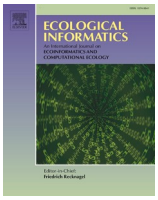
Ecological Informatics 77 (2023) 102257

Contents lists available at ScienceDirect

Ecological Informatics

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



Evaluating a tandem human-machine approach to labelling of wildlife in remote camera monitoring

Laurence A. Clarfeld [a,*], Alexej P.K. Sirén [a], Brendan M. Mulhall [b], Tammy L. Wilson [c], Elena Bernier [d], John Farrell [d], Gus Lunde [d], Nicole Hardy [d], Katherina D. Gieder [e], Robert Abrams [f], Sue Staats [g], Scott McLellan [h], Therese M. Donovan [i]

[a] Vermont Cooperative Fish and Wildlife Research Unit, University of Vermont, Burlington, VT, USA
[b] Texas State University, San Marcos, TX, USA
[c] U.S. Geological Survey, Massachusetts Cooperative Fish and Wildlife Research Unit, Department of Environmental Conservation, University of Massachusetts, Amherst, MA, USA
[d] University of Vermont, Burlington, VT, USA
[e] Vermont Department of Fish and Wildlife, Rutland, VT, USA
[f] U.S. Forest Service, Green Mountain National Forest, Manchester Center, VT, USA
[g] U.S. Forest Service, Green Mountain National Forest, Rochester, VT, USA
[h] Maine Dept of Inland Fisheries and Wildlife, Greenville, ME, USA
[i] U.S. Geological Survey, Vermont Cooperative Fish and Wildlife Research Unit, Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Camera trap
Data labeling
Machine learning
Trail camera
Wildlife monitoring
Bounding box

ABSTRACT

Remote cameras ("trail cameras") are a popular tool for non-invasive, continuous wildlife monitoring, and as they become more prevalent in wildlife research, machine learning (ML) is increasingly used to automate or accelerate the labor-intensive process of labelling (i.e., tagging) photos. Human-machine hybrid tagging approaches have been shown to greatly increase tagging efficiency (i.e., time to tag a single image). However, those potential increases hinge on the extent to which an ML model makes correct vs. incorrect predictions. We performed an experiment using a ML model that produces bounding boxes around animals, people, and vehicles in remote camera imagery (MegaDetector) to consider the impact of a ML model's performance on its ability to accelerate human labeling. Six participants tagged trail camera images collected from 12 sites in Vermont and Maine, USA (January–September 2022) using three tagging methods (one with ML bounding box assistance and two without assistance). We used a generalized linear mixed model to examine the influence of ML model performance and tagging method on tagging efficiency. We found that ML bounding boxes offer significant improvement in tagging efficiency when labelling data compared to unassisted tagging. Additionally, the time taken to label with bounding boxes was not statistically different from an unassisted tagging approach. However, we found that gains in efficiency are contingent on the ML algorithm's performance and that incorrect ML predictions, particularly the 4.2% false positive and 3.6% false negative predictions, can slow the tagging process compared to a non-hybrid approach. These findings indicate that although practitioners usually forgo the production of bounding boxes when selecting a data labelling process due to the increased effort, ML bounding box-assisted tagging can offer an efficient method for labeling. More broadly, ML-assisted data labelling offers an opportunity to accelerate the analysis of trail camera imagery, but an assessment of the ML model's performance can illuminate whether the hybrid-tagging approach is ultimately a help or hinderance.

1. Introduction

Remote cameras (aka trail cameras or camera traps) are a primary tool for studying many wildlife species (Burton et al., 2015; O'Connell et al., 2011) and have been used to observe animal behavior, monitor rare and endangered species, surveille invasive species, and build

* Corresponding author at: UVM George D. Aiken Forestry Sciences Laboratory, 705 Spear St, South Burlington, VT 05403, USA.
E-mail address: Laurence.Clarfeld@uvm.edu (L.A. Clarfeld).

https://doi.org/10.1016/j.ecoinf.2023.102257
Received 15 June 2023; Received in revised form 7 August 2023; Accepted 9 August 2023
Available online 10 August 2023

occupancy models to understand the distribution of populations across the landscape through time (Cove et al., 2021; Gilbert et al., 2020; Kays et al., 2020; Steenweg et al., 2017). Monitoring via remote cameras has grown in popularity due to a confluence of favorable circumstances. First, camera quality and reductions in cost have increased the effectiveness of monitoring with remote cameras compared to other methods (Wearn and Glover-Kapfer, 2019). Second, there have been rapid improvements in the technology for reading, writing, and storing the large volumes of digital data that are generated from remote monitoring (Mack, 2011). Third, some newer remote camera models allow for photos to be received wirelessly via cellular networks, reducing the number of costly site visits required to maintain cameras after deployment (Herrera et al., 2022). Consequently, the number of publications referencing "remote cameras" or "camera traps" have more than doubled in the past decade on Web of Science (accessed February 2023).

However, there are significant challenges to overcome in managing and interpreting the terabytes of images that can be produced by camera trapping. False triggers may occur when a camera's infrared sensor is triggered unintentionally by vegetation, sunlight, or patchy shade (Glover-Kapfer et al., 2019). The process of sorting through "empty" images can be time-consuming, potentially requiring hundreds of hours for large scale projects, and can take away from the already limited time needed to analyze noteworthy photos. After false triggers are removed, labelling images is a monumental task and manual labelling can quickly become a bottleneck that impedes timely dissemination of results. Data labelling typically involves a person reviewing images from the camera card in the order in which they were taken and annotating the species present in each image. Attributes such as the number of individuals present of each species, age class, sex, behavior, etc. may also be documented. Without labelled data, there is no way to translate monitoring images into meaningful information about wildlife.

Modern machine-learning (ML) methods can perform a variety of tasks to help meet these challenges. Determining whether an animal is present in an image is a binary classification task that can be performed via machine learning to filter out false detections from a dataset. Several methods have been used to predict whether an image is empty, including measurement of differences between consecutive images (Price Tack et al., 2016; Ren et al., 2013; Wei et al., 2020), convolutional neural networks (Tabak et al., 2019; Tabak et al., 2020), and ensemble learning (Yang et al., 2021). Alternatively, object detection models that localize target objects (animals) within an image can be used to filter empty images (Beery et al., 2019). In contrast to other techniques, this approach not only identifies empty images through the absence of target objects, but also indicates the location of each target object within an image (Fig. 1).

In addition to their utility in removing false detections, bounding boxes around target objects serve multiple functions in wildlife research. They can be used for counting animals in images (Torney et al., 2019) and aiding in distance estimation of animals from the camera (Hofmeester et al., 2017). These metrics allow researchers to go beyond simple presence/absence studies and begin to estimate abundance (Haucke et al., 2022; Johanns et al., 2022; Rowcliffe et al., 2008). Bounding boxes can also be used to infer animal movement, which can have important implications for interpreting behavior (Lopez-Marcano et al., 2021).

