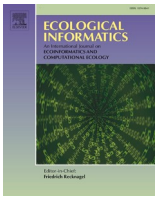
Ecological Informatics 81 (2024) 102578

Contents lists available at ScienceDirect

Ecological Informatics

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



A versatile, semi-automated image analysis workflow for time-lapse camera trap image classification

Gerardo Celis [a,*], Peter Ungar [a], Aleksandr Sokolov [b], Natalia Sokolova [b], Hanna Böhner [c], Desheng Liu [d], Olivier Gilg [e], Ivan Fufachev [b], Olga Pokrovskaya [b], Rolf Anker Ims [c], Wenbo Zhou [f], Dan Morris [g], Dorothee Ehrich [c]

[a] Department of Anthropology and Environmental Dynamics Program, University of Arkansas, Fayetteville, AR 72701, USA
[b] Arctic Research Station of the Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Zelenaya Gorka 21, Labytnangi, Russia
[c] UiT - The Arctic University of Norway, Department of Arctic and Marine Biology, 9019 Tromsø, Norway
[d] Department of Geography, The Ohio State University, Columbus, OH 43210, USA
[e] UMR 6249 Chrono-environnement, CNRS, Université de Franche-Comté, France & Groupe de Recherche en Ecologie Arctique, 16 rue de Vernot, Francheville, 25000 Besançon, France
[f] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA
[g] Google AI for Nature and Society, Mountain View, California, USA

ARTICLE INFO

Keywords:
Arctic wildlife monitoring
Deep learning
ResNet-50
MegaDetector
Time-lapse camera

ABSTRACT

Camera traps are a powerful, practical, and non-invasive method used widely to monitor animal communities and evaluate management actions. However, camera trap arrays can generate thousands to millions of images that require significant time and effort to review. Computer vision has emerged as a tool to accelerate this image review process. We propose a multi-step, semi-automated workflow which takes advantage of site-specific and generalizable models to improve detections and consists of (1) automatically identifying and removing low-quality images in parallel with classification into animals, humans, vehicles, and empty, (2) automatically cropping objects from images and classifying them (rock, bait, empty, and species), and (3) manually inspecting a subset of images. We trained and evaluated this approach using 548,627 images from 46 cameras in two regions of the Arctic: "Finnmark" (Finnmark County, Norway) and "Yamal" (Yamalo-Nenets Autonomous District, Russia). The automated steps yield image classification accuracies of 92% and 90% for the Finnmark and Yamal sets, respectively, reducing the number of images that required manual inspection to 9.2% of the Finnmark set and 3.9% of the Yamal set. The amount of time invested in developing models would be offset by the time saved from automation after 960 thousand images have been processed. Researchers can modify this multi-step process to develop their own site-specific models and meet other needs for monitoring and surveying wildlife, balancing the acceptable levels of false negatives and positives.

1. Introduction

Digital camera traps have become widely used for surveying and monitoring wildlife (Burton et al., 2015; Wearn and Glover-Kapfer, 2019). Camera traps are a non-invasive and relatively cost-effective method with many applications in ecology such as monitoring biodiversity (Oliver et al., 2023), investigating site occupancy (Hamel et al., 2013), estimating abundance (Stien et al., 2022), or studying species interactions (Rød-Eriksen et al., 2023). They make it realistic to obtain sufficient data to address ecological questions also for species that can be difficult to observe (e.g. Perera et al., 2022). However, trap arrays often generate thousands to millions of images requiring substantial effort to review manually. Computer vision offers the potential to significantly accelerate this image review process and is a rapidly developing field (e.g. Vélez et al., 2022, Morris, 2024).

Computer vision tools have been developed to facilitate different steps of image classification. A first step is often to remove empty images, here MegaDetector (Beery et al., 2019) or Machine Learning for Wildlife Image Classification 2 (MLWIC2, Tabak et al., 2020) are frequently used platforms. The next step is to classify and count animal

* Corresponding author.
E-mail address: gerardoc@uark.edu (G. Celis).

https://doi.org/10.1016/j.ecoinf.2024.102578
Received 18 May 2023; Received in revised form 21 February 2024; Accepted 24 March 2024
Available online 26 March 2024

species. A whole row of ready-made classifiers exist (Morris, 2024), however, existing classifiers focus in general on the fauna of a specific region, thus for example, MLWIC2 (Tabak et al., 2020) and CameraTrapDetectoR (Tabak et al., 2022) have been developed to classify North American species, the DeepFaune initiative aims at identifying the french fauna, WildID detects South African wildlife and the workflow developed by Böhner et al., 2023 aims specifically at registering Fennoscandian small rodents. Other workflows or platforms have been developed to allow users to train their own model (Mega Efficient Wildlife Classifier, Aandahl and Brook, 2024; Wildlife ML, Bothmann et al., 2023). This usually requires a large amount of images for training and often rather advanced computer skills. For a comprehensive list of available tools and options, see Morris (2024). However, the accuracy of computer vision still lags that of human annotators, particularly when images are derived from locations outside of a model's training domain, and several authors have emphasized the need for human review of computer vision results (Fennell et al., 2022; Schneider et al., 2020; Vélez et al., 2023).

The vast majority of camera traps are configured to use a motion sensor to trigger image capture when an animal is present in the camera's field of view (Böhner et al., 2023). However, in some cases, animals of interest may be too distant to trigger a motion sensor, or environmental conditions may result in an impractical number of false triggers, for instance during heavy snowfall; in these cases, a time-lapse protocol may be more appropriate (Hamel et al., 2013). Time-lapse camera trap datasets contain many more empty pictures than motion-triggered datasets; but they produce data in a more standardized form as the trigger behavior of motion sensors may vary quite substantially depending on species and other factors (Findlay et al., 2020). Moreover, time lapse protocols have the advantage of capturing small or distant animals in the camera's field of view that can be missed by motion sensors. This allows time-lapse cameras to capture as many as six times the number of animals recorded in motion trigger setups, but distant animals can be difficult for computer vision systems to detect (Leorna and Brinkman, 2022). A reliable method of identifying empty pictures is especially important for a workflow aimed at minimizing hands-on time required for analyzing time-lapse camera trap datasets. At the same time, to maintain data quality and maximize detection probabilities for animals that do not remain long at camera stations, it is important to minimize false negatives. Moreover, empty pictures should be distinguished from pictures with bad visibility or obstructed lenses to relate detections to observation effort (i.e., number of pictures per day) for downstream statistical analyses (Burton et al., 2015).

