.7 degrees for the Reconyx HF2X), we approximated the area sampled in the camera viewshed by each camera trigger type and image processing method using the equation for the area of a circle sector (i.e., area sampled = (maximum detection distance ^{2} X angular field of view of camera)/2).

3. Results

A total of 6224 images were included in the analysis (7.5% motion-triggered, 92.5% time-lapse) with 2862 images (46.0% of total) determined by manual human review as containing $\geq$ 1 animal (13.6% motion-triggered, 86.4% time-lapse) and 3362 images (54.0% of total) determined to not contain any animals (2.3% motion-triggered, 97.7% time-lapse) (Table 1). The proportion of images from each camera trigger type determined by manual human review as containing $\geq$ 1 animal was 83.7% for motion detection images and 42.9% for time-lapse images (Table 1). For motion detection images where animals were detected, caribou were in 99.5% and birds in 2.3% of images. For time-lapse triggered images where animals were detected, caribou were in 83.6% and birds in 18.1% of images, and $<$ 1.0% of images contained microtines, ground squirrels, or fox (Table 1). Examining FP and FN from the $>$ 0% MegaDetector confidence threshold, we did not find any images where manual human review missed an animal or determined an animal to be present which was absent. Therefore, we were confident that our manual human review process provided a quality truth reference for evaluating MegaDetector's object detection performance.

3.1. Manual human review vs. MegaDetector (Obj. 1)

For images recorded by the camera motion detection sensor, MegaDetector correctly labeled (i.e., TP + TN) between 94.6% (at the $>$ 0% detection threshold) and 95.7% (at the $\geq$ 75% detection threshold) of images (Fig. 2). For images recorded by the time-lapse trigger setting, MegaDetector correctly labeled between 55.7% (at the $>$ 0% detection threshold) and 61.6% (at the $\geq$ 75% detection threshold) of images (Fig. 2). Adjusting MegaDetector's object detection threshold resulted in a relatively small improvement for motion detection images compared to time-lapse triggered images (i.e., 1.1% and 5.9% increase in correctly labeled images, respectively) (Fig. 2). A summary of all class labels (e.g., TP, TN, FP, FTP, and FN) determined by comparing MegaDetector's predictions with labels from manual human review is presented in Table A1. Additionally, a summary of common performance metrics often used to describe computer vision model performance (e.g., recall, precision, F_1, Matthews Correlation Coefficient, etc.) is presented in Appendix Table A2.

Table 1
Image set characteristics of wildlife camera trap data from Arctic Alaska.

Animal	Total Number of Images with $\geq$ 1 Present	Images with $\geq$ 1 Present Based on Camera Trigger Type	
		Motion Detection	*Time-Lapse*
Caribou	2391 (38.4%)	380 (81.7%)	2011 (34.9%)
Bird	388 (6.2%)	2 (0.4%)	386 (6.7%)
Caribou & Bird	62 (1.0%)	7 (1.5%)	55 (1.0%)
Microtine	13 (0.2%)	0 (0.0%)	13 (0.2%)
Bird & Ground Squirrel	6 (0.1%)	0 (0.0%)	6 (0.1%)
Caribou & Ground Squirrel	1 (<0.1%)	0 (0.0%)	1 (<0.1%)
Fox	1 (<0.1%)	0 (0.0%)	1 (<0.1%)
None	3362 (54.0%)	76 (16.3%)	3286 (57.1%)
Total Images	6224	465 (7.5%)	5759 (92.5%)

3.2. Detection size and distance limits (Obj. 2)

Caribou were the only species with a sufficient sample to confidently identify the minimum detection pixel size (i.e., the minimum area in the image the animal must occupy to be detected) for the camera motion detection sensor, MegaDetector, and manual human reviewers. By examining images where animals were distributed at different distances from the camera, we found the minimum detection size limit of caribou was approximately 600px for the camera motion detector sensor, 60px for MegaDetector at the $>$ 0% confidence threshold, and 4px for manual human reviewers (Fig. 3). We estimated the maximum detection distance for caribou was approximately 29 m for the camera motion detection sensor, 222 m for MegaDetector with the $>$ 0% confidence threshold, and 2551 m for manual human reviewers (Fig. 4). Based on the angular field of view of the camera, these maximum detection distances correspond with the maximum area sampled for caribou of 273 m^2 for the camera motion detection sensor, 16,194 m^2 for MegaDetector with the $>$ 0% confidence threshold, and 2,141,651 m^2 (2.14 km^2) for manual human reviewers (Fig. 4).

4. Discussion

We found Microsoft AI for Earth's computer vision model MegaDetector performed exceptionally well compared to manual human review (94.6% and 95.7% correctly labeled images at the $>$ 0% and $\geq$ 75% detection thresholds, respectively) at detecting wildlife in images collected by the camera motion detection setting without any customization of the model or modifications to initial labels (Fig. 2). This finding is consistent with other evaluations of MegaDetector's performance in a variety of study systems (Fennell et al., 2022; Vélez et al., 2022). We identified the minimum detection size limit for MegaDetector was approximately ten times smaller than for the camera motion detection sensor (i.e., 60px for MegaDetector at the $>$ 0% confidence threshold and 600px for the camera motion detector sensor) (Fig. 3), resulting in the maximum detection distance and area sampled by MegaDetector being eight and 59 times greater, respectively, compared to the camera motion detection sensor (Fig. 4). These results suggest that MegaDetector is likely to perform well when evaluated on image sets generated by camera motion detection settings as captured wildlife are expected to be well within the detection limits of MegaDetector. These findings highlight the potential utility of integrating MegaDetector in camera trap image processing workflows for many camera trap studies using motion detection camera settings as there is only a minor difference in performance compared to manual human review (Fig. 2) with the potential of dramatically increasing efficiency (Beery et al., 2019; Fennell et al., 2022; Greenberg, 2020a; Norouzzadeh et al., 2021; Vélez et al., 2022). For example, Microsoft reports that a quality consumer laptop without dedicated/specialized hardware can process approximately 4000–10,000 images/day, while using their system, they can process approximately 3.8 million images/day using specialized hardware and distributing the workload over many processors (i.e., 16 NVIDIA V100 GPUs) (Microsoft and for Earth., 2020). Although MegaDetector's processing speed is dependent on the hardware it is run on and the efficiency of manual human review is influenced by reviewers' skill level and amount of scrutiny put into reviewing images, it is likely

