



Deep learning-based monitoring and spatiotemporal analysis of wildlife and human activities using camera trap imagery

Hao Chen¹ · Fang Qiu¹ · Umme Kulsum¹ · Caiyun Zhang² · Eve Bohnett³ · Haitao Lyu¹ · Shuang Tian¹ · Babu Ram Lamichhane⁴ · Douglas Stow⁵ · Rebecca Lewison⁶ · Li An⁷

Received: 1 May 2026 / Revised: 10 August 2026 / Accepted: 22 September 2026
© The Author(s) 2026

Abstract

Wildlife and human activities are key components of ecological systems, and understanding their spatiotemporal patterns is essential for evaluating human–wildlife interactions and informing conservation planning. Camera trap networks function as terrestrial remote sensing systems that enable continuous monitoring of ecological processes at fine spatiotemporal scales. In this study, camera trap imagery collected in Chitwan National Park (CNP), Nepal, and its surrounding zones, under visible light during daytime and infrared illumination at night, was organized into a multi-class dataset containing wildlife, humans, and domestic animals. To support the ecological interpretation of wildlife–human spatiotemporal patterns, deep learning–based object detection models (YOLOv5 and YOLOv11) were trained to extract animal and human occurrences from the imagery. The resulting detections were integrated into a spatiotemporal analysis framework to characterize wildlife and human activity patterns, map spatial co-occurrence across the landscape, and quantify temporal proximity between human–wildlife pairs. Among the evaluated models, YOLOv11s achieved the best detection performance, with a precision of 95.5%, recall of 92.9%, and mean Average Precision₅₀ (mAP₅₀) of 95.4%, indicating reliable performance in real-world scenarios. Spatiotemporal analysis revealed shorter temporal lags for human–rhino pairs compared to human–tiger pairs, and identified a clear spatial gradient of human–wildlife co-occurrence, with the highest intensity in the Buffer Zone, followed by the Corridor Forest and the CNP. These results demonstrate that integrating automated detection with spatiotemporal analysis transforms camera trap data into ecological information for scalable monitoring. The proposed framework identifies areas of elevated human–wildlife spatiotemporal proximity, helping prioritize targeted monitoring and landscape management.

Keywords Deep learning · Spatiotemporal analysis · Camera traps · Wildlife monitoring · Human–wildlife interactions

Introduction

Human activities frequently alter or destroy wildlife habitats, with illegal hunting, poaching, and habitat encroachment among the most detrimental impacts to wildlife populations and behavior (Ikeda et al. 2022; Lee et al. 2024; Lewis et al. 2021). Conversely, wildlife species also affect land use, management of agricultural practices, and human habitation dynamics in various ecological zones (Soulsbury et al., 2015). Human–wildlife interactions are heterogeneous in space and time, spanning a range from positive coexistence to negative encounters commonly referred to as human–wildlife conflict (Nyhus 2016). Monitoring wildlife

and human activities is therefore essential for understanding these interactions and plays a crucial role in effective wildlife conservation and sustainable ecological landscape management.

Camera traps are point-based imaging sensors that function as terrestrial remote sensing system, which have been widely employed for wildlife observation, contributing significantly to understanding human–wildlife interaction, biodiversity richness, and the conservation of endangered species (Caravaggi et al. 2017; Foca et al. 2025; Linkie and Ridout 2011; Pierini et al. 2024; Rowcliffe et al. 2008). These devices are typically equipped with visible-light sensors and infrared illumination for nighttime imaging, which

Extended author information available on the last page of the article

automatically gather data upon detecting any motion within their range. Unlike aerial and satellite observation methods, camera trap networks provide continuous day-and-night, non-invasive monitoring while minimizing disturbance to animals and their habitats. As shown in Fig. 1, camera traps collect rich contextual information, including recording timestamps and ambient temperature at the moment of data acquisition, and the GPS coordinates of each camera trap are recorded during deployment to enable spatial analysis of the collected data (Tanwar et al. 2021; Jasiulionis et al. 2023; Palmer et al. 2018).

By recording the timing of observations at fixed spatial locations, camera traps provide valuable data for spatiotemporal ecological analysis. These data allow researchers to estimate species abundance, analyze wildlife activity patterns, and investigate ecological processes such as predator–prey dynamics and human–wildlife interactions (Kukielka et al. 2013; Amburgey et al. 2021). Spatiotemporal information also provides opportunities to analyze species occurrence and co-occurrence patterns across landscapes, offering insights into how wildlife distributions and behaviors vary over space and time (Shamoon et al. 2018; Kavwele et al. 2022). However, manually processing the large volumes of images and videos captured by camera traps is time-consuming and presents substantial human and financial challenges (Christin et al. 2019). In addition, the presence of fast-moving animals with blurry imagery, partial body visibility, low-light conditions, and overlapping multiple targets further complicate the target detection process (Fig. 1c–e).

Deep learning, a subset of machine learning that uses multi-layered neural networks to learn hierarchical feature representations from data, has become a powerful tool for complex pattern recognition tasks. In recent years, deep

learning methods have shown significant advantages in both efficiency and accuracy for object detection, making them well-suited for analyzing large volumes of camera trap imagery in this project. Deep learning-based object detectors can generally be divided into two main categories: one-stage and two-stage. Two-stage detectors, including Faster Region-based Convolutional Neural Networks (Faster R-CNN) (Ren et al. 2016), Region-based Fully Convolutional Networks (R-FCN) (Dai et al. 2016), and Mask R-CNN (He et al. 2017), divide the detection process into two stages: region proposal and classification. The region proposal stage narrows the search space by selecting candidate regions from feature maps, whereas the classification stage determines the object category and refines the final detection results. In contrast, one-stage detectors, such as You Only Look Once (YOLO) (Redmon et al. 2016), Single Shot MultiBox Detector (SSD) (Liu et al. 2016), and RetinaNet (Lin et al. 2017), directly predict bounding boxes and class probabilities in a single step. These models typically offer higher processing speed, although sometimes at the expense of detection accuracy compared with two-stage methods.

The application of deep learning-based object detection to camera trap imagery has streamlined monitoring workflows, enhanced detection efficiency and accuracy, and substantially reduced operational costs (Whytock et al. 2021; Petso et al. 2022). Among two-stage detectors, Faster R-CNN began to be adopted for wildlife detection in camera trap data and has demonstrated great detection accuracy (Carl et al. 2020; Simoes et al., 2023). Compared with Faster R-CNN, R-FCN has emerged as a more practical architectural choice for large-scale animal detection, particularly in scenarios involving moving wildlife (Ibraheem et al. 2021).



Fig. 1 Representative camera trap images illustrating diverse observational conditions and target detection challenges in wildlife monitoring: (a) human activity, (b) multiple overlapping domestic goats, (c)

partial visibility of a rhino, (d) low-light detection of a spotted deer, (e) a fast-moving tiger with motion blur, and (f) a nighttime infrared-illuminated image of an elephant

Mask R-CNN, in contrast, has shown stronger suitability for instance segmentation tasks (Salau and Krieter 2020). On the other hand, one-stage detectors, such as the YOLO series, are designed for real-time detection, providing rapid processing speed and high efficiency while maintaining strong detection accuracy (Diwan et al. 2023; Ferrante et al. 2024). Multiple YOLO versions, including YOLOv3, YOLOv4, YOLOv5, and YOLOv10, have demonstrated strong potential in terms of both accuracy and efficiency for wildlife detection tasks (Bjerger et al. 2023; Leonid et al. 2023; Chalmers et al. 2025).

Deep learning has substantially improved the accuracy and efficiency of automated animal object detection, allowing massive volumes of camera trap images to be timely processed and generating abundant geospatial data for wildlife monitoring. Most prior studies have emphasized detection or classification performance, whereas relatively few have used automated camera trap detections for landscape-scale spatiotemporal assessment of wildlife and human activity patterns and their potential interactions.

To address these gaps, this study develops an integrated framework that uses deep learning to extract wildlife, domestic animal, and human detections from camera trap imagery and subsequently incorporates these detections into landscape-scale spatial and temporal analyses. An annotated multiclass dataset was constructed, and several YOLO-based object detection models were trained and evaluated to identify a reliable model for automated image processing. The selected model was then applied to camera trap imagery collected across Chitwan National Park and its surrounding landscapes. The resulting detections were incorporated into spatiotemporal analyses to characterize species activity patterns and quantify the shortest temporal lag between species across different landscapes, thereby enabling the identification and further investigation of areas with elevated human–wildlife spatiotemporal proximity.

