Remote Sensing in Ecology and Conservation
Open Access
ZSL

RESEARCH ARTICLE

Deep learning-based ecological analysis of camera trap images is impacted by training data quality and quantity

Peggy A. Bevan[1,2], Omiros Pantazis[1,3], Holly A.I. Pringle[1], Guilherme Braga Ferreira[1,4], Daniel J. Ingram[5], Emily K. Madsen[1,6], Liam Thomas[1,7], Dol Raj Thanet[8], Thakur Silwal[8], Santosh Rayamajhi[8], Gabriel J. Brostow[3], Oisin Mac Aodha[9] & Kate E. Jones[1]

[1]Centre for Biodiversity and Environment Research (CBER), Department of Genetics, Evolution and Environment, University College London, London, UK
[2]Institute of Zoology, Zoological Society of London, London, UK
[3]Department of Computer Science, University College London, London, UK
[4]Instituto Biotrópicos, Diamantina, Brazil
[5]Durrell Institute of Conservation and Ecology (DICE), School of Natural Sciences, University of Kent, Canterbury, UK
[6]Wildlife Conservation Research Unit, Department of Biology, University of Oxford, Oxford, UK
[7]RSK Wilding, Gloucestershire, UK
[8]Institute of Forestry, Tribhuvan University, Kathmandu, Nepal
[9]School of Informatics, University of Edinburgh, Edinburgh, UK

Keywords
activity patterns, camera traps, computer vision, deep neural networks, ecological metrics, occupancy, species richness

Correspondence
Peggy A. Bevan, Centre for Biodiversity and Environment Research (CBER), Department of Genetics, Evolution and Environment, University College London, London, UK.
E-mail: peggy.bevan.17@ucl.ac.uk

Funding Information
This research was funded by WWF-UK as part of the Biome Health Project. Peggy Bevan acknowledges support by the Natural Environment Research Council (NERC), United Kingdom (Grant Ref.: NE/S007229/1). Daniel J. Ingram acknowledges support from UK Research and Innovation (Future Leaders Fellowship, Grant Ref.: MR/W006316/1).

Peggy A. Bevan and Omiros Pantazis contributed equally to this work.

Editor: Prof. Nathalie Pettorelli
Associate Editor: Dr. Margarita Mulero-Pazmany

Received: 16 April 2025; Revised: 2 December 2025; Accepted: 4 December 2025

Abstract

Large image collections generated from camera traps offer valuable insights into species richness, occupancy, and activity patterns, significantly aiding biodiversity monitoring. However, the manual processing of these data sets is time-consuming, hindering analytical processes. To address this, deep neural networks have been widely adopted to automate image labelling, but the impact of classification error on key ecological metrics remains unclear. Here, we analyze data from camera trap collections in an African savannah (82,300 labelled images, 47 species) and an Asian sub-tropical dry forest (40,308 labelled images, 29 species) to compare ecological metrics derived from expert-generated species identifications with those generated by deep-learning classification models. We specifically assess the impact of deep- learning model architecture, the proportion of label noise in the training data, and the size of the training data set on three key ecological metrics: species richness, occupancy, and activity patterns. We found that predictions of species richness derived from deep neural networks closely match those calculated from expert labels and remained resilient to up to 10% noise in the training data set (mis-labelled images) and a 50% reduction in the training data set size. We found that our choice of deep-learning model architecture (ResNet vs. ConvNext-T) or depth (ResNet18, 50, 101) did not impact predicted ecological metrics. In contrast, species-specific metrics were more sensitive; less common and visually similar species were disproportionately affected by a reduction in deep neural network accuracy, with consequences for occupancy and diel activity pattern estimates. To ensure the reliability of their findings, practitioners should prioritize creating large, clean training sets and account for class imbalance across species over exploring numerous deep-learning model architectures.

doi: 10.1002/rse2.70052

Introduction

To track progress toward conservation targets (CBD, 2022), there is a pressing need for the intensification of biodiversity monitoring efforts at a global scale (Affinito et al., 2025; Gonzalez et al., 2023; Scharlemann et al., 2020). Steps toward standardized, large-scale monitoring are being introduced through the use of passive monitoring sensors that can scale up data collection efforts (Pimm et al., 2015; Pringle et al., 2025; Stephenson, 2020). Passive biodiversity monitoring sensors like camera traps (e.g., Lee et al., 2024), acoustic monitoring devices (e.g., Sethi et al., 2024), and satellite imagery (e.g., Wu et al., 2023) have allowed researchers to expand their ecological surveys both temporally and spatially with lower field costs and minimal environmental disturbance (Browning et al., 2017). Furthermore, the use of such devices provides standardized and reproducible survey methods and data formats that facilitate collaboration across projects and a network of sensors, slowly forming a global monitoring system (Blount et al., 2021; Kays et al., 2020; Steenweg et al., 2017; Wall et al., 2014).

Specifically, the utilization of camera traps has been beneficial for monitoring medium to large bodied terrestrial animals, primarily mammals (Burton et al., 2015; Fisher, 2023). These autonomous, motion-activated cameras can be used to collect a variety of ecological metrics such as occupancy (MacKenzie et al., 2002), abundance (Karanth & Nichols, 1998; Rowcliffe et al., 2008), and activity levels (Rowcliffe et al., 2014), which can be used to investigate complex interactions between wildlife, the environment, and human activity (Barcelos et al., 2022; Lee et al., 2024; Parsons et al., 2022) and to monitor the success of conservation interventions (Ferreira et al., 2020; Ferreira et al., 2023; Tobler et al., 2015). However, a major limitation of camera trap surveys is the data processing bottleneck, as millions of images need labelling (Duggan et al., 2021; Thomson et al., 2018). This bottleneck causes substantial delays in translating camera trap images into information that can be used in conservation efforts (Merkle et al., 2019).

Machine learning (ML) can increase the efficiency of camera trap analysis and speed up the extraction of ecological information. Deep neural networks have been applied to camera trap and other image data to tackle wildlife monitoring tasks such as locating and identifying species (Miao et al., 2021; Tabak et al., 2019; Willi et al., 2018), counting individuals (Norouzzadeh et al., 2021), classifying behavior (Kholiavchenko et al., 2024; Norouzzadeh et al., 2018) and estimating occupancy (Whytock et al., 2021). The creation of a general animal detector (MegaDetector) (Beery et al., 2019) has significantly improved efficiency by filtering out empty images, drastically reducing the number of images to be labelled (Fennell et al., 2022; Penn et al., 2024). Other works have exploited the context that accompanies camera trap images to improve species classification directly (Beery et al., 2020) or to learn representations in an unsupervised manner to reduce the number of labels required for training (Pantazis et al., 2021). Despite the growing number of ecological projects that utilize deep neural networks for image classification, evaluation is typically performed through metrics such as the total, or species level, classification accuracy. However, it remains untested whether classification accuracy of a deep-learning model is correlated with accuracy of downstream ecological metrics that the detection data are used for.

There is some evidence that ecological information obtained using deep neural networks is comparable to those generated by expert-labelled data. For example, Whytock et al. (2021) found that ML-generated species labels produced similar estimates of species richness, estimated occupancy and activity patterns as expert-labelled data. However, the data set focused on four species from central Africa, so the spatial and taxonomic generality of the findings are unclear. Practitioners developing deep-learning models for species classification must make a series of decisions with respect to the classification model and the training data set, often constrained by limited compute resources, time to annotate images, or ability to review existing labels. Therefore, even though it has been shown that models with deeper architectures with a higher number of parameters (He et al., 2016), large training data sets (Deng et al., 2009; Lin et al., 2014), and a low proportion of noise in the training data set (Rolnick et al., 2017; Sukhbaatar et al., 2015) benefit the classification accuracy of a deep neural network, it is unclear what the impact of such factors have on downstream ecological metrics.