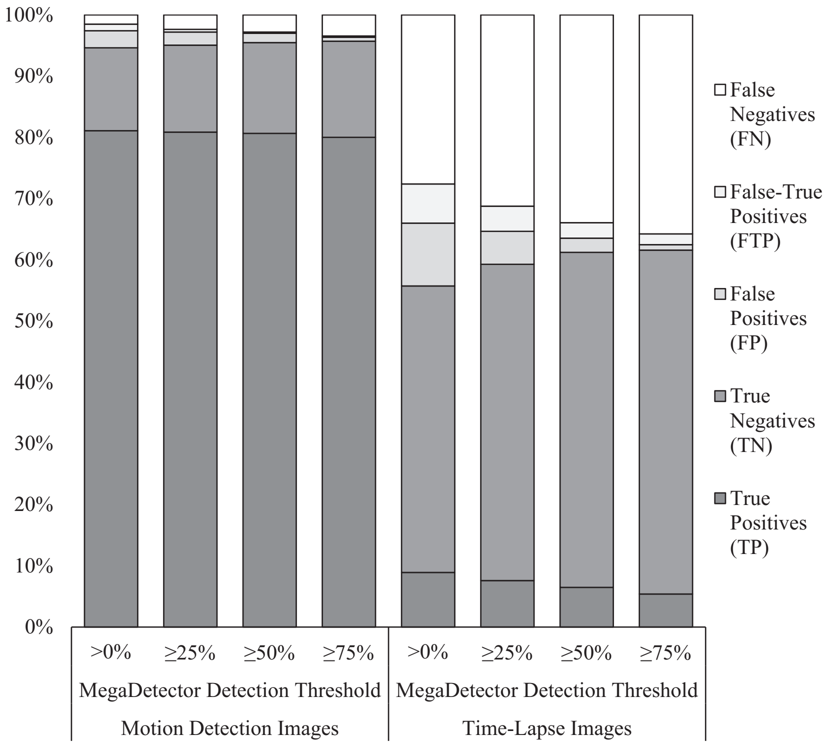


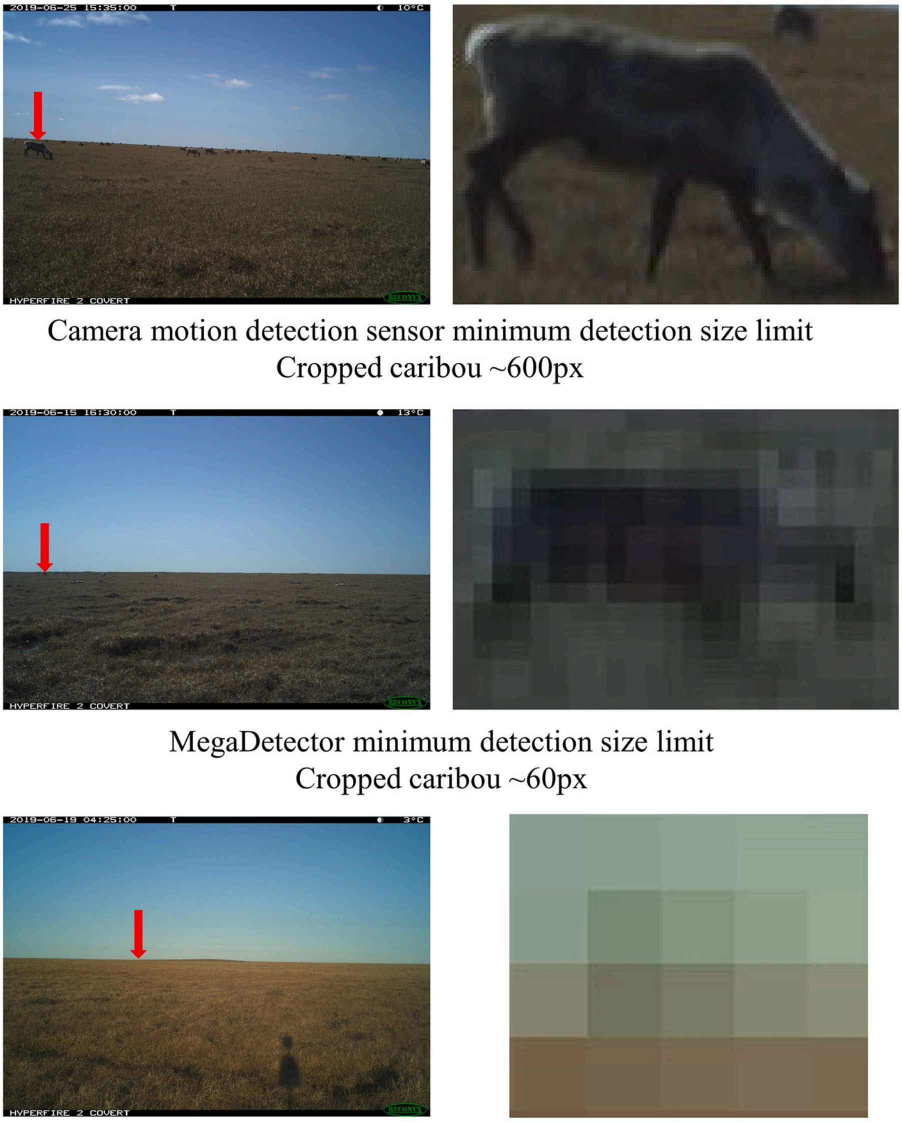
Fig. 2. Summary of the proportion of motion detection and time-lapse images correctly labeled with animal presence or absence by Microsoft AI for Earth MegaDetector at different object detection thresholds (i.e., value indicating the model's confidence in assigned labels). Correct labels (i.e., TP – true positives, TN – true negatives) are represented in darker shades and incorrect labels (i.e., FP – false positives, FTP – false-true positives, FN – false negatives) are represented in lighter shades and were determined from manual human review. *Note*: Specific values are presented in Table A1.

