Remote Sensing in Ecology and Conservation
Open Access

ZSL

RESEARCH ARTICLE

Automated visitor and wildlife monitoring with camera traps and machine learning

Veronika Mitterwallner[1], Anne Peters[2,3], Hendrik Edelhoff[4], Gregor Mathes[5], Hien Nguyen[4], Wibke Peters[4], Marco Heurich[2,3,6] & Manuel J. Steinbauer[1]

[1]Sport Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER) and Bayreuth Center of Sport Science (BaySpo), University of Bayreuth, Bayreuth, Germany
[2]Department of National Park Monitoring and Animal Management, Bavarian Forest National Park, Grafenau, Germany
[3]Chair of Wildlife Ecology and Management, Faculty of Environment and Natural Resources, University of Freiburg, Tennenbacher Straße 4, 79106, Freiburg, Germany
[4]Research Unit Wildlife Biology and Management, Bavarian State Institute for Forestry, Hans-Carl von Carlowitz Platz 1, 85354, Freising, Germany
[5]Paleontological Institute and Museum, University of Zurich, Zurich, Switzerland
[6]Institute for Forestry and Wildlife Management, Campus Evenstad, Inland Norway University for Applied Science, 2480, Koppang, Norway

Keywords
camera traps, human–wildlife interactions, machine learning, recreation ecology, wildlife ecology

Correspondence
Veronika Mitterwallne, Sport Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER) and Bayreuth Center of Sport Science (BaySpo), University of Bayreuth, Bayreuth, Germany. Tel: 0049921-55-3478; E-mail: veronika.mitterwallner@uni-bayreuth.de.

Funding Information
The Bavarian State Ministry for the Environment and Consumer Protection contributed to this work by funding the project 'Integrative evaluations of the effects of recreational use on wildlife as a basis for evidence-based visitor management', carried out in the Bavarian Forest National Park. The study in the Veldenstein Forest was financed by the Bavarian State Ministry of Agriculture and Forestry (grant I043). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 491183248. Funded by the Open Access Publishing Fund of the University of Bayreuth.

Editor: Marcus Rowcliffe
Associate Editor: Francesco Rovero

Received: 15 March 2023; Revised: 24 July 2023; Accepted: 4 August 2023

doi: 10.1002/rse2.367

Abstract

As human activities in natural areas increase, understanding human–wildlife interactions is crucial. Big data approaches, like large-scale camera trap studies, are becoming more relevant for studying these interactions. In addition, open-source object detection models are rapidly improving and have great potential to enhance the image processing of camera trap data from human and wildlife activities. In this study, we evaluate the performance of the open-source object detection model MegaDetector in cross-regional monitoring using camera traps. The performance at detecting and counting humans, animals and vehicles is evaluated by comparing the detection results with manual classifications of more than 300 000 camera trap images from three study regions. Moreover, we investigate structural patterns of misclassification and evaluate the results of the detection model for typical temporal analyses conducted in ecological research. Overall, the accuracy of the detection model was very high with 96.0% accuracy for animals, 93.8% for persons and 99.3% for vehicles. Results reveal systematic patterns in misclassifications that can be automatically identified and removed. In addition, we show that the detection model can be readily used to count people and animals on images with underestimating persons by −0.05, vehicles by −0.01 and animals by −0.01 counts per image. Most importantly, the temporal pattern in a long-term time series of manually classified human and wildlife activities was highly correlated with classification results of the detection model (Pearson's $r = 0.996$, $p < 0.001$) and diurnal kernel densities of activities were almost equivalent for manual and automated classification. The results thus prove the overall applicability of the detection model in the image classification process of cross-regional camera trap studies without further manual intervention. Besides the great acceleration in processing speed, the model is also suitable for long-term monitoring and allows reproducibility in scientific studies while complying with privacy regulations.

Introduction

Increasing human activities such as recreation can affect wildlife behaviour globally. Wildlife adapts to human presence by changing its habitat use, movement patterns and temporal activity (Gaynor et al., 2018; Naidoo & Burton, 2020; Tucker et al., 2018). Although evidence indicates that human disturbances can impact wildlife behaviour, these effects are potentially reversible (Nellemann et al., 2010; Shively et al., 2005). Moreover, knowledge on long-term consequences for populations and ecosystems is limited (Wilson et al., 2020). Assessing human–wildlife interactions on broad spatial and temporal scales is hence of importance to incorporate this information into wildlife management and conservation (Frank et al., 2019). As interactions are complex and hardly generalizable (Tablado & Jenni, 2017; Zimmermann et al., 2021), big data approaches such as the combination of camera trapping and machine learning will highly improve studies on the spatiotemporal interactions of human and wildlife activities.

In wildlife ecology, camera traps have proven to be a robust method to generate spatiotemporal data on multiple species with the advantage of being non-invasive, and cost-efficient (Burton et al., 2015; Caravaggi et al., 2017; Glover-Kapfer et al., 2019; Rowcliffe et al., 2014). Recently, this approach has further increased its applicability for research, wildlife management and conservation as machine learning models for wildlife classification on image data improved rapidly (Falzon et al., 2020; Norouzzadeh et al., 2020; Tabak et al., 2019). Computer vision algorithms reach high accuracies in animal species classification and outperform manual classification substantially on a temporal scale with, for instance, 2000 classified images per minute (Tabak et al., 2019). A key disadvantage is the restriction to species that were included and labelled while training the machine learning model resulting in low accuracies beyond those classes as well as troubles with untrained camera trap sites (Schneider et al., 2020). To overcome the site- and species-dependence issues, object detection models, which identify the location of an object on an image and classify those objects into basic categories may be favourable compared to mostly used image classifiers, which classify the entire image. Open-source object detection models, such as MegaDetector (Beery et al., 2019) are trained on millions of globally generated images to detect basic object classes, such as persons, animals or vehicles and have been used in multiple wildlife conservation programs worldwide.