By localizing target animals, object detection can allow downstream models to ignore the image background, reducing noise and potential bias (Beery et al., 2019; Norman et al., 2023). This is a common preprocessing step for ML models that attempt to identify individuals, which require an image to be cropped to a region of interest (Buehler et al., 2019; Crall et al., 2013; De Lorm et al., 2023). Image crops are also frequently used in species classification models (Li et al., 2022).

Machine learning models that produce bounding boxes can also be used in tandem with human tagging efforts. For example, the bounding boxes may help the human "tagger" improve the efficiency of data labeling. However, when used in tandem with human tagging efforts, the ML bounding boxes may help or hinder the efficiency of data labeling. Improved tagging efficiency with the assistance of ML bounding boxes may depend on whether the bounding box correctly identifies a target animal within an image (true positive) or not (false positive). In the case of true positives (Fig. 1 upper-left), previous assessments of other ML models have shown that ML can reduce the number of human observers

Fig. 1. Example model outputs from a machine learning model showing the bounding box around the target object (in red) for true positive (upper-left), false positive (upper-right), false negative (lower-left), and true negative (lower-right) predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

required to reach a consensus identification (Willi et al., 2019). However, the effect of false bounding boxes on human tagging efficiency has not been explored (Fig. 1 upper-right). In contrast to true and false positives, the lack of a ML bounding box results in a true negative classification when the ML model is correct (Fig. 1 lower-right) or a false negative classification the ML model fails to create a bounding box around a target object (Fig. 1 lower-left). Together, these four classification outcomes (true/false positives and negatives) are a proxy for the ML model's performance.

Without considering how classification outcome can increase efficiency for tandem human-machine tagging, practitioners of ML methods are left to guess whether a ML model is "good enough" to be of practical use. The need for understanding the interplay between ML system accuracy and realized time savings in hybrid computer-human labeling systems has been identified as an important direction for future work in the use of ML in camera trapping studies (Norouzzadeh et al., 2018). Ultimately, these types of evaluations help researchers use ML to shorten the gap between when data are collected and when the fully labelled dataset is available to inform management and conservation objectives.

We explored the effects of classification outcome on the ability of ML bounding boxes to speed up the labelling of remote camera images with bounding boxes and species-level identifications. This exploration accounted for the number of animals in the image and whether the image was part of a temporal sequence of images or not. The objectives of our study were to: (1) Perform an experiment to directly compare tagging efficiency of a human-ML hybrid tagging approach using a ML bounding box model (MegaDetector) vs. human tagging unassisted by ML; (2) Assess the ML model's performance via the observed classification outcomes from our experiment; and (3) Evaluate the effects of ML model performance on tagging efficiency, including an assessment of the overall utility of the ML bounding box model for accelerating the data labelling process.

2. Materials and methods

2.1. Objective 1: perform tagging efficiency experiment

The remote camera imagery was collected from 12 locations: seven in northern Maine (Maine Department of Inland Fisheries and Wildlife) and five from the Green Mountain National Forest in Vermont, USA between January and September 2022 (Fig. 2). The sites selected varied in the degree of canopy closure, from open sites with herbaceous vegetation to heavily forested sites with abundant coarse woody debris. All sites used a standardized monitoring protocol developed by Sirén et al. (2018) that included the placement of a turkey feather attached to a wooden stake in the camera's field-of-view to act as an attractant. All sites used cameras (Browning Recon Force Elite HP4) that were programmed to capture images when triggered by wildlife. Cameras were placed 1-2 m off the ground and programmed to record 1 image (no multi-shot/burst mode), with a 1 s resting time between triggers. Motion sensitivity was set to "normal" (60 ft), the triggering speed was set to "fast" (0.1 s), and the Infrared LED flash was set to "economy".

A subset of 100 consecutively captured photos were selected from each site (average range in days = 47), ensuring a balance between the number of images with animals vs. the number without (e.g., photo sets were checked to ensure they weren't all empty or all taken of a single,

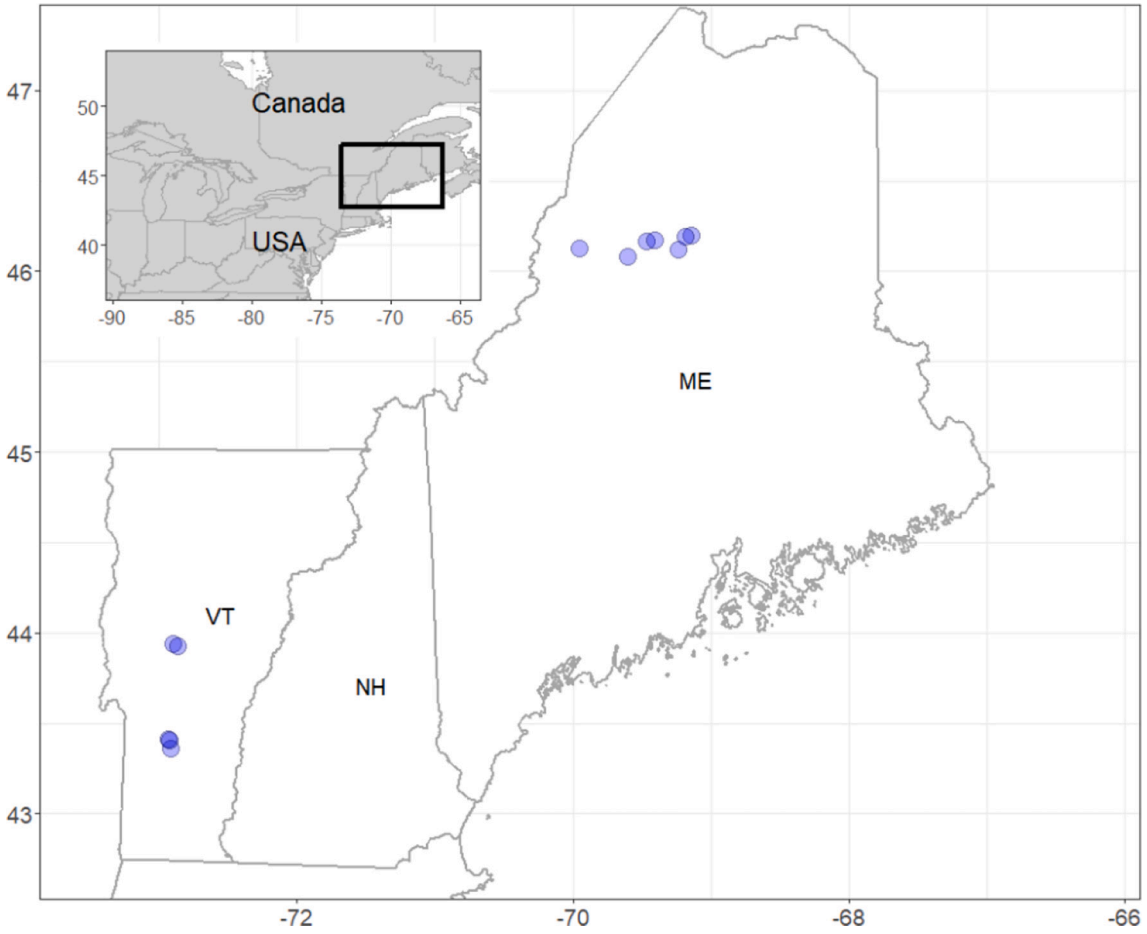


Fig. 2. A map showing twelve study sites, seven of which were in northern Maine (ME), two in the northern Green Mountain National Forest in Vermont (VT), and three in the southern Green Mountain National Forest (also in VT). (NH = New Hampshire).

stationary animal). The resultant data consisted of images that contained a variety of animal species and false triggers (empty images), with an approximate 50:50 split across all locations.