Methods

Study area

This study was conducted in Chitwan National Park (CNP) and the adjacent managed and multi-use landscapes that surround it (Fig. 2). CNP was established in 1973 as Nepal's first national park and was listed as a UNESCO World Heritage Site in 1984 due to its unique ecosystems of international significance (Mishra 1982). It covers an area of 932 km² in south-central Nepal and forms part of the Chitwan-Parsa-Valmiki Tiger Conservation Landscape where reducing human-wildlife conflict is a stated management objective (Bhattarai et al. 2019; Bhatt et al. 2023). The national park

contains a range of diverse ecosystems, including hills, lakes, flood plains, and rivers, which support endangered greater rhinos (One Horned Rhinoceros; *Rhinoceros unicornis*), tigers (Royal Bengal Tiger; *Panthera tigris tigris*), crocodiles (Gharial Crocodile; *Gavialis gangeticus*), and other endemic species (Dai et al. 2020; UNESCO, 2019). The regional climate is subtropical and is dominated by the southeast monsoon. From March to June, temperatures can reach as high as 43 °C. The monsoon season typically follows, lasting from late June until September. The mean annual rainfall in the park is approximately 2,150 mm.

Surrounding CNP is a Buffer Zone encompassing 750 km² comprising forest and other land uses, including cultivated areas. More than 273,977 people live within the Buffer Zone of CNP (Dangol et al. 2020), and many households depend on local natural resources for their livelihoods, collecting wood, grass and various herbs, obtained through both legal and illegal extraction. North of the Buffer Zone lie the Corridor Forest and Community Forest. The Corridor Forest connects the Terai Arc Landscape (TAL) and Chitwan Annapurna Landscape (CAL), which is the only remaining natural forest that allows the endangered greater one-horned rhino and Royal Bengal tiger access to refuge at higher elevations during monsoon seasons (Aryal et al. 2012; Lamichhane et al. 2018). Community Forests are managed through Nepal's community forest program, under which local Community Forest User Groups are granted rights to manage, protect, and sustainably use forest resources while supporting forest conservation and local socioeconomic benefits (Smith et al. 2023).

Data collection and preprocessing

Camera trap data were collected using motion-activated Bushnell Core DS-4 K cameras equipped with separate sensors for daytime and nighttime imaging. A total of 80 cameras were deployed, with 20 cameras placed in Chitwan National Park, 20 in the Buffer Zone, 20 in the Corridor Forest, and 20 in the Community Forest. The cameras were arranged using an approximately 1.5-km grid, although some locations were adjusted to nearby animal trails. Each camera was mounted approximately 50–80 cm above the ground, and no bait or lure was used. The geographic coordinates of each camera trap site were recorded before deployment. The cameras were configured to collect both still images and videos, with the date, time, and ambient temperature embedded in each record. Field staff checked the cameras and retrieved the stored data every two months.

For this study, all still images collected between February and July 2022 were used for model development and subsequent analysis. This period covered the dry and pre-monsoon seasons, during which the camera trap network

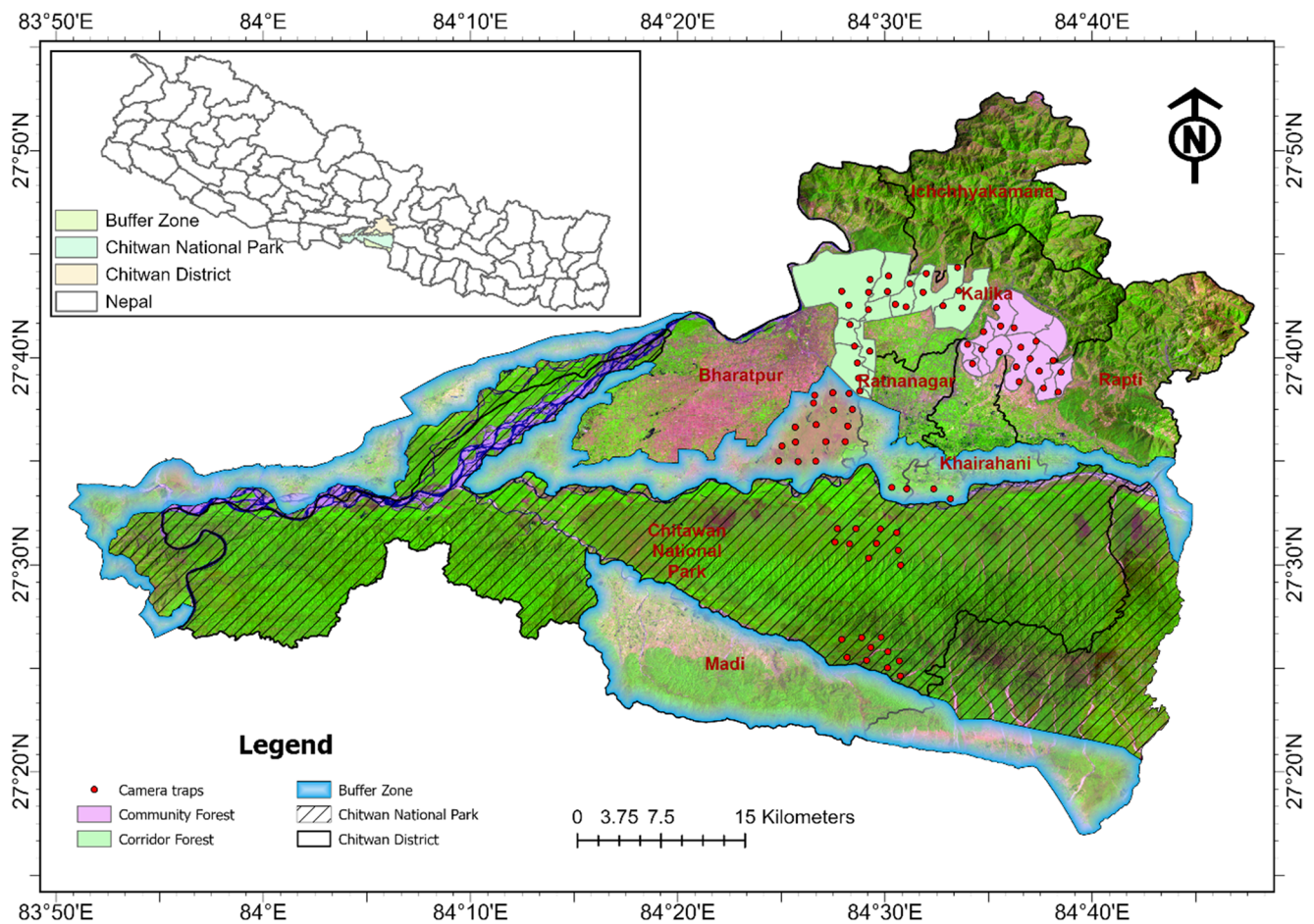


Fig. 2 Chitwan National Park and its surrounding zones (False color image of Landsat 8 collected on April 5, 2022. Red: band 6, 1560–1660 nm; Green: band 5, 845–885 nm; Blue: band 4, 630–680 nm)

Table 1 Numbers of annotated images and object instances in the camera trap dataset

Class	Tiger	Rhino	Elephant	Sambar Deer	Spotted Deer	Rhesus Macaque	Human	Domestic Goat
Images	1211	2733	1776	3639	2145	2220	1842	1120
Instances	1211	2805	1884	3931	4996	2997	2081	3042

provided consistent observations across the four landscape zones. Although the camera trap survey recorded a broader range of wildlife, six focal species were selected for annotation and model development based on their detection frequency, conservation importance, and relevance to human–wildlife interactions. These species included spotted deer (*Axis axis*), sambar deer (*Rusa unicolor*), rhesus macaques (*Macaca mulatta*), greater one-horned rhinos (*Rhinoceros unicornis*), Royal Bengal tigers (*Panthera tigris tigris*), and Asian elephants (*Elephas maximus*). Humans (*Homo sapiens*) and domestic goats (*Capra hircus*) were additionally included as indicators of human presence and associated activities.

Image labelling was conducted using the Roboflow platform, which facilitated precise image annotation and streamlined the export process for deep learning model

development (Roboflow 2025). The original annotated images were divided into training, validation, and test subsets in proportions of 70%, 20%, and 10%, respectively. This allocation provided sufficient data for model training and validation across multiple model configurations while retaining a separate held-out test set for final performance assessment. To reduce class imbalance while preventing data leakage, random rotation augmentation was applied to underrepresented classes, including tiger, rhino, and elephant, within the training subset. The numbers of annotated images and object instances in the camera trap dataset are shown in Table 1. Here, an image refers to a camera trap photograph containing at least one annotated target, whereas an instance refers to one individually annotated object, represented by a bounding box, within an image. In total, the dataset contained 16,686 images across the

eight target classes. All images were resized to 640×640 pixels to standardize the input dimensions, reduce computational cost, and meet the input requirements of the detection models.