This paper presents a multi-tool solution that is specifically tailored to analyzing time-lapse camera trap datasets using site-specific models in conjunction with generalized models. It provides a guided example of a multi-step workflow for semi-automated classification of images from camera traps using a personal computer.

Our approach combines training custom site-specific models that can be adapted to a new context in a flexible way with a highly performant openly available model, MegaDetector (Beery et al., 2019). Specifically, our approach consists of (1) identifying high-quality images, for which there is no model we are aware of currently available, (2) separating empty images from images with animals, humans, or vehicles, (3) cropping out detected objects from images and classifying them by object type (rock, bait, empty and species), and (4) manually inspecting a selection of images. We investigate trade-offs between false negatives and manual reviewing time, and we evaluate the benefit of several enhancements to the typical MegaDetector workflow.

Because arctic ecosystems are at present rapidly changing under the impact of climate change and increasing human activity (e.g., Ims et al., 2013), there is an urgent need for thorough monitoring of important arctic wildlife species such as carnivores. Camera traps are a well-suited non-invasive method that can be deployed relatively easily in remote areas (Hamel et al., 2013). Consequently, we demonstrate the proposed workflow by applying it to two long-term programs from the Arctic monitoring changes in the predator/scavenger community in the Yamalo-Nenets Autonomous District, Russia, and Finnmark County, Norway. Our datasets consist of time-lapse images taken at bait stations in the late winter, a time at which frequent snowfalls make the use of motion sensors difficult.

2. Workflow

The multi-step, semi-automated workflow proposed here (Fig. 1) is adapted from Böhner et al., 2023, including pre-processing of images, model training, classification, manual quality checks, and final data formatting. Specifically, we build on the results of Rigoudy et al. (2022) and Fennell et al. (2022), who combined MegaDetector with manual classification and custom-trained models. The workflow consists of the following two classification steps in addition to pre-processing of images and final manual inspection, quality control, and data formatting (Fig. 1).

2.1. Classification 1 – Image quality and animal presence/absence

As an initial classification step, we categorized images by quality and Animal presence/absence in parallel using two models. Separating empty images from low-quality images is important to quantify the observation effort (i.e., the number of high-quality images per day) for downstream modeling of detections. We trained a custom model that could classify images as *Bad* (low quality) or *Good* (high quality).

For animal detection on all images, we used MegaDetector v5.0[1] (Beery et al., 2019), which detects animals, humans, and vehicles in each image. In our case, MegaDetector was more accurate than other products considered for detecting animals and minimizing the number of false negatives. It further allows cropping individual detection from images, thus facilitating counting and species identification. Combining the results from both models, images were separated low quality images, high quality empty images and images potentially containing an animal.

2.2. Classification 2 – Reducing false positives

We used the bounding boxes of each object detected in the images potenitally containing an animal to crop parts with pixels that contain an object (Fig. S1). We cropped only those classified by the model as animals and applied a custom model to identify each crop into different categories. This step greatly reduced the number of false positives, as crops containing stones or other artifacts could be sorted out by the custom model.

2.3. Final step – data formatting and quality check

Quality control is an important part of every automated image classification workflow (Böhner et al., 2023), and applying an automatic classification workflow to a new dataset requires particular care. Optimally, in the case of a multi-annual monitoring program, a workflow should be validated by applying it to a year or season of data that has not been used for its development. As species identification depends in large part on the number of images available for training and the complexity of the community in a specific study, it may be necessary to manually check all images with animals.

3. Materials and methods

3.1. Camera trap setup and data collection

Images were obtained from monitoring programs of the tundra carnivore scavenger guild in two low Arctic regions: "Yamal" (the

[1] https://github.com/agentmorris/MegaDetector

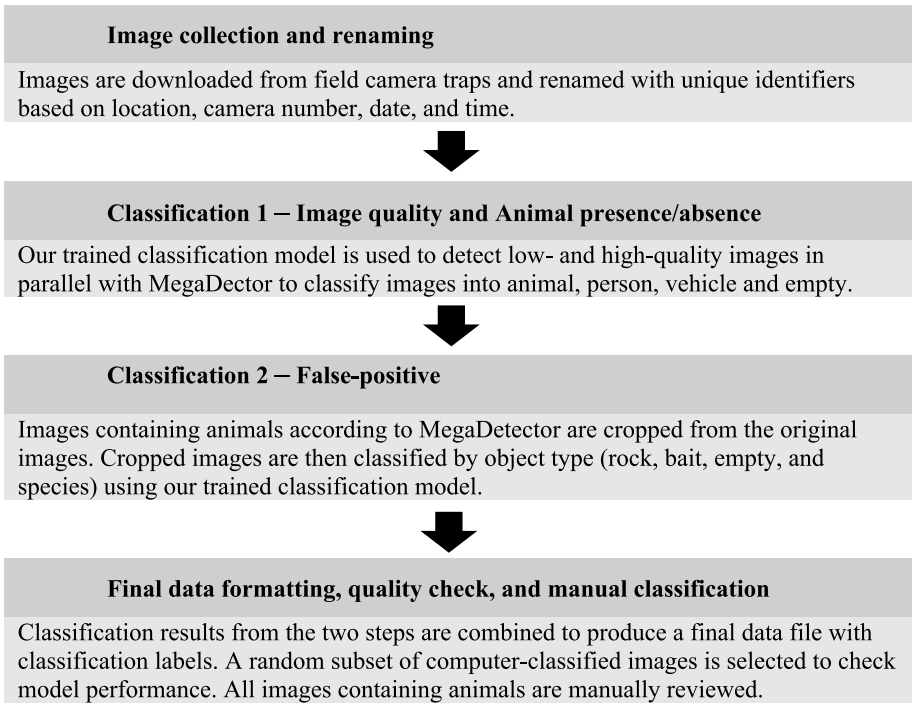


Fig. 1. Time-lapse camera trap workflow. Data preparation and model steps are adapted from Böhner et al., 2023.