Here, we analyze camera trap data from two ecosystems, African Savannah (Maasai Mara, Kenya) and Asian sub-tropical dry forest (Terai region, Nepal) to compare ecological metrics derived from expert-generated image labels with those generated by a trained deep neural network. We specifically assess the impact of neural network model architecture and depth, training data set size, and proportion of noise in the training data set on producing three key ecological metrics: species richness, occupancy, and activity patterns. It is expected that as these manipulations reduce the classification accuracy, the resulting ecological metrics will deviate further from those produced from expert-labelled data. We expect there may be some robustness in species richness and occupancy, as these only require one positive detection per survey occasion to contribute to the metric. However, activity

patterns may be impacted more strongly by a reduction in model accuracy due to the high temporal resolution in the underlying detections. We also explore the relationship between conventional ML evaluation metrics (Top-1 Accuracy, Precision, Recall, and F1 score) and accuracy of ecological metrics. Given the shortage of time and resources typically associated with a conservation project, it is not realistic for practitioners to optimize for every parameter. Through our analysis, we aim to shed light on the corresponding impact such factors have on the accuracy of downstream ecological analysis to aid practitioners in their decision making.

Methods

Camera trap data

We collected camera trap data from two ecosystems, African Savannah (Maasai Mara, Kenya) and sup-tropical dry forest (Terai region, Nepal). The field sites were part of the Biome Health Project, a field-based study system investigating the impact of human pressures and conservation interventions on biodiversity (https://www.biomehealthproject.com/). Each field site covers a gradient of anthropogenic pressure, but the type of pressure varies between ecosystems (Ingram et al., 2021). Both survey sites were set up using the same survey design. At each field site, un-baited Browning Dark Ops 2017 cameras were deployed evenly across a grid of 2 km^2 cells. A single camera was placed as close as possible to the centroid of each survey grid cell and were not biased towards trails or roads. Cameras were attached to a tree or a post at a height of ca. 50 cm, were operational 24 h/day with a 1 s delay between sequential triggers.

Maasai Mara camera traps (MMCT)

Data were collected from 176 camera traps deployed in four protected areas in the Maasai Mara, south-western Kenya: the Mara Triangle (MT), Mara North Conservancy (MN), Olare-Motorogi Conservancy (OMC), and the Naboisho Conservancy (NB), which each have different restrictions on livestock grazing and human activities. The data were collected throughout October and November 2018 and contains images from 47 species, or groups of species. The data used in this analysis were collected between October 5th and November 29th, 2018 and have previously been published in (Connolly et al., 2025). Human infrastructure data were obtained from Klaassen and Broekhuis (2018), and used to calculate the shortest distance from each camera trap station to human development, including settlements, bomas, towns, dams, and agriculture (Connolly et al., 2025).

Bardia camera traps (BCT)

148 cameras were deployed across three contiguous areas under different land management regimes in south-western Nepal: Bardia National Park (NP), the Buffer Zone (BZ), and outside the Buffer Zone (OBZ). These three areas vary in the level of restriction of human activities and development. The survey area therefore covers a gradient of pressure in the form of increasing habitat fragmentation, human density and agricultural activities. The data used in this analysis were collected between February 13 and April 16, 2019 and have previously been published in Ferreira et al. (2023).

Data labelling

We labelled both camera trap data sets by identifying species in each image using the Visual Object Tagging Tool (VOTT) (Microsoft, 2023). Before labelling, images were systematically sampled by using a set time interval of five minutes for MMCT and one minute for BCT, to avoid labelling the same event multiple times (Connolly et al., 2025; Ferreira et al., 2023). The time intervals differed between data sets due to environmental differences between the two ecosystems; the Masai Mara is dominated by large herds of herbivores which consistently triggered the camera, creating an image data set an order magnitude larger than the BCT data set for the same survey effort. During labelling, bounding boxes were drawn around each animal, vehicle, or human present in a photo. Images that were hard to identify were shared amongst the labelling team for a second opinion. After labelling, tagging accuracy for each species was checked by randomly sampling 10% of images per species, and any species with poor sampling accuracy in the sample (>3% error rate) were entirely relabelled. This ensured that both manually labelled data sets are highly accurate and contained minimal errors. To account for the fact that some species were under-represented within our collected data, relevant and visually similar species were grouped together where necessary, resulting in a list of labels that consists of either species or species groups (Table S1; Table S2).

For deep neural network training, the labelled images from each data set were split into subsets for model training, model validation and testing model performance. The data set was split temporally, not accounting for class, to ensure each subset had even spatial coverage across the survey area. Due to the shorter collection times of these data sets (2 months for MMCT; 1 month for BCT), seasonality did not need to be considered when creating these subsets. The MMCT data set was split into 53,102 images for model training, 5879 images for

validation, and 23,319 images for testing model performance. The BCT data set was split into 28,210 training, 3119 validation, and 8979 test images. For the final ecological analyses, we performed a series of filtering steps by first applying a minimum threshold of 20 expert labels per species in the test data set, as this allowed a reasonable classification accuracy to be quantified. We also removed domestic species, birds and any species groups that contained combinations of visually similar species (e.g., small cat category in BCT contained domestic cat and jungle cat). This resulted in 20 mammal species from the MMCT data set and 8 mammal species from the BCT data set (Table S3).

Ecological analysis

Species richness

Species richness was measured as the count of wild species observed at each camera trap location (a single camera trap deployment) over the entire survey period, using species detections from either the expert-generated labels or labels predicted by a deep neural network.

Occupancy

We adopted a multi-species occupancy framework to estimate occupancy while accounting for imperfect detection (Dorazio et al., 2006). Given the differences between the two study areas, we implemented slightly different occupancy models for the MMCT and BCT data sets (see Supplementary Text 1 and 2 for model specifications). To quantify the impact of deep neural network-based image classification on ecological responses, we investigated the effect of variables that had a strong influence on occupancy according to the model results: shortest distance to human infrastructure in the MMCT data set and management regime in BCT data set (following Ferreira et al., 2023). We ran both occupancy models on the species detections from either expert-generated labels or labels predicted by a deep neural network and extracted the species-specific model coefficients with 95% credible intervals from the posterior distributions.

The model coefficients represent the effect of a variable on occupancy, *i.e.*, the response of zebra occupancy to distance from human infrastructure. For the MMCT occupancy model, we extracted the raw coefficients of distance to human infrastructure as the response, as this is a continuous metric. For the BCT occupancy model, "management regime" is a categorical variable, so we calculated the difference in occupancy probability between national park (NP) and outside buffer zone (OBZ) as the response to changing management regime.

Activity patterns

We estimated the diel activity pattern of each species by fitting a circular kernel density function using the activity R package (version 1.3.4) (Rowcliffe, 2023). The kernel density function calculates the probability a species is active at each moment over a 24-hour period. To avoid large biases in estimates, only species that had ≥20 detections were included in the analysis (Rowcliffe et al., 2014). For each species, we calculated activity patterns from detections from expert-labelled data and the deep neural network, and compared them by calculating the bootstrapped overlap coefficient of the two activity patterns using the overlap R package (version 0.3.4) (Meredith & Ridout, 2023). The resulting overlap coefficient ranges from 0 to 1, where a value of 1 means perfect overlap between the two activity patterns. In this case, a higher overlap value indicates high ecological accuracy of the deep neural network-generated labels when compared to the expert labelled data.