Deep learning models

We used YOLO to automate detection of wildlife, humans, and domestic animals in camera trap imagery. Among the YOLO series, YOLOv5 is one of the most widely used owing to its efficiency and flexibility, while YOLOv11 (Khanam and Hussain 2024) represents the most recent advancement in the series at the time of this study. YOLOv5 optimizes a weighted training loss that combines classification, objectness, and bounding-box regression, and applies Non-Maximum Suppression (NMS) during inference to remove overlapping boxes, enabling accurate object detection in cluttered or partially occluded scenes. It follows an anchor-based design, in which each potential object location begins with a set of predefined reference boxes (anchors) of different sizes and shapes, and the model learns how to adjust these boxes to fit actual objects. YOLOv11 adopts an anchor-free detection approach that removes predefined anchor boxes and directly predicts object locations from multi-scale feature maps, thereby simplifying the model architecture while improving robustness across different object sizes and scene conditions. In practice, YOLOv11 represents a later generation within the YOLO family, incorporating more recent design advances while retaining the core YOLO detection framework.

YOLOv5 and YOLOv11 both provide five network configurations: nano (n), small (s), medium (m), large (l), and extra-large (x), which adjust network depth and width to balance detection performance against runtime and memory requirements. These configurations use the same overall detection framework but apply different architectural scaling factors, with the suffix indicating the relative model size from the most compact n variant to the largest x variant. In this study, several YOLO-based object detection models were trained and evaluated, including YOLOv5n, YOLOv5s, and YOLOv5m, YOLOv11n, YOLOv11s, and YOLOv11m. All models were trained in Python using the Ultralytics YOLO framework based on PyTorch, with a batch size of 32 for 150 epochs on an NVIDIA RTX 3090 GPU with 24 GB of dedicated memory.

The detection models were evaluated using standard performance metrics (described below) to identify the most suitable approach for the camera trap dataset. The best-performing model was then applied to the full collected data to automatically detect human and wildlife. The resulting detection outputs were subsequently used in spatial and temporal analyses to characterize relative detection-intensity

patterns, assess spatial co-occurrence, and quantify temporal proximity between human–wildlife pairs.

Evaluation metrics

To evaluate the performance and reliability of the object-detection component of the proposed framework, we used four standard metrics: precision, recall, mean Average Precision50 (mAP50), and mAP50–95 (Ibraheam et al. 2021). During training, model performance was monitored on the validation set, while final evaluation was conducted on the fully annotated test set. Predicted bounding boxes were compared with ground-truth annotations following standard YOLO evaluation protocols. A prediction was considered a true positive (TP) if it matched a ground-truth object of the same class at or above the specified Intersection over Union (IoU) threshold; otherwise, it was counted as a false positive (FP). Ground-truth objects that were not matched by any prediction were counted as false negatives (FN). Precision measures the proportion of predicted detections that are correct, whereas recall, also referred to as sensitivity, measures the proportion of ground-truth instances correctly detected by the model. Precision and recall were calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

We reported mAP at two levels of stringency. mAP50 is the mean average precision over all classes at an IoU threshold of 0.50, indicating how well the model detects objects with reasonably good overlap. In contrast, mAP50–95 is the mean average precision averaged over IoU thresholds from 0.50 to 0.95 at intervals of 0.05, providing a more stringent measure of both detection performance and localization precision.

Data analysis

Spatial analysis

Spatial patterns were characterized using the relative abundance index (RAI) and kernel density estimation (KDE). For each species and camera site, RAI was calculated as the number of independent camera trap detections divided by the total number of camera trap days, multiplied by 100 (Rovero et al., 2009):

$$RAI = \frac{\text{Number of independent detections}}{\text{Total number of camera trap day}} \times 100 \quad (3)$$

RAI provided an effort-standardized measure of relative detection frequency across camera stations. Independent camera trap detections were defined by treating consecutive detections of the same target class at the same camera trap site within a 30-min interval as a single detection event, thereby reducing repeated counting of the same individual or group during a single visit.

The site-specific RAI values were subsequently used as weights in KDE to transform the discrete camera-sites observations into continuous surfaces of relative detection intensity. Camera trap sites with higher RAI values contributed more strongly to the estimated surfaces. Locations with greater concentrations of detections yielded higher estimated intensity values, which decreased with distance from the camera trap sites and approached zero at the search radius. The RAI-weighted KDE value at location (x, y) was calculated as:

$$\hat{f}(x, y) = \frac{1}{h^2} \sum_{i=1}^n RAI \left[\frac{3}{\pi} \left(1 - \left(\frac{dist_i}{h} \right)^2 \right)^2 \right] \quad (4)$$

where n is the number of camera trap sites, h denotes the kernel bandwidth (search radius), and $dist_i$ is the distance between camera trap site i and location (x, y) . Only camera trap sites within the bandwidth ($dist_i < h$) contributed to the density estimation.

For the spatial co-occurrence analysis, the site-specific RAI values were aggregated into two composite activity classes. At each camera trap site, the RAI values of humans and domestic goats were summed to represent human activity, whereas the RAI values of tigers and rhinos were summed to represent dangerous wildlife activity. Deer and rhesus macaques pose no direct danger to humans or livestock, while wild elephants could not be distinguished in the camera trap imagery from domestic elephants, which are frequently used for transportation in Nepal (Szydlowski 2022). These species were therefore excluded from the subsequent spatial co-occurrence analysis.

The aggregated RAI values of the two composite classes were then used as weights in separate KDE analyses to generate continuous surfaces of human activity and dangerous wildlife activity. Each KDE surface was normalized to a range of 0–1 using min–max normalization. To assess spatial co-occurrence, the two normalized KDE surfaces were multiplied cell by cell. Higher product values indicate locations where relatively high human activity and dangerous wildlife activity spatially coincide. The resulting surface was used to identify areas of elevated human–wildlife spatial proximity. RAI calculation, RAI-weighted KDE, surface normalization, and spatial co-occurrence mapping were conducted using ArcGIS Pro 3.3.

Temporal analysis

Model-derived detections were aggregated into daily presence–absence records at each camera trap site. For each target class, a site was assigned a value of presence for a calendar day if the target was detected at least once on that day; otherwise, it was assigned absence. Thus, multiple images or repeated triggers of the same target class at the same site on the same day contributed only one daily presence record, reducing pseudoreplication from burst sequences and repeated observations (Ridout et al., 2009; Hamel et al. 2013; Hu et al. 2025).

Daily aggregation was selected because the analysis aimed to assess broad site-level temporal proximity across the six-month survey period and among landscape zones. Original timestamps were retained and used to calculate exact sub-daily intervals for selected same-day human–wildlife detection events.

Temporal proximity was assessed for human–tiger and human–rhino pairs using the shortest temporal lag, in days, between their daily presence records at the same camera trap site. For species a and species b at site s , the temporal lag between two daily detections was defined as:

$$\Delta d_{ij}^s = |d_{a,i}^s - d_{b,j}^s| \quad (5)$$

where $d_{a,i}^s$ denotes the i -th day on which species a was detected at camera trap site s , and $d_{b,j}^s$ denotes the j -th day on which species b was detected at the same site. Based on these pairwise differences, the shortest temporal lag between the two species at site s was defined as:

$$D_{ab}^s = \min_{i,j} (\Delta d_{ij}^s) = \min_{i,j} |d_{a,i}^s - d_{b,j}^s| \quad (6)$$

where D_{ab}^s represents the minimum number of days between any daily detection of species a and species b at the same camera trap site. A value of 0 indicates that the two species were detected at the same site on the same calendar day, whereas larger values indicate greater temporal separation. This measure was used to quantify temporal proximity between species after detections had been aggregated into daily presence records.

The shortest temporal-lag distributions for both species pairs were also compared across the major landscape types using box-whisker plots. Only the Buffer Zone, Chitwan National Park, and the Corridor Forest were included in this comparison because no tigers or rhinos were detected in the Community Forest. To further characterize temporal proximity at finer resolution, detection records from selected camera-trap sites in the regions identified through the spatial co-occurrence analysis were reviewed. The exact time

lag between human and dangerous-wildlife detections was then calculated. The temporal analyses in this section were performed using custom Python scripts.

Results

Deep learning model evaluation

Detailed training convergence curves of the evaluated models are provided in Fig. S1. Overall, the YOLO models showed rapid performance gains during the initial stages of training, followed by gradual stabilization.