Yamalo-Nenets Autonomous District, Russia) and "Finnmark" (Finnmark County, Norway). In Yamal, ten cameras were deployed at one site, Erkuta (68.2° N, 69.1° E), and in Finnmark, 36 cameras were spread across five sites (70–71° N, 25–30° E; Table 1; Killengreen et al., 2012). Cameras were activated from the end of February to early April. Data used in this study were collected from 2016 to 2022.

We used RECONYX® cameras (RapidFire, HyperFire and HyperFire 2, Holmen, WI, USA) placed on a permanently fixed metal pole at 30–50 cm above the snow surface. Cameras were painted in white and equipped with external batteries. In Finnmark, each camera station was baited with a ca 15 kg block of frozen slaughterhouse remains of reindeer (tendons, entrails, small meat fragments). In Yamal, frozen pelvis bones of reindeer with 1–2 kg of meat were mounted on a metal pole placed 2–5 m north of the camera. Cameras were programmed to take a picture every 5 min (no motion sensor). After 2–3 weeks of deployment, baits were replaced if needed, and memory cards were collected and replaced until the end of the observation period.

Initially, all images were reviewed manually by trained observers using the software MapView Professional (RECONYX ®) and separated into low-quality images (*Bad*) that were out of focus or obstructed (snow/ice in front of the lens or snowstorms), and high-quality images (*Good*) where an animal could have been detected. *Good* images were classified by animal presence/absence (*Animal*), and species and number of animal(s), when present. A total of 2,288,351 images were annotated in Finnmark (2016–2021) and 656,750 images in Yamal (2017–2022; Table 1). Most images from both locations were classified as *Good* (>83%). In Finnmark, *Bad* images represented 18.1% and in Yamal *Bad* images represented 8.7% (Table 2). At least one animal was detected in 6.9% of all images from Finnmark and in 2% of the images in Yamal. Twelve species of mammals and birds were documented in both locations, although community structure differed (Table 3). In Finnmark, the most common species was the raven (*Corvus corax*), appearing in 121,409 images, followed by the red fox (*Vulpes vulpes*) in 14,903 images. In Yamal, the most common species was the Arctic fox (*Vulpes lagopus*), appearing in 5017 images, followed by the magpie (*Pica pica*) in 4269 images.

3.2. Image quality classification: Training dataset and model training

Using the manual classifications, we randomly selected images from each site, camera, and year, to obtain ~15,000 images of *Bad* quality and ~ 57,000 images of *Good* quality for each location (Finnmark 2016–2018 and 2020–2021 and Yamal 2017–2021). These images were then reexamined by GC and DE, and any misclassified images were removed or reclassified. We also excluded marginal images (e.g., partly blurred images, images where an animal is only visible with a tail in a

Table 1
The number of cameras deployed in the field for each site and the total number of images available for workflow development (model training, validation, and testing; see supporting information for details) from Finnmark, Norway (2016–2018, 2020–2021) and Yamal, Russia (2017–2021), and an independent test set from years not used in training (Finnmark 2019, and Yamal 2022).

Region	Site	Number of cameras	Training/ validation images	Test images
Finnmark	Komagdalen	8	461,407	75,824
	Vestre Jakobselv	7	367,483	79,807
	Stjernevann	5	284,994	53,555
	Ifjordfjellet	8	401,039	58,652
	Gaissene	8	426,970	78,620
Yamal	Erkuta	10	535,910	120,840

Table 2
Total number of images per classification group as assessed manually in Finnmark and Yamal (N), together with median and mean (standard deviation) percentage of images for each individual camera trap per year. The total dataset (workflow development and independent validation) comprised 36 cameras at 5 sites for 6 years in Finnmark and 9 or 10 cameras for 6 years in Yamal.

Location	Class ID	N	Median [%]	Mean (SD) [%]
Finnmark	Bad	402,409	14.5	18.1 (14.5)
	Good Animal	150,532	6.5	6.9 (3.5)
	Good Empty	1,735,410	78.5	75.7 (13.5)
Yamal	Bad	55,420	5.4	8.7 (9.4)
	Good Animal	14,133	1.3	2.0 (1.7)
	Good Empty	587,197	93.4	90.1 (9.3)

Table 3
The total number of individuals or crops for each species assessed manually in Finnmark (2016–2021) and Yamal (2017–2022).

Class ID	Finnmark	Yamal	Included in model
Moose - *Alces alces*	66	0	Yes
Golden eagle - *Aquila chrysaetos*	5168	0	Yes
Snowy owl - *Bubo scandiacus*	69*	2	No
Dog – *Canis familiaris*	0	19	No
Raven - *Corvus corax*	121,409	246	Yes
Hooded crow - *Corvus cornix*	1011	38	Yes
Wolverine *Gulo gulo*	1103	171	Yes
White-tailed eagle - *Haliaeetus albicilla*	1474	0	Yes
Human – *Homo sapiens*	152	1044	No
Ptarmigan – *Lagopus* spp.**	0	131	Yes
Mountain hare – *Lepus timidus*	0	1677	Yes
Magpie - *Pica pica*	1	4269	Yes
Reindeer - *Rangifer tarandus*	3143	679	Yes
Arctic fox - *Vulpes lagopus*	2341	5017	Yes
Red fox - *Vulpes vulpes*	14,903	1037	Yes

* All snowy owl images were from the test set (2019), this species was thus not used to train the model.

** Most ptarmigan observed in Erkuta are willow ptarmigan (*Lagopus lagopus*), but rock ptarmigan (*Lagopus muta*) occur as well. It is difficult to identify the species reliably on camera trap pictures. Both species are also present in Finnmark, but they were not recorded systematically in that data set because the focus was on predator monitoring.

corner etc.), as high-quality training data are important for model training (Böhner et al., 2023). In particular, images of animals at large distances (e.g., appearing as points on the horizon) that could be identified by humans only because they moved in and out of frame were excluded from model training. The resultant data subsets (46,491 images for Finnmark and 33,889 for Yamal; Table S1) were randomly divided into 92% to be used for model training, 8% for validation of the trained model.