Deep neural network experiments

To investigate the impact of deep neural networks on downstream ecological metrics, three experiments were run, each of which manipulated a different aspect of the training pipeline. We varied the underlying model architecture, the size of the training set, and the proportion of noise (incorrect labels) within the data set. Except for the model architecture experiment, the baseline model utilized for each experiment is a ResNet50 CNN (He et al., 2016). This model is commonly used as a baseline in machine-learning experiments. All experimental deep neural network models were sourced from the PyTorch library (Paszke et al., 2019). Across all experiments we utilized transfer learning (Yosinski et al., 2014), where neural network models were initialized from weights obtained via pre-training on ImageNet (Deng et al., 2009). This approach has demonstrated benefits for biodiversity monitoring by improving model accuracy (Willi et al., 2018). The model training was conducted on crops of each animal image (determined by the bounding box drawn in the labelling process) given that this approach has shown to benefit model accuracy (Beery et al., 2018; Gadot et al., 2024; Norman et al., 2023). We trained the classifier on all classes from the training set (Table S1; Table S2), but we only make predictions for the subset of classes that belong to the list of species on which the ecological analysis is conducted (Table S3). For evaluation of trained deep neural networks, we predicted labels on a held-out test set and applied a 70% confidence threshold across all experiments, as filtering out uncertain

labels has been shown to improve the robustness of the resulting ecological analysis (Whytock et al., 2021). The deep neural network predictions are then treated in the same way as expert-labelled data sets and used to calculate the ecological metrics described above as well as conventional ML evaluation metrics.

Impact of deep-learning classification model

To examine the impact that model architecture and depth have on downstream ecological metrics, four model types were compared. These were three ResNet (He et al., 2016) models with varying depth, ResNet18, 50, and 101. We chose to use the ResNet architecture because of its widespread use in computer vision research on camera trap data (Beery et al., 2018; Willi et al., 2018). The inclusion of three ResNet models at multiple depths allows us to explore the impact of using deeper models. In addition, a ConvNeXt-T model was used. This CNN uses a modified version of a ResNet architecture that takes inspiration from state-of-the-art vision transformer (ViT) models to achieve a higher performance with almost half the number of parameters as ResNet101 (Liu et al., 2022). The ConvNeXt-T model has similar performance to many ViT variants with similar parameter counts (Pucci et al., 2023; Vishniakov et al., 2024).

Impact of training set label noise

To investigate the impact of label noise (incorrect labels) on downstream ecological metrics, we created six versions of each training set (MMCT and BCT) with varying levels of label error, from 1% to 50% of examples mis-labelled. For a realistic simulation of label errors within each species, the iNaturalist citizen science platform (iNaturalist, 2023) was used to retrieve the three species that were most commonly misidentified as the original species that also exist in our data (Tables S4 and S5). In the case that three species were not available from iNaturalist, another species from the data set was randomly selected.

Impact of training set size

To investigate the impact of training set size on downstream ecological metrics, seven versions of the training set were created for each data set (MMCT and BCT) where the number of labels for each species was varied, from 100% of the original training set to 1% of the original training set.

Correlation between machine-learning evaluation and ecological metrics

To describe the relationship between deep neural network accuracy and ecological metric accuracy, we measured the correlation between classification error and ecological accuracy. To quantify "ecological accuracy," we took the absolute difference between the species-level occupancy coefficients measured from expert-generated labels and deep neural network-generated labels. For activity patterns, 1 minus the species-level overlap value was used, so that for both metrics, a 0 value equates to perfect prediction, i.e., no deviation between expert-labelled data and classifier prediction. To quantify classification error, we use four metrics commonly utilized to evaluate deep neural network classification performance: Top-1 Accuracy.

(proportion of correct classifications among all images), Precision (proportion of true positives among all positive predictions), Recall (proportion of true positive predictions out of all actual positives, including false negatives), and F1 Score (harmonic mean of Precision and Recall). To test correlation with ecological metrics, we use the error rates of these metrics, i.e., Top-1 Error, Precision Error, Recall Error, and F1 Error, which are calculated as 1 − metric.

Results

Species richness

We found that species richness predicted from DL-generated labels was robust to different model architectures and model depths but was impacted by high levels of noise in the training data and reduction in training set size, particularly in the BCT data set which is relatively smaller (Fig. 1a,d). Each of the four DL models accurately estimated the number of species present at each location compared to expert-labelled data. Although the error was slightly higher in the BCT data set than the MMCT data set, this did not appear to be driven by any particular DL model. In contrast, when there is more than 10% label noise in the data set, species richness is underestimated (Fig. 1b,e). Reduced training set size also had an impact on underestimating species richness, but to different degrees between the data sets. For the MMCT data set, an impact of reduced training set on predicting species richness is clear at 10% of the original size (Fig. 1c). However, for the BCT data set, this impact is clear from 25% and below (Fig. 1f).

Occupancy modelling

The results of occupancy modelling showed that species' labels predicted from deep neural networks can recover some species-specific responses to environmental covariates, even with high levels of manipulation to the training set (Fig. 2a). However, responses of less common and visually similar species were not predicted consistently. In