Particularly studies in recreation ecology seeking to understand complex interactions of species and humans in space and time, would benefit from such an approach as large spatiotemporal data from wildlife and human activities can be analysed simultaneously. So far, assessments of human–wildlife interactions using camera traps were limited by data protection regulations, that is, personal rights, and costly manual classification processes (Lupp et al., 2021; Miller et al., 2017; Reilly et al., 2017). Using automated object detection would significantly reduce temporal and financial efforts in this field and additionally facilitate the compliance with data protection regulations. MegaDetector has been shown to detect humans with a high precision of 99% and animals with 82% precision, which resulted in a 500% increase in processing speed (Fennell et al., 2022). Despite these advantages, the establishment of a widely usable approach by coupling camera trap data and open-source object detection models on wildlife and human activity (Fennell et al., 2022; Staab et al., 2021) requires a detailed assessment of the methodological restrictions and bottlenecks. Investigating general patterns in human–wildlife interactions needs long-term, large-scale and cross-regional study designs resulting in the use of differing camera trap models and multiple human classifiers. Different site and trail conditions, varying seasons (e.g. snow heights and vegetation heights) and recreational activities (e.g. mountain biking, hiking and skiing) as well as fluctuating staff in research and camera trap maintenance tend to increase the complexity of the underlying data and might impose biases to the detection models. Nevertheless, a determination of error sources, which decrease the accuracy of automated detection has not been conducted yet. Such an approach would help to define site-specific workflows to increase the accuracy of datasets generated with object detection models and hence the output of subsequent studies in wildlife and recreation ecology.

Here, we present and evaluate a methodological approach for automated visitor and wildlife monitoring in multiple recreational areas using camera traps and automated object detection. We test the object detection model MegaDetector on 352 426 images generated from 159 different off- and on-trail camera traps in three study regions in Bavaria, Germany. This is to our knowledge, the largest dataset of classified human and wildlife activities tested for automated object detection since so far mostly image classifiers were used. Here, we used an object detection model as image classifier to overcome the problem of having different backgrounds due to varying camera trap sites and environmental conditions. We specifically evaluate (a) the accuracy of MegaDetector following Fennell et al. (2022) while discriminating between object class, detection confidence, study areas and camera trap sites and (b) the performance of MegaDetector in counting objects on camera trap images. In addition, we (c) identify systematic misclassifications of the object detection model and (d) test the detection results for long-term time series of activity patterns as well as diurnal activity patterns, two typical analyses in wildlife and recreation ecology, which have yet only been addressed for wildlife species (Whytock et al., 2021).

Materials and Methods

Camera trap data and classification process

Camera trap images were collected in three areas in Bavaria: the Fichtelgebirge, the Veldensteiner Forst and the Bavarian Forest National Park (Fig. 1). In the nature park Fichtelgebirge, 72 879 images from 10 camera traps running from December 2019 to September 2020 were used for the study. For the Bavarian Forest National Park (BFNP), 269 051 images were used from 143 camera traps deployed from November 2020 to July 2021. In the nature park Veldensteiner Forst, six camera traps were deployed from July 2020 to January 2022 and recorded 10 496 images. In total, we used 352 426 images for this analysis, 229 100 contained humans and/or vehicles and 114 937 contained animals.

We used different camera trap models including Reconyx HyperFire2, CuddebackC2 and CuddebackG at on-trail as well as off-trail locations. The cameras were set to continuously record 1–3 images per trigger with no delay in high-sensitivity mode. We checked the cameras every 4–8 weeks to replace image data storage and batteries. Camera traps were installed in a way that human privacy rights are conserved. We either directed the camera traps to the lower body parts only, manually added a blurring filter to the camera trap and/or blackened the bounding boxes around human detections prior to manual classification in order to meet privacy regulations. The latter was done using the artificial intelligence software Amazon Rekognition (Amazon Web Services, 2023), since an automated anonymization of humans was not available via MegaDetector yet.

Wildlife ecologists and students manually classified the image data into detections of 19 wildlife species, human outdoor activities and vehicles and identified empty images (Table S1). The counts per class were recorded for each image using the tagging software xnviewmp (version 1.0) and the server-based software TRAPPER (Bubnicki et al., 2016). The same camera trap images were classified into animals, persons, vehicles and empty images by the object detection algorithm MegaDetector (version 4.1, Beery et al., 2019), running locally on a desktop (Intel i9-9000 series CPU, 64 GB RAM and an NVIDIA Quadro RTX 4000 GPU).

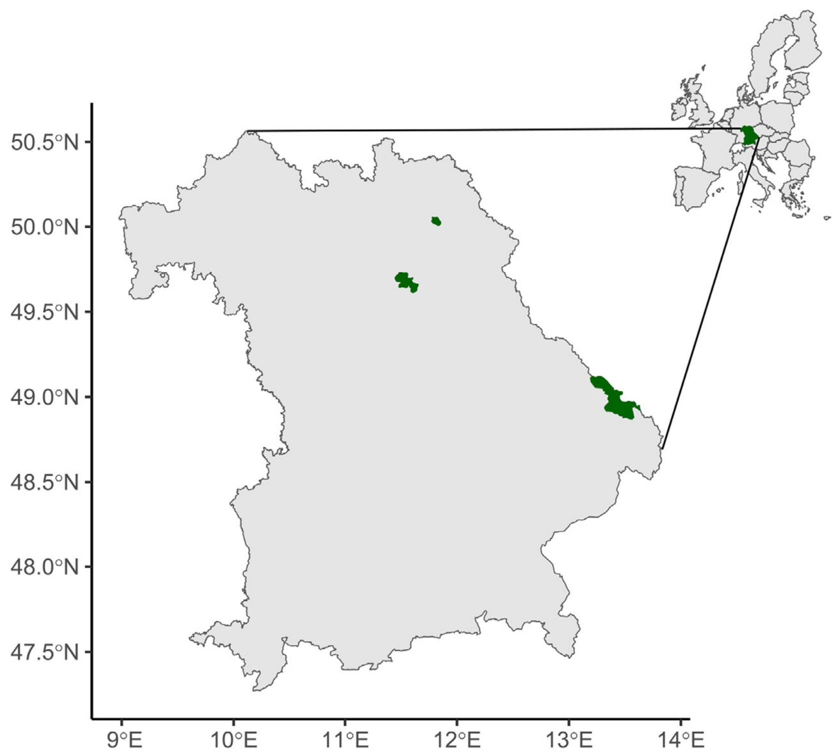

Figure 1. Map of Bavaria in Germany with the three study areas Bavarian Forest National Park, Veldensteiner Forst and Fichtelgebirge (from bottom to top).