Separate two-class models were trained for Finnmark and Yamal using the *keras* package in R (Allaire and Chollet, 2023) with a TensorFlow backend (Allaire and Tang, 2023). Preliminary trials showed that region-specific models performed better. The ResNet-50 architecture, a convolutional neural network that is 50 layers deep (He et al., 2015), was used to train the models with 55 epochs (number of times the algorithm goes through the entire training data set) and a batch size of 64 (number of samples to work through before updating model parameters) with a one-cycle learning rate (hyperparameter controlling model response to estimated error each time the model weights are updated) policy with a minimum of 0.000001 and a maximum of 0.001 (Smith, 2018).

We used the *keras* image_data_generator function for image augmentation, which included random assignment of the following: rotation 0–40°, width and height shift range of 20%, shear range 0–0.2 rad, zoom range 0–0.2 scalar range, a horizontal flip and a fill mode with the nearest pixel.

We trained and validated the image quality classifier on a laptop (MacBook Pro, M1 Pro 8-core central processing unit (CPU), 14-core graphics processing unit (GPU), 16GB RAM), using the GPU rather than CPU for data processing. GPUs are optimized for complex imaging tasks and, in our case, outperform CPUs by ~7×.

After evaluating a range of confidence thresholds for each of the two classes, we found that the best results were obtained by using a 0.95 threshold for the *Bad* class in the image quality model. Images that are below this threshold are considered to be of adequate quality for further review. Furthermore, any image that contains an animal detection according to MegaDetector is considered for further review, regardless of the output of the image quality model.

3.3. Animal detection with MegaDetector

Two versions of MegaDetector are available, trained on slightly different datasets: MegaDetector v5.0a (MDv5a) and v5.0b (MDv5b). Both versions of MegaDetector were applied to all images. MegaDetector also provides two optional enhancements that can be combined with either model version:

1. MegaDetector normally resizes each image to be 1280 pixels wide prior to detecting objects. The *tiling* feature instead breaks each image into overlapping 1280-pixel by 1280-pixel "tiles", runs MegaDetector independently on each tile, and combines the results.
2. The *test-time augmentation* (TTA) feature makes several copies of each image and applies a different transformation to each copy prior to detecting objects, then combines the results.

To our knowledge, this is the first evaluation of the impact of tiling and TTA on MegaDetector's accuracy.

We found that detection of animals was slightly better using MDv5a than MDv5b, and that tiling and TTA with MDv5a further enhanced detection (Fig. S3) (see Table S2 for settings). Tiling helped detect animals at a distance and also those less conspicuous in the snow (white hares and arctic foxes). Test-time augmentation was also helpful for detection of less conspicuous animals, and especially for those under low light or night conditions (Fig. S4). We merged all detections from the MDv5a results with tiling and the MDv5a results with TTA which provided the lowest number of false negatives (Table S3); all subsequent analysis of MegaDetector results is based on this merged set of detections. We used a confidence threshold of 0.1 for all three MegaDetector categories (animal, person, vehicle).

To reduce the number of false-positive detections, MegaDetector has a post-processing tool for identifying detections that occur in the same location in many images from the same camera, which are often rocks or sticks, but may also be sleeping or stationary animals. Consequently, this tool is semi-automated: a human reviewer examines one example of each detection, along with a grid showing each instance of that detection (Fig. S5). The repeat detection elimination (RDE) tool was applied to the merged output (see Table S2 for settings). We found that ~8 k tiled images take about 1.5 h to review, which reduced 99,509 animal detections to 36,886 from our merged output (Table S4, Figs. S3 & S6).

MegaDetector was run on a Windows PC with two Nvidia RTX 4090 GPUs.

3.4. False-positive classification: Training dataset and model training

In the previous sections, we primarily referred to MegaDetector as a tool for categorizing images. MegaDetector also predicts the location of each object within the image, in the form of a bounding box around each object. For each image that MegaDetector identified as containing one or more animals, objects were cropped from the images using MegaDetector's predicted bounding boxes, and those crops were used to train a model for false positive classification. We include all classes that had >50 images from the combined sites (Finnmark and Yamal). For the 12 species classes and 4 non-animal classes (baits, rocks and empty) we retained (Table 3), we obtained 42,591 image crops to train the model, and 3746 for validation at each object detection class (Table S5). The classes used for training included empty, rock, and bait in addition to animal species, as one of the aims of this classification step was to further reduce the number of false positive detections. The animal false positive classification model was trained using the ResNet-50 architecture with the same approach as the image quality model described above. Each crop was assigned to a class, obtaining the maximum confidence value from the model without any threshold. The results from this model were combined with the detection confidence obtained from MegaDetector. For images classified as containing an animal with a confidence ≥0.35, the maximum confidence value from the animal false positive model was chosen only among species predictions. This allowed us to reduce the number of false positives without a large impact on false negatives.

3.5. Workflow performance

Workflow performance was assessed using data sets representing a separate year of data from each site (2019 for Finnmark and 2022 for Yamal, Table 1; hereafter 'test data sets'). Although cameras were placed at the same location every year, the background within the site varied both within and between years (e.g., snow cover, lighting, exact camera positioning), creating distinct image sets (Fig. S2). This added complexity to the images allowed us to test our workflow (Fig. 1) under "real-world" conditions. This approach corresponds to the situation of long-term monitoring programs, where new image datasets are obtained annually and should be classified with a procedure developed based on available data from previous years (Böhner et al., 2023).

Performance was measured in terms of accuracy, precision, recall, and F1 metrics as defined in Table 4 using the *caret* R package (Kuhn, 2008). For the test data, we also compared the number of days and time of day with detection of each species between the workflow results and the manual scoring, in addition to the picture-by-picture performance evaluation. Indeed, daily or time of day detections are often used in downstream analyses of camera trap data for ecological analyses (Hamel et al., 2013; Rød-Eriksen et al., 2023).