far more efficient to process images with AI tools rather than manual human review in most cases (Fennell et al., 2022). However, the tradeoffs between efficiency and accuracy may not be warranted for smaller datasets that are more manageable for manual human review.

While MegaDetector performed excellent on motion detection images, we found performance was substantially poorer compared to manual human review (55.7% and 61.6% correctly labeled images at the > 0% and ≥ 75% detection thresholds, respectively) (Fig. 2) when evaluated using images generated by the camera time-lapse setting. This result is likely due to the minimum detection size limit of manual human reviewers being 15 times smaller compared to MegaDetector (i.e., 4px for manual human reviewers and 60px for MegaDetector at the > 0% confidence threshold) (Fig. 3). This corresponds with the maximum detection distance and area sampled for caribou being 11 and 132 times greater, respectively, when manual human review was used to process images compared to MegaDetector (Fig. 4). These results support our predictions and indicate why there was a greater discrepancy in performance (i.e., proportion of correctly labeled images) between MegaDetector and manual human review for time-lapse compared to motion detection images (i.e., MegaDetector detected smaller objects than the camera motion detection sensor, and manual human reviewers detected smaller objects than MegaDetector) (Fig. 2). Additionally, these results indicate that the predominant method used to generate images in camera trap studies (i.e., camera motion detection) underutilizes both computer vision and manual human reviewers' capacity to identify wildlife in images, ultimately restricting the potential insight that can be gained from model evaluations and camera trapping in general. While the apparent size of an object in an image has been suggested to influence the likelihood of detection (Beery et al., 2018; Greenberg, 2020a; Norouzzadeh et al., 2018), to our knowledge, our study represents the first to quantify this limitation, providing a performance metric that may be more generalizable to different model evaluations and study systems.

In our system, we found camera motion detection and time-lapse trigger settings generated image sets with considerably different characteristics (Table 1), contributing to the differences we observed in MegaDetector's performance compared to manual human review (Fig. 2). For example, we found time-lapse images contributed six times greater total number of images containing ≥1 animal and captured several animal groups that were missed by the camera motion detection sensor (Table 1). Although there were very few images captured by time-lapse of the animal groups missed by motion detection (i.e., microtine, ground squirrel, and fox) (Table 1), this information may be particularly important for assessments such as occupancy, species richness, or species diversity (Burton et al., 2015; Sollmann, 2018; Wearn and Glover-kapfer, 2017). This reveals the potential added value of using time-lapse settings in open landscapes where targeted animals can be observed beyond the camera motion detection range (e.g., tundra, treeless alpine, prairie grasslands, or deserts) or for studies targeting small or well-insulated species that are unlikely to trigger the camera motion detection sensor (Apps and Mcnutt, 2018; Driessen et al., 2017; Hamel et al., 2013). We note, however, that our camera installation was intended to primarily capture large terrestrial mammals (i.e., cameras ~1 m above the ground near the shoulder height of caribou), and as such, detection of small animals near the ground by motion detection was unlikely. We also found image sets differed in the proportion of images containing ≥1 animal (i.e., ratio of detections:non-detection), with motion detection images having almost two times greater proportion of animal images compared to time-lapse images (i.e., 83.7% and 42.9%, respectively) (Fig. 2). These results suggest motion detection settings in our system were more efficient with respect to battery life and data storage with the tradeoff of contributing far fewer detections overall as the result of having a dramatically reduced maximum detection distance and area sampled compared to time-lapse images (Table 1, Fig. 4). The ideal camera settings (e.g., motion detection sensitivity, time-lapse interval length) to maximize the proportion of images containing animals is likely case specific (i.e., conditional on target species, landscape, battery life/data storage, etc.) and may not have been reflected in the image collection protocols used in this study. Future research may attempt to identify the tradeoffs between these image collection considerations and animal detection rates for different study systems to inform optimal case-specific image collection protocols.

As our results demonstrated performance of an automated computer



Camera motion detection sensor minimum detection size limit
Cropped caribou ~600px

MegaDetector minimum detection size limit
Cropped caribou ~60px

Manual human review minimum detection size limit
Cropped caribou ~4px

Fig. 3. Example of the minimum detection size limit in pixels (px) of caribou for the camera motion detection sensor, MegaDetector, and manual human reviewers. *Note*: the identity of the object detected by manual human reviewers could only be confidently determined to be a caribou given the low diversity of large terrestrial species in this study region and by reviewing consecutive images and observing the behavior of the object.