The performance metrics of the trained YOLO models on the validation dataset are presented in Table 2. Inference time refers to the average time required for a model to process a single image and generate detection outputs, whereas parameters indicate the total number of learnable weights in the network. YOLOv11 consistently outperformed YOLOv5 at equivalent network sizes. For example, YOLOv11n achieved a precision of 95.0%, a recall of 90.9%, a mAP50 of 96.1%, and a mAP50-95 of 79.4%, whereas YOLOv5n attains a precision of 93.4%, a recall of 88.8%, a mAP50 of 94.6%, and a mAP50-95 of 74.1%. Among the evaluated models, YOLOv11s achieved the highest precision and mAP50, whereas YOLOv11m achieved the highest recall and mAP50-95. However, YOLOv11s required only 3.4 ms per image and contained 9.41 million parameters, compared with 8.0 ms and 20.03 million parameters for YOLOv11m. Therefore, YOLOv11s provided the best balance between detection performance and computational efficiency and was selected for subsequent analysis.

The selected YOLOv11s model was subsequently evaluated on 1,145 held-out test images using an IoU threshold of 0.5 and a confidence threshold of 0.5. These images were not seen by the model during training or validation and were reserved exclusively for final performance assessment. The model achieved an overall precision of 95.5%, a recall of 92.9%, and a mAP50 of 95.4%, indicating reliable generalization for the subsequent spatiotemporal analyses. As shown in Fig. 3, the model successfully detected targets under both favorable imaging conditions (Fig. 3a–d) and

challenging conditions (Fig. 3e–n). Representative incorrect detections are shown in Fig. 3o–p, including ambiguous multiple predictions and a missed target near the image boundary, reflecting persistent challenges in automatic object detection.

Spatial pattern analysis

The trained YOLOv11s model was then applied to a total of 57,037 camera trap images acquired in the study areas between February and July 2022 to automatically detect all relevant classes. Species-specific RAI-weighted kernel density maps derived from these detections reveal pronounced spatial variation and clustering (Fig. 4).

Rhinos showed the highest relative detection intensity along the northern boundary of CNP and within the Buffer Zone, with additional detections extending into the Corridor Forest (Fig. 4a). Elephants were detected primarily within CNP, with more limited occurrence in the Buffer Zone (Fig. 4b). Deer, the most frequently detected group, occurred at nearly all camera trap sites except those in the Community Forest. Their relative detection intensity was highest in the Buffer Zone and lower inside CNP (Fig. 4c). Rhesus macaques showed the broadest distribution, with the highest relative detection intensity in the Corridor Forest and northern Buffer Zone and additional presence in the Community Forest (Fig. 4d). Tigers were detected across the southern and northern parts of the study area, with the highest relative detection intensity occurring primarily inside CNP and additional presence in the Buffer Zone (Fig. 4e).

The spatial patterns of human-associated activity are illustrated in Fig. 4f and g through detections of humans and domestic goats. Human detections were concentrated mainly in the Corridor Forest, followed by the Community Forest and northern Buffer Zone. Similarly, domestic goats detections were concentrated in the Corridor Forest and Community Forest but were absent from the Buffer Zone.

Spatial co-occurrence analysis

The normalized RAI-weighted KDE surfaces of dangerous wildlife and human activities are shown in Fig. 5a and b,

Table 2 Validation performance and computational efficiency of the YOLO models. The suffixes n, s, and m indicate nano, small, and medium variants, respectively

Models	Precision	Recall	mAP50	mAP50-95	Inference (ms)	Parameters (M)
YOLOv5n	0.934	0.888	0.946	0.741	3.6	1.77
YOLOv5s	0.948	0.904	0.953	0.770	5.1	7.03
YOLOv5m	0.948	0.927	0.957	0.788	12.0	20.88
YOLOv11n	0.950	0.909	0.961	0.794	1.6	2.58
YOLOv11s	<i>0.962</i>	0.923	<i>0.967</i>	0.813	3.4	9.41
YOLOv11m	0.958	<i>0.931</i>	0.965	<i>0.819</i>	8.0	20.03

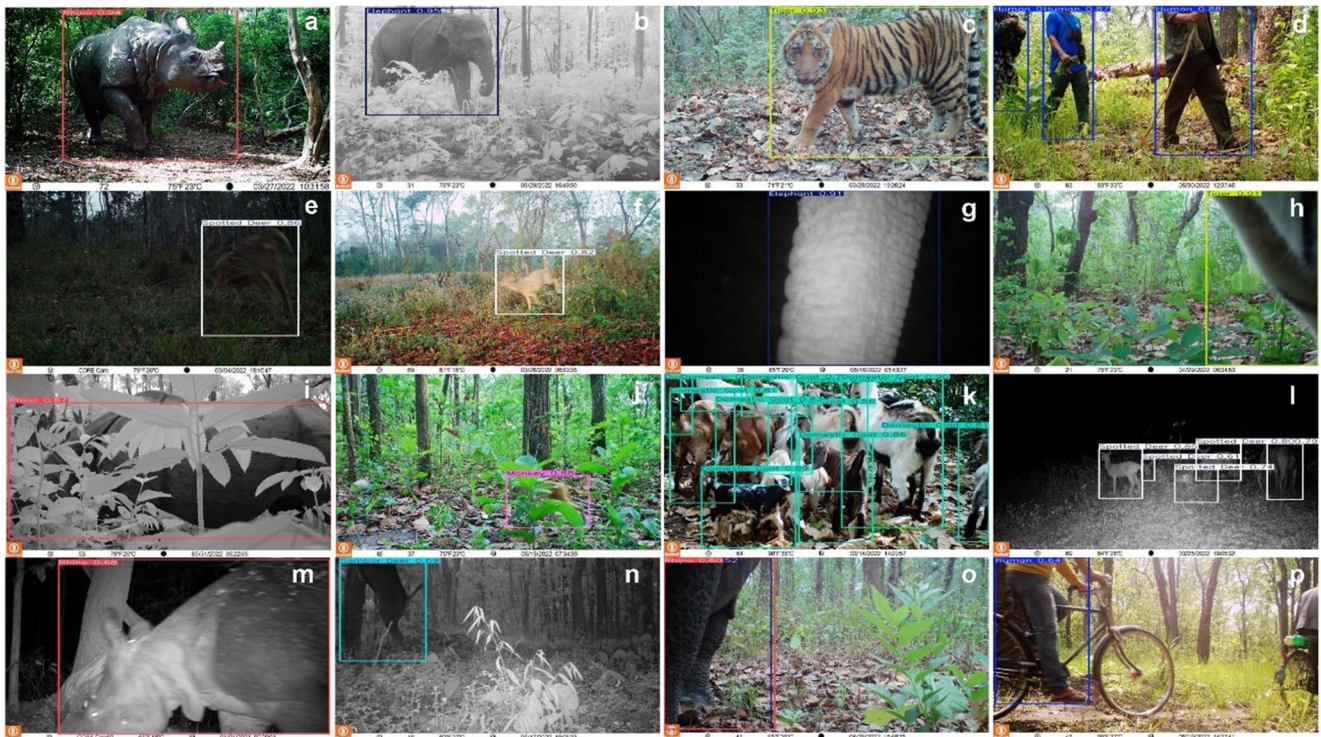


Fig. 3 Representative YOLOv11s detection results on the test dataset: (a–d) favorable conditions, (e–n) challenging conditions, and (o–p) incorrect detections

respectively. Figure 5a reveals that dangerous wildlife was primarily distributed within CNP and the Buffer Zone, with some extension into the Corridor Forest. In contrast, Fig. 5b shows that human activities were concentrated mainly in the Corridor Forest and Community Forest, with additional activities detected in the Buffer Zone and limited activity within CNP.

Figure 5c presents the product of the two normalized density surfaces used to assess spatial co-occurrence. Based on the intensity of the product surface, four regions with elevated spatial co-occurrence were delineated by red rectangles. Region 1 showed the highest co-occurrence intensity, followed by Region 2, and both were located in the Buffer Zone near the boundary of CNP. Region 3 was located within CNP near its southern boundary adjacent to the Buffer Zone. Region 4 was located primarily in the southern part of the Corridor Forest and extended into the Buffer Zone.