Analysis

We determined the performance of the object detection model by comparing the results with the manual classification. First, we compared each image file, summarized results into a confusion matrix of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) and calculated accuracy $(TP + TN)/\ (TP + TN + FP + FN)$, precision $(TP)/(TP + FP)$, recall $(TP)/(TP + FN)$ and specificity $(TN)/(TN + FP)$ (Fennell et al., 2022) for the object classes person, animal and vehicle. Here, accuracy is the most suitable metric to report the performance of MegaDetector in detecting and counting the target object classes per frame, which is why we focus on reporting our results for accuracy. For other analysis, however, and for enabling future comparisons, the results for precision, recall and specificity are reported in the supplements (Figure S1; Table S2). We visualized the performance of the object detection model using boxplots in order to identify variances amongst study regions and camera trap sites. The performance of the object detection model for counting objects per image frame was evaluated by calculating the difference in detected counts per object class between manual and automated detection. We further summed detections on a daily level and calculated Pearson correlation coefficient of manual and automated detections. To test the suitability of automated object detection results for ecological analyses, we calculated kernel densities of the activities of animals and persons and determined the overlap coefficient of those two classes (Niedballa et al., 2016). All analyses were performed for the confidence thresholds 50, 60, 70, 80, 90 and 95 per cent to evaluate the best confidence threshold in dependence of the object class and using R Statistical Software (version 4.2.1; R Core Team, 2022) and camtrapR R package (Niedballa et al., 2016).

Results

Overall, the object detection model accurately detected the object classes animal (96% for the 95 per cent confidence threshold), person (93.8% for a 95 per cent confidence threshold) and vehicle (99.3% for a 95 per cent confidence threshold). Results varied across study areas with the lowest detection accuracy being 84.3% for animals at Veldensteiner Forst (95 per cent confidence threshold) and the highest accuracy with 99.7% for vehicles in the Bavarian Forest (95 per cent confidence threshold). Similarly, detection accuracy highly differed within study areas (Fig. 2; Table S2). In the Fichtelgebirge, model accuracy for detecting persons ranged between 52% (95 per cent confidence) and 97.9% (50 per cent confidence), whereas detection accuracy of animals and vehicles was more consistent across camera trap sites (animals: 86.5–98.6%; vehicles: 83.9–99.8%). Likewise, the detection accuracy for animals and vehicles was slightly more consistent than for persons in the Bavarian Forest (animal: 77.5–99.6%; person: 68.9–100%; vehicles: 90.5–100%). Contrarily, the variance of the detection accuracy of animals (72.2–96.6%) was greater compared to persons and vehicles (persons: 78.7–95.4%; vehicle: 94–100%) in the Veldensteiner Forst. Across all images, a 95 per cent confidence threshold has the highest detection accuracy for the object classes animal and vehicle, whereas for persons,

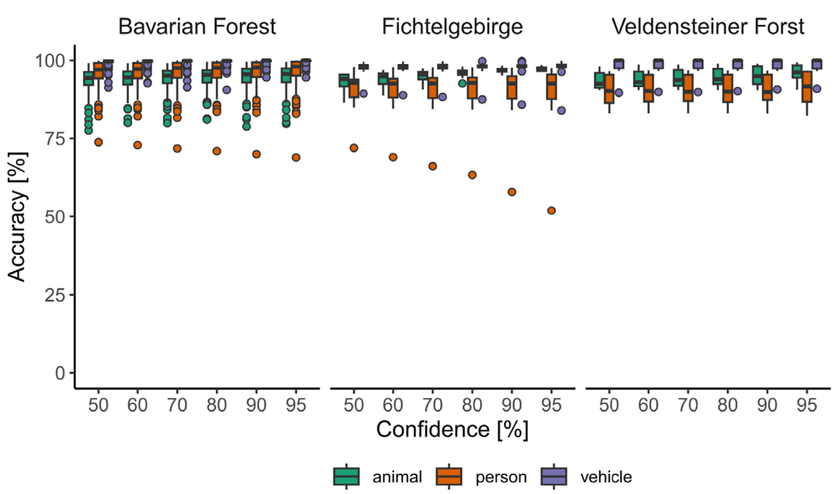

Figure 2. Object detection model accuracy for detecting the object classes animal (green), person (orange) and vehicle (violet). The accuracy, which is the proportion of correctly detected images, is shown per camera trap location and study area and calculated for the confidence thresholds 50, 60, 70, 80, 90 and 95 per cent.

the highest accuracy was achieved with a 50 per cent confidence threshold (93.98% accuracy, 93.83% accuracy for 95 per cent confidence) (Fig. 2). Similarly, precision for all object classes was highest for a 95 per cent confidence threshold and reached 97.4% for animals, 99.6% for persons and 94.1% for vehicles, whereas recall was highest at a 50 per cent confidence threshold (animal: 91.2%; person: 91.4%; vehicle: 89.7%) (Figure S1; Table S2).

Overall, detection accuracy slightly differed between images taken at night and daylight. Given a 95 per cent confidence threshold, accuracy was 96.8% at day and 93.4% at night for animals. Performance in detecting persons at day was lower (92.0% accuracy) than during the night (98.9% accuracy), whereas there was no difference in detection performance between day and night time for vehicles (99.0%; 99.9%). Compared to manual detection, the object detection model was on average barely underestimating the object counts for animals (−0.01). However, for individual images, the count difference between the automated and manual detection was high, for example, ranged between −29 and +7 for an image with a flock of chaffinches. The best count was achieved at a confidence threshold of 90 per cent (Fig. 3). For persons, the object detection model underestimated counts by on average −0.05 counts per image (min = −8; max = +9) and performed best at a low confidence threshold of 50 per cent. Vehicle object counts were only slightly underestimated by −0.01 (−6 to +4), and the results marginally varied between confidence thresholds (Fig. 3).

The object detection model continually detected some patterns on camera trap images inaccurately. For instance, false positive detections commonly occurred on blurry pictures when the camera trap was moved for maintenance, when vegetation is moving in sunlight as well as branches and cones in the shape of animals (Fig. 4A). However, detections were similarly identified as false positives when the manual detection was incorrect (Fig. 4A). The detection model frequently misclassified fast moving objects which are blurry, such as running shoes or bicycle wheels, or a head of hair close to the camera trap as animals and smartphones or backpacks were detected as vehicles (Fig. 4B). Particularly persons on a bicycle were repeatedly not detected by the object detection model when only small parts of the human body are visible on the image, resulting in false negatives (Fig. 4C). Additionally, one bicycle was occasionally detected as two vehicles (Fig. 4C).

