Mammal Review

SPECIAL ISSUE Surveying and Monitoring Mammals in a Changing World: Lessons Learned Across Ecological and Spatial Scales

PRACTICE INSIGHTS OPEN ACCESS

Using Drones for Large Mammal Monitoring

Chaim Chai Elchik[1] | Michael Hudson[2] | Sam Buckland[2] | Steven Longmore[3] | Serge Wich[4]

[1]Lancaster Environment Centre, Lancaster University, Lancaster, Lancashire, UK | [2]Durrell Wildlife Conservation Trust, Trinity, Jersey, Channel Islands | [3]Astrophysics Research Institute, Liverpool John Moores University, Liverpool, UK | [4]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

Correspondence: Chaim Chai Elchik (c.elchik123@gmail.com)

Received: 5 February 2026 | **Revised:** 18 June 2026 | **Accepted:** 19 June 2026

Keywords: BoT-SORT | computer vision | deep learning | drone videos | object detection | object tracking | re-identification | wildlife conservation | YOLO

ABSTRACT

Introduction: The emergence of drones has profoundly transformed the field of large mammal monitoring, offering unprecedented capabilities for data collection on both terrestrial and aquatic species but challenges remain for data collection and analyses.

Aims: In this Practice Insights article we provide a succinct review of drone types, their sensors and main applications to monitor large terrestrial mammals. We focus on obtaining key ecological metrics through using drones: presence/absence, population density and overall abundance, including spatial distribution. We also review machine learning efforts to automate animal detection and tracking, as well as the statistical methods to estimate animal density.

Methods: We provide a new method to determine animal locations from RGB videos obtained with drones. The geolocation method shows promising results with an accuracy of between 1.58 and 3.1 m on a small test sample, which is an improvement over those from crewed aircraft.

Discussion: Challenges for drone usage in terms of data collection and analyses remain but both are becoming increasingly more user friendly. There are still challenges such as the lack of open-source deep learning models to detect mammals and classify them but we expect this to be solved in the coming years. In addition the analytical workflows to move from such detections and classifications will likely become more user friendly as well.

1 | Introduction

Effective conservation of large mammals relies on cost-effective field data collection methods and statistical inference frameworks that provide accurate and precise estimates of ecological metrics such as density (Buckland et al. 2015; MacKenzie et al. 2017; Piel et al. 2022). Historically, wildlife data have been collected using a range of ground-based technologies, including direct visual surveys, camera trapping, acoustic monitoring and physical capture systems (Cordier et al. 2022; Hoefer et al. 2023; Hoffmann et al. 2010; Plumptre 2000; Romairone et al. 2018). These technologies are often paired with statistical inference methods such as distance sampling (Buckland et al. 2001), capture–mark–recapture models (Amstrup et al. 2010; Otis et al. 1978) and occupancy modelling (MacKenzie et al. 2002, 2017) to estimate abundance and density. Although widely used, these approaches have several limitations. They typically offer limited spatial coverage, require substantial field effort and can cause disturbance to target animals, potentially biasing observations (Beery et al. 2018; Burton et al. 2015; Marques et al. 2013; Stevenson et al. 2015; Wearn and Glover-Kapfer 2019). In addition, operational costs associated with large-scale deployments can be substantial (Davis et al. 2020; Nichols and Williams 2006). Crewed aerial surveys have provided broader spatial coverage for wildlife monitoring for decades, but they are vulnerable to observer subjectivity and fatigue—detection probability can decline markedly after hours of continuous surveying—and operational risks, including the safety hazards of aviation work where aircraft accidents account for job-related wildlife field fatalities; they also impose substantial costs and airstrip and aircraft availability is often limited (Delisle et al. 2022; Ransom 2012). Moreover, conventional aerial surveys relying on human observers often fail to detect nocturnally active, camouflaged or canopy-obscured animals because of limited visibility, poor contrast or forest cover—all of which bias results in dense habitat contexts (Clement et al. 2022; Delisle et al. 2022; Jachmann 2002; Ransom 2012).

To complement current survey methods, drones have rapidly emerged as a new tool to monitor wildlife (Chabot and Bird 2015; Wich and Koh 2018). Drones offer a unique suite of advantages, including elevated aerial perspectives, enhanced manoeuvrability to either follow animals and/or fly in hilly terrain and the capacity for non-intrusive data collection that minimizes disturbance to wildlife when guidelines are followed (Cantu de Leija et al. 2023; Mulero-Pázmány et al. 2017). They can be equipped with a variety of sensors that can capture high-resolution visual-spectrum imagery, thermal infrared image data and even audio recordings (Linchant et al. 2015; Seymour et al. 2017). Their ability to cover large areas rapidly, especially when multiple drones are deployed simultaneously, often at reduced operational costs and with improved safety compared to traditional crewed aircraft, positions them as a highly advantageous tool in contemporary wildlife research (Christie et al. 2016; Hodgson et al. 2018). A particularly compelling additional benefit of drone surveys compared to crewed surveys is the provision of a permanent, verifiable digital record of sightings, which allows repeated review and validation of data, which is not possible with data collected by humans directly. To effectively monitor large mammal populations, researchers rely on several key ecological metrics such as presence/absence, density and abundance. In this review, we explore the usage of drones for estimating these metrics.