Images were manually labelled by six "taggers" with and without the aid of ML bounding boxes, as described next. We used the ML model, MegaDetector, a free object detection model created by Microsoft, to place bounding boxes in camera trap data (Beery et al., 2019). MegaDetector is an object detection model developed for identifying people, animals, and vehicles from remote camera imagery (Beery et al., 2019). The model outputs include a bounding box around each target object detected, the predicted classification, and a confidence score for the prediction (Fig. 1). Trained on several hundred thousand images from a wide range of locations and contexts, it performs well and consequently has been adopted into multiple analytical tools (Ahumada et al., 2020; Cove et al., 2021; Hendry and Mann, 2017) and workflows by dozens of conservation organizations around the world (Beery et al., 2019). Because MegaDetector does not provide species-level classifications, it is frequently used to remove false triggers by placing bounding boxes around potential targets; if no targets are found, the image is likely a false trigger. Removal of blank images via MegaDetector can increase tagging efficiency 8 times that of a fully manual workflow (Fennell et al., 2022).

Tagging was performed using a customized version of the image tagging application from the R package AMMonitor, which provides a streamlined interface for tagging remote camera imagery (Balantic and Donovan, 2020). The tagging application allowed users to put a bounding box around each animal (or human) and indicate its taxon. The images were tagged by 6 coauthors on this paper (BM, EB, GL, NH, JF, APKS) familiar with wildlife species identification. The level of tagging experience varied between taggers, but all had been trained in use of the AMMonitor image tagging application on separate images prior to this study.

The tagging experiment included three treatment groups representing different tagging protocols; each image was subjected to all three treatments by different taggers:

(1) Bounding Boxes, Assisted: In this treatment, taggers identified animals/humans from monitoring photos, adding bounding boxes around each individual, with the ML model's predicted bounding boxes pre-populated into the tagging application. For this treatment, we ran the ML model (MegaDetector model version MDv5a) on all images prior to the experiment on a computer (Dell Latitude 7420 with an Intel i7-1165G7 CPU at a speed of 2.80 GHz with 16 GB of RAM). Any ML model detections with a confidence score of 0.1 or higher were included in this treatment group.
(2) Bounding Boxes, Unassisted: Like the previous treatment, taggers identified animals/humans from monitoring photos, adding bounding boxes around each individual, but without the assistance of ML bounding boxes.
(3) No Bounding Boxes, Unassisted: In this treatment, taggers identified animals/humans from monitoring photos, but did not add bounding boxes around each individual (and without ML bounding box assistance). This treatment represents the standard approach used by studies to tag camera data.

Taggers were randomly assigned to one of the three treatment groups for each location so that all images from a given location received each treatment twice. Taggers were instructed to open the tagging application, select a location, and label all images from each location in a single, uninterrupted sitting. The image tagging application's state varied according to treatment. For treatment 1 (Bounding Boxes, Assisted) only, the tagging application prepopulated the ML bounding boxes that the tagger could either use or delete/replace by drawing a new bounding box. In cases where a human or vehicle was detected, the "Human" species label was also applied to the image. Otherwise, taggers could add a species label. Images with no animals present were tagged as "empty" without the need to apply a bounding box. For treatments 2 (Bounding Boxes, Unassisted) and 3 (No Bounding Boxes, Unassisted), the image tagging application presented the image without ML bounding boxes.

The tagging application was configured to store the exact time at which the user toggled between images, allowing the duration spent looking at each image to be directly calculated. Upon completion of tagged images from a given location, taggers were instructed to save all results to a csv file and pause briefly before continuing to the next location.

2.2. Objective 2: evaluate the ML bounding box model's performance

To evaluate the performance of the ML bounding box model, we considered classification outcome (true/false positives and negatives) as our key predictor variable. Together, these values form a confusion matrix, which is a standard tool in evaluating classification models. The classification outcome for each image was determined relative to the labels provided by each tagger in treatment 1 (the only treatment in which the ML model's bounding boxes were displayed). The treatment 1 labels were assumed to be the "ground-truth" because a tagger who believes an animal is absent (or present) in an image will behave the same way whether or not an animal is truly absent; thus the time per tagging event is dependent on the tagger's actions rather than the true classification of the image.

2.3. Objective 3: evaluate the effects of ML bounding boxes on tagging efficiency

We fit a generalized linear mixed model (GLMM), including fixed and random effects, to evaluate the impacts of ML bounding boxes on tagging efficiency. Tagger efficiency was the outcome variable in our model and is represented by the total time required to tag a single image under a specific treatment. As these times are left-bounded by zero, a Gamma family and log link function were used. The model was (Eq. 1).

$$\text{t\_diff} \sim \text{treatment}^* \ \text{class\_outcome} + \text{num\_animals} + \text{in\_seq} + (1 \mid \text{site}^* \ \text{tagger}) \quad (1)$$

Fixed effects included the interaction between the ML model performance and tagging treatment. The ML model performance was measured by the classification outcomes (class_outcome) described in the introduction, namely, whether the image was a true positive bounding box, a true negative with no bounding box, a false positive bounding box, and false negative (lack of bounding box around an identified animal or person). The tagging treatment included (1) Bounding Boxes, Assisted; (2) Bounding Boxes, Unassisted; and (3) No Bounding Boxes, Unassisted. Because the effect of treatment on tagging efficiency may depend on the class outcome, the model specification included an interaction effect between classification outcome and treatment group.

Additional fixed effects included (a) the number of animals in an image (e.g., creating bounding boxes for 2 animals is expected to take longer than for 1 animal) and (b) whether an image was "in sequence" or not (0/1). We defined an image to be "in sequence" if it had the same species (or absence of species) and had visually similar background conditions to the image that immediately preceded it. We defined consecutive images to have visually similar background when they have the same environmental conditions. For example, a moose (*Alces alces*) that lingers in the camera's field-of-view for many consecutive frames or a long series of empty frames with the same lighting would be considered in-sequence. Examples of images that are not "in sequence" are when one frame has an animal and the next does not, or when one frame is in daytime conditions and the next was taken after dark. While this definition is somewhat subjective, in practice the determination was typically obvious. Whether an image is in sequence is important because

these images can be tagged more quickly based on the context of the prior image and occur frequently in real-world datasets.