vision model can be influenced by characteristics of the image set used in evaluation (Fig. 2), the detection limits (i.e., minimum detection size, maximum detection distance/area) we identified may offer camera trap users more generalizable and/or practical information to guide planning and implementation of study designs. For example, given a particular study system (i.e., target species, landscape, field equipment, data collection and processing strategy, etc.), understanding the maximum detection distance may inform the location and placement of cameras in the field (e.g., proximity to trails/travel corridors, bait/lure, etc.) and/or help in determining whether potential obstructions should be removed or avoided to reduce the possibility of occlusion (i.e., animals blocked by tall vegetation, bushes, trees, rocks, or other objects) (Apps and Mcnutt, 2018). Also, understanding the maximum area sampled may help decide which camera trap model(s) to use and how many to budget for, inform how to distribute cameras on the landscape, and/or provide spatial context by which to interpret data (e.g., were there a greater number of animal detections at a particular location because it was more used/preferred by wildlife or because it sampled a larger area compared to another) (Burton et al., 2015; Driessen et al., 2017; Rovero et al., 2013; Wearn and Glover-kapfer, 2017). As many analytical methods to estimate wildlife metrics from camera trap data benefit from or require estimating the distance to detected wildlife and the associated area sampled in the camera viewshed (Gilbert et al., 2020; Moeller et al., 2018; Rowcliffe et al., 2008; Sollmann, 2018), our study provides an example of how this information can be generated and how choices of camera settings and image processing methods can dramatically impact the data available to inform study objectives. While our study and results focused on caribou in the vast and open landscape of the Arctic, our approach and emphasis in identifying and describing the impacts of camera settings and image processing methods on animal detection limits has universal merit for all camera trap studies in any ecosystem. Similar evaluations would contribute to making more informed and intentional data collection and processing decisions across camera trap applications.

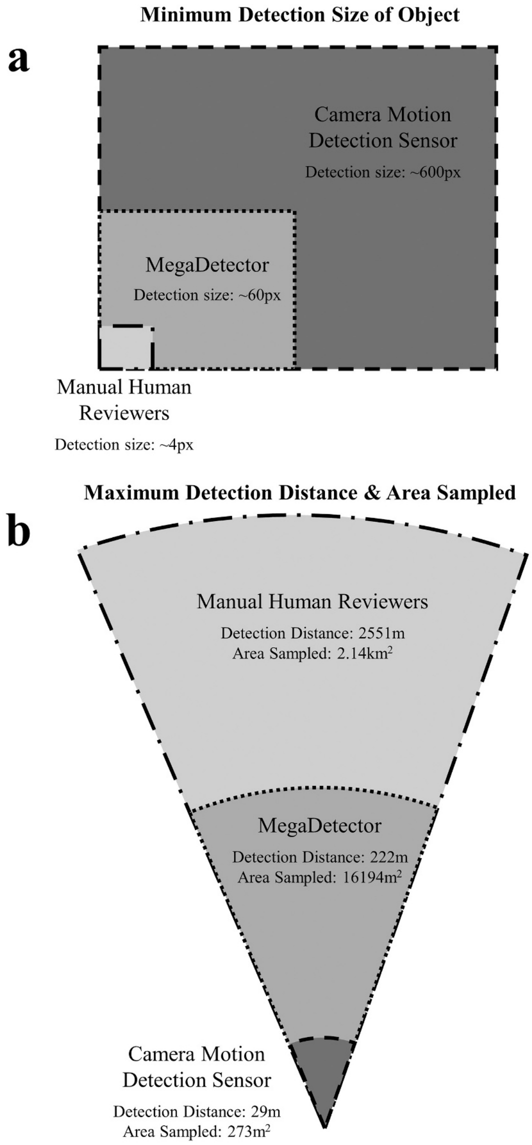


Fig. 4. Conceptual model of the (a) minimum detection size limits and the associated (b) maximum detection distance and area sampled for caribou of the camera motion detection sensor, MegaDetector, and manual human reviewers. *Note*: the differences between size, distance, and area are not to scale.

4.1. Limitations and considerations

While we gained valuable insights into the potential impacts of using different methods to collect and process camera trap images, there are several important things to consider when interpreting our results. First, our evaluation of MegaDetector's performance only relates to the most common first step of processing camera trap images (i.e., separating images with/without animals). Supplemental information such as how many individuals are present in images is likely of interest to many camera trap researchers for metrics such as relative abundance estimates, density estimates, or demographics (Burton et al., 2015; Sollmann, 2018; Wearn and Glover-kapfer, 2017). However, given animals captured in our images were often in large, tightly packed groups, the exact number of individuals contained in the image was often unclear which made evaluating MegaDetector's individual bounding boxes around detected animals somewhat ambiguous as they often contained (or partially contained) several individuals (Fig. A2). Therefore, consolidating MegaDetector's output to binary image labels (i.e., ≥1 animal detected or no animals detected) was the most conservative and sensible way to evaluate MegaDetector's performance in this system.

Second, we note that absolutely no customization of MegaDetector or editing of the image labels it produced were made for our evaluation, meaning our performance results were based on the initial raw output from MegaDetector and are inextricably linked to the specific characteristics of the image set we used in evaluations (e.g., ratios of detections:non-detections, animal species captured, habitat and topography in landscape, etc.). Performance metrics could likely be improved by adding a manual human review stage in the workflow (e.g., correcting obvious mislabels) and/or by making customizations to MegaDetector by including additional training data from the particular study system in which it will be used (Beery et al., 2018, 2019; Greenberg, 2020a; Norouzzadeh et al., 2021; S. Schneider et al., 2020; Vélez et al., 2022). This points to the iterative process of improving AI model performance which is likely to continue to improve as more training data are made available and new model advancements are evaluated and modified. We encourage prospective AI users to carry out similar evaluations as performed in this study to gain more relevant information on model performance specific to their study systems. However, we anticipate the minimum detection size limits we identified are more generalizable to different study systems provided the model is well trained on a particular object class (i.e., under ideal conditions, MegaDetector is likely to detect animals when they occupy at least 60px in an image) (Fig. 3). Other image characteristics such as complex vegetation, poor lighting, low image resolution, and unique camera angles are likely to decrease MegaDetector's overall performance (Beery et al., 2018; Greenberg, 2020a).