The effectiveness of drones for mammal monitoring is linked to various platforms and their integrated sensor technologies (Chabot and Bird 2015; Wich and Koh 2018). The different drone types can be used for specific monitoring objectives, ranging from hovering and slow flying over small areas to flying long-distance large-scale surveys (Christie et al. 2016; Linchant et al. 2015). There are three main drone platforms. Multirotor: These platforms are characterized by multiple rotors that provide vertical take-off and landing (VTOL) and hovering capability. Their vertical flying agility makes them ideally suited for close-up monitoring (often in combination with zoom options on the camera), capturing stable images and videos, and excellent opportunities to follow animals. Multirotors can also be used for flying longitudinal transects or grid patterns to survey mammals over relatively small areas (exact surface area covered depends on flying height and overlap of flight pattern) (Barnas et al. 2020; Hodgson et al. 2016). Fixed-wing: These drones resemble small aircraft and are optimized for efficient horizontal flight over extensive areas. Their extended endurance and flight speed compared to multirotors are particularly beneficial for large-scale surveys. But less so for detailed searches in small areas or flying low, which could also lead to reduced overlap when flying at higher speed (Seymour et al. 2017; Wich and Koh 2018). Hybrid VTOL: These drones combine the most advantageous features of both multirotor (VTOL) and fixed-wing (efficient horizontal flight) designs. They offer the convenience of vertical take-off and landing without requiring runways, coupled with extended endurance. This versatility makes them particularly effective for large-scale surveys that occur in areas without opportunities to take off and land on a runway or grassy area as well as to combine the opportunity to fly horizontally and hover if need be (Chabot and Bird 2015).

Drones primarily use two sensor types to monitor mammals: standard RGB (visual-spectrum) cameras and thermal infrared (TIR) sensors, which detect the body heat emitted by animals. These sensors can be deployed either separately or together on a single drone via dual-sensor payloads, enabling both visual identification and thermal detection within the same survey. RGB cameras deliver much higher spatial resolution but require sufficient light, making them less effective during night or low-contrast conditions. In contrast, TIR sensors excel at detecting mammals under low-light or nocturnal conditions, such as within dense forest canopies or when visual contrast is minimal. For instance, thermal-equipped drones have successfully detected orangutans and proboscis monkeys at dawn or dusk, situations where RGB imagery fails due to poor lighting or camouflage, and have been applied across a range of species (Burke, Rashman, Longmore, et al. 2019; Kays et al. 2019; Wich and Koh 2018). While RGB cameras provide higher spatial resolution, TIR sensors enable surveys at night or under low-light conditions where RGB imagery is ineffective. More recently, TIR sensors have also been used in combination with onboard spotlights to guide detection, followed by RGB-based species identification under low-light conditions (Kays et al. 2019).

The effectiveness of drone-based mammal monitoring is influenced by flight parameters, which directly impact data quality and detectability. The main factor is flight height as flying higher will lead to a larger camera footprint but a coarser resolution of the images which can influence detectability. But because flying higher generally reduces disturbance to animals there is a clear balance between flying as high as possible to reduce disturbance while still obtaining the required data. A higher flight speed allows for faster coverage of larger areas, but increased speed can negatively impact image quality by introducing motion blur, which can hinder accurate animal identification and counting, especially when animals are moving (Mulero-Pázmány et al. 2014; Whitworth et al. 2022). Adjusting the camera angle can also affect detectability. A nadir (directly downward) camera angle is often preferred as it maximizes the detectability of animals and facilitates accurate body measurements of species by minimizing distortion. Conversely, tilting the camera somewhat towards the horizon (e.g., −45°) allows for coverage of larger observation areas and can potentially facilitate detecting animals before they might get disturbed (Linchant et al. 2015; Pinel-Ramos et al. 2025). Time of day, and associated sun angle, has also been shown to influence detectability in RGB drone imagery (Jones et al. 2023; Whitworth et al. 2022). For surveys using drones equipped with thermal infrared sensors the time of day and season is also an important factor that influences detectability due to their influence on the temperature contrast between animals and vegetation. Flights conducted near sunrise have generally been recommended due to the ground and vegetation being at its coolest then, which leads to a higher thermal contrast (Burke, Rashman, Wich, et al. 2019; Israel 2011). Surveys are generally more effective in seasons with minimal vegetation cover to reduce the occlusion of thermal signatures, which can otherwise obscure animals from view. Similarly, RGB surveys can be more effective during leaf-off seasons if the aim is to monitor mammals that otherwise would be occluded by leaves (Linchant et al. 2015).