Daily activity patterns of animals, persons and vehicles detected by the object detection model were overall highly correlated with daily manual detections (Pearson's $r = 0.996$, $p < 0.001$). While daily detections were slightly over- or underestimated depending on the confidence threshold and object class (Fig. 5), the distribution of daily detections of automated (95 per cent confidence) and manual classification overlapped by 95.4%.

Analyses of the diurnal activity pattern as kernel density curves of all manually detected animals and persons were of high conformity with the patterns of automatically (95 per cent confidence) detected animal and human diurnal activities (Fig. 6). Consequently, the activity overlap coefficients of both approaches were almost equivalent (manual: 0.42, automatically: 0.41).

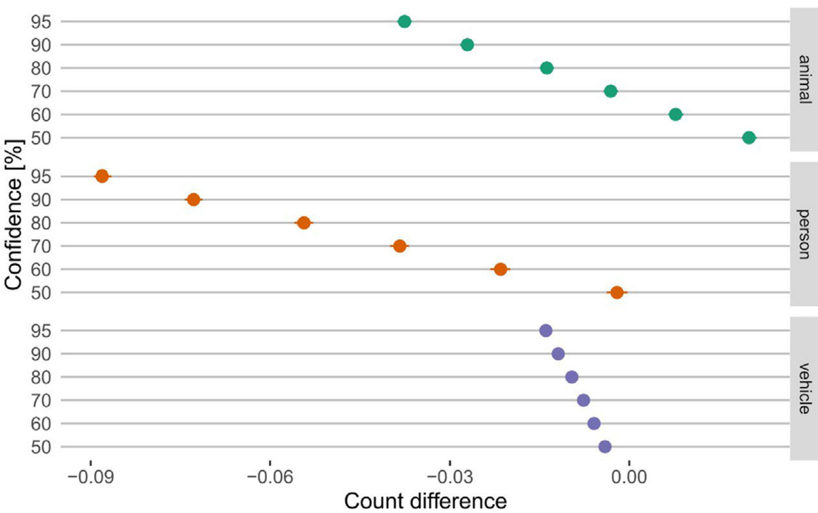

Figure 3. The average difference and confidence intervals in counts per detection of the object detection model and the manual detection for the object classes animal (green), person (orange) and vehicle (violet) and the confidence thresholds 50, 60, 70, 80, 90 and 95 per cent across all camera trap locations and study areas.

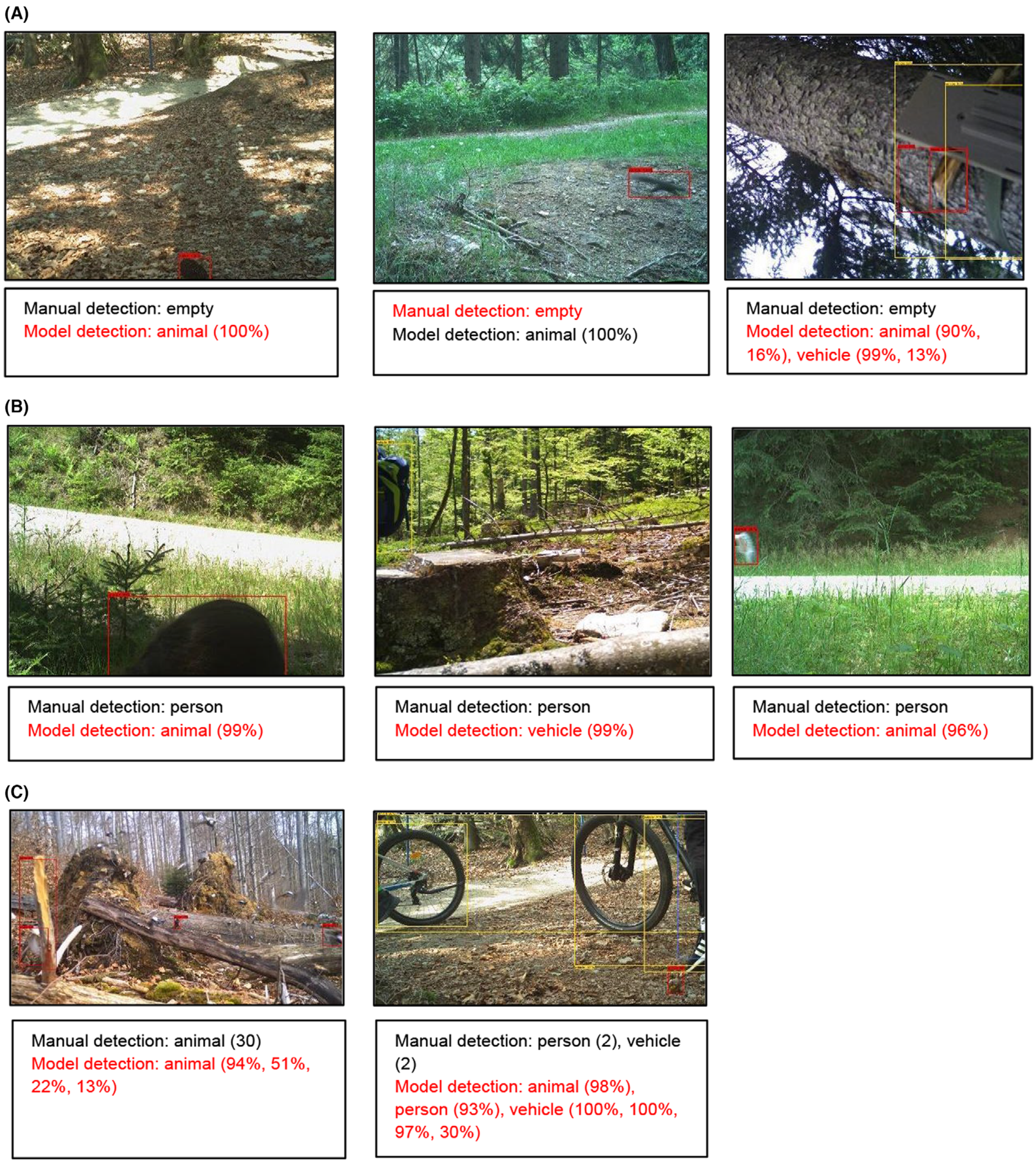

Figure 4. Examples of incorrectly labelled images with (A) false positive detections of the object detection model or the human classifier, (B) false classes detected by the object detection model and (C) false counts of the detected classes by the object detection model.