Random effects were included to account for dependencies among the twelve sites and six taggers. Sites varied from each other in a variety of ways that could increase or decrease tagging efficiency. For example, sites with an abundance of coarse woody debris present more challenging conditions for spotting animals compared to open sites. The species composition between sites also varied, with some having more challenging species to spot or identify than others. Likewise, taggers varied in their level of tagging expertise, each having their own strengths and weaknesses. Given the potential interplay between tagger ability and site characteristics, a cross-factor random effect between tagger and site was used in the final model (Eq. 1).

We used the R package glmmTMB (Magnusson et al., 2017) to fit the model, and the R package DHARMa (Hartig, 2022) to evaluate the fit of the GLMM. The DHARMa package uses simulation-based dispersion and outlier tests to assess goodness-of-fit. Beta coefficients and standard errors from the fitted GLMM were then used to create confidence intervals for each permutation of classification outcome and treatment, as discussed in the following section.

3. Results

3.1. Objective 1: perform tagging efficiency experiment

Of the 7200 instances of taggers labelling an image (6 taggers × 1200 images), 3.6% were excluded from analyses for a variety of reasons: (1) The first and last ($N = 24$) images tagged by a single tagger in a single sitting were excluded to avoid edge effects associated with starting and stopping the experiment; (2) Tagging events that required over 30 s ($N = 104$) were also dropped, as these were considered longer than expected without some extraneous explanatory factors; (3) Some tagging events were omitted due to human error ($N = 95$) (e.g., forgetting to include a bounding box but identifying present species for treatment 1); and, (4) Cases where the ML model produced a bounding box in the wrong location in images where animals were present ($N = 48$) could be considered to have classification outcome of both a false positive and a false negative and were excluded since this case was not well-enough represented in our dataset to draw any meaningful conclusions. The final dataset consisted of 6936 tagging events under the 3 treatments.

The images in the final dataset spanned 564 trap days between January 7 and September 14, 2022. About half of the 1200 images collected did not contain any animals. Those with animals represented 20 different species of mammals and birds. The majority (72%) of images were "in sequence" and, of the images with animals, 96.0% contained one individual, 3.9% contained two individuals, and 0.1% contained three.

3.2. Objective 2: evaluate the ML bounding box model performance

Overall, the ML bounding box model was 92.2% accurate on the given data, with 45.2% true positive classifications, 47.0% true negative, 4.2% false positive and 3.6% false negative. These outcomes are shown in the confusion matrix (Fig. 3). Several additional performance metrics can be derived from the observed classification outcomes, including precision (the number of true positives relative to true and false positives) and recall (the number of true positives relative to true positives and false negatives). The performance was well-balanced between precision and recall, with 92.6% and 91.5%, respectively. The classes in the dataset were also well-balanced, with 49.6% of images containing at least one animal and 51.4% with no animals (empty).

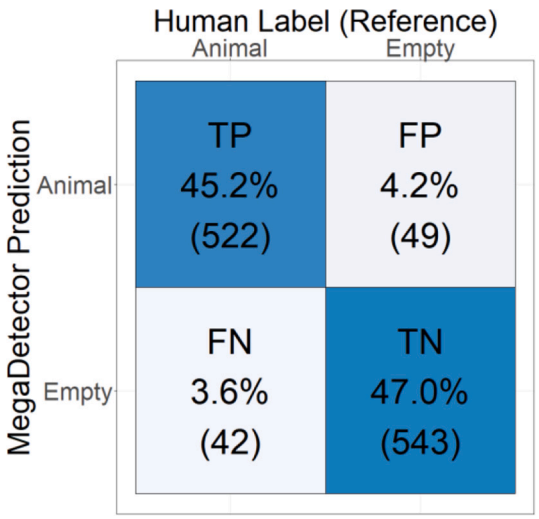


Fig. 3. Confusion matrix showing the percentage of true positive (upper-left quadrant), false positive (upper-right quadrant), false negative (lower-left quadrant), and true negative (lower-right quadrant) classification outcomes observed during the experiment.

3.3. Objective 3: evaluate the effects of ML model performance on tagging efficiency

The GLMM model was fitted on 6936 observations using the glmmTMB R package. The standard deviations of the random effects in our model were 0.13 for site ($N = 12$), 0.21 for tagger ($N = 6$), and 0.34 for site:tagger ($N = 72$). Goodness-of-fit was measured using several tests from the DHARMa R package. The outlier test considers the likelihood of an observation being outside the simulation envelope and indicated that outliers occur at expected frequencies ($p = 0.499$). The dispersion test performs a simulation-based assessment for under- and over-dispersion and found no significant deviations from expected values ($p = 0.152$).

The GLMM model estimated a significant difference in the time required to tag an image based on all combinations of the key covariates, treatment group and classification outcome, for all combinations except treatment 3 (no bounding boxes, unassisted) with the class outcome "false positive" (Table 1). Treatments 2 (bounding boxes, unassisted) and 3 (no bounding boxes, unassisted) both had a positive coefficient when the class_outcome was true positive and true negative, indicating an increased amount of time required to tag an image. The same treatments had a negative coefficient when class_outcome was false negative or false positive, indicating a decreased amount of time required to tag an image. The opposite pattern was observed for treatment 1 (bounding boxes, assisted), which had a positive coefficient for class_outcome of false positive and false negative, and a negative coefficient for class_outcome of true positive and true negative. The variable "in_seq" (in sequence) had a negative coefficient, indicating "in sequence" images took less time to tag. The variable "num_animals" (number of animals) had a positive coefficient, indicating tagging time increased for images with more than 1 animal (Table 1).

We used the beta coefficients from the fitted GLMM model (Table 1) to generate predicted values for the time required to tag each image by tagger treatment type and classification outcome while trying to control for all other factors (Fig. 4). In generating predicted values, we set all random effects to zero and assumed predictions were not in-sequence in all cases. The number of animals in each image were assumed to be zero for true negative and false positive classification outcomes and one for true positive and false negative classification outcomes. A 95% confidence interval for these predictions was derived from the beta coefficient standard error, allowing us to consider how each combination of tagging method × classification outcome affects the time required to tag

Table 1
Beta coefficients and confidence intervals for GLMM model parameters. The model fixed effects include: Treatments 1 (bounding boxes, assisted), 2 (bounding boxes, unassisted), and 3 (no bounding boxes, unassisted); classification outcomes (class_outcome) of true positive (tp), true negative (tn), false positive (fp) and false negative (fn); the number of animals in the image (num_animals); and, whether an image is "in-sequence" (in_seq). The model intercept represents treatment 1 and classification outcome of false negative (fn). For each fixed effect, the beta estimates, standard error, z-value, probability, and 95% confidence intervals are shown.