By far, the most common usage of drones for mammal surveys is to determine the presence of species by surveying areas. This approach has been used for a variety of terrestrial and aquatic mammals through either direct sightings or signs (e.g., nests and burrows) (Wich and Koh 2018). Ideally, these presence detections would be compared with other survey techniques. Such comparisons have been conducted for orangutan nests using RGB cameras (Wich et al. 2015), spider monkeys using TIR (Spaan et al. 2019) and koalas using TIR compared to traditional spotlight surveys (Witt et al. 2020). Despite its importance for species distribution modelling (Franklin and Miller 2009) there has been little focus on using drones to estimate absences except for a recent study on chimpanzee nests (Wich et al. 2023).

Estimating the density and abundance of mammal populations is crucial for effective conservation. Drone data have been used with several statistical methods to estimate density and abundance. Distance sampling has commonly been used in crewed aircraft surveys of animal species where observers directly estimate the distance between the aircraft and animals so that perpendicular distances can be estimated (Buckland et al. 2004). For drone surveys the same method has been used but then instead of observers directly estimating the perpendicular distance, the distance between the midline of images and the animals has been measured on that are acquired from the drone. This method has, however, not been used often in combination with drone image data (Delaney et al. 2025; Espíndola et al. 2023; Wohlfahrt et al. 2025). N-mixture models are a common hierarchical modelling approach used to estimate population abundance from data, particularly when detection probability is imperfect. These models estimate detection probability by assessing the variation in counts among temporally or spatially replicated surveys and estimate abundance based jointly on the variation in counts among sites and average detection probability. They are robust to the issue of imperfect detection, a common challenge in wildlife surveys and have been successfully used in combination with drone data (Brack et al. 2023; Ryan et al. 2025). Looking ahead, coordinated teams of drones offer exciting opportunities for density estimation approaches that have not yet been formally explored. While a single drone flown twice along offset paths has recently been used to emulate a double-observer design (Campbell et al. 2026), the formal extension of mark–recapture distance sampling (MRDS) frameworks to simultaneous multi-drone configurations has, to our knowledge, not yet been implemented. Such an approach could involve one drone acting as a primary observer and a second following directly behind along the same transect with a sufficient lag to ensure independence of detections, enabling estimation of perception bias and accounting for individuals missed by the primary platform (Laake and Borchers 2004). Beyond paired platforms, larger groups of drones following sequential transect paths over the same area within a short time window could facilitate spatial capture–recapture analyses, provided individual recognition is possible. We anticipate that these approaches will develop into a new research frontier in drone-based wildlife monitoring.

Recent advances in machine learning have greatly enhanced the automation of wildlife monitoring from drone footage (Elchik, Wich, and Burger 2025). In particular, deep learning–based object detectors, such as the You Only Look Once (YOLO) (Khanam and Hussain 2024; Wang and Liao 2024) series, have demonstrated strong accuracy and real-time performance in detecting large mammals across varying distances, lighting conditions and occlusion levels (Aliane 2025; Dave et al. 2023; Du et al. 2025; Duporge et al. 2025; Elchik, Karimi Nejadasl, et al. 2025; Elchik, Wich, and Burger 2025; Fang et al. 2024; Fonod et al. 2025; Mou et al. 2023; Povlsen et al. 2024). These models can be trained or fine-tuned using datasets derived from drone video frames, which are often pre-processed through subsampling strategies to reduce redundancy and speed up annotation.

In practice, detection workflows involve extracting frames from high-resolution drone videos and annotating them with bounding boxes around animal subjects. Tools like Roboflow (2025) can accelerate this process by leveraging pre-trained models to generate initial bounding box suggestions, which can then be refined by human reviewers. Once trained, detection models produce spatial coordinates and confidence scores for each identified animal, creating structured data that can be used for downstream analysis.

This approach enables scalable processing of large video datasets and facilitates more efficient identification of individual animals across diverse visual and environmental conditions. When combined with object detection, drone footage and tracking algorithms enable the continuous monitoring of animal movements over time. Multi-object trackers such as BoT-SORT (Bag of Tricks for SORT) (Aharon et al. 2022; Fonod et al. 2025) associate detections across frames using both motion prediction and appearance-based features, enabling persistent identity tracking even during brief occlusions or rapid movements. This temporal continuity is critical for accurate individual counts and for analysing group-level dynamics.

In previous applications, including those targeting large terrestrial mammals, this combination of detection and tracking has enabled not only automated species counting but also the derivation of behavioural classification and metrics such as individual and group movement speed, cumulative travel distance and spatial overlap between individuals (Elchik, Karimi Nejadasl, et al. 2025; Elchik, Wich, and Burger 2025; Kholiavchenko et al. 2024). These previous applications also allowed for the identification of dynamic clusters and subgroup formations, illustrating fission–fusion behaviour within observed animal groups. Such integrated workflows reduce the need for frame-by-frame manual analysis and enable the scalable extraction of high-resolution spatiotemporal data on large mammal populations.