4. Results

4.1. Model performance on test data

4.1.1. Image quality models

The image quality models had high accuracy for the test data sets both in Finnmark (0.977) and Yamal (0.959), with higher precision (0.920) and recall (0.910) for low-quality class in Finnmark and Yamal (0.764 and 0.779, respectively) (Table 5, Fig. S7).

4.1.2. MegaDetector

All results presented in this subsection refer to the merged detections from the MDv5a results with tiling and the MDv5a results with TTA, with a confidence threshold of 0.1.

There were 123 images classified as *Bad* that included animals, but this was reduced to 31 after including animal detection using MegaDetector. After eliminating all *Bad* images and excluding images in which MegaDetector predicted an animal and a human in the same image (1165 images Finnmark, 434 Yamal; Fig. S8), because these did not occur in the manual classification. For Finnmark, MegaDetector had an overall accuracy of 0.890. For animals, MegaDetector had a precision of 0.514 and recall of 0.992 (Table 6). The empty class had a precision of 0.999 and 0.880 recall. A total of 48,704 (14.0%) images were classified as having animals present, but approximately half of those were empty (Fig. S8). Excluding false positives (assuming that animal images would be reviewed manually), the total number of days with detection of an animal per camera station was similar to that of manual classification, with 12 individual camera station of the 138 underestimating by one camera day in most cases for arctic fox, red fox, wolverine, raven, and reindeer (Fig. S9). The detection frequency for each hour of the day was also very similar between manual review and MegaDetector predictions, with no directional bias by time of day (Fig. S11). The model thus results in an acceptably low level of false negatives randomly distributed in time.

For Yamal, MegaDetector had an accuracy of 0.906. The animal class had a low precision of 0.206, but a relatively high recall of 0.862 (Table 6, Fig. S8), whereas the empty class had a precision of 0.996 and a recall of 0.908. Excluding the false positive animal images, 6 of the 32 individual species camera detections for all camera stations were underestimated (mostly by one day) for willow ptarmigan, mountain hare, and magpie (Fig. S10). The detection frequency for each hour of the day was also very similar between manual review and MegaDetector predictions, with no particular bias to any specific time of day, for all species except willow ptarmigan and mountain hare (Fig. S12). For willow ptarmigans, MegaDetector predicted more detections between 5 and 10 h and very few detections during evenings than human reviewers. MegaDetector predicted fewer mountain hare detections during mid-day than human reviewers.

4.1.3. False positive model

A total of 83,941 image crops were created from MegaDetector results for Finnmark. Forty-nine images were of new classes that were not included in the model (human, snowy owl, black-backed gull). After excluding these, the accuracy for the false positive model in Finnmark was 0.919. The model was very precise at classifying cropped images as "No animal" (>0.985; Table 7), with only one animal image misclassified as empty (Fig. S13) and there were 408 false positives.

For Yamal, a total of 15,276 image crops were created. Nine images were of animals not included in the trained model (snowy owl and ptarmigan). After excluding these, the overall accuracy for the false positive model in Yamal was 0.898. The model was precise at classifying image crops that did not contain an animal (0.985 precision for the "no animal" class; Table 7) with 128 false positives.

4.1.4. All automated steps combined

After combining all model predictions (image quality, MegaDetector, and false positive) and manual inspections we obtain an overall classification accuracy of 0.942 for Finnmark and 0.926 for Yamal. Including the false positive classification model reduced the animal false positives created by MegaDetector from 23,593 to 6242 images for the Finnmark data set and from 6990 to 2534 images for the Yamal data set (Figs. S8 & S14), but at a cost of 132 and 45 false negatives, respectively. This

Table 4
Definitions of model performance metrics based on "caret" R package, based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Metric	Equation	Definition
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$	Proportion of correct predictions in the whole data set.
Precision	$\frac{TP}{TP+FP}$	The proportion of images that a model classified as a specific category C that are actually category C.
Recall	$\frac{TP}{TP+FN}$	The proportion of images that are actually a specific category C that the a model classified as C.
F1	$\frac{2*precision*recall}{precision+recall}$	Weighted average of precision and recall.

Table 5
Performance of the image quality model on the test data.

Location	Id	Precision	Recall	F1
Finnmark	Bad	0.920	0.910	0.915
	Good	0.986	0.988	0.987
Yamal	Bad	0.764	0.779	0.771
	Good	0.979	0.977	0.978

Table 6
MegaDetector performance on the test data. Excludes *Bad* images with no MegaDetetor confidence below 0.35 and images in which MegaDetector predicted animals and humans, because these did not occur in the manual classification.

Location	Class id	Precision	Recall	F1
Finnmark	Animal	0.514	0.992	0.677
	Empty	0.999	0.880	0.936
	Human	0.0009	0.381	0.018
Yamal	Animal	0.206	0.862	0.332
	Empty	0.996	0.908	0.950
	Human	0.046	0.426	0.084

Table 7
False positive model performance on the test data. To estimate model performance metrics, classes that were exclusively in the manual assessment or model output were not included. The "No animal" class combines the Empty, Bait, Bait_yamal, and Rock model classes.

	Finnmark			Yamal		
Class ID	Precision	Recall	F1	Precision	Recall	F1
No animal	0.985	0.860	0.919	0.985	0.834	0.904
Golden eagle	0675	0.899	0.771	–	–	–
White-tailed eagle	0.623	0.812	0.706	–	–	–
Raven	0.921	0.985	0.952	0.04	1.00	0.079
Hooded crow	0.941	0.888	0.914	–	–	–
Magpie	–	–	–	0.964	0.908	0.936
Mountain hare	–	–	–	0.684	0.181	0.286
Reindeer	0.084	0.375	0.138	–	–	–
Wolverine	0.612	0.719	0.661	0.013	0.600	0.027
Arctic fox	0.310	0.781	0.444	0.683	0.693	0.688
Red fox	0.557	0.944	0.701	0.012	0.833	0.025

equates to animal detection of 4806 images or 3.9% of the total images from camera traps in Yamal, and 32,208 or 9.2% in Finnmark (Fig. 2). These results are higher than the mean animal detection rates obtained by manual inspection – 2.0% for Yamal and 6.9% for Finnmark (Table 2) due to the remaining false positives.