Parameter(s)	Estimate	Std. Error	z value	Pr(>|z|)	2.50%	97.50%
(Intercept)	2.60	0.14	18.74	2.34E-78	2.33	2.87
treatment: 2	−0.25	0.12	−2.04	4.18E-02	−0.50	−0.01
treatment: 3	−1.12	0.13	−8.94	4.07E-19	−1.36	−0.87
class_outcome: fp	0.33	0.11	2.88	3.97E-03	0.10	0.55
class_outcome: tn	−1.06	0.09	−11.30	1.33E-29	−1.24	−0.88
class_outcome: tp	−1.21	0.08	−15.37	2.63E-53	−1.36	−1.05
num_animals	0.62	0.05	12.61	1.90E-36	0.53	0.72
in_seq	−0.85	0.02	−45.85	0.00E+00	−0.89	−0.82
treatment/class_outcome: 2/fp	−0.95	0.14	−6.65	2.94E-11	−1.22	−0.67
treatment/class_outcome: 3/fp	−0.28	0.14	−1.94	5.19E-02	−0.55	2.24E-03
treatment/class_outcome: 2/tn	0.30	0.11	2.74	6.05E-03	0.09	0.52
treatment/class_outcome: 3/tn	1.13	0.11	10.19	2.22E-24	0.91	1.34
treatment/class_outcome: 2/tp	1.15	0.11	10.63	2.23E-26	0.94	1.36
treatment/class_outcome: 3/tp	0.79	0.11	7.29	3.08E-13	0.58	1.00

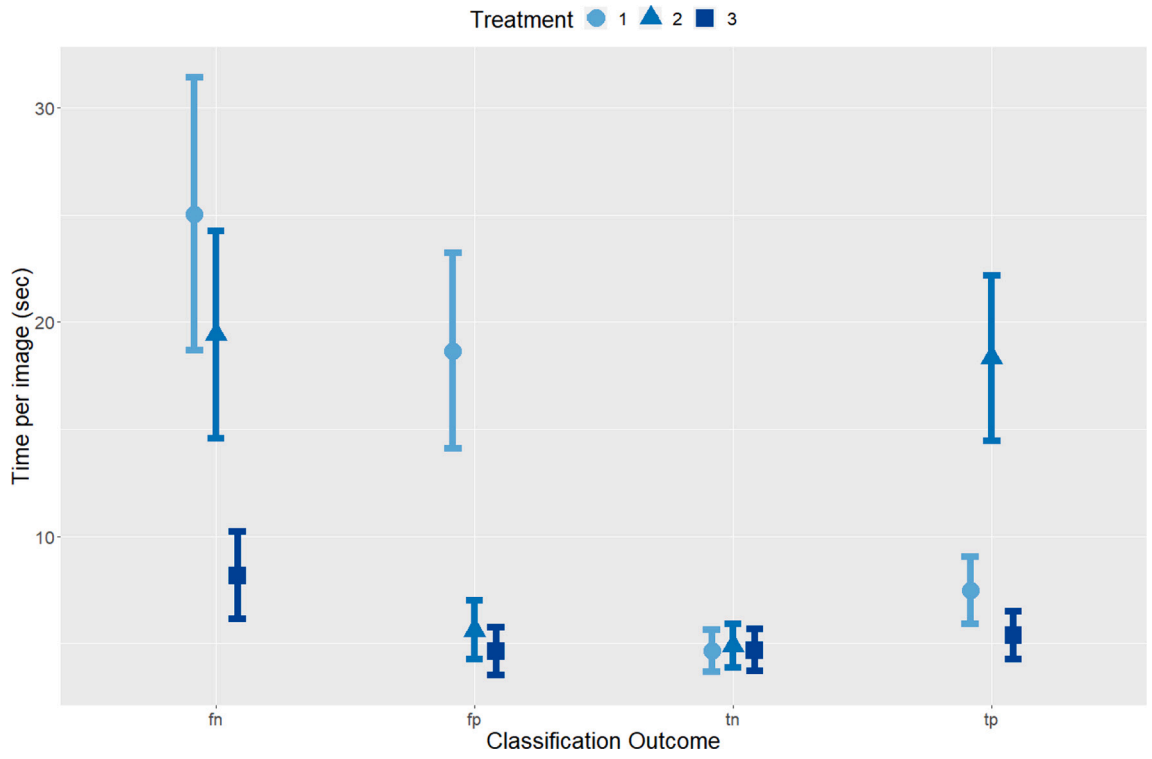


Fig. 4. Model estimates, with 95% confidence intervals, of tagging time per image based on treatments 1 (bounding boxes, assisted), 2 (bounding boxes, unassisted), and 3 (no bounding boxes, unassisted) and classification outcomes true positive (tp), true negative (tn), false positive (fp), and false negative (fn). Random effects were set to zero (population-level estimates) and the fixed effect "in sequence" was assumed to be 0 for true negatives and false positives, and the number of animals per image (integer) was assumed to be 0 for "empty" images (true negatives and false positives) and 1 otherwise.

an image.

When the ML model classification outcome contained a false negative prediction, treatments 1 (bounding boxes, assisted) and 2 (bounding boxes, unassisted) required significantly more time to tag an image than treatment 3 (no bounding boxes, unassisted). False positive model outcomes for treatment 1 resulted in significantly longer tagging times than treatments 2 and 3. For true negative classification outcome, all treatment groups performed similarly, and for true positives treatment 2 required more tagging time than treatments 1 and 3. A direct comparison, including expected tagging time per image and confidence intervals for each permutation of treatment group and classification outcome, is shown in Fig. 4.

The overall expected tagging time per image was calculated as a weighted average for each treatment group, given the observed relative frequencies of each classification outcome (Table 1). The point estimates for mean tagging time per image for treatments 1, 2, and 3 were 7.3, 11.5, and 5.1 s, respectively. Confidence intervals show treatment 2 (bounding boxes, unassisted) to be significantly slower than the other two treatments (Fig. 5). There was no statistical evidence that the time required per image for treatment 3 (no bounding boxes, unassisted) differed from treatment 1 (bounding boxes, assisted).

4. Discussion

We investigated how a hybrid human-ML tagging approach, using bounding boxes produced by the ML model (MegaDetector), influenced

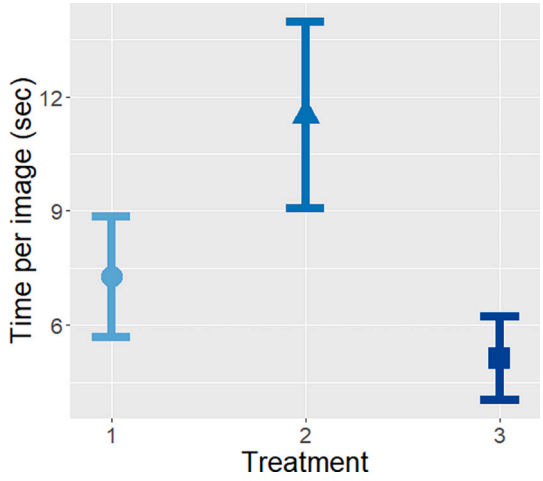


Fig. 5. Model estimates, with 95% confidence intervals, of tagging time per image based on treatments 1 (bounding boxes, assisted), 2 (bounding boxes, unassisted), and 3 (no bounding boxes, unassisted), under the assumption that classification outcomes occur at the frequencies observed in the experiment, of 45.2% true positive, 47.0% true negative, 4.2% false positive, and 3.6% false negative.

the efficiency of tagging remote camera data. Overall, we found that tagging images with ML bounding boxes is nearly twice as fast as tagging with bounding boxes, unassisted. Tagging images with only the species label (without bounding boxes) was the fastest tagging method, although not significantly different relative to tagging with ML bounding boxes. Our results highlight the value of combining ML-based approaches with manual tagging to increase the speed at which camera data is processed, which can ultimately allow for faster analyses.