2 | Methods

This study implements a modular computational framework designed to automate data collection for ecological distance sampling using drones. We propose three distinct workflows to accommodate different data availability and processing needs: (1) a zero-shot detection pipeline for imagery using an open-vocabulary model (CountGD), where user-defined class labels or species names are provided as text inputs to guide detection without additional training, (2) a fine-tuned detection pipeline for imagery using custom-trained models (YOLO), and (3) a spatiotemporal pipeline for video data integrating detection (YOLO), tracking and telemetry synchronization (Figure 1).

2.1 | Data Acquisition and Pre-Processing

The framework is sensor-agnostic but requires specific metadata regarding the camera's interior orientation parameters. For this study, data were acquired using a DJI Mavic 3T equipped with a high-resolution CMOS sensor. Critical sensor parameters, specifically the sensor width ($W_{sensor} = 17.3\,mm$) and height ($H_{sensor} = 13.0\,mm$), are currently configured for the Mavic 3T. However, to accommodate future or bespoke UAV platforms, these parameters are exposed as configuration variables, allowing the pipeline to calibrate the intrinsic matrix for any sensor provided the physical dimensions and focal length are known.

Accurate georeferencing relies on the precise synchronization of visual data with onboard telemetry (Hansen and Figueiredo 2024). The extraction methodology varies depending on the input modality. For static imagery, we utilize the exiftool (Harvey 2025) utility to parse EXIF metadata, extracting the GNSS coordinates (latitude, longitude), absolute altitude (AMSL), relative altitude (AGL), gimbal orientation (yaw, pitch, roll) and focal length. Conversely, for the video-based workflow, telemetry is derived from the SubRip Subtitle (SRT) sidecar files generated by DJI drones. We parse the SRT file to map specific frame counts to discrete telemetry packets, ensuring that the camera's pose and location are rigorously synchronized with every individual video frame.

2.2 | Object Detection and Tracking Architectures

We employ three distinct computer vision strategies to identify targets within the raw data.

For scenarios lacking labelled training data, we utilized the CountGD framework (Amini-Naieni et al. 2024). This approach leverages GroundingDINO (Liu et al. 2023), a transformer-based open-set object detector. This allows the user to query the image using natural language prompts (e.g., 'Elephant', 'Tree') without prior model training, similar to approaches used by Lalgudi et al. (2025). The model outputs bounding boxes and confidence scores based on semantic alignment between the text prompt and visual features.

In situation with training data available, we employed the YOLO architecture (Khanam and Hussain 2024; Wang and Liao 2024). The YOLO architecture was selected due to its prevalence in animal detection and drone-based applications (Aliane 2025; Dave et al. 2023; Du et al. 2025; Duporge et al. 2025; Elchik, Karimi Nejadasl, et al. 2025; Elchik, Wich, and Burger 2025; Fang et al. 2024; Fonod et al. 2025; Mou et al. 2023; Povlsen et al. 2024). For the validation phase of this study, the model was fine-tuned on a custom dataset of human subjects serving as proxy targets; however, the architecture remains highly adaptable and can be retrained on any custom dataset to detect