5. Discussion

Our workflow correctly classified, on average, 91% of images into *Bad*. While it may at first seem to be an easy task to exclude bad-quality images from the data set, the number of *Bad* images can be large (on average 18% in Finnmark and 8.7% in Yamal, Table 2), vary by site, and change quickly depending on environmental conditions. Nevertheless, separating good images from bad images is particularly important for analyses that consider sampling effort, such as relative abundance indices (Burton et al., 2015) and exclusion of periods when the camera cannot determine the presence or absence of an animal. We are unaware of any other available models that are better able to parse good from bad images.

The detection of images with and without animals was 91% for our time-lapse cameras, similar to what other researchers have reported using MegaDetector for cameras using motion sensor triggers. We found that MegaDetector's test-time augmentation, tiling, and repeat detection elimination tools improved detection for animals with time-lapse triggers. It could detect smaller objects in images than previously reported resolution (60px for Reindeer; Leorna and Brinkman, 2022). For example, the smallest reindeer detected in our images was 18px, and the cropped image was correctly classified by our false positive model. This enhanced detection is attributable to the tiling of images, which improves identification of small objects, but some detections can be duplicates (Ünel et al., 2019) when an animal spans two or more tiles. Therefore, downstream use, such as counting the number of individuals from MegaDetector crops (Mitterwallner et al., 2023; Wang et al., 2022), must be considered cautiously, as it may overestimate the number of individuals.

Fals positive classification of all images with animals was 77% accurate when compared with a manually derived classification. These results are promising, though further work is needed to improve accuracy. Although MegaDetector's repeat detection elimination tool (RDE) helped reduce the number of false positives, using our false positive model, which included classes of species, baits, rocks or empty, we could reduce the number of images with false positives even further.

We obtained for the final portion of the workflow an animal recall accuracy of 0.985 for Finnmark and 0.843 for Yamal, with false-negative rates for animals of 1.4% and 15%, respectively within the range of what other studies have found (Clarfeld et al., 2023) but are dependent on the confidence threshold used (Bothmann et al., 2023). Assuming that all pictures where animals were detected by our workflow would be reviewed manually, something we would recommend given the performance of the present false positive classification model for other classes, this would reduce the number of images that require manual inspection to 9.2% of the total number of images to review in Finnmark and 3.9% in Yamal. Implementing this procedure could, therefore, save a great deal of time and effort associated with manual inspection/classification of imagery. Other computer-assisted workflows have have shown to reduce the processing time of image classification load by as much as 5× to 13× depending on the tasks (Fennell et al., 2022; Henrich

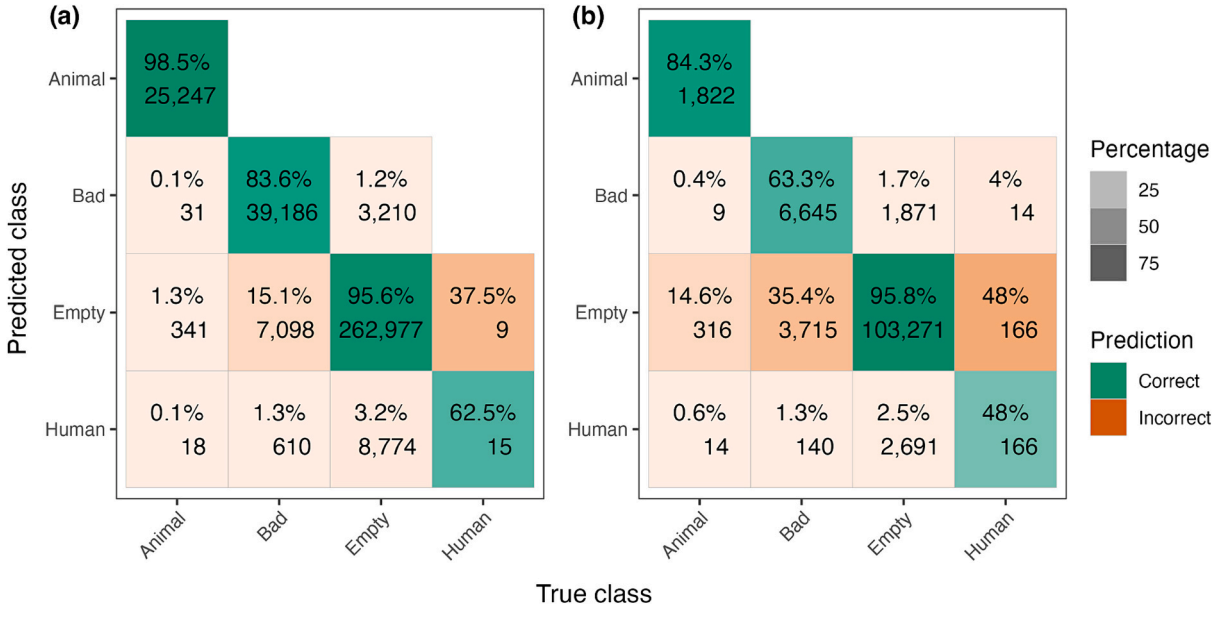


Fig. 2. Confusion matrix for the results of the complete workflow applied to the test data, which represents a full season of data not used in model development at (a) Finnmark and (b) Yamal. False positive model results were aggregated in the "Animal" class. The percentage and number of images with correct (diagonal, green) and incorrectly predicted classes (off-diagonal) are displayed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

et al., 2023). Our workflow reduces the load by ~62 h, however the amount of time required to develop the site-specific models and workflow took ~160 h. It would take approximately 960 k processed images to recover the time invested in developing the workflow, which makes sense for long-term projects where the initial investment of time is recouped over the life of the project.