There are multiple benefits to using bounding boxes to enhance human tagging. First, this hybrid ML-human tagging approach greatly reduced the burden of procuring bounding boxes. Because manually adding bounding boxes is labor intensive, this task is often forgone in favor of simply indicating the species present without localizing them within the image (Norouzzadeh et al., 2021). Further, our results suggest that the hybrid ML-human approach is not significantly slower than traditional species-only labelling in efficiency.

The use of bounding boxes in wildlife research is still somewhat limited, but has been growing over the past two decades, especially when applied towards species classification (Neupane et al., 2022). Training classification models based on crops of animals from an image have been shown to be effective at allowing models to focus on the target object and not overfit on the background (Beery et al., 2019; Norman et al., 2023). In some past studies, bounding boxes for training models could be procured through manual annotation (Yu et al., 2013). As automated bounding box generation was introduced to the species classification pipeline, the quality of those boxes was recognized as a factor that could impact model training and performance (Chen et al., 2014). More recently, the ML bounding box model MegaDetector has been incorporated into several species-classification models (Bothmann et al., 2023; Cunha et al., 2023). Validating bounding boxes through manual annotation is needed to ensure the cropped images used to train classification models are accurate, and so understanding the trade-offs in effort for labelling bounding boxes with a variety of methods is increasingly important.

Additionally, if we consider the use of ML models as a pre-filtering method for removing blank images, the effective time required for tagging false negatives becomes zero, and ML-assisted tagging with bounding boxes could be even faster than tagging without bounding boxes. However, the increased efficiency from using a ML model to pre-filter empty images would come at a cost (e.g., in our application of MegaDetector, throwing away the 3.6% of false negatives predicted to be empty image that contained an animal or object).

Despite these benefits, the favorable outcomes of ML bounding box assisted tagging assume that our observed model performance (Fig. 3) is representative and may not be applicable in all circumstances. In our study, precision and recall rates were 92.6% and 91.5%, respectively. Direct comparisons with other studies are difficult due to differences in the chosen confidence threshold and the version of the ML bounding box model that we used (MegaDetector MDv5a), but one study that examined MegaDetector V4.1 reported precision as high as 99% for a threshold of 0.75 on motion-detected images and as low 35% on time lapse images with a threshold of 0.0 (Leorna and Brinkman, 2022). Likewise, that study reported recall ranging from 98% for motion-detected images with a threshold of 0.0 to 13% for time lapse images with a threshold of 0.75. Correct predictions (true positives and negatives) have the potential to greatly accelerate the tagging process. However, when ML models make a mistake (false positives and negatives), it can be costly. When an animal is present in an image, a ML model's failure to locate it (false negative) causes the time required for tagging to quadruple for the ML-assisted treatment. Conversely, when no animal is present in an image, tagging time triples when an ML model fails to recognize it (false positive).

False negative predictions have been shown to be more likely when animals encountered are small or far from the camera, such as if often the case in timelapse mode compared to triggered images (Leorna and Brinkman, 2022). Several factors may also affect a ML model's false positive rate, including elements of the image background that could be confused for animals. For example, 12 of 49 (24.5%) false positive predictions in our experiment were from bounding boxes placed around the turkey feather that is included as an attractant in the camera field-of-view. Had the camera protocol differed and the feather not been placed in the camera's field-of-view, the false positive rate would likely have been lower.

One of the most important factors influencing the false positive and false negative rates is also the most controllable: the threshold for the ML confidence score for determining which predicted bounding boxes to consider. When used to automatically cull empty images, the threshold might be set quite low, to minimize the number of false negatives and avoid loss of images with animals. This increase in recall typically results in a loss of precision, and so the false positive rate will be relatively higher. Tuning of the confidence threshold to optimize performance in ML-classification pipelines has been recommended (Bothmann et al., 2023), and could likewise be of benefit in a hybrid image labelling approach. Given the role of classification outcome in the effectiveness of an ML model to accelerate tagging, and the heterogeneity of performance in different circumstances, a pre-assessment of the ML model is important for informing the appropriate confidence threshold and in determining its utility.

There are some caveats of our statistical analysis that are worth noting. Internet speed (for rendering images from the web) and personal device performance were not measured and hence could not be included as covariates. However, taggers all self-reported that they experienced no issues with computing tools and any impacts would be at least partially captured by the random effect of "tagger". Additionally, identifying animals may take varying amounts of time based on combinations of the species present, image quality, and environmental conditions. These factors were not incorporated directly into our statistical model due to either the small sample sizes of their representative classes or our inability to explicitly measure them. Some were at least partially captured through correlation with other parameters that we did model. One example of this is background complexity (e.g., the amount of coarse woody debris). Since the background is more-or-less constant at each location, this feature is at least partially accounted for by "site".

The species present in an image are also partially captured by other features. For example, moose (*Alces alces*) are large and obvious and frequently linger in the camera field-of-view, captured in many

consecutive frames. Small species like tree squirrels (*Sciurus* sp.) are more likely to move quickly in and out of the frame or become obscured by vegetation. In this way, the species present may correlate with the "in sequence" feature and be captured in that way by our statistical model. Ultimately, we observed significant associations between our key fixed effects.

5. Conclusion

Ultimately, we show that ML model performance can affect the reliability of using ML in camera trapping research. In our study, the relative occurrence of each classification outcome can tip the scales in favor of or against its use as an aid in producing bounding boxes. While it may seem obvious that model performance can impact effectiveness, quantifying the relative gains from ML via a pre-assessment can inform performance thresholds based on a desired gain in tagging efficiency.

As ML-techniques become more pervasive in the field of remote wildlife monitoring, further study of how ML model performance impacts their utility could become increasingly important. Several open-source species-level classification models have recently been released (Böhner et al., 2022; Norouzzadeh et al., 2018; Tabak et al., 2019; Tabak et al., 2020; Tabak et al., 2022; Vecvanags et al., 2022; Whytock et al., 2021) which can similarly be used in tandem with human tagging to improve tagging efficiency (Willi et al., 2019). Yet, species-level classification models can also be geographically constrained and underperform when applied on "out-of-sample" data from locations or in contexts that differ from the original training data (Tabak et al., 2019). Understanding the role of model performance on tagging efficiency for species-detection could prove useful in determining in what circumstances these models are effective.