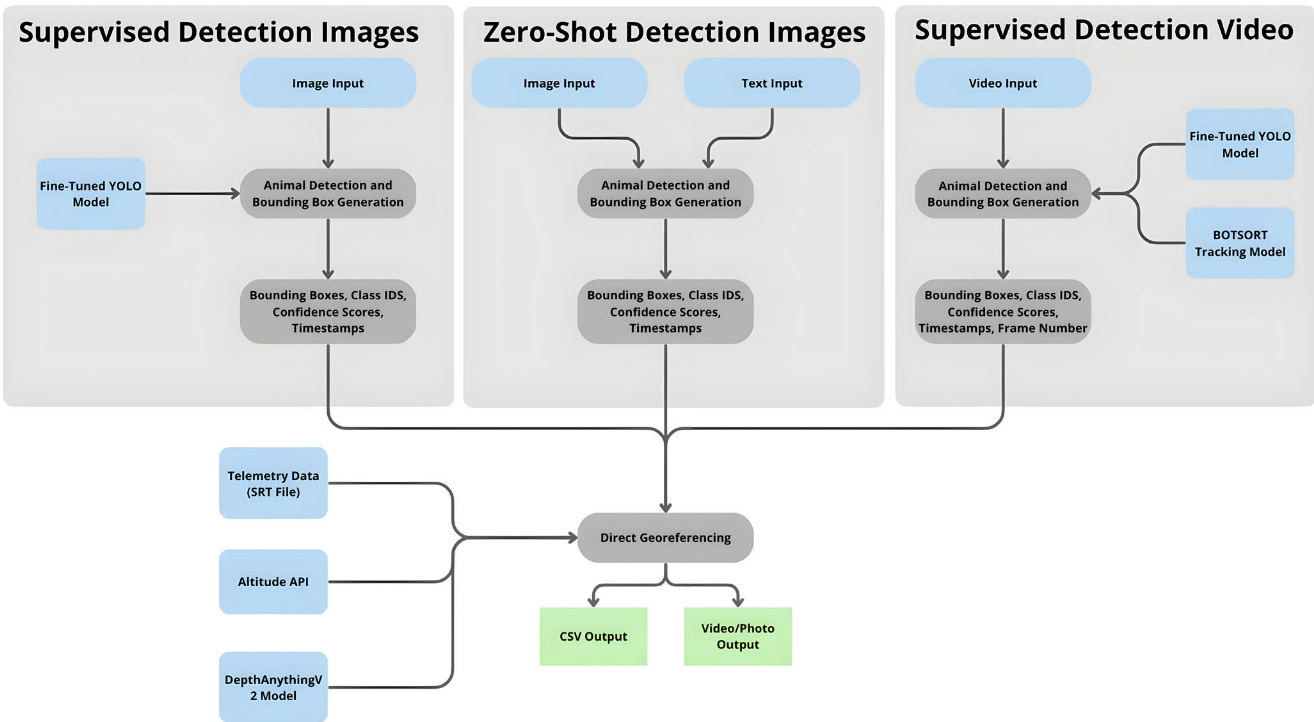

FIGURE 1 | Schematic overview of framework.

specific target species. This architecture offers rapid inference speeds suitable for high-volume video processing, outputting bounding boxes, class IDs and confidence scores.

To address the issue of double-counting animals in video transects, we integrated the BoT-SORT tracker (Aharon et al. 2022; Fonod et al. 2025). Unlike single-image processing, the video pipeline assigns a unique track_id to each detected object and maintains its identity across consecutive frames. The system logic ensures that georeferencing is performed only once per track, typically at the first high-confidence detection. This prevents repeated detections of the same individual across multiple frames from inflating abundance estimates.

2.3 | Monocular Depth Estimation and Metric Scaling

To locate objects in 3D space from 2D inputs, we employed the Depth-Anything-V2-Large-hf model (Yang et al. 2024). This model generates a dense relative depth map for each input frame. These depth values are unitless and represent only relative depth relationships within the scene.

To convert relative depth into metric space, we use drone-derived above-ground-level (AGL) altitude as an external reference. The AGL is computed from GNSS altitude corrected using digital elevation model (DEM) data.

The absolute scale of the scene is estimated by assuming that the ground surface corresponds to the lower bound of the relative depth distribution. The relative depth map is then rescaled such that this ground reference aligns with the known AGL altitude, enabling metric reconstruction of scene geometry.

This procedure allows pixel coordinates and relative depth estimates to be expressed in a unified metric coordinate system for subsequent georeferencing.

We calculate the precise above-ground-level (AGL) altitude (H_{agl}) at the moment of capture. To account for terrain undulation, the pipeline queries the OpenTopoData API (Walker 2020) to retrieve the ground elevation (H_{dem}) at the drone's current GPS location.

$$H_{\text{agl}} = H_{\text{absolute}} - H_{\text{dem}}$$

If the DEM query fails, the system falls back to the drone's barometric relative altitude sensor.

As monocular depth estimation provides only relative, unitless depth values, we rescale the depth map using drone-derived above-ground-level (AGL) altitude as an external reference for metric conversion. We calculate a scale factor, S, by relating the AGL to the mean pixel intensity of the central region of the relative depth map ($\bar{d}_{\text{centre}}$). The absolute metric depth map (D_{abs}) is derived as:

$$D_{\text{abs}} = D_{\text{rel}} \times \left(\frac{H_{\text{agl}}}{\bar{d}_{\text{centre}}} \right)$$

2.4 | Direct Georeferencing Pipeline

We implement a direct georeferencing algorithm to transform 2D pixel coordinates into geodetic coordinates (latitude, longitude) (Correia et al. 2022; Gabrlik 2015; Li et al. 2011; Wang et al. 2016; Watts et al. 2024; Zhou, Jia, et al. 2025; Zhou, Tang, et al. 2025; Zhao et al. 2019).

First, we construct the camera intrinsic matrix, K, which maps 3D camera coordinates to the 2D image plane. To ensure the framework supports new or non-standard cameras, K is computed dynamically using the provided sensor dimensions (W_{sensor}, H_{sensor}) and the real-time focal length (F) extracted from telemetry:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

where c_x and c_y represent the optical centre (half the image width and height in pixels), and the focal lengths in pixel units (f_x,f_y) are derived as:

$$f_x = \frac{F \times W_{\text{pixels}}}{W_{\text{sensor}}}, \quad f_y = \frac{F \times H_{\text{pixels}}}{H_{\text{sensor}}}$$

We then construct the rotation matrix, R_{gimbal}, from the gimbal's Euler angles (yaw ψ, pitch θ, roll ϕ). For a detected object centred at pixel (u, v) with an estimated metric depth Z (retrieved from D_{abs}), we back-project the point into the camera coordinate system:

$$\mathbf{P}_{\text{cam}} = Z \cdot K^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

The vector is then rotated into the North-East-Down (NED) coordinate frame:

$$\mathbf{P}_{\text{ned}} = R_{\text{gimbal}} \cdot R_{\text{cam2body}} \cdot \mathbf{P}_{\text{cam}}$$

where R_{cam2body} represents the transformation from the optical axis to the drone body frame.