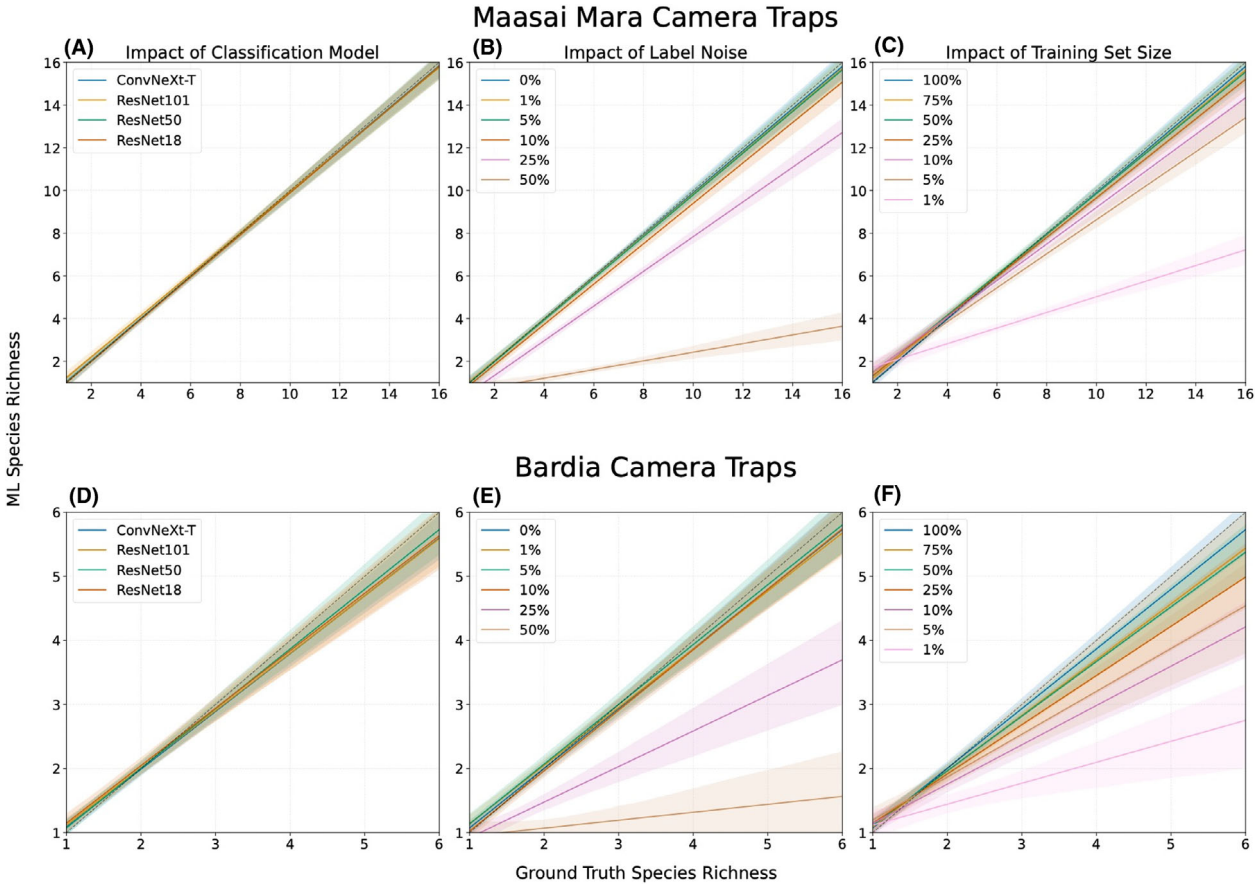

Figure 1. Observed species richness calculated with a variety of deep neural network architectures and training settings across two data sets. The Y-axis corresponds to the observed species richness per camera trap location using labels predicted by a deep neural network and the X-axis corresponds to species richness calculated using labels provided by experts. The lines in each plot correspond to a linear fit of the calculated richness across camera trap locations and a diagonal line that goes through the origin corresponds to the perfect match between the two axes. Shaded areas show 95% confidence intervals for the linear regression model.

the MMCT occupancy model derived from expert-generated labels, four species had posterior estimates indicating higher occupancy at sites further from human infrastructure, 13 species showed little evidence of an effect, and three species has estimates suggesting lower occupancy (Fig. 2a). Posterior distributions for elephant (n = 549 training images) and zebra (n = 4424 training images) consistently indicated positive effects across all DL models, even with up to 10% label noise and 10% of the original training set size. These effects were not evident for eland (n = 407 training images) and inconsistent for topi (n = 1440 training images). Some species had coefficients with 95% credible intervals overlapping zero in occupancy models from expert-generated labels (e.g., hippopotamus) yet showed strong posterior evidence for positive effects in many of the DL-predicted data sets. This pattern was more common for species with negative coefficients, such as vervet monkey (n = 382 training images) and Grant's gazelle (n = 543 training images). Grant's gazelle is visually similar to Thomson's gazelle (n = 5580 training images), which consistently showed posterior estimates indicating a negative effect across multiple manipulations, suggesting detections of these two species were conflated by our algorithms.

Occupancy results from the BCT data set were slightly more consistent overall. Posterior estimates from the expert-generated labels indicated three species had higher occupancy in NP (chital, grey langur and sambar), four species had little evidence of a difference between NP and OBZ, and one species had higher occupancy in OBZ (nilgai) (Fig. 2b). The responses of chital (n = 8715 training images) and grey langur (n = 391 training images) were consistently recovered in occupancy models from DL-generated labels, even with high levels of label noise and up to 5% of the original training set (Fig. 2b). However, the response of sambar (n = 265 training images)

Figure 2. Responses of occupancy to environmental pressures predicted from expert-generated (manual) labels and deep neural network-generated labels with various manipulations in the training pipeline. Plots show results of multi-species occupancy model predicting the species-specific occupancy response of mammals to (a) distance to human infrastructure in the Masai Mara, Kenya, where a positive effect means higher occupancy further away from infrastructure and (b) to management regime in Bardia, Nepal, where a positive effect means occupancy is higher in the National Park (NP) than in unprotected land (OBZ). Each tile shows the direction of the effect indicated by posterior estimates, and grey tiles show cases where the 95% credible interval overlapped zero, suggesting little evidence of an effect. The X-axis represents a range of manipulations to model architecture and training set.

was less consistently recovered, and nilgai (n = 185 training images), one of the least common species in the data set, was never correctly recovered. Notably, the posterior support of a positive response in sambar was missed by our baseline model (ResNet50) but was occasionally recovered by our deep neural networks under heavy manipulation. Further exploration found that these confusing responses were likely to be a false signal: sambar classification accuracy declined with increasing manipulations and the range of species misclassified as sambar increased, suggesting false positives were driving spurious occupancy patterns. Similar effects were seen for barking deer and wild boar.

Activity patterns

The predictions from all four DL models produced a similar range of activity pattern overlap with the expert-labelled data, with most overlap coefficients for the species ranging between 0.8 and 1 (Fig. 3a,d). Even though ConvNext-T performs well for most of the species, on the BCT data set it detected fewer species when compared to the ResNet models; one species was dropped from analysis due to lack of detections (Fig. 3d). Average overlap in the MMCT data set is consistent with changing model architecture, but there is a clear trend of reducing overlap with model architecture in the BCT data set, with the ResNet18 performing worst out of the four models (Fig. 3d). Accuracy of activity patterns were robust to a certain level of noise, with reductions in number of species and overall accuracy beyond 10% noise in both data sets (Fig. 3b,e). The activity patterns were also robust to a 50% drop in training set size (Fig. 3c, f). Displaying the predicted activity patterns of a subset of species showed that model manipulations mainly caused an overestimation of diurnal predictions compared to expert-labelled data (Fig. S1; Fig S2).