Temporal proximity analysis

Figure 6 shows the frequency distributions of the shortest temporal lags for the two species pairs. For both pairs, the distributions were concentrated at low values, indicating that humans and these dangerous wildlife species were often detected at the same camera trap site within short temporal intervals. Human-tiger pairs (Fig. 6a) displayed a broader

distribution with a longer right tail, indicating a higher frequency of longer temporal lags. The frequencies of 0-, 1- and 2-day lags were 8, 16, 22, respectively, giving a total of 46 detections within the first 3 day-lag classes. In contrast, human-rhino pairs (Fig. 6b) were more tightly concentrated at shorter lags of less than 22 days, with most detections occurring within the lower day intervals. The frequencies of 0-, 1- and 2-day lags were 12, 21, and 20, respectively, giving a total of 53 detections for the first three day-lag classes. Overall, these patterns suggest that human-rhino pairs tend to occur closer together in time, whereas human-tiger pairs exhibit longer temporal spacing across the camera sites.

The shortest temporal lag distributions for both species pairs across major landscape types are shown as box-whisker plots in Fig. 7. In the Buffer Zone, shortest temporal lags were concentrated at low values (less than 20 days) for both pair types, while the distribution for human-tiger pairs was more dispersed than that for human-rhino pairs. In Chitwan National Park, only three human-tiger pairs were recorded, and their shortest temporal lags were three, 15, and 41 days. In the Corridor Forest, the contrast between the two pair types was most pronounced. Human-tiger pairs showed the largest median and the widest interquartile spread, indicating greater temporal separation in detection intervals, whereas human-rhino pairs remained concentrated at lower values. Overall, the figure reveals that shortest temporal lags were generally lower and more tightly clustered

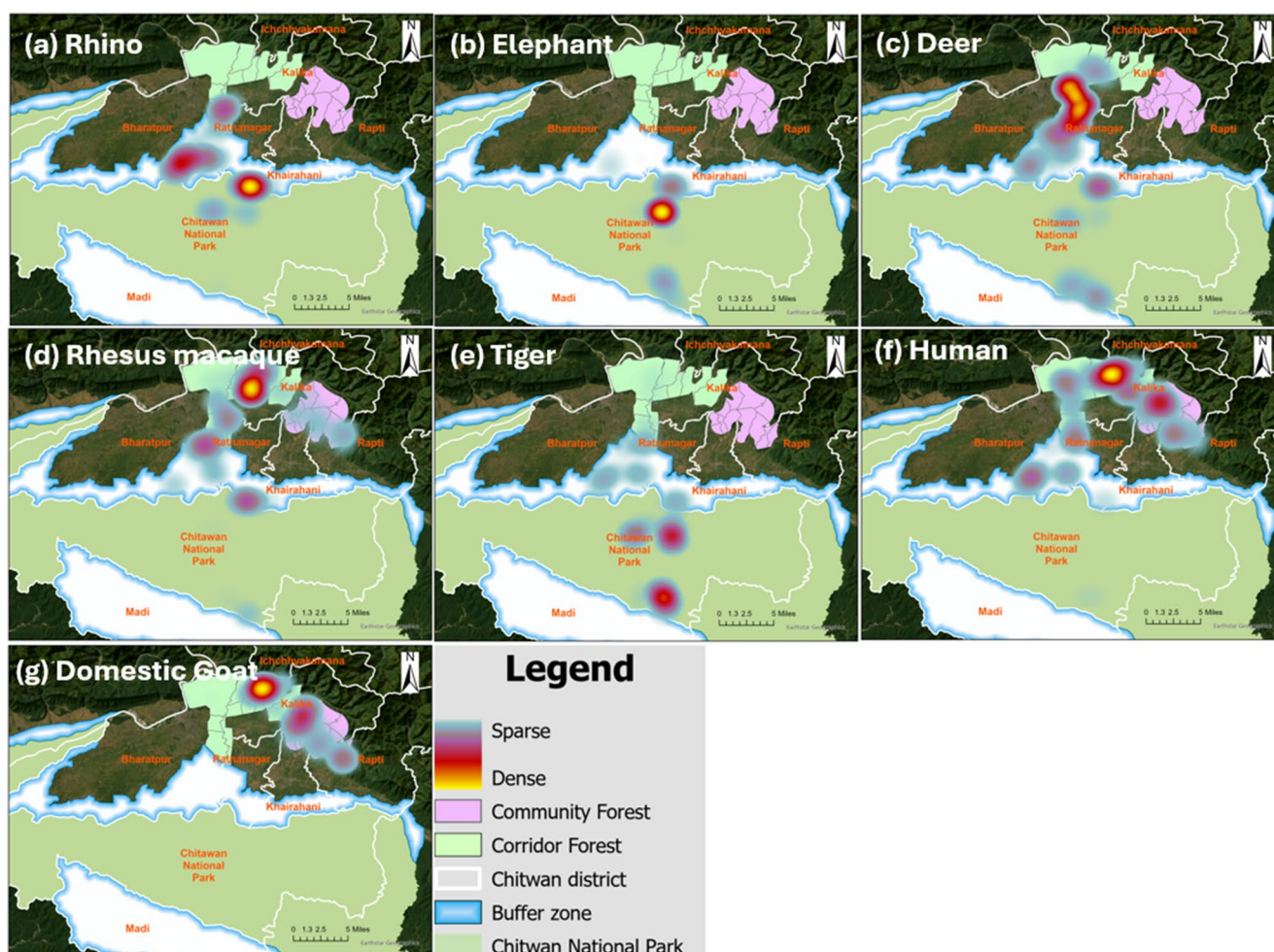


Fig. 4 Species-specific RAI-weighted kernel density maps of relative detection intensity

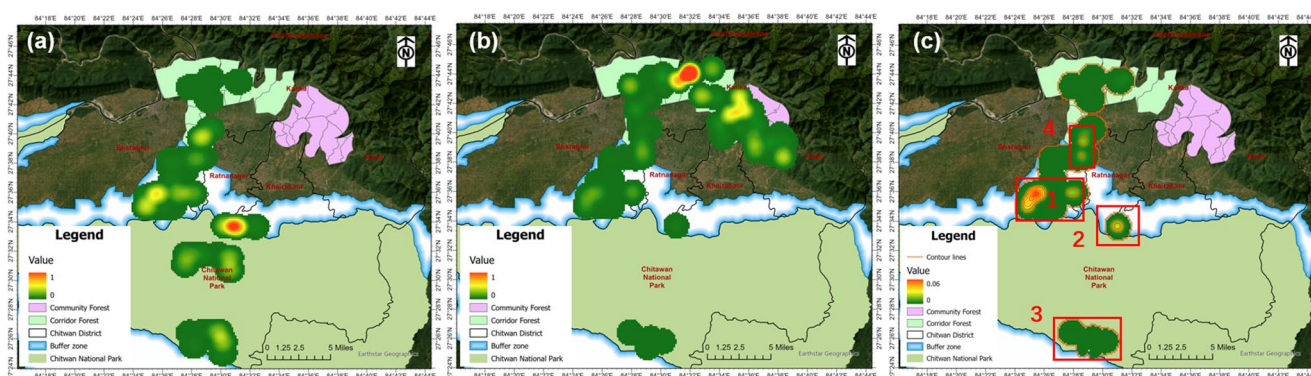


Fig. 5 Normalized RAI-weighted KDE surfaces of (a) dangerous wildlife, (b) human activity, and (c) the product of the two normalized surfaces. Regions of elevated spatial co-occurrence are delineated by red rectangles

for human–rhino pairs, while human–tiger pairs exhibited greater temporal spread. Across the three landscape types, human–wildlife pairs were temporally closest in the Buffer Zone, followed by the Corridor Forest and then Chitwan National Park, consistent with the spatial co-occurrence pattern shown in Fig. 5c.

Representative cases

Representative examples from the four regions identified through the spatial co-occurrence analysis are shown in Fig. 8, illustrating close temporal proximity between human activity and wildlife presence.