The resulting vector provides the North and East offset in meters from the drone's position. Using the GeoPy (Esmukov and GeoPy Contributors 2023) library, we calculate the destination coordinates using the Vincenty inverse formula, originating from the drone's GNSS location.

2.5 | Evaluation Methodology

To validate the direct georeferencing pipeline, we conducted a series of field experiments using human subjects as proxies for wildlife targets. The experiments were designed to assess the system's precision under varying kinematic conditions. Ground truth data was acquired using the *GNSS Logger* application (Google LLC 2024) by Google, running on commercial smartphones carried by the subjects to log high-frequency GNSS coordinates.

The field tests were conducted in Hoylake in the Northwest of the UK during daylight hours on 8 October 2025. The drone operated at an approximate relative altitude (AGL) of 80 m and an absolute altitude (AMSL) of 120 m. To rigorously test the robustness of the geolocation algorithm, we devised two distinct testing scenarios:

1. Dynamic-Static: The UAV translates along the XY-plane while the subject remains stationary.
2. Dynamic-Dynamic: Both the UAV and the subject translate simultaneously along linear trajectories.

To quantify the system's performance, we calculated the geodesic distance between the predicted coordinates derived from the computer vision pipeline and the ground truth coordinates recorded by the GNSS logger. The error δ (in meters) was computed using the geodesic method on the WGS-84 ellipsoid:

$$\delta = \text{Geodesic}\left(\mathbf{P}_{\text{predicted}}, \mathbf{P}_{\text{ground truth}}\right)$$

where $\mathbf{P}$ represents the coordinate tuple (latitude, longitude).

3 | Results

3.1 | Experimental Results

The experimental validation assesses the accuracy of the proposed direct georeferencing method across two distinct scenarios: a stationary subject (human) with a moving drone (Test 1) and a moving subject with a moving drone (Test 2).

The results, summarized in Tables 1 and 2, demonstrate a positional error ranging from 0.27 to 5.87 m. Positional error is computed as the Euclidean distance between the predicted georeferenced location and the ground-truth GPS coordinates of the subject.

In Test 1 (stationary subject), the mean positional error was 3.10 m (SD = 1.87 m, n = 8). Errors showed considerable variability, reflecting sensitivity to scene and viewpoint changes during flight. In Test 2 (moving subject), the mean positional error was 1.58 m (SD = 1.05 m, n = 3), although this result is based on a limited sample size.

TABLE 1 | Drone Test 1: Stationary subject, moving drone. The subject remained at a fixed location (53.38485, −3.17296) throughout the test.

Measurement ID	Projected GPS	Error (m)
Projected Point 1	53.38488, −3.17303	5.868
Projected Point 2	53.38486, −3.17299	2.883
Projected Point 3	53.38486, −3.17293	2.019
Projected Point 4	53.38484, −3.17295	0.603
Projected Point 5	53.38485, −3.17298	1.663
Projected Point 6	53.38488, −3.17302	5.363
Projected Point 7	53.38487, −3.17301	4.671
Projected Point 8	53.38484, −3.17294	1.729
Average		3.099

3.2 | Hypothetical Pipeline Results

To facilitate distance sampling analysis, the pipeline models the flight path as a linear transect (Chang et al. 2025). For every detected object, we calculate the perpendicular distance from the object's geolocated coordinate to the transect line defined by the drone's bearing (Rahman et al. 2023).

The system generates two primary outputs:

1. Visual Output: Annotated images (or video files) visualizing the bounding boxes, unique IDs and projected transect lines for verification (Figure 2).
2. Statistical Output: A structured CSV file containing the Image ID, Object ID, Class, computed GPS coordinates, time of detection and the perpendicular distance to the transect line in the required format for distance sampling in R (Table 3).

4 | Discussion

The emergence of drones has added a novel method to support large mammal monitoring, offering a cost-effective alternative to traditional ground-based and crewed aerial surveys. While ground-based methods will remain essential for many species (e.g., Kays et al. 2019), they are often limited by narrow spatial coverage, significant labour requirements and the risk of

TABLE 2 | Drone test 2: Moving subject, moving drone. Both the subject and the projected point coordinates vary per instance.

Subject GPS (truth)	Projected GPS	Error (m)
53.38486, −3.17296	53.38484, −3.17295	1.96
53.38480, −3.17287	53.38480, −3.17287	0.27
53.38474, −3.17277	53.38476, −3.17279	2.51
Average		1.58

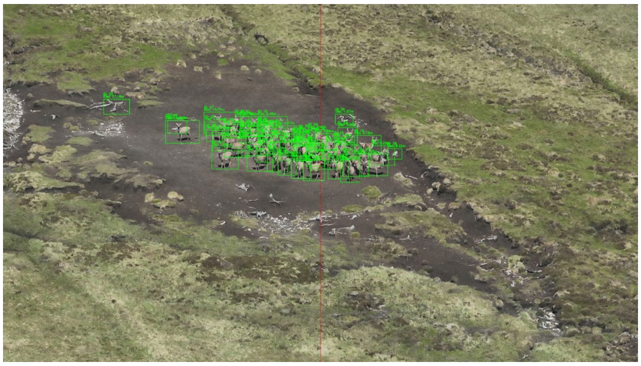

FIGURE 2 | Example overlay output for red deer in Scotland.