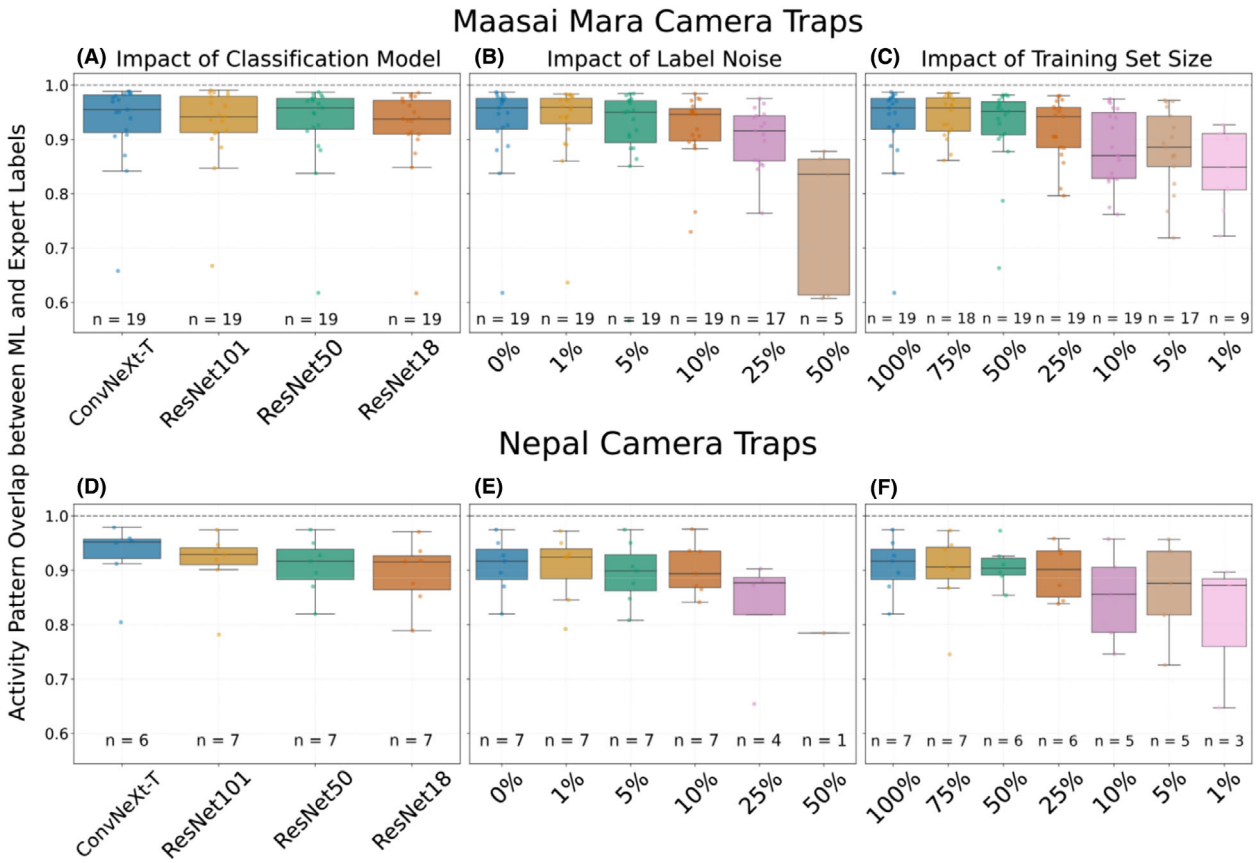

Figure 3. The overlap coefficient of species activity patterns produced from expert-generated labels compared to deep neural network-generated labels with variety of manipulations in the training pipeline. The box plots describe the mean, upper and lower quartile of the overlap coefficients calculated for each species in each data set. The Y-axis corresponds to the activity overlap coefficient, where a value of 1 represents perfect agreement between the activity patterns calculated from expert-generated labels and DL-generated labels. In cases where the deep neural network did not predict ≥ 20 detections of a species, this overlap calculation was dropped from the aggregate, represented in the n under each boxplot.

Correlation between machine-learning evaluation and ecological metrics

Classification accuracy of each deep neural network reduced with increasing manipulations, as expected (Table S6). However, we found that the accuracy of ecological metrics derived from deep neural network-generated labels is not always directly correlated with performance of the deep neural network. In most cases, the error in predicting ecological metrics is quite low, for example when looking at the impact of model, the majority of ecological accuracy values are below 0.4 (0 = perfect accuracy) (Fig. 4a,d). Overall, the accuracy of activity pattern predictions is more strongly correlated with deep neural network performance than the accuracy of occupancy (Fig. 4). Classification model and reducing training set size had a weak positive correlation with occupancy accuracy, but when looking at the impact of increasing noise in the data set on occupancy predictions, there is no correlation with model performance (Fig. 4b, e).

Discussion

Our study shows that deep neural network species classifiers can serve to estimate ecological metrics with reasonable accuracy, particularly for community-level measures like species richness. However, estimates were less reliable for rarer or visually similar species. The utilization of two camera trap data sets from two very different biomes, African savannah (MMCT) and Asian sub-tropical dry forest (BCT), each with a high diversity of medium and

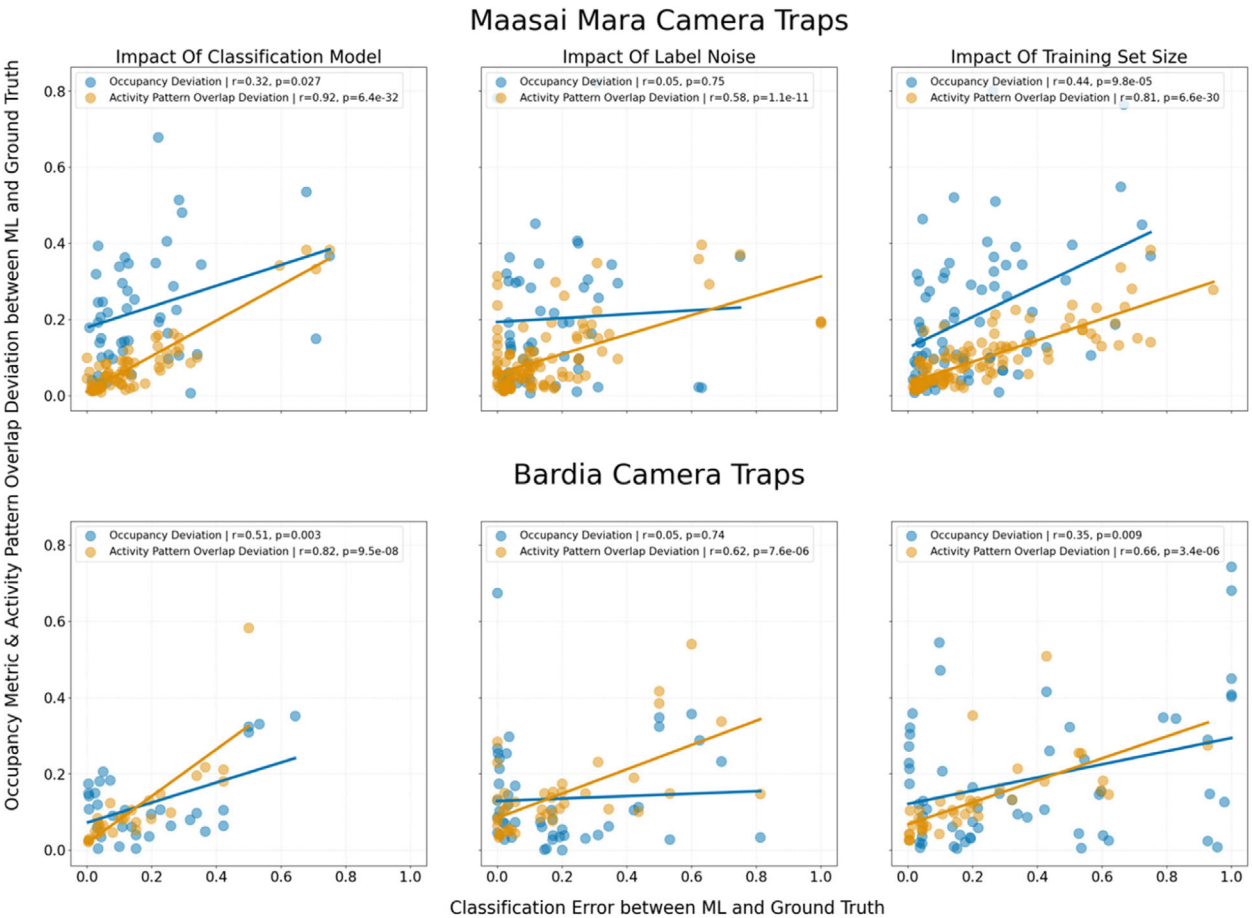

Figure 4. Correlation between ecological metric accuracy and deep neural network classification accuracy. Ecological metric accuracy is calculated as the absolute difference between metrics calculated from neural network-generated labels and expert-generated labels. The Y-axis corresponds to the absolute difference between occupancy coefficient estimate (blue lines) or activity pattern overlap (orange lines). The X-axis shows the classification error of each neural network. For each experiment, the individual manipulations are pooled together to show the overall trend, for example, the different DL models used in the Impact of Classification Model experiment are not differentiated. The results cover 20 and 8 species from MMCT data set and BCT data set, respectively.

large mammals increases the generality of our findings. Despite differences between ecosystems, background, data set size, and species composition, our findings were broadly consistent across data sets. Using two data sets with standardized survey design and rich metadata allowed us to control for survey effort and site-specific context, enabling robust analyses of occupancy and activity patterns that would have been difficult with benchmark data sets such as iWildCam (Beery et al., 2021), illustrating the value in using real, location-specific data when testing how machine learning can support biodiversity monitoring.