6. Conclusion

The proposed semi-automatic workflow for classifying camera trap images is a robust method for identifying high-quality images, identifying images that contain animals, and reducing the number of false positives. Our workflow detected low-quality images and those with animals within the ranges of those detected by manual classification. The false positive classification step reduced the number of false positive animal detections generated by MegaDetector. Although the false positive model reduced the number of false positives, we recommend that users manually review images with animals because the model was not sufficiently accurate to rely solely on computer vision for species classification (hence our description of our workflow as "semi-automated").

We provide code ((https://github.com/gerlis22/CameraTrap.git) for this multi-step process so that researchers can create their own site-specific models and modify it to meet their needs for monitoring and surveying wildlife. Because our workflow is subdivided into several steps, it is flexible and can be adapted to various situations. The initial classification step could, for instance, be modified to include a classification into pictures with and without bait in addition to quality, or with and without snow, depending on the study's aims.

Author contributions

GC, DE, AS, NS, HB, RAI, IF, and OP conceived the ideas and designed methodology; DE, AS, NS, IF, and OP collected the data; GC, DM, WZ, and HB were involved in software, validation, formal analysis; RAI, DL, and PSU project administration and funding acquisition; GC led the writing of the manuscript and visualization. All authors contributed critically to the drafts and gave final approval for publication.

CRediT authorship contribution statement

Gerardo Celis: Conceptualization, Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Peter Ungar:** Funding acquisition, Project administration, Writing – review & editing, Resources, Writing – original draft. **Aleksandr Sokolov:** Conceptualization, Data curation, Funding acquisition, Project administration, Writing – review & editing. **Natalia Sokolova:** Conceptualization, Data curation, Writing – review & editing. **Hanna Böhner:** Conceptualization, Methodology, Software, Writing – review & editing. **Desheng Liu:** Funding acquisition, Writing – review & editing. **Olivier Gilg:** Writing – review & editing. **Ivan Fufachev:** Data curation, Writing – review & editing. **Olga Pokrovskaya:** Data curation, Writing – review & editing. **Rolf Anker Ims:** Conceptualization, Funding acquisition, Investigation, Methodology, Writing – review & editing. **Wenbo Zhou:** Software, Validation, Writing – review & editing. **Dan Morris:** Writing – review & editing. **Dorothee Ehrich:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing.

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

The original pictures and data from Finnmark and Yamal can be obtained from the authors upon request. The Finnmark data is available from the Climate-Ecological Observatory for Arctic Tundra (COAT) database (doi:10.48425/6jytn3ga) The code and detailed instructions for the workflow are on GitHub (https://github.com/gerlis22/CameraTrap.git).

Acknowledgments

This research was made possible by funding from Navigating the New Arctic – National Science Foundation (awards GC, PSU DE-2126796, VI-2126792, JZ-2126794, DL-2126798). The collection of the Norwegian data set was conducted within the Climate-Ecological Observatory for Arctic Tundra (COAT) and financed by the Norwegian Environment Agency. Jan Erik Knutsen, Berit Gaski, and the State Nature Surveillance in Vadsø and Tana carried out the field work in Norway. AS, NS, IF, and OP were supported by the Russian Ministry of Science and Higher Education program "Terrestrial ecosystems of northwestern Siberia: assessment of the modern transformation of the communities" No. 122021000089-9. Collection the Russian data set was supported by the Governor of Yamal-Nenets autonomous okrug. We thank the families Laptander, Serotetto, and Vanuy'to who helped us in all years in challenging early spring conditions in Yamal, and Vyacheslav Osokin for invaluable contribution to field work. Numerous assistants contributed to the manual scoring of the images, notably Torunn Moe, Stijn Hofhuis, Dag A. H. Olsen, and Fanny Berthelot.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102578.

References

Aandahl, Z., Brook, B., 2024. MEWC - Mega Efficient Wildlife Classifier [WWW document]. URL. https://github.com/zaandahl/mewc?tab=readme-ov-file.

Allaire, J.J., Chollet, F., 2023. keras: R interface to "Keras." R package version 2.13.0. https://CRAN.R-project.org/package=keras.

Allaire, J.J., Tang, T., 2023. tensorflow: R interface to "Tensorflow." R package version 2.14.0. https://CRAN.R-project.org/[ackage=tensorflow.

Beery, S., Morris, D., Yang, S., 2019. Efficient pipeline for camera trap image review. Arxiv. https://doi.org/10.48550/arxiv.1907.06772.

Böhner, H., Kleiven, E.F., Ims, R.A., Soininen, E.M., 2023. A semi-automatic workflow to process camera trap images from small mammal camera traps. Ecol. Inform. 76, 102150.. https://doi.org/10.1016/j.ecoinf.2023.102150.

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. 77, 102231 https://doi.org/10.1016/j.ecoinf.2023.102231.

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T., Bayne, E., Boutin, S., 2015. REVIEW: 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.

Clarfeld, L.A., Sirén, A.P.K., Mulhall, B.M., Wilson, T.L., Bernier, E., Farrell, J., Lunde, G., Hardy, N., Gieder, K.D., Abrams, R., Staats, S., McLellan, S., Donovan, T.M., 2023. Evaluating a tandem human-machine approach to labelling of wildlife in remote camera monitoring. Ecol. Inform. 77, 102257 https://doi.org/10.1016/j.ecoinf.2023.102257.

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.

Findlay, M.A., Briers, R.A., White, P.J.C., 2020. Component processes of detection probability in camera-trap studies: understanding the occurrence of false-negatives. Mammal Res. 65, 167–180. https://doi.org/10.1007/s13364-020-00478-y.

Hamel, S., Killengreen, S.T., Henden, J., Eide, N.E., Roed-Eriksen, L., Ims, R.A., Yoccoz, N.G., 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.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. Arxiv. doi:10.48550/arxiv.1512.03385.

Henrich, M., Burgueño, M., Hoyer, J., Haucke, T., Steinhage, V., Kühl, H.S., Heurich, M., 2023. A semi-automated camera trap distance sampling approach for population density estimation. Remote Sens. Ecol. Conserv. https://doi.org/10.1002/rse2.362.