Discussion

We show that the used open-source object detection model is an effective tool for analysing camera trap data of wildlife and human activities. The overall detection performance is highly accurate for the object classes animal, person and vehicle, indicating great improvements in processing large image data amounts by reducing time and

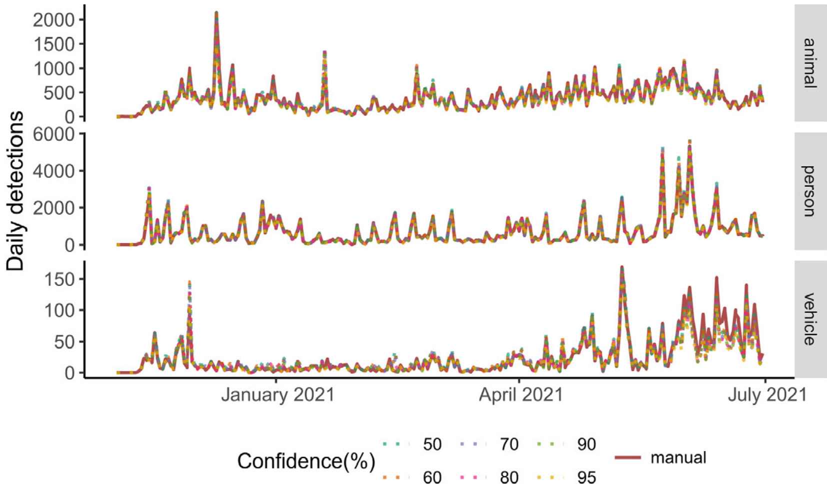

Figure 5. Daily detections of manually and automatically (model confidence threshold 50, 60, 70, 80, 90 and 95 per cent) detected object classes animal, person and vehicle on the camera trap image data in the Bavarian Forest National Park.

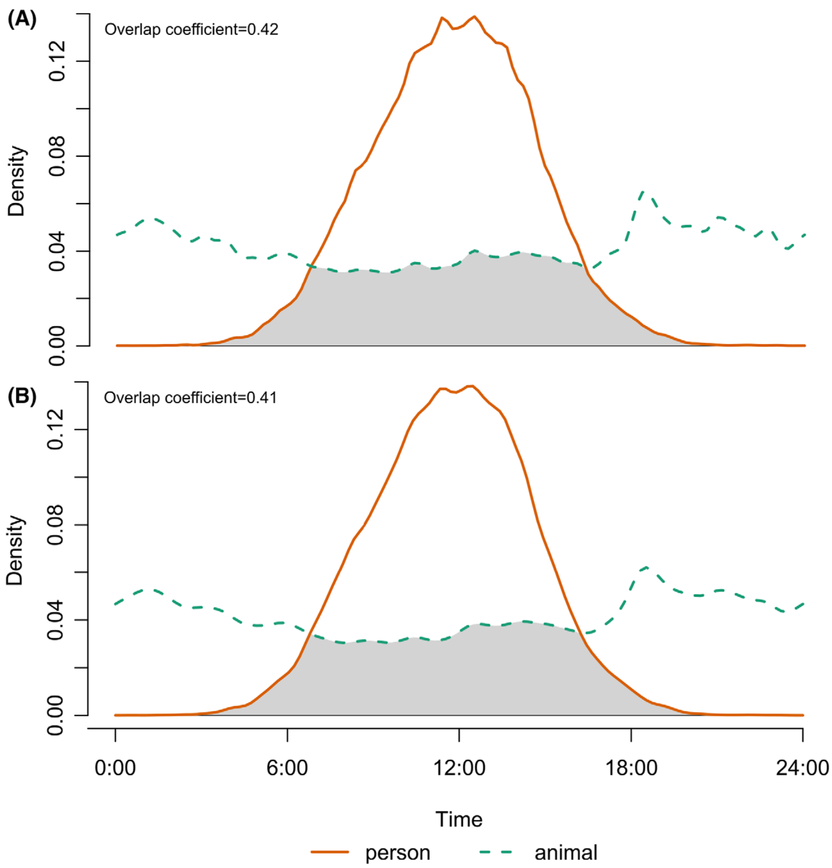

Figure 6. Kernel densities of diurnal activities of (A) manually and (B) automatically (95 per cent confidence threshold) detected object classes animal and person on all camera trap image data of the study and the underlying activity overlap coefficients (Niedballa et al., 2016).

cost efforts. Although misclassification of objects other than animals, persons or vehicles as well as of repeatedly generated blurry images during maintenance of camera traps may result in reduced detection performance, we found that they occur systematically and are hence excludable. Therefore, we consider object detection models for assisting in the image classification process a highly beneficial tool for studies in recreation or wildlife ecology.

In general, object detection models allow a strong reduction in manual labour while classification of objects is highly accurate and independent from geographical regions (Beery et al., 2019; Fennell et al., 2022; Norouzzadeh et al., 2018; Tabak et al., 2022). Empty images or images with specific objects can be excluded rapidly and hence, the time for the image classification process can be significantly reduced (Beery et al., 2019). Likewise, our results indicate that within cross-regional study designs with differing camera trap models and technical equipment as well as slightly differing data handling, the overall detection accuracy remains high, while accelerating classification speed from 200–500 images to 15 000–20 000 images per hour. This proves the applicability of detection models for realistic scenarios of camera trap studies in wildlife and recreation ecology. Particularly for recreation ecological research questions, where human activities are of similar interest as wildlife activity, using automated object detection models, such as MegaDetector is beneficial. It enables the processing of large amounts of images collected on recreational trails in a short time (Fennell et al., 2022), while keeping compliance with data protection regulations (Sharma et al., 2020).

Moreover, if the types of recreational activities commonly carried out in a region are known, detections based on object detection models have the potential to additionally differentiate these activities from each other. For example, if hiking and biking are the most common activities, a differentiation into hiker and biker is possible as detected vehicles and persons on the same image should refer to biking while images with exclusively persons being detected should represent hikers. This is a strong advantage of using camera traps and automated object detection for investigating recreational activities in natural systems (Lupp et al., 2021), since other methods used for generating large spatiotemporal data of human activities such as pressure sensors or infrared sensors are limited in differentiating between activities (Cessford & Muhar, 2003; Staab et al., 2021). However, depending on the common human activity and the proximity of cameras to human settlements, images with detections of a person plus vehicle may need to be further investigated, since also objects such as buggies and wheelchairs were classified as person plus vehicle. If motorized activities are allowed, humans on scooters, quads, etc., add further complexity to interpreting such classification results. Likewise, images with detections of both animals and persons may indicate hikers with dogs but also horseback riders, in case the latter activity is allowed in the study region. For camera trap studies in wildlife ecology, detection models have the potential to drastically reduce the classification effort of, for example, large carnivores. Carnivores are known to preferably use linear features such as recreational trails (Bojarska et al., 2020; Newton et al., 2017), which is why camera traps are often placed along these linear features to increase capture rates, thereby unintentionally generating large amounts of image data with persons or vehicles. By employing the output of object detection models to filter these, in this example, unnecessary object classes and the vast majority of empty images, manual classification time can be reduced.