TABLE 3 | CSV file example output snippet.

Image ID	Drone Lat	Drone Lon	Obj ID	Class	Obj Lat	Obj Lon	Perp. Dist (m)
TestImage	56.84110	−4.14881	1	0	56.84202	−4.14733	4.1682
TestImage	56.84110	−4.14881	2	0	56.84214	−4.14710	6.1748
TestImage	56.84110	−4.14881	3	0	56.84216	−4.14730	4.8798
TestImage	56.84110	−4.14881	4	0	56.84206	−4.14721	6.3957

disturbing target animals. In contrast, drones provide an elevated perspective that allows rapid coverage of expansive and often inaccessible terrain. An additional advantage of using drones is the capture of permanent and verifiable data. Unlike human observers in crewed aircraft, who are prone to fatigue that can lead to estimation errors (Garcia-Garin et al. 2020; Schlossberg et al. 2016), drone-captured imagery provides a digital record that can be reviewed manually by humans or automatically by computer models. In this Practice Insights article, we have outlined the current state of large mammal monitoring using drones, identified existing limitations and proposed a new methodology supported by experimental results. Here, we discuss the implications of this proposed framework for automating the pathway from raw imagery to statistical input. We synthesize the technical constraints of drone sensors, the practical advancements in continuous georeferencing and challenges that remain.

4.1 | Sensor Capabilities and Technical Challenges

The effectiveness of drone monitoring for large mammals is largely dictated by the choice of sensor and the environmental context. Standard RGB cameras offer high spatial resolution for detailed identification, but are limited by lighting conditions, camouflage and vegetation cover (Kellenberger et al. 2018; Pinel-Ramos et al. 2025). In contrast, Thermal Infrared (TIR) sensors detect the body heat of animals, excelling in low-light or nocturnal conditions and proving effective for detecting species like orangutans and even smaller spider monkeys through forest canopies if these are not too dense (Burke, Rashman, Longmore, et al. 2019; Spaan et al. 2019). Despite these advances, several challenges remain. Occlusion from dense vegetation, particularly in evergreen forests, leads to non-detection even with TIR sensors (Beaver et al. 2020; Pagacz and Witczuk 2023). Large mammals that blend into the environment may remain difficult to detect even in open areas with RGB imagery during daylight (Kellenberger et al. 2018). Furthermore, flight parameters such as speed, camera angle and altitude can affect data quality; high flight speeds may introduce motion blur, oblique camera angles can reduce detectability and increased altitude reduces spatial resolution while expanding ground footprint (Pinel-Ramos et al. 2025; Wich et al. 2023).

Statistical methods that explicitly account for imperfect detection, such as N-mixture models, can help mitigate bias arising from missed detections by estimating detection probability and latent abundance from repeated observations (Brack et al. 2023; Ryan et al. 2025). However, these models do not address observation-level constraints such as occlusion, motion blur or reduced detectability due to sensor limitations or flight altitude.

4.2 | Georeferencing Precision and Metric Scaling

The results of our experimental trials indicate that the proposed framework achieves high positional accuracy for ecological monitoring. The georeferencing error ranged from 0.27 to 5.87 m. In traditional aerial wildlife surveys, positional information is often not recorded as continuous geospatial coordinates but instead discretised into coarse distance classes, for example 40 m distance bins in strip-transect analyses (Alisauskas and Conn 2019; Schnupp et al. 2013). In contrast, the proposed framework provides continuous meter-scale geolocation of individual detections. The average error of 3.10 m in stationary tests and 1.58 m in dynamic tests demonstrates that integrating drone telemetry with monocular depth estimation via Depth-Anything-V2 yields accurate and consistent geospatial localisation of detections. This level of precision enables a substantially finer spatial resolution of wildlife observations than is typically achievable in conventional aerial survey workflows. It is important to note that the reported error may partially reflect uncertainty in the validation data. Ground truth positions were obtained using consumer-grade GNSS devices, which introduce inherent positional noise. Similarly, drone GNSS and inertial measurement systems are subject to drift and sensor noise, meaning that a portion of the measured error likely reflects baseline instrumentation uncertainty rather than errors in the computer vision pipeline alone.