Third, we acknowledge that the actual task being completed was inherently different between MegaDetector and manual human reviewers. For example, MegaDetector attempts to find and identify objects within a single image as opposed to manual human reviewers often using a series of images, and the changes between consecutive images, to find and identify objects (i.e., manual human reviewers used more information from the data to draw conclusions). We found when we played the 5 min interval time-lapse imagery at rapid frame rates in TimeLapse2 (up to 15 frames/s), the still images closely resembled video making objects in consecutive images appear as movement rather than still objects on an individual image. We found detecting the “motion” of objects much easier and consistent than identifying objects from still images (Fleming and Tracey, 2008; Greenberg et al., 2019). While the task at hand may not be directly comparable between MegaDetector and manual human reviewers, our intent was to identify and describe characteristics that may reduce or limit the performance of MegaDetector (e.g., minimum object detection size), and thinking about how these challenges were overcome by manual human review may help stimulate thinking and development about how to improve automated computer vision methods (e.g., using image series vs still images or converting time-lapse images into video).

Lastly, while we identified and discussed several impacts of using different camera trigger settings and image processing methods as they relate to limits for detecting wildlife in images, we note that reported estimates are solely based on reviewing caribou in images and are likely accompanied with some level of error. Future studies may attempt to expand the detection limits to a wider variety of animals and in various study systems to generate estimates with greater precision and generalizability. We suggest researchers carefully examine the characteristics of their study system in regard to what data processing challenges they may encounter based on their data collection methods, target species, and landscape and what information they need to address study objectives and consider and weigh tradeoffs between efficiency and accuracy (Fennell et al., 2022; Greenberg, 2020a; Huang et al., 2017; Norouzzadeh et al., 2018; S. Schneider et al., 2018, 2020; Thomson et al., 2018; Tuia et al., 2022; Vélez et al., 2022).

5. Conclusions

In this study, we identified some of the strengths and limitations of using AI tools compared to manual human review for detecting wildlife in camera trap images. We found Microsoft AI for Earth's state-of-the-art object detection model MegaDetector performed exceptionally well at detecting wildlife in camera trap images triggered by the camera motion detection sensor, however, performance was substantially worse compared to manual human review for time-lapse triggered images (Fig. 2). We found differences in detection limits (i.e., minimum detection size, maximum detection distance/area) provided more generalizable and/or practical information to interpret findings and anticipate these metrics will help camera trap users make more well informed decisions about whether a particular computer vision model may be suitable for their specific study conditions (Fig. 4). While there are many performance measures commonly used to evaluate automated object detection models (e.g., those reported in Table A2), greater dialogue among model developers and camera trap users may contribute to greater use of metrics with more practical use, interpretation, and generalizability such as the detection limits identified in this study. Also, as we found performance was strongly influenced by the characteristics of the data being evaluated, we strongly encourage future studies provide detailed descriptions of the image sets generated and used in evaluations to aid in interpretation (e.g., how the images were collected, which animals were captured, proportions of detections:non-detections, characteristics of the landscape, etc.) (Meek et al., 2014; Scotson et al., 2017).

AI tools provide an appealing solution for Big Data challenges associated with many camera trap studies and we anticipate they will continue to improve as more training data are incorporated in model development. To facilitate this progress, however, inevitably more painstaking manual human review of images will need to occur. While no one approach for collecting and processing wildlife camera trap data is likely to be optimal for all study systems, our study provides an example of how choices of camera trigger settings and image processing methods can dramatically influence data available to draw insights from and the importance of making well-informed and intentional decisions about which methods to use based on study specific conditions.

Data availability

Labeled images and data used in this study are available through the National Science Foundation's Arctic Data Center and can be found at https://doi.org/10.18739/A2J38KJ9R (Leorna, 2022).

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.

Acknowledgements

We thank the National Science Foundation for providing funding for this study [Office of Polar Programs, Arctic System Science Program Award # 1839192]. We thank technicians Michael Apperson and Katja Diekgers for their many hours of manually reviewing and labeling images. We thank Dan Morris and the Microsoft AI for Earth Team for their collaboration in running our images through MegaDetector and for what we anticipate will be an instrumental tool for the advancement of camera trap studies in wildlife conservation. We sincerely appreciate the dedication to providing an open-sourced model that continues to improve and stimulate thinking and dialogue among wildlife professionals and computer scientists.

Appendix A. Appendix

Table A1
Microsoft AI for Earth MegaDetector model predictions for detection of wildlife in camera trap images from Arctic Alaska.

Class Label		Motion Detection Images				Time-Lapse Images			
		MegaDetector Detection Threshold				*MegaDetector Detection Threshold*			
		> 0%	*≥ 25%*	*≥ 50%*	*≥ 75%*	*> 0%*	*≥ 25%*	*≥50%*	*≥ 75%*
Correct Predictions	True Positives (TP)	377 (81.1%)	376 (80.9%)	375 (80.6%)	372 (80.0%)	514 (8.9%)	437 (7.6%)	373 (6.5%)	311 (5.4%)
	True Negatives (TN)	63 (13.5%)	66 (14.2%)	69 (14.8%)	73 (15.7%)	2695 (46.8%)	2976 (51.7%)	3152 (54.7%)	3235 (56.2%)
Incorrect Predictions	False Positives (FP)	13 (2.8%)	10 (2.2%)	7 (1.5%)	3 (0.6%)	591 (10.3%)	310 (5.4%)	134 (2.3%)	51 (0.9%)
	False-True Positives (FTP)	5 (1.1%)	2 (0.4%)	1 (0.2%)	1 (0.2%)	368 (6.4%)	237 (4.1%)	145 (2.5%)	102 (1.8%)
	False Negatives (FN)	7 (1.5%)	11 (2.4%)	13 (2.8%)	16 (3.4%)	1591 (27.6%)	1799 (31.2%)	1955 (33.9%)	2060 (35.8%)

Note: Values indicate the number of images in the corresponding category with the column proportions presented in parentheses and are represented graphically in Fig. 2.