In addition to the significant acceleration of the classification process when using detection models, we also found that the MegaDetector detection model, by detecting each individual object in an image, generally performs well at counting objects on images. For more general assessments of human–wildlife interactions, the accuracy of detection models in detecting and counting animals, persons and vehicles is sufficient as it only slightly over- or underestimates counts. However, in rare events where, for instance, a blurry flock of birds is passing or where single individuals in a group of animals or humans are partly covered by others, counts are less precise due to difficulties in detection of and differentiation between objects. The assessment of more detailed spatiotemporal patterns of visitors or wildlife in natural areas (Arnberger & Hinterberger, 2003; Ladle et al., 2017) may require a more precise group size estimate. Furthermore, object detection models are often limited to broad categories (person, animal and vehicle), while wildlife studies usually address questions on the species level. In this case, we suggest applying a combined approach of initially using MegaDetector to filter out empty pictures, and for example, pictures of human activities, followed by either manual classification and counting of remaining wildlife images on the species level or using additional models for automated species or activity identification (Redmon et al., 2016; Rigoudy et al., 2022, Fig. 7). Depending on the specific research question, it might be relevant to additionally classify more detailed attributes such as sex, age or behaviour of the animal, which is not applicable yet when using open source models (Vélez et al., 2022). Additionally, it is worth mentioning that MegaDetector as well as other machine learning models are continuously improving with increasing training data, and therefore, the version (version 4) used in this study might not be the current one in future studies.

A combined approach of camera traps and automated object detection enables the long-term collection of human and wildlife presence–absence data on a large spatial scale. For assessing variation in activity patterns over time or species responses to habitat change or human disturbances, camera trap data is commonly analysed in time-series analysis or kernel densities of diurnal activities. Here, we prove that the classification results of the object detection model were extremely consistent with manual classifications and both approaches were highly correlated in these typical analyses. This is in line with the results for calculating species richness, activity patterns and occupancy of wildlife species when using a machine learning model for automated labelling (Whytock et al., 2021). For the analysis presented here, we suggest choosing a confidence threshold higher than 90 per cent, when high precision is needed for ecological studies. However, this leads to losses of detections when the object detection model was unconfident due to, for instance, suboptimal positioning of the object in the image. A strategy to reduce this deficiency would be to adjust the confidence threshold according to the study specific requirements to lower levels. Studies targeting, for example, species richness or occupancy may need to reduce the confidence threshold in order to achieve higher recall and not miss animal detections.

Particularly for cross-regional camera trap studies with large image datasets, the application of automated object detection models has the advantage to standardize classification processes to a certain extent. Manual classification of large datasets requires multiple observers, resulting in an inter-observer bias and misidentification of mammal species (Zett et al., 2022). Such mistakes can be reduced by categorizing observations in certain classes using an automated object detection model prior to manual classification. In addition, automated blackboxing of humans on camera trap images to keep in compliance with data protection increases miscounting of individuals by human classifiers as one bounding box may cover multiple persons walking next to each other. Conversely, automated object detection models serve as a standardized method to protect the privacy of human individuals photographed by camera traps (Sharma et al., 2020), enabling the use of the images of human activities for recreational purposes studies. Using detection model classification results, images with humans are easily removed from the dataset to protect privacy rights. Besides, this study shows that misclassification mostly results from blurry parts of a human without faces recorded on the image. If a further investigation of images classified by an object detection model is needed, for example, to determine outdoor activities in more detail, we recommend the use of filters such as instance segmentation,

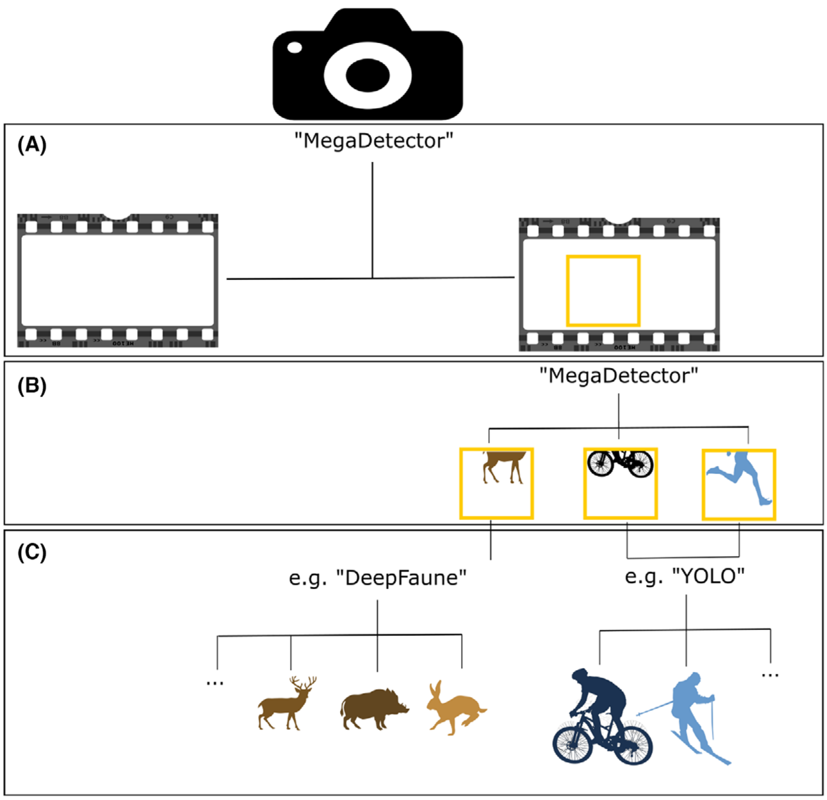

Figure 7. Hierarchal workflow for automated visitor and wildlife detection on camera trap images by using different models for the classification levels (A) elimination of empty images (B) detection of the object classes animal, vehicle and person using MegaDetector (Beery et al., 2019) and (C) classify into wildlife species and human activities using, for instance, DeepFaune or YOLO (Desai et al., 2022; Redmon et al., 2016; Rigoudy et al., 2022).

which only blur humans in pictures instead of completely covering the part of the picture the human bounding box is in (Desai et al., 2022; Kirillov et al., 2020). Further advances in automated image processing may additionally enable the identification of similar individuals on a sequence of images using object tracking (Wojke et al., 2017).