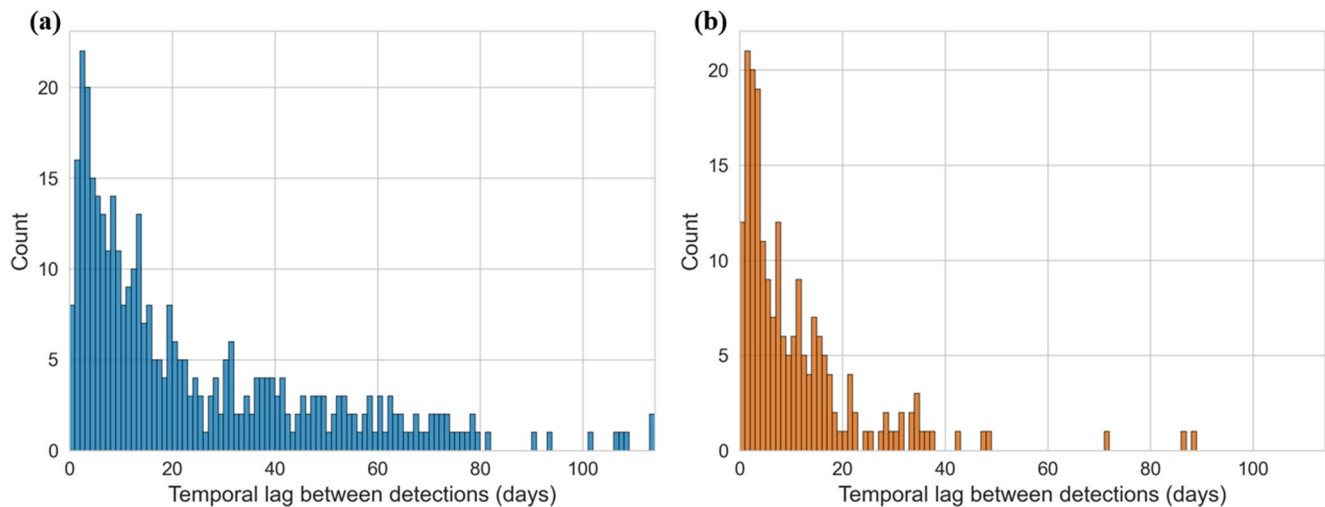


Fig. 6 Frequency distributions of the shortest temporal lag in days for (a) human-tiger pairs and (b) human-rhino pairs

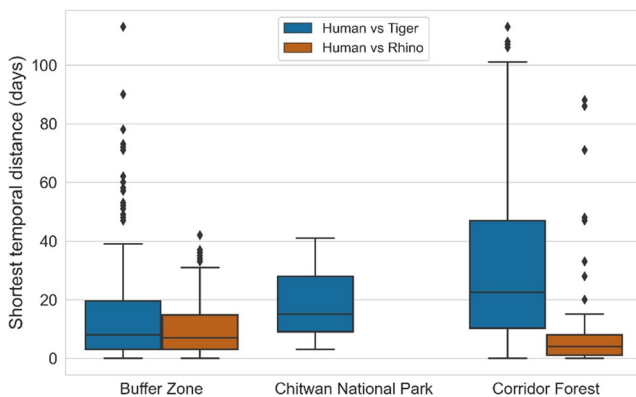


Fig. 7 Box-whisker plot distributions of the shortest temporal lag in days for human-tiger and human-rhino pairs across the Buffer Zone, Chitwan National Park, and Corridor Forest

Figure 8a shows two local residents gathering foliage and wood for household within Region 1 of the Buffer Zone, which was visited by a tiger 7 h later. In Fig. 8b, the camera trap captured images of armed individuals at dusk in Region 2 of the Buffer Zone, followed by a rhino passing the same location approximately 30 min later. In Fig. 8c, three individuals were recorded walking through Region 3 within CNP, which had been visited by a tiger 3 days earlier. In Fig. 8d, a person is observed carrying a bundle of cut vegetation and a sickle through Region 4 in the Corridor Forest, which had been visited by a rhino three hours earlier.

Discussion

Spatial patterns and co-occurrence

The spatial patterns of relative detection intensity for the target classes reflect differences in their use of protected and

human-influenced landscapes (Fig. 4). The pattern of rhinos was consistent with reports of occasional rhino incursions into agricultural areas and associated crop damage (Fig. 4a; Bhandari et al. 2022). Elephant activity suggested comparatively stronger spatial confinement to protected habitats during the study period (Fig. 4b). The absence of deer from the Community Forest may indicate avoidance of areas with higher levels of human and livestock activity (Fig. 4c). The broad spatial pattern of rhesus macaques was consistent with their affinity for landscapes in closer proximity to human activity (Fig. 4d).

The spatial concentration of human presence corresponded closely with that of domestic goats in the Community Forest and the northern part of the Corridor Forest, which is expected because domestic-livestock grazing in these landscapes is permitted only with authorization (Fig. 4f and g). The spatial correspondence between deer and tiger concentrations highlights potential predator-prey dynamics, while their proximity to human settlements may increase the likelihood of livestock depredation or human-carnivore encounters if tigers follow deer into these areas (Fig. 4c and e). The clustering of rhesus macaques near human presence may similarly indicate a greater likelihood of crop raiding and other nuisance behaviors (Fig. 4d and f).

The elevated spatial co-occurrence of dangerous wildlife and human activities in Regions 1 and 2 may be associated with rhinos crossing the river separating CNP from the Buffer Zone to seek fresh grass or agricultural crops, as well as tigers following deer into the Buffer Zone (Fig. 5c). In Region 3, the detected human presence may reflect unauthorized entry by local residents into the national park. Such overlap may increase the likelihood of direct encounters between humans and large mammals, potentially resulting in wildlife-related injuries and subsequent retaliatory



Fig. 8 Detection results of human activities and wildlife at the same camera trap sites: (a) BZF10 in the Buffer Zone, Region 1, (b) BZF17 in the Buffer Zone, Region 2, (c) CNP15 in the CNP, Region 3, and (d) CRF03 in the Corridor Forest, Region 4

actions against wildlife. Region 4 represents an area where regulated livelihood activities are permitted, and the frequent presence of rhinos together with the occasional occurrence of tigers may pose a threat to human safety.

Temporal proximity

The numerous 0-day lags observed for both human–tiger and human–rhino pairs indicate that humans and dangerous

wildlife were frequently detected at the same camera trap sites on the same calendar day, reflecting elevated temporal proximity (Fig. 6). Together with the landscape-level comparison in Fig. 7, these findings complement the spatial co-occurrence results and further identify the Buffer Zone as the landscape with the greatest human–wildlife spatiotemporal proximity (Fig. 5c).

The selected cases in Fig. 8 provide finer-resolution evidence of this proximity. In Fig. 8a, the close spatiotemporal proximity between local residents and a tiger raises the risk of sudden human encounters with tigers at short range, with limited escape opportunities in dense forest terrain where visibility and reaction time are constrained. In Fig. 8b, the immediate temporal proximity between armed individuals and a rhino raises serious concerns about direct human–wildlife encounters, possibly associated with illegal hunting activity, and further underscores the value of camera trap monitoring for detecting and mitigating threats to wildlife.

In Fig. 8c, human intrusion from the southern boundary into this area (Fig. 4f), which was also frequently used by tigers (Fig. 4e), is likely to pose a substantial risk to the safety of the intruders. In Fig. 8d, the short interval between the person collecting vegetation and the rhino illustrates the potential for close human–wildlife encounters in areas where livelihood activities overlap with wildlife presence. These findings help explain patterns of human–wildlife encounters and may support mitigation strategies aimed at reducing wildlife-related fatalities and injuries in CNP. According to the CNP information office, 11 people were killed during fiscal year 2024–2025 (July 16, 2024–July 15, 2025), including 8 by rhinos and 1 by a tiger. In addition, 41 individuals were injured in wildlife attacks, of whom 32 sustained serious injuries and nine suffered minor injuries (TRN Online, 2025).

Limitations and future research

RAI was used to standardize detections by operational camera trap effort and compare relative wildlife and human activity among camera trap sites. Because RAI represents relative detection frequency rather than population abundance, density, or occupancy, the resulting RAI-weighted KDE surfaces should be interpreted as patterns of relative detection intensity and spatial co-occurrence rather than true species distributions or probabilities of direct interaction. Nevertheless, this approach was appropriate for the objective of the present study, which was to characterize broad patterns of wildlife and human activity across the monitored landscape zones and identify areas of elevated spatial co-occurrence. The camera traps were deployed using an approximately 1.5-km grid, although some locations were adjusted to nearby animal trails. Accordingly, the resulting

patterns characterize relative wildlife and human activity around the monitored camera trap sites rather than providing unbiased estimates for unsampled portions of the landscape. The automated deep-learning model also produced occasional incorrect and missed detections under challenging conditions, particularly for partially visible targets and animals located near image boundaries (Fig. 3o–p).

Future research could focus on the Buffer Zone, where the highest human–wildlife spatiotemporal proximity was identified, and examine monthly changes using the existing camera trap records. Monthly RAI-weighted KDE surfaces could be generated for human activity and dangerous-wildlife activity to identify seasonal shifts in spatial co-occurrence. With additional data, the training dataset could be expanded to include more challenging examples, and the model could be further trained to improve its performance. Targeted manual verification could also be applied to detections used in fine-scale case analyses or management decision-making.