Table A2
Summary of performance metrics of Microsoft AI for Earth MegaDetector for detecting wildlife in camera trap images from Arctic Alaska.

Performance Measure	Equation	Motion Detection				Time Lapse				Overall			
		MegaDetector Detection Threshold				*MegaDetector Detection Threshold*				*MegaDetector Detection Threshold*			
		> 0	*≥ 25*	*≥ 50*	*≥ 75*	*> 0*	*≥ 25*	*≥ 50*	*≥ 75*	*> 0*	*≥ 25*	*≥ 50*	*≥ 75*
TPR (True Positive Rate, Recall)	TP/(TP + FN)	98%	97%	97%	96%	24%	20%	16%	13%	36%	31%	28%	25%

(*continued on next page*)

Table A2 (*continued*)

Performance Measure	Equation	Motion Detection				Time Lapse				Overall			
		MegaDetector Detection Threshold				*MegaDetector Detection Threshold*				*MegaDetector Detection Threshold*			
		> 0	*≥ 25*	*≥ 50*	*≥ 75*	*> 0*	*≥ 25*	*≥ 50*	*≥ 75*	*> 0*	*≥ 25*	*≥ 50*	*≥ 75*
TNR (True Negative Rate, Selectivity)	TN/(FP + TN)	78%	85%	90%	95%	74%	84%	92%	95%	74%	84%	92%	95%
FPR (False Positive Rate, fall-out)	FP/(FP + TN)	22%	15%	10%	5%	26%	16%	8%	5%	26%	16%	8%	5%
FNR (False Negative Rate, miss rate)	FN/(FN + TP)	2%	3%	3%	4%	76%	80%	84%	87%	64%	69%	72%	75%
PPV (Positive Predictive Value, Precision)	TP/(TP + FP)	95%	97%	98%	99%	35%	44%	57%	67%	48%	59%	72%	81%
NPV (Negative Predictive Value)	TN/(TN + FN)	90%	86%	84%	82%	63%	62%	62%	61%	63%	63%	62%	61%
FDR (False Discovery Rate)	FP/(FP + TP)	5%	3%	2%	1%	65%	56%	43%	33%	52%	41%	28%	19%
FOR (False Omission Rate)	FN/(FN + TN)	10%	14%	16%	18%	37%	38%	38%	39%	37%	37%	38%	39%
ACC (Accuracy)	(TP + TN)/(TP + TN + FP + FN)	95%	95%	95%	96%	56%	59%	61%	62%	59%	62%	64%	64%
ERR (Error Rate)	(FP + FN)/(TP + TN + FP + FN)	5%	5%	5%	4%	47%	42%	40%	39%	44%	40%	37%	36%
F1 Score (Harmonic mean between TPR and PPV)	(2*TP)/((2*TP) + FP + FN))	97%	97%	97%	97%	29%	27%	25%	22%	41%	41%	40%	38%
MCC (Matthews Correlation Coefficient)	((TP*TN)-(FP*FN))/(SQRT((TP + FP)*(TP + FN)*(TN + FP)*(TN + FN)))	81%	82%	84%	86%	−2%	5%	12%	16%	10%	18%	26%	29%

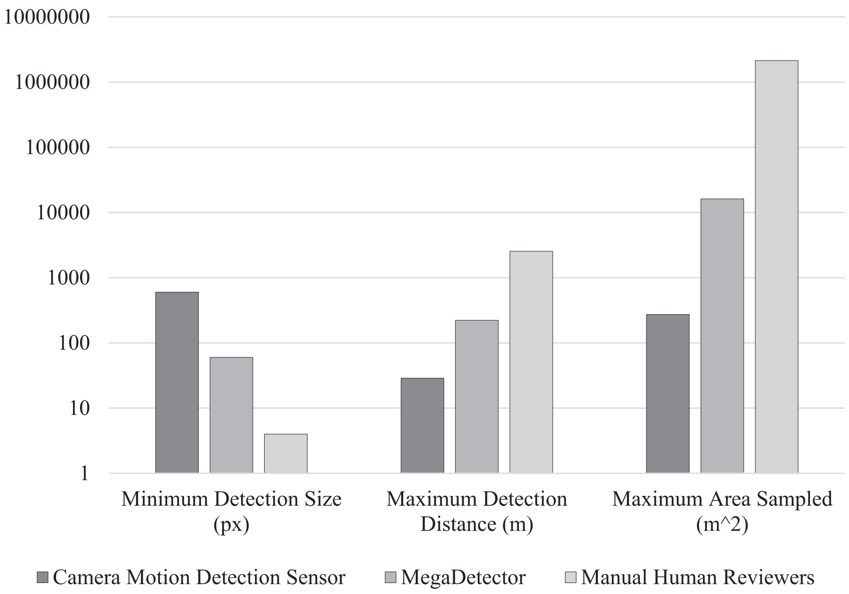


Fig. A1. Summary of detection limits and associated area sampled for caribou for the camera motion detection sensor, MegaDetector, and manual human reviewers. (Supplement to Figs. 3 & 4).