While the bare use of automated object detection models for the classification of large camera trap data is still limited to broader object classes, we conclude that the integration of automated object detection in a hierarchical classification approach is highly beneficial. Beside a great acceleration in image processing speed and a subsequent reduction of costs involved, it has the potential to standardize the classification process across studies and keeps the compliance with data protection regulations. For detailed analyses in recreation and wildlife ecology, where precise species classification and counting of individuals as well as additional attribute identification is necessary, a subsequent manual classification is still unavoidable. However, as data collection is rapidly increasing worldwide and likewise, software as well as hardware is improving quickly (Glover-Kapfer et al., 2019; Norouzzadeh et al., 2020), reliable models for more detailed classification and broad use in ecology are certain.

Acknowledgments

The Bavarian State Ministry for the Environment and Consumer Protection contributed to this work by funding the project 'Integrative evaluations of the effects of recreational use on wildlife as a basis for evidence-based visitor management', carried out in the Bavarian Forest National Park. The study in the Veldenstein Forest was financed by the Bavarian State Ministry of Agriculture and Forestry (grant I043). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 491183248. Funded by the Open Access Publishing Fund of the University of Bayreuth. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

V.M., A.P., W.P., M.H. and M.S. conceived the study; V.M., A.P., H.E., H.N. and G.M. designed the methodology, generated and processed the data and V.M. analysed and visualized the data, with helpful support by A.P.; V.M. led the writing with main contribution of A.P. and helpful comments and edits from all co-authors.

References

Amazon Web Services. (2023) *Amazon Rekognition* [software]. Available at: https://aws.amazon.com/de/rekognition/ (accessed 22 August 2023).

Arnberger, A. & Hinterberger, B. (2003) Visitor monitoring methods for managing public use pressures in the Danube Floodplains National Park, Austria. *Journal for Nature Conservation*, **11**(4), 260–267. Available from: https://doi.org/10.1078/1617-1381-00057

Beery, S., Morris, D., Yang, S., Simon, M., Norouzzadeh, A. & Joshi, N. (2019) Efficient pipeline for automating species ID in new camera trap projects. *Biodiversity Information Science and Standards*, **3**, e37222. Available from: https://doi.org/10.3897/biss.3.37222

Bojarska, K., Sulich, J., Bachmann, S., Okarma, H., Theuerkauf, J. & Gula, R. (2020) Opportunity and peril: how wolves use a dense network of forest roads. *Mammalian Biology*, **100**(2), 203–211. Available from: https://doi.org/10.1007/s42991-020-00014-0

Bubnicki, J.W., Churski, M. & Kuijper, D.P.J. (2016) TRAPPER: an open source web-based application to manage camera trapping projects. *Methods in Ecology and Evolution*, **7**(10), 1209–1216. Available from: https://doi.org/10.1111/2041-210X.12571

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T. et al. (2015) Review: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, **52**(3), 675–685. Available from: https://doi.org/10.1111/1365-2664.12432

Caravaggi, A., Banks, P.B., Burton, A.C., Finlay, C.M.V., Haswell, P.M., Hayward, M.W. et al. (2017) A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, **3**(3), 109–122. Available from: https://doi.org/10.1002/rse2.48

Cessford, G. & Muhar, A. (2003) Monitoring options for visitor numbers in national parks and natural areas. *Journal for Nature Conservation*, **11**, 240–250.

Desai, K., Misra, I. & Johnson, J. (2022) *Scaling up instance segmentation using approximately localized phrases*. Available at: https://bmvc2022.mpi-inf.mpg.de/0648.pdf (accessed 22 August 2023).

Falzon, G., Lawson, C., Cheung, K.-W., Vernes, K., Ballard, G.A., Fleming, P.J.S. et al. (2020) ClassifyMe: a field-scouting software for the identification of wildlife in camera trap images. *Animals*, **10**(1), 58. Available from: https://doi.org/10.3390/ani10010058

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. *Global Ecology and Conservation*, **35**, e02104. Available from: https://doi.org/10.1016/j.gecco.2022.e02104

Frank, B., Glikman, J.A. & Marchini, S. (Eds.). (2019) *Human-wildlife interactions: turning conflict into coexistence*. Cambridge, UK: Cambridge University Press.

Gaynor, K.M., Hojnowski, C.E., Carter, N.H. & Brashares, J.S. (2018) The influence of human disturbance on wildlife nocturnality. *Science*, **360**(6394), 1232–1235. Available from: https://doi.org/10.1126/science.aar7121

Glover-Kapfer, P., Soto-Navarro, C.A. & Wearn, O.R. (2019) Camera-trapping version 3.0: current constraints and future priorities for development. *Remote Sensing in Ecology and Conservation*, **5**(3), 209–223. Available from: https://doi.org/10.1002/rse2.106

Kirillov, A., Wu, Y., He, K. & Girshick, R. (2020) *PointRend: image segmentation as rendering*. (arXiv:1912.08193). *arXiv*. http://arxiv.org/abs/1912.08193

Ladle, A., Avgar, T., Wheatley, M. & Boyce, M.S. (2017) Predictive modelling of ecological patterns along linear-feature networks. *Methods in Ecology and Evolution*, **8**(3), 329–338. Available from: https://doi.org/10.1111/2041-210X.12660

Lupp, G., Kantelberg, V., Förster, B., Honert, C., Naumann, J., Markmann, T. et al. (2021) Visitor counting and monitoring in forests using camera traps: a case study from Bavaria (Southern Germany). *Land*, **10**(7), 736. Available from: https://doi.org/10.3390/land10070736

Miller, A.B., Leung, Y.-F. & Kays, R. (2017) Coupling visitor and wildlife monitoring in protected areas using camera traps. *Journal of Outdoor Recreation and Tourism*, **17**, 44–53. Available from: https://doi.org/10.1016/j.jort.2016.09.007