Ims, R.A., Ehrich, D., Forbes, B.C., Huntley, B., Walker, D.A., Wookey, P.A., Berteaux, D., Bhatt, U.S., Bråthen, K.A., Edwards, M.E., Epstein, H.E., Forchhammer, M.C., Fuglei, E., Gauthier, G., Gilbert, S., Leung, M., Menyushina, I.E., Ovsyanikov, N., Post, E., Raynolds, M.K., Reid, D.G., Schmidt, N.M., Stien, A., Sumina, O.I., van der Wal, R., 2013. Terrestrial ecosystems. In: Arctic Biodiversity Assessment: Status and Trends in Arctic Biodiversity. The Conservation of Arctic Flora and Fauna (CAFF), pp. 384–440.

Killengreen, S.T., Strømseng, E., Yoccoz, N.G., Ims, R.A., 2012. How ecological neighbourhoods influence the structure of the scavenger guild in low arctic tundra. Divers. Distrib. 18, 563–574. doi:10.1111/j.1472-4642.2011.00861.x.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Soft. 28 (5), 1–26. https://doi.org/10.18637/jss.v028.i05.

Leorna, S., Brinkman, T., 2022. Human vs. machine: detecting wildlife in camera trap images. Ecol. Inform. 72, 101876 https://doi.org/10.1016/j.ecoinf.2022.101876.

Mitterwallner, V., Peters, A., Edelhoff, H., Mathes, G., Nguyen, H., Peters, W., Heurich, M., Steinbauer, M.J., 2023. Automated visitor and wildlife monitoring with camera traps and machine learning. Remote Sens. Ecol. Conserv. https://doi.org/10.1002/rse2.367.

Morris, D., 2024. Everything I know about ML and Camera Traps [WWW document]. URL. https://agentmorris.github.io/camera-trap-ml-survey.

Oliver, R.Y., Iannarilli, F., Ahumada, J., Fegraus, E., Flores, N., Kays, R., Birch, T., Ranipeta, A., Rogan, M.S., Sica, Y.V., Jetz, W., 2023. Camera trapping expands the view into global biodiversity and its change. Philos. Trans. R. Soc. B 378, 20220232. https://doi.org/10.1098/rstb.2022.0232.

Perera, P., Karawita, H., Jayasinghe, C., 2022. The applicability of camera trap data to monitor the cryptic Indian pangolin (Manus crassicaudata) populations: a survey from a tropical lowland rainforest in Southwest Sri Lanka. Glob. Ecol. Conserv. 34, e02046 https://doi.org/10.1016/j.gecco.2022.e02046.

Rigoudy, N., Benyoub, A., Besnard, A., Birck, C., Bollet, Y., Bunz, Y., Carriburu, J.C., Caussimont, G., Chetouane, E., Cornette, P., Delestrade, A., Backer, N.D., Dispan, L., Brah, M.L., Duhayer, J., Elder, J.-F., Fanjul, J.-B., Fonderflick, J., Froustey, N., Garel, M., Gaudry, W., Gerard, A., Gimenez, O., Hemery, A., Hemon, A., Jullien, J.-M., Malafosse, I., Marginean, M., Prunet, V., Rabault, J., Randon, M., Regnier, A., Ribiere, R., Ricci, J.-C., Ruette, S., Sentilles, J., Siefert, N., Smith, B., Terpereau, G., Thuiller, W., Uzal, A., Vautrain, V., Dussert, G., Spataro, B., Miele, V., Chamaille-Jammes, S., 2022. The DeepFaune initiative: a collaborative effort towards the automatic identification of the French fauna in camera-trap images. Biorxiv. https://doi.org/10.1101/2022.03.15.484324, 2022.03.15.484324.

Rød-Eriksen, L., Killengreen, S.T., Ehrich, D., Ims, R.A., Herfindal, I., Landa, A.M., Eide, N.E., 2023. Predator co-occurrence in alpine and Arctic tundra in relation to fluctuating prey. J. Anim. Ecol. 92, 635–647. https://doi.org/10.1111/1365-2656.13875.

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, 3503–3517. https://doi.org/10.1002/ece3.6147.

Smith, L.N, 2018. A disciplined approach to neural network hyper-parameters: part 1 - learning rate, batch size, momentum, and weight decay. Arxiv. https://doi.org/10.48550/arxiv.1803.09820.

Stien, J., Stien, A., Tveraa, T., Rød-Eriksen, L., Eide, N.E., Killengreen, S.T., 2022. Estimating abundance in unmarked populations of Golden eagle (*Aquila chrysaetos*). Ecol. Solut. Évid. 3 https://doi.org/10.1002/2688-8319.12170.

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., Iannarilli, F., Erb, J., Brook, R. K., Davis, A.J., Lewis, J., Walsh, D.P., Beasley, J.C., VerCauteren, K.C., Clune, J., Miller, R.S., 2020. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. Ecol. Evol. 10, 10374–10383. https://doi.org/10.1002/ece3.6692.

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. https://doi.org/10.1101/2022.02.07.479461, 2022.02.07.479461.

Ünel, F.Ö., Özkalaycı, B.O., Çığla, C., 2019. The power of tiling for small object detection. In: 2019 IEEECVF Conf. Comput. Vis. Pattern Recognit. Work. (CVPRW), pp. 582–591. https://doi.org/10.1109/cvprw.2019.00084.

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. Arxiv. https://doi.org/10.48550/arxiv.2202.02283.

Vélez, J., McShea, W., Shamon, H., Castiblanco-Camacho, P.J., Tabak, M.A., Chalmers, C., Fergus, P., Fieberg, J., 2023. An evaluation of platforms for processing camera-trap data using artificial intelligence. Methods Ecol. Evol. 14, 459–477. https://doi.org/10.1111/2041-210x.14044.

Wang, Y., Zhang, Y., Feng, Y., Shang, Y., 2022. Deep learning methods for animal counting in camera trap images. In: 2022 IEEE 34th Int. Conf. Tools Artif. Intell. (ICTAI), pp. 939–943. https://doi.org/10.1109/ictai56018.2022.00143.

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 https://doi.org/10.1098/rsos.181748.