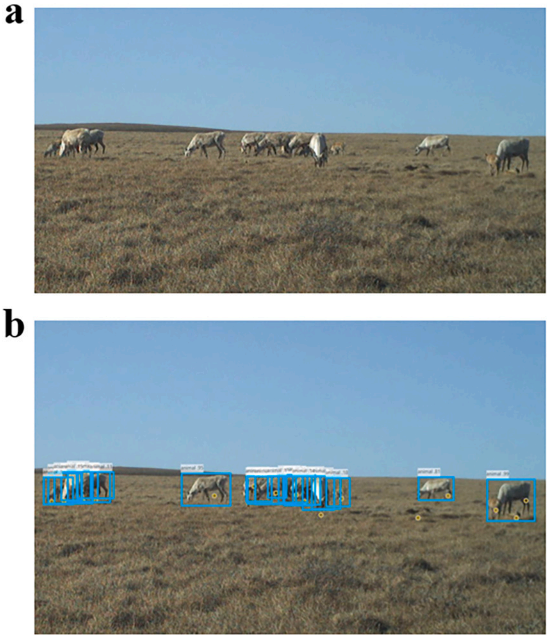


Fig. A2. Example of caribou group (a) cropped from original image and (b) labeled by MegaDetector (blue bounding boxes, $n = 21$) and manual human review (yellow dots, $n = 16$). *Note*: the exact number of individual caribou contained within the image is uncertain so the accuracy of MegaDetector's individual labels could not be determined. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Apps, P.J., Mcnutt, J.W., 2018. How camera traps work and how to work them. Afr. J. Ecol. 56, 702–709. https://doi.org/10.1111/aje.12563.

Beery, S., van Horn, G., Pietro, P., 2018. Recognition in terra incognita. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), European Conference on Computer Vision. Springer, Cham, pp. 472–489. https://beerys.github.io/CaltechCameraTraps/.

Beery, S., Morris, D., Yang, S., 2019. Efficient Pipeline for Camera Trap Image Review. https://arxiv.org/pdf/1907.06772.pdf.

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T., Bayne, E., Boutin, S., 2015. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. J. Appl. Ecol. 52 (3), 675–685. https://doi.org/10.1111/1365-2664.12432.

Caravaggi, A., Banks, P.B., Burton, A.C., Finlay, C.M.V., Haswell, P.M., Hayward, M.W., Rowcliffe, M.J., Wood, M.D., 2017. A review of camera trapping for conservation behaviour research. Remote Sens. Ecol. Conserv. 3 (3), 109–122. https://doi.org/10.1002/rse2.48.

Chen, M., Mao, S., Liu, Y., 2014. Big data: a survey. Mob. Netw. Appl. 19 (2), 171–209. https://doi.org/10.1007/s11036-013-0489-0.

Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. Methods Ecol. Evol. 10 (10), 1632–1644. https://doi.org/10.1111/2041-210X.13256.

Driessen, A., Michael, M., Peter, J., 2017. Animal detections vary among commonly used camera trap models. Wildl. Res. 44 (4), 291–297.

Farley, S.S., Dawson, A., Goring, S.J., Williams, J.W., 2018. Situating ecology as a big-data science: current advances, challenges, and solutions. BioScience 68 (8), 563–576. https://doi.org/10.1093/biosci/biy068.

Fennell, M., Beirne, C., Burton, A.C., 2022. Use of object detection in camera trap image identification: assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology. Glob. Ecol. Conserv. 35, e02104 https://doi.org/10.1016/j.gecco.2022.e02104.

Fleming, P.J.S., Tracey, J.P., 2008. Some human, aircraft and animal factors affecting aerial surveys: how to enumerate animals from the air. Wildl. Res. 35 (4), 258–267. https://doi.org/10.1071/WR07081.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2015. How “big data” can make big impact: findings from a systematic review and a longitudinal case study. Int. J. Prod. Econ. 165, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031.

Gilbert, N.A., Clare, J.D.J., Stenglein, J.L., Zuckerberg, B., 2020. Abundance estimation of unmarked animals based on camera-trap data. Conserv. Biol. 35 (1), 88–100. https://doi.org/10.1111/cobi.13517.

Glover-kapfer, P., Soto-navarro, C.A., Wearn, O.R., 2019. Camera-trapping version 3.0: current constraints and future priorities for development. Remote Sens. Ecol. Conserv. 5 (3), 209–223. https://doi.org/10.1002/rse2.106.

Greenberg, S., 2020a. Automated Image Recognition for Wildlife Camera Traps: Making it Work for You. http://grouplab.cpsc.ucalgary.ca/grouplab/uploads/Publications/Publications/2020-08-ImageRecognitionCameraTraps.pdf.

Greenberg, S., 2020b. Timelapse: An Image Analyser for Camera Traps. https://saul.cpsc.ucalgary.ca/timelapse/pmwiki.php?n=Main.HomePage.

Greenberg, S., 2022. Timelapse Image Recognition Guide. https://saul.cpsc.ucalgary.ca/timelapse/uploads/Guides/TimelapseImageRecognitionGuide.pdf.

Greenberg, S., Godin, T., Whittington, J., 2019. Design patterns for wildlife-related camera trap image analysis. Ecol. Evol. 9, 13706–13730. https://doi.org/10.1002/ece3.5767.

Hamel, S., Killengreen, S., Henden, J., Eide, N., Roed-eriksen, L., Ims, R., Yoccoz, N., 2013. Towards good practice guidance in using camera-traps in ecology: influence of sampling design on validity of ecological inferences. Methods Ecol. Evol. 4, 105–113. https://doi.org/10.1111/j.2041-210x.2012.00262.x.

Hariri, R.H., Fredericks, E.M., Bowers, K.M., 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. J. Big Data 6 (1). https://doi.org/10.1186/s40537-019-0206-3.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., Murphy, K., 2017. Speed/accuracy trade-offs for modern convolutional object detectors. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3296–3297. http://arxiv.org/abs/1611.10012.

Leorna, S., 2022. Arctic Alaska Camera Trap Project (Microsoft AI for Earth MegaDetector Evaluation Image Subset), Arctic coastal plain, Alaska, USA, May - September. 2019. Arctic Data Center. https://doi.org/10.18739/A2J38KJ9R.