Conclusion

This study demonstrates that deep learning–based object detection can effectively support the ecological interpretation of wildlife–human spatiotemporal patterns from camera trap imagery. Among the evaluated models, YOLOv11s achieved the best overall performance and produced reliable detection results for the subsequent spatial and temporal analyses.

Further spatiotemporal analysis of the detection outputs revealed clear patterns of human–wildlife co-occurrence across the study area. Humans and dangerous wildlife were frequently detected at the same camera trap sites, with co-occurrence highest in the Buffer Zone, followed by the Corridor Forest and then Chitwan National Park, whereas no co-occurrence was observed in the Community Forest. The temporal analysis further showed that human–rhino pairs generally occurred within shorter and more tightly clustered temporal lags than human–tiger pairs. Examination of selected original camera trap records allowed these daily temporal patterns to be evaluated at finer temporal resolution.

Overall, the findings indicate that automated camera trap detection can contribute not only to efficient wildlife monitoring but also to landscape-scale ecological assessment by providing spatially and temporally explicit evidence of species occurrence, spatial co-occurrence, and temporal proximity. In the context of the Chitwan study area, the trained model can be directly deployed for operational automated analysis of subsequent camera trap imagery through batch inference pipelines.

The dataset will be made available publicly to support related research on wildlife monitoring and human–wildlife

interaction analysis. By covering diverse landscape contexts, including national park, Buffer Zone, Corridor Forest, and Community Forest, together with both daytime and nighttime camera trap imagery, it provides a valuable benchmark for developing and evaluating automated detection methods under varied environmental conditions. The trained model will also be released and can be directly applied in other areas where the target species are similar. If new target species are introduced, the model can also be efficiently adapted through fine-tuning using a small locally annotated dataset with the same data structure. The proposed framework therefore offers practical value for conservation planning by helping identify areas where management attention and mitigation efforts may be most needed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10344-026-02156-x>.

Acknowledgements We are grateful to Chitwan National Park, particularly the protected area managers and staff, for their support during fieldwork. We also thank Michelle Tang for assistance with proofreading and refining the manuscript to enhance its clarity and readability.

Author contributions Hao Chen: Writing – original draft, Writing – review & editing, Software, Methodology, Investigation, Data curation, Visualization, Conceptualization. Fang Qiu: Writing – review & editing, Methodology, Conceptualization, Project administration, Investigation, Funding acquisition. Umme Kulsum: Methodology, Visualization. Caiyun Zhang: Methodology, Conceptualization, Writing – review & editing. Eve Bohnett: Data curation, Conceptualization, Writing – review & editing. Haitao Lyu: Data curation, Visualization. Shuang Tian: Data curation, Conceptualization. Babu Ram Lamichhane: Data curation, Conceptualization. Douglas Stow: Conceptualization, Project administration, Investigation, Writing – review & editing. Rebecca Lewison: Conceptualization, Project administration, Investigation. Li An: Conceptualization, Project administration, Investigation, Writing – review & editing.

Funding This research was supported by the National Science Foundation (NSF) of the United States under Grant No. BCS-1826839.

Data availability Data will be made available on request.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amburgey SM, Yackel Adams AA, Gardner B, Hostetter NJ, Siers SR, McClintock BT, Converse SJ (2021) Evaluation of camera trap-based abundance estimators for unmarked populations. *Ecol Appl* 31:e02410. <https://doi.org/10.1002/eap.2410>
- Aryal A, Brunton D, Pandit R, Shrestha TK, Lord J, Koirala RK, Raubenheimer D (2012) Biological diversity and management regimes of the northern Barandabhar Forest Corridor: an essential habitat for ecological connectivity in Nepal. *Trop Conserv Sci* 5(1):38–49. <https://doi.org/10.1177/194008291200500105>
- Bhandari S, Adhikari B, Baral K, Subedi SC (2022) Greater one-horned rhino (*Rhinoceros unicornis*) mortality patterns in Nepal. *Glob Ecol Conserv* 38:e02189. <https://doi.org/10.1016/j.gecco.2022.e02189>
- Bhatt TR, Castley JG, Sims-Castley R, Baral HS, Chauvenet ALM (2023) Connecting tiger (*Panthera tigris*) populations in Nepal: identification of corridors among tiger-bearing protected areas. *Ecol Evol* 13(5):e10140. <https://doi.org/10.1002/ece3.10140>
- Bhattarai BR, Wright W, Morgan D, Cook S, Baral HS (2019) Managing human–tiger conflict: lessons from Bardia and Chitwan National Parks, Nepal. *Eur J Wildl Res* 65(3):34. <https://doi.org/10.1007/s10344-019-1270-x>
- Bjerge K, Alison J, Dyrmann M, Frigaard CE, Mann HM, Høye TT (2023) Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustain Transform* 2(3):e0000051. <https://doi.org/10.1371/journal.pstr.0000051>
- Caravaggi A, Banks PB, Burton AC, Finlay CM, Haswell PM, Hayward MW, Wood MD (2017) A review of camera trapping for conservation behaviour research. *Remote Sens Ecol Conserv* 3(3):109–122. <https://doi.org/10.1002/rse2.48>
- Carl C, Schönfeld F, Profft I, Klammer A, Landgraf D (2020) Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *Eur J Wildl Res* 66(4):62. <https://doi.org/10.1007/s10344-020-01404-y>
- Chalmers C, Fergus P, Wich S, Longmore SN, Walsh ND, Oliver L, Warrington J, Quinlan J, Appleby K (2025) AI-driven real-time monitoring of ground-nesting birds: a case study on curlew detection using YOLOv10. *Remote Sens* 17:769. <https://doi.org/10.3390/rs17050769>
- Christin S, Hervet É, Lecomte N (2019) Applications for deep learning in ecology. *Methods Ecol Evol* 10(10):1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Dai J, Li Y, He K, Sun J (2016) R-FCN: object detection via region-based fully convolutional networks. In: *Advances in neural information processing systems*, vol 29
- Dai J, Roberts DA, Stow DA, An L, Hall SJ, Yabiku ST, Kyriakidis PC (2020) Mapping understory invasive plant species with field and remotely sensed data in Chitwan, Nepal. *Remote Sens Environ* 250:112037. <https://doi.org/10.1016/j.rse.2020.112037>
- Dangol D, Ghimire A, Bhattarai SB (2020) Human–elephant conflict in the buffer zone of Chitwan National Park, Nepal. *Nepal J Zool* 4(1):36–43. <https://doi.org/10.3126/njz.v4i1.30671>
- Diwan T, Anirudh G, Tembhurne JV (2023) Object detection using YOLO: challenges, architectural successors, datasets, and applications. *Multimed Tools Appl* 82(6):9243–9275. <https://doi.org/10.1007/s11042-022-13644-y>
- Ferrante GS, Vasconcelos Nakamura LH, Sampaio S, Filho GPR, Meneguette RI (2024) Evaluating YOLO architectures for detecting road-killed endangered Brazilian animals. *Sci Rep* 14(1):1353. <https://doi.org/10.1038/s41598-024-52054-y>
- Foca JM, Visscher DR, Becker M, Boyce MS (2025) Camera trapping for density estimation: comparing the TIFC model to aerial surveys for multiple ungulate populations. *Environ Monit Assess* 197(10):1129. <https://doi.org/10.1007/s10661-025-14581-7>