Deep neural network experiments

We found that that the choice of image classification model architecture or model depth has very little impact on the resulting ecological findings. Even though the utilization of deeper (He et al., 2016) or better performing (Liu et al., 2022) model architectures results in increased classification performance, we observe that CNN architectures that may be considered lower performing or outdated, such as a ResNet50, can still produce accurate predictions of species richness or occupancy for some species. Indeed, more advanced neural networks architectures than ResNet or ConvNext-T are now available. However, the purpose of this study was to assess the relative influence of model depth and architecture type on downstream ecological analyses. This insight can help practitioners to allocate available resources more efficiently, as there may not be a need for large models with expensive computational needs.

Our experiments show that most ecological metrics calculated from DL-predicted labels maintained a high similarity with expert-labelled data, even with up to 10% noise in the training set—a pattern that is common for deep neural network approaches (Drory et al., 2018; Rolnick et al., 2017). Above levels of 10% noise, reduced prediction confidence caused by noisy training labels means that more labels are dropped when a 70% confidence threshold is applied, resulting in apparent non-detection of rarer species. This created a bias in the results presented, as the measure of metric accuracy does not account for missing species. The loss of rarer or more cryptic species from a data set represents a wider problem of class imbalance that is regularly seen in camera trap data sets and can disproportionately affect detection of less common species (Schneider et al., 2020). Additionally, CNNs are more resilient to noise that is uniformly spread across the data set, compared to noise that is concentrated (Drory et al., 2018). A potential hindrance to our experimental design is the choice to mis-label images uniformly across species, when real labelling error would likely be focused on a particular species or part of the data set (*e.g.*, nocturnal images only).

The accuracy of the key ecological metrics generally held up against reductions in the training set size (*e.g.*, up to 50% of the available data dropped) across both data sets. We note that we applied a temporal data split to maximize applicability to this data set, when a spatial split might be more conventional (Norman et al., 2023). In that scenario, reducing the training set size could more strongly impact generalization to new sites. Past research has succeeded in building accurate classification models when training with millions of images, a scenario that might not be possible in most wildlife monitoring projects given the time-consuming and demanding nature of the manual image annotation step (Norouzzadeh et al., 2018). Thus, knowing where to stop within the labelling phase is a crucial decision given the time-consuming nature of camera trap image annotation. The emergence of efficient methods within the biodiversity monitoring domain such as active learning (Norouzzadeh et al., 2021), self-supervised learning (Pantazis et al., 2021), or large vision-language models (Pantazis et al., 2022) claim to reduce the need for large labelled sets. Whilst it is clear from our results that large training sets will always improve classifier accuracy, using these cutting-edge methods will further reduce the need for image labelling.

Even though our results suggest a small amount of label noise or reduced training set size is acceptable when predicting community-level metrics such as species richness, we observed a disproportionate impact on less common species for species-specific metrics. Our occupancy model results showed that classification error led to spurious responses of certain species to human infrastructure in Kenya or protected area management in Nepal. All analyses showed that certain species were dropped completely due to a lack of detections. This demonstrates how misclassifications caused by lower performing deep neural networks could obscure ecological patterns and misinform conservation decisions if interpreted as reliable signals. Our findings highlight the need to consider class imbalance when reporting on ecological analysis from DL-generated labels. Addressing such problems could involve species-specific confidence thresholds, statistical methods that account for false positives and false negatives (Katsis et al., 2025; Royle & Link, 2006), or developing nested classification models that focus on a subsets of animals (Mulero-Pázmány et al., 2025).

Correlation between conventional neural network evaluation and ecological metrics

We observe that the accuracy of deep–learning based ecological analysis does not always correlate strongly with

conventional ML evaluation metrics, but this varied with model manipulation and ecological metric. For example, adding noise to training labels degrades ML classification performance but there is no analogous impact on estimating species' occupancy responses to anthropogenic pressure, although a weak correlation appeared when reducing training set size. Classification error had a greater impact on activity pattern accuracy. This may be due to the resolution of detections needed for each analysis. For species richness and occupancy, detection frequency at a CT site within the detection window does not affect the metric, allowing for a degree of classification error. Predicting activity patterns requires several detections over the 24 h cycle, and any loss, *e.g.*, less detections at night, strongly affects interpretation. Data set characteristics will also impact interpretation: the MMCT data set is dominated by grazers, many of which are visually similar with overlapping ecology. This could not be said for the BCT data set, where species show a diversity of responses to changing management regime (Ferreira et al., 2023), making accuracy of responses more sensitive to label accuracy. It is important for practitioners to understand this impact when choosing analysis tools and to remain transparent in the DL methods used.

Conclusion

This study presents an end-to-end evaluation of deep-learning models trained under different settings based on metrics relevant to downstream ecological tasks. Our results provide clarity on the robustness of such models against a variety of typical design decisions related to DL model training and highlight areas where caution is warranted, for example interpreting species-specific responses, particularly for rare or visually similar species. Ultimately, our findings aim to empower practitioners with limited access to high computing power or specialist knowledge to build effective tools for conservation. Future research in this field should focus on enhancing accessibility, ensuring that deep-learning tools can be widely adopted and applied by the global conservation community.

Acknowledgments

This research was funded by WWF-UK as part of the Biome Health Project. Peggy Bevan acknowledges support by the Natural Environment Research Council (NERC), United Kingdom (Grant Ref.: NE/S007229/1). We thank Miranda Jones, Sarah Carroll, Georgia Cronshaw, Stratton Hatfield, Alex Rabeau, Taras Bains, and Liam Patullo for their help with image tagging. Thank you to Naresh Khanal, Prabin Poudel, Thomas Luypaert, Tilak BK, and several students from Tribhuvan University who helped with data collection. Research permits were granted by the Nepali Ministry of Forests and Environment and the Kenyan National Commission for Science, Technology, and Innovation (NACOSTI) (Ref. no: NACOSTI/P/18/61494/23703). Research permit was also granted by the Kenya Wildlife Services (Ref.: KWS/BRM/5001). We acknowledge WWF Kenya for logistical support and Olare Motorogi, Naboisho, Mara North, and the Mara Triangle conservancies and their teams in Kenya for access and field support. Daniel J. Ingram acknowledges support from UK Research and Innovation (Future Leaders Fellowship, Grant Ref.: MR/W006316/1). This article is written in memory of Dr. Ben Collen and Professor Dame Georgina Mace who both sadly passed since the inception of the Biome Health Project.

AUTHOR CONTRIBUTIONS

Peggy A. Bevan: Conceptualization; investigation; writing – original draft; methodology; visualization; writing – review and editing; formal analysis. **Omiros Pantazis:** Conceptualization; investigation; writing – original draft; methodology; visualization; writing – review and editing; formal analysis. **Holly Pringle:** Conceptualization; investigation; methodology; writing – review and editing. **Guilherme Braga Ferreira:** Conceptualization; methodology; writing – review and editing; supervision; project administration. **Daniel J. Ingram:** Conceptualization; funding acquisition; writing – review and editing; supervision; project administration. **Emily Madsen:** Investigation; writing – review and editing; conceptualization; project administration. **Liam Thomas:** Investigation; writing – review and editing; conceptualization; project administration. **Dol Raj Thanet:** Investigation; writing – review and editing. **Thakur Silwal:** Investigation; writing – review and editing. **Santosh Rayamajhi:** Conceptualization; investigation; funding acquisition; writing – review and editing; supervision. **Gabriel Brostow:** Investigation; methodology; supervision; writing – review and editing. **Oisin Mac Aodha:** Conceptualization; methodology; supervision; writing – review and editing. **Kate E. Jones:** Conceptualization; investigation; funding acquisition; writing – review and editing; supervision; project administration.

Data Availability Statement

Data have been made available for review through this restricted link: https://zenodo.org/records/13372531. We plan to archive them in Zenodo or LILA BC after publication of this work. Similarly, our code is available for review in the following anonymized repository and will

be released publicly in GitHub upon acceptance: https://anonymous.4open.science/r/ml_ecological_metrics-9F54/README.md.