Author contributions

Laurence A. Clarfeld conceived the ideas and designed the methodology. Alexej P.K. Sirén and T.L. Wilson designed the camera trapping arrays and Robert Abrams, Sue Staats, and Scott McLellan, Alexej P.K. Sirén also maintained the cameras and collected the field data. Alexej P. K. Sirén, Brendan Mulhall, Elena Bernier, John Farrell, Gus Lunde, and Nicole Hardy were taggers for the image tagging experiment. Laurence A. Clarfeld and Therese M. Donovan performed the statistical analyses. Laurence A. Clarfeld and Therese M. Donovan led the writing of the manuscript, with additional contributions by Alexej P.K. Sirén. All authors contributed critically to the drafts and gave final approval for publication.

Funding

This work was supported by the U.S. Geological Survey: Grant # USGS - G21AC10001.

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 are available on the USGS ScienceBase repository. https://doi.org/10.5066/P9FGUQEZ (Clarfeld et al., 2023).

Acknowledgements

We would like to thank Sarah Bassing for her review and two anonymous reviewers for their feedback on drafts of this manuscript. Funding was provided by the U.S. Geological Survey. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The Vermont Cooperative Fish and Wildlife Research Unit is jointly supported by the U.S. Geological Survey, University of Vermont, Vermont Department of Fish and Wildlife, US Fish and Wildlife Service, and Wildlife Management Institute.

References

Ahumada, J.A., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T.G., Palmer, J., Schuttler, S., Zhao, J.Y., Jetz, W., et al., 2020. Wildlife insights: a platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. Environ. Conserv. 47, 1–6.

Balantic, C., Donovan, T., 2020. AMMonitor: remote monitoring of biodiversity in an adaptive framework with r. Methods Ecol. Evol. 11, 869–877. https://doi.org/10.1111/2041-210X.13397.

Beery, S., Morris, D., Yang, S., 2019. Efficient Pipeline for Camera Trap Image Review arXiv preprint arXiv:1907.06772.

Böhner, H., Kleiven, E.F., Ims, R.A., Soininen, E.M., 2022. A semi-automatic workflow to process camera trap images in R. bioRxiv 2010–2022.

Bothmann, L., Wimmer, L., Charrakh, O., Weber, T., Edelhoff, H., Peters, W., Nguyen, H., Benjamin, C., Menzel, A., 2023. Automated wildlife image classification: an active learning tool for ecological applications. Ecol Inform 102231. https://doi.org/10.1016/j.ecoinf.2023.102231.

Buehler, P., Carroll, B., Bhatia, A., Gupta, V., Lee, D.E., 2019. An automated program to find animals and crop photographs for individual recognition. Ecol Inform 50, 191–196. https://doi.org/10.1016/j.ecoinf.2019.02.003.

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, 675–685. https://doi.org/10.1111/1365-2664.12432.

Chen, G., Han, T.X., He, Z., Kays, R., Forrester, T., 2014. Deep convolutional neural network based species recognition for wild animal monitoring. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 858–862. https://doi.org/10.1109/ICIP.2014.7025172.

Clarfeld, L., Donovan, T., Siren, A., Mulhall, B., Bernier, E., Farrell, J., Lunde, G., Hardy, N., Abrams, R., Staats, S., McLellan, S., 2023. Evaluating a tandem human-machine approach to labelling of wildlife in remote camera monitoring: U.S. Geological Survey data release. https://doi.org/10.5066/P9FGUQEZ.

Cove, M.V., Kays, R., Bontrager, H., Bresnan, C., Lasky, M., Frerichs, T., Klann, R., Lee, T. E., Crockett, S.C., Crupi, A.P., Weiss, K.C.B., Rowe, H., Sprague, T., Schipper, J., Tellez, C., Lepczyk, C.A., Fantle-Lepczyk, J.E., LaPoint, S., Williamson, J., Fisher-Reid, M.C., King, S.M., Bebko, A.J., Chrysafis, P., Jensen, A.J., Jachowski, D.S., Sands, J., MacCombie, K.A., Herrera, D.J., van der Merwe, M., Knowles, T.W., Horan, R.V., Rentz, M.S., Brandt, L.R.S.E., Nagy, C., Barton, B.T., Thompson, W.C., Maher, S.P., Darracq, A.K., Hess, G., Parsons, A.W., Wells, B., Roemer, G.W., Hernandez, C.J., Gompper, M.E., Webb, S.L., Vanek, J.P., Lafferty, D.J.R., Bergquist, A.M., Hubbard, T., Forrester, T., Clark, D., Cincotta, C., Favreau, J., Facka, A.N., Halbur, M., Hammerich, S., Gray, M., Rega-Brodsky, C.C., Durbin, C., Flaherty, E.A., Brooke, J.M., Coster, S.S., Lathrop, R.G., Russell, K., Bogan, D.A., Cliché, R., Shamon, H., Hawkins, M.T.R., Marks, S.B., Lonsinger, R.C., O'Mara, M.T., Compton, J.A., Fowler, M., Barthelmess, E.L., Andy, K.E., Belant, J.L., Beyer, D.E., Kautz, T.M., Scognamillo, D.G., Schalk, C.M., Leslie, M.S., Nasrallah, S.L., Ellison, C. N., Ruthven, C., Fritts, S., Tleimat, J., Gay, M., Whittier, C.A., Neiswenter, S.A., Pelletier, R., DeGregorio, B.A., Kuprewicz, E.K., Davis, M.L., Dykstra, A., Mason, D. S., Baruzzi, C., Lashley, M.A., Risch, D.R., Price, M.R., Allen, M.L., Whipple, L.S., Sperry, J.H., Hagen, R.H., Mortelliti, A., Evans, B.E., Studds, C.E., Sirén, A.P.K., Kilborn, J., Sutherland, C., Warren, P., Fuller, T., Harris, N.C., Carter, N.H., Trout, E., Zimova, M., Giery, S.T., Iannarilli, F., Higdon, S.D., Revord, R.S., Hansen, C.P., Millspaugh, J.J., Zorn, A., Benson, J.F., Wehr, N.H., Solberg, J.N., Gerber, B.D., Burr, J.C., Sevin, J., Green, A.M., Şekercioğlu, Ç.H., Pendergast, M., Barnick, K.A., Edelman, A.J., Wasdin, J.R., Romero, A., O'Neill, B.J., Schmitz, N., Alston, J.M., Kuhn, K.M., Lesmeister, D.B., Linnell, M.A., Appel, C.L., Rota, C., Stenglein, J.L., Anhalt-Depies, C., Nelson, C., Long, R.A., Jo Jaspers, K., Remine, K.R., Jordan, M.J., Davis, D., Hernández-Yáñez, H., Zhao, J.Y., McShea, W.J., 2021. SNAPSHOT USA 2019: a coordinated national camera trap survey of the United States. Ecology 102. https://doi.org/10.1002/ecy.3353.

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), pp. 230–237. https://doi.org/10.1109/WACV.2013.6475023.

Cunha, F., dos Santos, E.M., Colonna, J.G., 2023. Bag of tricks for long-tail visual recognition of animal species in camera-trap images. Ecol Inform 76, 102060. https://doi.org/10.1016/j.ecoinf.2023.102060.