Leorna, S., Brinkman, T., Fullman, T., 2022. Estimating animal size or distance in camera trap images: photogrammetry using the pinhole camera model. Methods Ecol. Evol. 00, 1–12. https://doi.org/10.1111/2041-210X.13880.

Meek, P.D., Ballard, G., Claridge, A., Kays, R., Moseby, K., Sanderson, J., Swann, D.E., Tobler, M., Townsend, S., 2014. Recommended guiding principles for reporting on camera trapping research. Biodivers. Conserv. 23, 2321–2343. https://doi.org/10.1007/s10531-014-0712-8.

Miao, Z., Gaynor, K.M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M.S., Mcinturff, A., Bowie, R.C.K., Nathan, R., Yu, S.X., Getz, W.M., 2019. Insights and approaches using deep learning to classify wildlife. Sci. Rep. 1–9. https://doi.org/10.1038/s41598-019-44565-w. November 2018.

Microsoft AI for Earth, 2020. MegaDetector. https://github.com/microsoft/CameraTraps/blob/master/megadetector.md#megadetector-overview.

Moeller, A.K., Lukacs, P.M., Horne, J.S., 2018. Three novel methods to estimate abundance of unmarked animals using remote cameras. Ecosphere 9 (8). https://doi.org/10.1002/ecs2.2331.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. U. S. A. 115 (25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115.

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J., 2021. A deep active learning system for species identification and counting in camera trap images. Methods Ecol. Evol. 12, 150–161. http://arxiv.org/abs/1910.09716.

Philip Chen, C.L., Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Inf. Sci. 275, 314–347. https://doi.org/10.1016/j.ins.2014.01.015.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. https://arxiv.org/pdf/1506.01497.pdf.

Rovero, F., Zimmermann, F., Berzi, D., Meek, P., 2013. "Which camera trap type and how many do I need?" A review of camera features and study designs for a range of wildlife research applications. Hystrix 24 (2), 148–156. https://doi.org/10.4404/hystrix-24.2-6316.

Rowcliffe, J.M., Carbone, C., 2008. Surveys using camera traps: are we looking to a brighter future? Anim. Conserv. 11 (3), 185–186. https://doi.org/10.1111/j.1469-1795.2008.00180.x.

Rowcliffe, J.M., Field, J., Turvey, S.T., Carbone, C., 2008. Estimating animal density using camera traps without the need for individual recognition. J. Appl. Ecol. 45 (4), 1228–1236. https://doi.org/10.1111/j.1365-2664.2008.01473.x.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089.

Schneider, S., Taylor, G.W., Kremer, S., 2018. Deep learning object detection methods for ecological camera trap data. In: Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018, pp. 321–328. https://doi.org/10.1109/CRV.2018.00052.

Schneider, S., Greenberg, S., Taylor, G.W., Kremer, S.C., 2020. Three critical factors affecting automated image species recognition performance for camera traps. Ecol. Evol. 10 (7), 3503–3517. https://doi.org/10.1002/ece3.6147.

Scotson, L., Johnston, L.R., Iannarilli, F., Wearn, O.R., Mohd-azlan, J., Wong, W.M., Gray, T.N.E., Dinata, Y., Suzuki, A., Willard, C.E., 2017. Best practices and software for the management and sharing of camera trap data for small and large scales studies. Remote Sens. Ecol. Conserv. 3 (3), 158–172. https://doi.org/10.1002/rse2.54.

Sollmann, R., 2018. A gentle introduction to camera-trap data analysis. Afr. J. Ecol. 56 (4), 740–749. https://doi.org/10.1111/aje.12557.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci. Data 2. https://doi.org/10.1038/sdata.2015.26.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., di Salvo, P.A., Lewis, J.S., White, M.D., Teton, B., Beasley, J.C., Schlichting, P.E., Boughton, R.K., Wight, B., Newkirk, E.S., Ivan, J.S., Odell, E.A., Brook, R.K., Miller, R.S., 2019. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol. Evol. 10 (4), 585–590. https://doi.org/10.1111/2041-210X.13120.

Thomson, R., Potgieter, G., Bahaa-el-din, L., 2018. Closing the gap between camera trap software development and the user community. Afr. J. Ecol. 56, 721–739. https://doi.org/10.1111/aje.12550.

Trolliet, F., Huynen, M., Vermeulen, C., Hambuckers, A., 2014. Use of camera traps for wildlife studies. A review. Biotechnol. Agron. Soc. Environ. 18 (3), 446–454. https://doi.org/10.1016/0308-0161(78)90006-6.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I.D., van Horn, G., Crofoot, M.C., Stewart, C.V., Berger-Wolf, T., 2022. Perspectives in machine learning for wildlife conservation. In: Nature Communications, vol. 13, Issue 1. https://doi.org/10.1038/s41467-022-27980-y. Nature Research.

Vecvanags, A., Aktas, K., Pavlovs, I., Avots, E., Filipovs, J., Brauns, A., Done, G., Jakovels, D., Anbarjafari, G., 2022. Ungulate detection and species classification from camera trap images using RetinaNet and faster R-CNN. Entropy 24. https://doi.org/10.3390/e24030353.

Vélez, J., Castiblanco-Camacho, P.J., Tabak, M.A., Chalmers, C., Fergus, P., Fieberg, J., 2022. Choosing an Appropriate Platform and Workflow for Processing Camera Trap Data using Artificial Intelligence. http://arxiv.org/abs/2202.02283.

Wearn, O.R., Glover-kapfer, P., 2017. Camera-Trapping for Conservation: A Guide to Best-Practices. https://www.wwf.org.uk/sites/default/files/2019-04/CameraTraps-WWF-guidelines.pdf.

Young, S., Rode-Margono, J., Amin, R., 2018. Software to facilitate and streamline camera trap data management: a review. Ecol. Evol. 8 (19), 9947–9957. https://doi.org/10.1002/ece3.4464.