- Hamel S, Killengreen ST, Henden JA, Eide NE, Røed-Eriksen L, Ims RA, Yoccoz NG (2013) Towards good practice guidance in using camera traps in ecology: influence of sampling design on validity of ecological inferences. *Methods Ecol Evol* 4:105–113. <https://doi.org/10.1111/j.2041-210X.2012.00262.x>
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>
- Hu W, Li X, Onditi KO, Wang H, Song W, Hu Z, Pu C, Pu M, Xiong Y, Yang Y, Zhu S, Yin L, He J, Song L, Li H, Gao Y, Bleisch WV, Jiang X (2025) Spatiotemporal distribution patterns of large and medium-sized mammals in a biodiversity hotspot: implications for conservation. *Biol Conserv* 301:110863. <https://doi.org/10.1016/j.biocon.2024.110863>
- Ibraheem M, Li KF, Gebali F, Sielecki LE (2021) A performance comparison and enhancement of animal species detection in images with various R-CNN models. *AI* 2:552–577. <https://doi.org/10.3390/ai2040034>
- Ikeda T, Higashide D, Shichijo T (2022) Impact of human disturbance in Japan on the distribution and diel activity pattern of terrestrial mammals. *J Nat Conserv* 70:126293. <https://doi.org/10.1016/j.jnc.2022.126293>
- Jasiulionis M, Stirke V, Balčiauskas L (2023) The distribution and activity of the invasive raccoon dog in Lithuania as found with country-wide camera trapping. *Forests* 14(7):1328. <https://doi.org/10.3390/f14071328>
- Kavwele CM, Torney CJ, Morrison TA, Fulford S, Masolele MM, Masoy J, Hopcraft JGC (2022) Non-local effects of human activity on the spatial distribution of migratory wildlife in Serengeti National Park, Tanzania. *Ecol Solut Evid* 3(3):e12159. <https://doi.org/10.1002/2688-8319.12159>
- Khanam R, Hussain M (2024) YOLOv11: an overview of the key architectural enhancements. <https://doi.org/10.48550/arXiv.2410.17725>
- Kukielka E, Barasona JA, Cowie CE, Drewe JA, Gortazar C, Cotarelo I, Vicente J (2013) Spatial and temporal interactions between livestock and wildlife in south central Spain assessed by camera traps. *Prev Vet Med* 112:213–221. <https://doi.org/10.1016/j.prevetmed.2013.08.008>
- Lamichhane S, Gurung A, Pokheral CP, Rayamajhi T, Gotame P, Regmi PR, Lamichhane BR (2018) First record of the dhole *Cuon alpinus* in Barandabhar Corridor Forest, Chitwan, Nepal. *J Threat Taxa* 10(1):11243–11244. <https://doi.org/10.11609/jott.3533.10.1.11243-11244>
- Lee SXT, Amir Z, Moore JH, Gaynor KM, Luskin MS (2024) Effects of human disturbances on wildlife behaviour and consequences for predator–prey overlap in Southeast Asia. *Nat Commun* 15(1):1521. <https://doi.org/10.1038/s41467-024-45905-9>
- Leonid TT, Kanna H, VJ CC, Hamritha AS, Lokesh C (2023) Human–wildlife conflict mitigation using YOLO algorithm. In: *Proceedings of the 8th International Conference on Science, Technology, Engineering and Mathematics (ICONSTEM)*, pp. 1–7. <https://doi.org/10.1109/ICONSTEM56934.2023.10142629>
- Lewis JS, Spaulding S, Swanson H, Keeley W, Gramza AR, VandeWoude S, Crooks KR (2021) Human activity influences wildlife populations and activity patterns: implications for spatial and temporal refuges. *Ecosphere* 12(5):e03487. <https://doi.org/10.1002/ecs2.3487>
- Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988. <https://doi.org/10.1109/ICCV.2017.324>
- Linkie M, Ridout MS (2011) Assessing tiger–prey interactions in Sumatran rainforests. *J Zool* 284(3):224–229. <https://doi.org/10.1111/j.1469-7998.2011.00801.x>
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. In: *Leibe B (ed) Computer Vision – ECCV 2016. Lecture notes in computer science*, vol 9905. Springer, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
- Mishra HR (1982) Balancing human needs and conservation in Nepal's Royal Chitwan Park. *Ambio* 11(5):246–251
- Nyhus PJ (2016) Human–wildlife conflict and coexistence. *Annu Rev Environ Resour* 41:143–171. <https://doi.org/10.1146/annurev-environ-110615-085634>
- Palmer MS, Swanson A, Kosmala M, Arnold T, Packer C (2018) Evaluating relative abundance indices for terrestrial herbivores from large-scale camera trap surveys. *Afr J Ecol* 56(4):791–803. <https://doi.org/10.1111/aje.12566>
- Petso T, Jamisola RS Jr, Mpoeleng D (2022) Review on methods used for wildlife species and individual identification. *Eur J Wildl Res* 68(1):3. <https://doi.org/10.1007/s10344-021-01549-4>
- Pierini SE, Simoncini MS, Larriera A, Guarascio VFF, Scarpa LJ, Piña CI (2024) Camera traps detect the maned wolf preying on broad-snouted caiman eggs. *Glob Ecol Conserv* 51:e02916. <https://doi.org/10.1016/j.gecco.2024.e02916>
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39(6):1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Ridout MS, Linkie M (2009) Estimating overlap of daily activity patterns from camera trap data. *J Agric Biol Environ Stat* 14:322–337. <https://doi.org/10.1198/jabes.2009.08038>
- Roboflow (2025) Computer vision tools for developers and enterprises. <https://roboflow.com>. Accessed Mar 2025
- Rovero F, Marshall AR (2009) Camera trapping photographic rate as an index of density in forest ungulates. *J Appl Ecol* 46(5):1011–1017. <https://doi.org/10.1111/j.1365-2664.2009.01705.x>
- Rowcliffe JM, Field J, Turvey ST, Carbone C (2008) Estimating animal density using camera traps without the need for individual recognition. *J Appl Ecol* 45(5):1228–1236. <https://doi.org/10.1111/j.1365-2664.2008.01473.x>
- Salau J, Krieter J (2020) Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting. *Animals* 10:2402. <https://doi.org/10.3390/ani10122402>
- Shamoon H, Cain S, Shanas U, Bar-Massada A, Malihi Y, Shapira I (2018) Spatio-temporal activity patterns of mammals in an agro-ecological mosaic with seasonal recreation activities. *Eur J Wildl Res* 64:35. <https://doi.org/10.1007/s10344-018-1195-5>
- Simões F, Bouveyron C, Precioso F (2023) DeepWILD: wildlife identification, localization and estimation on camera trap videos using deep learning. *Ecol Inf* 75:102095. <https://doi.org/10.1016/j.ecoinf.2023.102095>
- Smith AC, Humi K, Fox J, Van Den Hoek J (2023) Community forest management led to rapid local forest gain in Nepal: a 29-year mixed methods retrospective case study. *Land Use Policy* 126:106526. <https://doi.org/10.1016/j.landusepol.2022.106526>
- Soulsbury CD, White PC (2015) Human–wildlife interactions in urban areas: a review of conflicts, benefits, and opportunities. *Wildl Res* 42(7):541–553. <https://doi.org/10.1071/WR14229>
- Szydlowski M (2022) Elephants in Nepal: correlating disease, tourism, and welfare. *J Appl Anim Welf Sci* 25(2):126–138. <https://doi.org/10.1080/10888705.2022.2028628>
- Tanwar KS, Sadhu A, Jhala YV (2021) Camera trap placement for evaluating species richness, abundance, and activity. *Sci Rep* 11(1):23050. <https://doi.org/10.1038/s41598-021-02459-w>

TRN Online (2025) 11 killed in wild animal attacks in a year. The Rising Nepal. Available at: <https://risingnepaldaily.com/news/65605> Accessed Mar 2026

UNESCO (2019) World Heritage List. Available at: <http://whc.unesco.org/en/list/284> Accessed Mar 2026

Whytock RC, Świeżewski J, Zwerts JA, Bara-Słupski T, Koumba Pambo AF, Rogala M, Abernethy KA (2021) Robust ecological

analysis of camera trap data labelled by a machine learning model. *Methods Ecol Evol* 12(6):1080–1092. <https://doi.org/10.1111/2041-210X.13576>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Hao Chen¹ · Fang Qiu¹ · Umme Kulsum¹ · Caiyun Zhang² · Eve Bohnett³ · Haitao Lyu¹ · Shuang Tian¹ · Babu Ram Lamichhane⁴ · Douglas Stow⁵ · Rebecca Lewison⁶ · Li An⁷

✉ Fang Qiu
ffqiu@utdallas.edu

Hao Chen
hao.chen2@utdallas.edu

Umme Kulsum
Umme.Kulsum@utdallas.edu

Caiyun Zhang
czhang3@fau.edu

Eve Bohnett
evebohnnett@ufl.edu

Haitao Lyu
Haitao.Lyu@utdallas.edu

Shuang Tian
Shuang.Tian@utdallas.edu

Babu Ram Lamichhane
baburam@ntnc.org.np

Douglas Stow
stow@sdsu.edu

Rebecca Lewison
rlewison@sdsu.edu

Li An
anli@auburn.edu

¹ Geospatial Information Sciences, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA

² Department of Geosciences, Florida Atlantic University, Boca Raton, FL 33431, USA

³ Department of Landscape Architecture, University of Florida, Gainesville, FL 32611, USA

⁴ Wildlife Conservation and Research Endeavour (WILD CARE) Nepal, Pulchowk, Lalitpur 44700, Nepal

⁵ Department of Geography, San Diego State University, San Diego, CA 92182, USA

⁶ Department of Biology Ecology, San Diego State University, San Diego, CA 92182, USA

⁷ College of Forestry, Wildlife and Environment, Auburn University, Auburn, AL 36849, USA