De Lorm, T., Horswill, C., Rabaiotti, D., Ewers, R., Groom, R., Watermeyer, J., Woodroffe, R., Fund, A.W.C., 2023. Optimising the automated recognition of individual animals to support population monitoring. Ecol Evol. 13 (7), e10260.

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.

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, 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, 209–223.

Hartig, F., 2022. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.

Haucke, T., Kühl, H.S., Hoyer, J., Steinhage, V., 2022. Overcoming the distance estimation bottleneck in estimating animal abundance with camera traps. Ecol Inform 68, 101536.

Hendry, H., Mann, C., 2017. Camelot–intuitive software for camera trap data management. BioRxiv 203216.

Herrera, D.J., Dixon, J.D., Cove, M.V., 2022. Long-term monitoring reveals the value of continuous trapping to curtail the effects of free-roaming cats in protected island habitats. Glob Ecol Conserv 40, e02334. https://doi.org/10.1016/j.gecco.2022.e02334.

Hofmeester, T.R., Rowcliffe, J.M., Jansen, P.A., 2017. A simple method for estimating the effective detection distance of camera traps. Remote Sens Ecol Conserv 3, 81–89.

Johanns, P., Haucke, T., Steinhage, V., 2022. Automated distance estimation for wildlife camera trapping. Ecol Inform 70, 101734.

Kays, R., Arbogast, B.S., Chris, M.B., Boone, H.M., Bowler, M., Burneo, S.F., Cove, M.V., Hansen, C.P., Jansen, P.A., Kolowski, J.M., Knowles, T.W., Guimarães, M., Lima, M., Millspaugh, J., Mcshea, W.J., Pacifici, K., Parsons, A.W., Pease, B.S., Rovero, F., Santos, F., Schuttler, S.G., Sheil, D., Si, X., 2020. An empirical evaluation of camera trap study design: How many, how long and when? Methods Ecol. Evol. 11, 700–713. https://doi.org/10.1111/2041-210X.13370.

Leorna, S., Brinkman, T., 2022. Human vs. machine: detecting wildlife in camera trap images. Ecol Inform 72, 101876.

Li, X., Tian, H., Piao, Z., Wang, G., Xiao, Z., Sun, Y., Gao, E., Holyoak, M., 2022. cameratrapR: an R package for estimating animal density using camera trapping data. Ecol Inform 69, 101597. https://doi.org/10.1016/j.ecoinf.2022.101597.

Lopez-Marcano, S., Jinks, L., Buelow, C.A., Brown, C.J., Wang, D., Kusy, B., Ditria, E., Connolly, R.M., 2021. Automatic detection of fish and tracking of movement for ecology. Ecol Evol 11, 8254–8263.

Mack, C.A., 2011. Fifty years of Moore's law. IEEE Trans. Semicond. Manuf. 24, 202–207.

Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., van Bentham, K., Bolker, B., Brooks, M., Brooks, M.M., 2017. Package 'glmmtmb.'. In: R Package Version 0.2. 0.

Neupane, S.B., Sato, K., Gautam, B.P., 2022. A literature review of computer vision techniques in wildlife monitoring. IJSRP 16, 282–295.

Norman, D.L., Bischoff, P.H., Wearn, O.R., Ewers, R.M., Rowcliffe, J.M., Evans, B., Sethi, S., Chapman, P.M., Freeman, R., 2023. Can CNN-based species classification generalise across variation in habitat within a camera trap survey? Methods Ecol. Evol. 14, 242–251.

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. 115, E5716–E5725.

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. https://doi.org/10.1111/2041-210X.13504.

O'Connell, A.F., Nichols, J.D., Karanth, K.U., 2011. Camera Traps in Animal Ecology: Methods and Analyses. Springer Japan, Tokyo. https://doi.org/10.1017/CBO9781107415324.004.

Price Tack, J.L., West, B.S., McGowan, C.P., Ditchkoff, S.S., Reeves, S.J., Keever, A.C., Grand, J.B., 2016. AnimalFinder: A semi-automated system for animal detection in time-lapse camera trap images. Ecol Inform 36, 145–151. https://doi.org/10.1016/j.ecoinf.2016.11.003.

Ren, X., Han, T.X., He, Z., 2013. Ensemble video object cut in highly dynamic scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

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, 1228–1236.

Sirén, A.P.K., Somos-Valenzuela, M., Callahan, C., Kilborn, J.R., Duclos, T., Tragert, C., Morelli, T.L., 2018. Looking beyond wildlife: Using remote cameras to evaluate accuracy of gridded snow data. Remote Sens Ecol Conserv 4, 375–386. https://doi.org/10.1002/rse2.85.

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., Townsend, S.E., Carbone, C., Rowcliffe, J.M., Whittington, J., Brodie, J., Royle, J.A., Switalski, A., Clevenger, A.P., Heim, N., Rich, L.N., 2017. Scaling up camera traps: monitoring the planet's biodiversity with networks of remote sensors. Front. Ecol. Environ. 15, 26–34. https://doi.org/10.1002/FEE.1448.

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., et al., 2019. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol. Evol. 10, 585–590.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Newton, E.J., Boughton, R.K., Ivan, J.S., Odell, E.A., Newkirk, E.S., Conrey, R.Y., Stenglein, J., et al., 2020. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. Ecol Evol 10, 10374–10383.

Tabak, M.A., Falbel, D., Hamzeh, T., Brook, R.K., Goolsby, J.A., Zoromski, L.D., Boughton, R.K., Snow, N.P., VerCauteren, K.C., Miller, R.S., 2022. CameraTrapDetectoR: Automatically detect, classify, and count animals in camera trap images using artificial intelligence. bioRxiv 2022.

Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M., Hopcraft, G.C., 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods Ecol. Evol. 10, 779–787.

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, 353.

Wearn, O.R., Glover-Kapfer, P., 2019. Snap happy: camera traps are an effective sampling tool when compared with alternative methods. R. Soc. Open Sci. 6, 181748.

Wei, W., Luo, G., Ran, J., Li, J., 2020. Zilong: a tool to identify empty images in camera-trap data. Ecol Inform 55, 101021.

Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Słupski, T., Koumba Pambo, A.F., Rogala, M., Bahaa-el-din, L., Boekee, K., Brittain, S., Cardoso, A.W., et al., 2021. Robust ecological analysis of camera trap data labelled by a machine learning model. Methods Ecol. Evol. 12, 1080–1092.

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., Fortson, L., 2019. Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol. Evol. 10, 80–91.

Yang, D.-Q., Tan, K., Huang, Z.-P., Li, X.-W., Chen, B.-H., Ren, G.-P., Xiao, W., 2021. An automatic method for removing empty camera trap images using ensemble learning. Ecol Evol 11, 7591–7601.

Yu, X., Wang, J., Kays, R., Jansen, P.A., Wang, T., Huang, T., 2013. Automated identification of animal species in camera trap images. EURASIP J Image Video Process 2013. https://doi.org/10.1186/1687-5281-2013-52.