References

Affinito, F., Butchart, S.H.M., Nicholson, E., Hirsch, T., Williams, J.M., Campbell, J.E. et al. (2025) Assessing coverage of the monitoring framework of the Kunming-Montreal Global Biodiversity Framework and opportunities to fill gaps. *Nature Ecology & Evolution*, **9**(7), 1280–1294. https://doi.org/10.1038/s41559-025-02718-3

Barcelos, D., Vieira, E.M., Pinheiro, M.S. & Ferreira, G.B. (2022) A before−after assessment of the response of mammals to tourism in a Brazilian national park. *Oryx*, **56**(6), 854–863. https://doi.org/10.1017/s0030605321001472

Beery, S., Agarwal, A., Cole, E., Birodkar, V. (2021). The iWildCam 2021 competition dataset. *arXiv*. https://doi.org/10.48550/arXiv.2105.03494

Beery, S., Morris, D., & Yang, S. (2019). Efficient pipeline for camera trap image review. *arXiv*, arXiv:1907.06772. http://arxiv.org/abs/1907.06772

Beery, S., Van Horn, G. & Perona, P. (2018) Recognition in terra incognita. In: Ferrari, V., Hebert, M., Sminchisescu, C. & Weiss, Y. (Eds.) *Computer Vision - ECCV 2018. Proceedings of the European Conference on Computer Vision (ECCV 2018), Lecture Notes in computer science*. Cham, Switzerland: Springer, p. 11220. https://doi.org/10.1007/978-3-030-01270-0_28

Beery, S., Wu, G., Rathod, V., Votel, R. & Huang, J. (2020) Context r-cnn: Long term temporal context for per-camera object detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. WA, USA: Seattle, pp. 13075–13085. https://doi.org/10.1109/CVPR42600.2020.01309

Blount, J.D., Chynoweth, M.W., Green, A.M. & Şekercioğlu, Ç.H. (2021) Review: COVID-19 highlights the importance of camera traps for wildlife conservation research and management. *Biological Conservation*, **256**, 108984. https://doi.org/10.1016/j.biocon.2021.108984

Browning, E., Gibb, R., Glover-Kapfer, P. & Jones, K.E. (2017) Passive acoustic monitoring in ecology and conservation. In: *WWF Conservation Technology Series 1(2)*. United Kingdom: WWF-UK, Woking.

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. https://doi.org/10.1111/1365-2664.12432

CBD. (2022) *Decision adopted by the conference of the parties to the convention on biological diversity 15/4*. Kunming-Montreal Global Biodiversity Framework. https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf

Connolly, E., Pringle, H.A.I., Pantazis, O., Ferreira, G.B., Madsen, E.K., Ingram, D.J. et al. (2025) Sustainable cattle management by communities supports African wildlife. *bioRxiv*. https://doi.org/10.1101/2025.10.09.681397

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. & Li, F.F. (2009) ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), Miami, FL, USA*, pp. 248–255. https://doi.org/10.1109/cvpr.2009.5206848

Dorazio, R.M., Royle, J.A., Söderström, B. & Glimskär, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, **87**(4), 842–854. https://doi.org/10.1890/0012-9658(2006)87[842:ESRAAB]2.0.CO;2

Drory, A., Ratzon, O., Avidan, S. & Giryes, R. (2018) The resistance to label noise in K-NN and DNN depends on its concentration. *arXiv*. https://doi.org/10.48550/arXiv.1803.11410

Duggan, M.T., Groleau, M.F., Shealy, E.P., Self, L.S., Utter, T.E., Waller, M.M. et al. (2021) An approach to rapid processing of camera trap images with minimal human input. *Ecology and Evolution*, **11**(17), 12051–12063. https://doi.org/10.1002/ece3.7970

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. https://doi.org/10.1016/j.gecco.2022.e02104

Ferreira, G.B., Collen, B., Newbold, T., Oliveira, M.J.R., Pinheiro, M.S., Pinho, F.F. et al. (2020) Strict protected areas are essential for the conservation of larger and threatened mammals in a priority region of the Brazilian Cerrado. *Biological Conservation*, **251**, 108762.https://doi.org/10.1016/j.biocon.2020.108762

Ferreira, G.B., Thomas, L., Ingram, D.J., Bevan, P.A., Madsen, E.K., Thanet, D.R. et al. (2023) Wildlife response to management regime and habitat loss in the Terai Arc Landscape of Nepal. *Biological Conservation*, **288**, 110334. https://doi.org/10.1016/j.biocon.2023.110334

Fisher, J.T. (2023) Camera trapping in ecology: A new section for wildlife research. *Ecology and Evolution*, **13**(3), e9925.https://doi.org/10.1002/ece3.9925

Gadot, T., Istrate, Ş., Kim, H., Morris, D., Beery, S., Birch, T. et al. (2024) To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. *IET Computer Vision*, **18**(8), 1193–1208. https://doi.org/10.1049/cvi2.12318

Gonzalez, A., Vihervaara, P., Balvanera, P., Bates, A.E., Bayraktarov, E., Bellingham, P.J. et al. (2023) A global biodiversity observing system to unite monitoring and guide action. *Nature Ecology & Evolution*, **7**(12), 1947–1952. https://doi.org/10.1038/s41559-023-02171-0

He, K. , Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *arXiv*. https://doi.org/10.48550/arXiv.1512.03385

iNaturalist. (2023). iNaturalist. https://www.inaturalist.org

Ingram, D.J., Ferreira, G.B., Jones, K.E. & Mace, G.M. (2021) Targeting conservation actions at species threat response thresholds. *Trends in Ecology & Evolution*, **36**(3), 216–226. https://doi.org/10.1016/j.tree.2020.11.004

Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using photographic captures and recaptures. *Ecology*, **79**(8), 2852–2862. https://doi.org/10.1890/0012-9658(1998)079[2852:EOTDII]2.0.CO;2

Katsis, L.K.D., Rhinehart, T.A., Dorgay, E., Sanchez, E.E., Snaddon, J.L., Doncaster, C.P. et al. (2025) A comparison of statistical methods for deriving occupancy estimates from machine learning outputs. *Scientific Reports*, **15**(1), 14700. https://doi.org/10.1038/s41598-025-95207-3

Kays, R., Arbogast, B.S., Baker-Whatton, M., Beirne, C., Boone, H.M., Bowler, M. et al. (2020) An empirical evaluation of camera trap study design: How many, how long and when? *Methods in Ecology and Evolution*, **11**(6), 700–713. https://doi.org/10.1111/2041-210X.13370

Kholiavchenko, M., Kline, J., Kukushkin, M., Brookes, O., Stevens, S., Duporge, I. et al. (2024) Deep dive into KABR: a dataset for understanding ungulate behavior from in-situ drone video. *Multimedia Tools and Applications*, **84**(21), 24563–24582. https://doi.org/10.1007/s11042-024-20512-4

Klaassen, B. & Broekhuis, F. (2018) Living on the edge: Multiscale habitat selection by cheetahs in a human-wildlife landscape. *Ecology and Evolution*, **8**(15), 7611–7623. https://doi.org/10.1002/ece3.4269

Lee, S.X.T., Amir, Z., Moore, J.H., Gaynor, K.M. & Luskin, M.S. (2024) Effects of human disturbances on wildlife behaviour and consequences for predator-prey overlap in Southeast Asia. *Nature Communications*, **15**(1), 1521. https://doi.org/10.1038/s41467-024-45905-9