Naidoo, R. & Burton, A.C. (2020) Relative effects of recreational activities on a temperate terrestrial wildlife assemblage. *Conservation Science and Practice*, **2**(10), 271. Available from: https://doi.org/10.1111/csp2.271

Nellemann, C., Vistnes, I., Jordhøy, P., Støen, O.-G., Kaltenborn, B.P., Hanssen, F. et al. (2010) Effects of recreational cabins, trails and their removal for restoration of reindeer winter ranges. *Restoration Ecology*, **18**(6), 873–881. Available from: https://doi.org/10.1111/j.1526-100X.2009.00517.x

Newton, E.J., Patterson, B.R., Anderson, M.L., Rodgers, A.R., Vander Vennen, L.M. & Fryxell, J.M. (2017) Compensatory selection for roads over natural linear features by wolves in northern Ontario: implications for caribou conservation. *PLoS One*, **12**(11), e0186525. Available from: https://doi.org/10.1371/journal.pone.0186525

Niedballa, J., Sollmann, R., Courtiol, A. & Wilting, A. (2016) camtrapR: an R package for efficient camera trap data management. *Methods in Ecology and Evolution*, **7**(12), 1457–1462. Available from: https://doi.org/10.1111/2041-210X.12600

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N. & Clune, J. (2020) A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, **12**(1), 150–161. Available from: https://doi.org/10.1111/2041-210X.13504

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C. et al. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(25), E5716–E5725. Available from: https://doi.org/10.1073/pnas.1719367115

R Core Team. (2022) *R: a language and environment for statistical computing* [software]. R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (accessed 21 April 2023).

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016) You only look once: unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

Reilly, M.L., Tobler, M.W., Sonderegger, D.L. & Beier, P. (2017) Spatial and temporal response of wildlife to recreational activities in the San Francisco Bay ecoregion. *Biological Conservation*, **207**, 117–126. Available from: https://doi.org/10.1016/j.biocon.2016.11.003

Rigoudy, N., Benyoub, A., Besnard, A., Birck, C., Bollet, Y., Bunz, Y. et al. (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

Rowcliffe, J.M., Kays, R., Kranstauber, B., Carbone, C. & Jansen, P.A. (2014) Quantifying levels of animal activity using camera trap data. *Methods in Ecology and Evolution*, **5**(11), 1170–1179. Available from: https://doi.org/10.1111/2041-210X.12278

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

Sharma, K., Fiechter, M., George, T., Young, J., Alexander, J.S., Bijoor, A. et al. (2020) Conservation and people: towards an ethical code of conduct for the use of camera traps in wildlife research. *Ecological Solutions and Evidence*, **1**(2), 12033. Available from: https://doi.org/10.1002/2688-8319.12033

Shively, K.J., Alldredge, A.W. & Phillips, G.E. (2005) Elk reproductive response to removal of calving season disturbance by humans. *Journal of Wildlife Management*, **69**(3), 1073–1080. Available from: https://doi.org/10.2193/0022-541X(2005)069[1073:ERRTRO]2.0.CO;2

Staab, J., Udas, E., Mayer, M., Taubenböck, H. & Job, H. (2021) Comparing established visitor monitoring approaches with triggered trail camera images and machine learning based computer vision. *Journal of Outdoor Recreation and Tourism*, **35**, 100387. Available from: https://doi.org/10.1016/j.jort.2021.100387

Tabak, M.A., Falbel, D., Hamzeh, T., Brook, R.K., Goolsby, J.A., Zoromski, L.D. et al. (2022) CameraTrapDetectoR: automatically detect, classify, and count animals in camera trap images using artificial intelligence. [preprint]. *bioRXiv*. https://doi.org/10.1101/2022.02.07.479461

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P. et al. (2019) Machine learning to classify animal species in camera trap images: applications in ecology. *Methods in Ecology and Evolution*, **10**(4), 585–590. Available from: https://doi.org/10.1111/2041-210X.13120

Tablado, Z. & Jenni, L. (2017) Determinants of uncertainty in wildlife responses to human disturbance: modulators of wildlife response to recreation. *Biological Reviews*, **92**(1), 216–233. Available from: https://doi.org/10.1111/brv.12224

Tucker, M.A., Böhning-Gaese, K., Fagan, W.F., Fryxell, J.M., Van Moorter, B., Alberts, S.C. et al. (2018) Moving in the Anthropocene: global reductions in terrestrial mammalian movements. *Science*, **359**(6374), 466–469. Available from: https://doi.org/10.1126/science.aam9712

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:2202.02283). *arXiv*. http://arxiv.org/abs/2202.02283

Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Słupski, T., Koumba Pambo, A.F., Rogala, M. et al. (2021) Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods in Ecology and Evolution*, **12**(6), 1080–1092. Available from: https://doi.org/10.1111/2041-210X.13576

Wilson, M.W., Ridlon, A.D., Gaynor, K.M., Gaines, S.D., Stier, A.C. & Halpern, B.S. (2020) Ecological impacts of human-induced animal behaviour change. *Ecology Letters*, **23**(10), 1522–1536. Available from: https://doi.org/10.1111/ele.13571

Wojke, N., Bewley, A. & Paulus, D. (2017) Simple online and realtime tracking with a deep association metric (arXiv:1703.07402). *arXiv*. http://arxiv.org/abs/1703.07402

Zett, T., Stratford, K.J. & Weise, F.J. (2022) Inter-observer variance and agreement of wildlife information extracted from camera trap images. *Biodiversity and Conservation*, **31**(12), 3019–3037. Available from: https://doi.org/10.1007/s10531-022-02472-z

Zimmermann, A., Johnson, P., de Barros, A.E., Inskip, C., Amit, R., Soto, E.C. et al. (2021) Every case is different: cautionary insights about generalisations in human-wildlife conflict from a range-wide study of people and jaguars. *Biological Conservation*, **260**, 109185. Available from: https://doi.org/10.1016/j.biocon.2021.109185

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Classes and subclasses of the classification processes.
Table S2. Accuracy, precision, recall and specificity of the automated classification using MegaDetector for the object classes animal, person and vehicle and for the confidence thresholds 50, 60, 70, 80, 90 and 95 per cent.
Figure S1. Results for (A) precision and (B) recall of the automated classification via MegaDetector for the object classes animal (green), person (orange) and vehicle (violet and calculated for the confidence thresholds 50, 60, 70, 80, 90 and 95 per cent.