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. et al. (2014) Microsoft COCO: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T. (Eds.) *Computer Vision - ECCV 2014. ECCV 2014 Lecture Notes in Computer Science*, Vol. **8693**. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T. & Xie, S. (2022) A convnet for the 2020. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11976–11986. https://doi.org/10.48550/arXiv.2201.03545

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:Esorwd]2.0.Co;2

Meredith, M., & Ridout, M. S.2023. overlap: Estimates of coefficient of overlapping for animal activity patterns. https://cran.r-project.org/package=overlap

Merkle, J.A., Anderson, N.J., Baxley, D.L., Chopp, M., Gigliotti, L.C., Gude, J.A. et al. (2019) A collaborative approach to bridging the gap between wildlife managers and researchers. *The Journal of Wildlife Management*, **83**(8), 1644–1651.https://doi.org/10.1002/jwmg.21759

Miao, Z., Liu, Z., Gaynor, K.M., Palmer, M.S., Yu, S.X. & Getz, W.M. (2021) Iterative human and automated identification of wildlife images. *Nature Machine Intelligence*, **3**(10), 885–895.https://doi.org/10.1038/s42256-021-00393-0

Microsoft. 2023. *Visual object tagging tool*. https://github.com/microsoft/VoTT

Mulero-Pázmány, M., Hurtado, S., Barba-González, C., Antequera-Gómez, M.L., Díaz-Ruiz, F., Real, R. et al. (2025) Addressing significant challenges for animal detection in camera trap images: a novel deep learning-based approach. *Scientific Reports*, **15**(1), 16191. https://doi.org/10.1038/s41598-025-90249-z

Norman, D., Bischoff, P.H., Wearn, O.R., Ewers, R.M., Rowcliffe, J.M., Evans, B. et al. (2023) Can CNN-based species classification generalise across variation in habitat within a camera trap survey? *Methods in Ecology and Evolution*, **14**(1), 242–251. https://doi.org/10.1111/2041-210x.14031

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N. & Clune, J. (2021) A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, **12**(1), 150–161. 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. https://doi.org/10.1073/pnas.1719367115

Pantazis, O., Brostow, G., Jones, K.E. & Mac Aodha, O. (2021) Focus on the positives self-supervised learning for biodiversity monitoring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montréal, Canada. 10583–10592. https://doi.org/10.1109/ICCV48922.2021.01041

Pantazis, O. Brostow, G. Jones, K. E. & Mac Aodha, O. (2022) SVL-adapter: Self-supervised adapter for vision-language pretrained models. *arXiv*. https://doi.org/10.48550/arXiv.2210.03794

Parsons, A.W., Kellner, K.F., Rota, C.T., Schuttler, S.G., Millspaugh, J.J. & Kays, R.W. (2022) The effect of urbanization on spatiotemporal interactions between gray foxes and coyotes. *Ecosphere*, **13**(3), e3993. https://doi.org/10.1002/ecs2.3993

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G. et al. (2019) Pytorch: An imperative style, high-performance deep learning library. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada*, Vol. **721**, pp. 8026–8037. https://doi.org/10.48550/arXiv.1912.01703

Penn, M.J., Miles, V., Astley, K.L., Ham, C., Woodroffe, R., Rowcliffe, M. et al. (2024) Sherlock—A flexible, low-resource tool for processing camera-trapping images. *Methods in Ecology and Evolution*, **15**(1), 91–102. https://doi.org/10.1111/2041-210x.14254

Pimm, S.L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z. et al. (2015) Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution*, **30**(11), 685–696.https://doi.org/10.1016/j.tree.2015.08.008

Pringle, S., Dallimer, M., Goddard, M.A., Le Goff, L.K., Hart, E., Langdale, S.J. et al. (2025) Opportunities and challenges for monitoring terrestrial biodiversity in the robotics age. *Nature Ecology & Evolution*, **9**(6), 1031–1042. https://doi.org/10.1038/s41559-025-02704-9

Pucci, R. Kalkman, V. J. & Stowell, D. 2023 Comparison between transformers and convolutional models for fine-grained classification of insects. *arXiv*. https://doi.org/10.48550/arXiv.2307.11112

Rolnick, D., Veit, A., Belongie, S. & Shavit, N. (2017). Deep learning is robust to massive label noise. *arXiv*. Preprint arXiv:1705.10694.

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

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

Rowcliffe, M. (2023). activity: Animal activity statistics - R package. https://cran.r-project.org/package=activity

Royle, J.A. & Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**(4), 835–841. https://doi.org/10.1890/0012-9658(2006)87[835:Gsomaf]2.0.Co;2

Scharlemann, J.P.W., Brock, R.C., Balfour, N., Brown, C., Burgess, N.D., Guth, M.K. et al. (2020) Towards understanding interactions between Sustainable Development Goals: The role of environment–human linkages. *Sustainability Science*, **15**, 1573–1584.https://doi.org/10.1007/s11625-020-00799-6

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. https://doi.org/10.1002/ece3.6147

Sethi, S.S., Bick, A., Chen, M.-Y., Crouzeilles, R., Hillier, B.V., Lawson, J. et al. (2024) Large-scale avian vocalization detection delivers reliable global biodiversity insights. *Proceedings of the National Academy of Sciences*, **121**(33), e2315933121. https://doi.org/10.1073/pnas.2315933121

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C. et al. (2017) Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, **15**(1), 26-34. https://doi.org/10.1002/fee.1448

Stephenson, P.J. (2020) Technological advances in biodiversity monitoring: applicability, opportunities and challenges. *Current Opinion in Environmental Sustainability*, **45**, 36–41.https://doi.org/10.1016/j.cosust.2020.08.005

Sukhbaatar, S., Bruna, J., Paluri, M., Bourdev, L. & Fergus, R. (2015) Training convolutional networks with noisy labels. In*International Conference on Learning Representations (ICLR), San Diego, CA, USA*. https://doi.org/10.48550/arXiv.1406.2080

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.https://doi.org/10.1111/2041-210X.13120

Thomson, R., Potgieter, G.C. & Bahaa-el-din, L. (2018) Closing the gap between camera trap software development and the user community. *African Journal of Ecology*, **56**(4), 721–739. https://doi.org/10.1111/aje.12550

Tobler, M.W., Zúñiga Hartley, A., Carrillo-Percastegui, S.E., Powell, G.V.N. & Lukacs, P. (2015) Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. *Journal of Applied Ecology*, **52**(2), 413–421. https://doi.org/10.1111/1365-2664.12399

Vishniakov, K. Shen, Z. & Liu, Z. 2024 ConvNet vs transformer, supervised vs CLIP: Beyong ImageNet accuracy ICML'24. Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria.

Wall, J., Wittemyer, G., Klinkenberg, B. & Douglas-Hamilton, I. (2014) Novel opportunities for wildlife conservation and research with real-time monitoring. *Ecological Applications*, **24**(4), 593–601. https://doi.org/10.1890/13-1971.1

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 l earning model. *Methods in Ecology and Evolution*, **12**(6), 1080–1092.https://doi.org/10.1111/2041-210X.13576

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A. et al. (2018) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, **10**(1), 80–91. https://doi.org/10.1111/2041-210x.13099

Wu, Z., Zhang, C., Gu, X., Duporge, I., Hughey, L.F., Stabach, J.A. et al. (2023) Deep learning enables satellite-based monitoring of large populations of terrestrial mammals across heterogeneous landscape. *Nature Communications*, **14** (1), 3072. https://doi.org/10.1038/s41467-023-38901-y

Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014) How transferable are features in deep neural networks? The 28th Annual Conference on *Neural Information Processing Systems (NIPS 2014), Montréal, Canada*. https://doi.org/10.48550/arXiv.1411.1792

Supporting Information

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

Data S1 Supplementary Information.