Crucially, the accuracy of the metric scaling relies on specific operational assumptions regarding geometry and terrain. First, the framework calculates the scaling factor based on the drone's altitude relative to the ground, assuming a planar projection where the camera sensor is parallel to the terrain (nadir view). Consequently, the pipeline is strictly designed for top-down imagery; oblique or high-angle footage (e.g., 45°) will result in invalid scaling factors and incorrect distance estimations. Second, the method assumes a relatively flat terrain or the use of terrain-following flight modes. If the drone maintains a fixed altitude relative to the take-off point while flying over uneven topography (e.g., hills or steep slopes), the actual height above ground level (AGL) will fluctuate. Without real-time AGL adjustment (e.g., via LiDAR or radar altimetry), the scaling factor derived from the initial altitude will become inaccurate, leading to systematic errors in perpendicular distance calculations.

Regarding the capabilities of the computer vision model, Depth Anything V2 significantly overcomes limitations seen in previous architectures. It is explicitly trained to handle complex scenes and fine-grained details, such as thin vegetation structures, which are critical for detecting animals in cover. Furthermore, unlike earlier depth estimation models that struggled with transparency, V2 demonstrates high robustness on reflective and transparent surfaces, including water bodies and ice, as well as in adverse styles involving fog or snow. However, while the model is highly robust, users should exercise caution in environments that deviate significantly from the training distribution or lack distinct visual cues. Although the model leverages large-scale unlabeled real images to bridge the domain gap, extremely homogeneous textures (e.g., featureless sand dunes or white-out snow conditions) may still present challenges for monocular depth inference relying on visual context.

4.3 | Framework Advances

The modularity of the framework addresses the diverse needs of the conservation community. The zero-shot (CountGD) pipeline is particularly useful as it allows for the detection of species not included in the model's training label set, thereby bypassing the need for labour-intensive labelled datasets and task-specific model training. For surveys targeting species included in the supervised training classes of deep learning models, the supervised pipeline (YOLOv11) provides rapid inference speeds required for large-scale video processing. This dual approach allows the framework to be applied across a wide range of large mammal species.

To reduce the possibility of double counting of the same individual we applied the BoT-SORT tracker to automate this process. By assigning unique identity tags to objects across frames, the system ensures that each individual is only counted once for abundance estimates, even if it remains in view for an extended period. This structured output, generated as a CSV containing GNSS coordinates and perpendicular distances, is formatted for immediate use in statistical models such as Distance Sampling. By automating the pipeline from raw video to statistical input, we effectively eliminate the manual data-processing bottleneck that currently hinders many conservation efforts.

4.4 | Challenges

Despite the current usage and future potential of drones and deep learning methods, challenges remain. Even though drone costs have decreased they are generally still in the order of a few thousand US dollars for multirotor systems and more for fixed-wing and hybrid (VTOL) drones. These costs combined with a general lack of a clear set of maintenance guidelines and costs from companies selling drones means that for organizations using drones it is difficult to budget for purchasing and maintenance costs for drone operations. For conservation coverage of large areas is essential but the flight times, particularly for multirotor drones, are still limited and can be decreased further in very warm environments under tropical forest canopies. For fixed-wing drones flight durations are better and due to the flying speed larger areas can be covered. However, regulations often restrict flying to a horizontal distance of 500 m which is small compared to the areas that need to be covered. Exemptions are possible but depend on national and local regulations (Wich and Koh 2018). In addition, pilots need to be trained and often certified by the authorities which is an added cost to organizations.

5 | Conclusion and Future Work

Drones are a relatively novel survey tool for mammal surveys which can complement other survey methods. Their usage is being facilitated by a growing use of deep learning models to automate detection of animals in the imagery which is reducing the time required for analyses. Here we presented a next step in which perpendicular distances are automatically calculated to support the workflow from drone imagery to statistical analyses. This is particularly important as detection probability could decrease towards the edge of images due to factors such as vignetting, geometric perspective, occlusion, radial blur and chromatic aberration, particularly in situations where the camera is not pointing down but under an angle. Future development should focus on improving performance in high-occlusion environments and validating the system in a wider range of mammal species to further refine detection and tracking accuracy. As drone technology and computer vision continue to evolve, integrated frameworks like the one presented here will be vital for providing the timely, accurate data required for effective global conservation.

Author Contributions

Chaim Chai Elchik: conceptualization, methodology, software, data curation, formal analysis, writing – original draft, visualization. **Michael Hudson:** writing – review and editing, funding acquisition. **Sam Buckland:** writing – review and editing, funding acquisition. **Steven Longmore:** funding acquisition, writing – review and editing, methodology. **Serge Wich:** conceptualization, methodology, funding acquisition, writing – original draft, writing – review and editing.

Funding

LJMU's School for Biological and Environmental Sciences provided funding for the data collection on deer at Dalnacardoch.

Ethics Statement

All data collection followed the LJMU ethical protocols.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Aharon, N., R. Orfaig, and B.-Z. Bobrovsky. 2022. "BoT-SORT: Robust Associations Multi-Pedestrian Tracking." arXiv. https://arxiv.org/abs/2206.14651.

Aliane, N. 2025. "Drones and AI-Driven Solutions for Wildlife Monitoring." *Drones* 9, no. 7: 455. https://doi.org/10.3390/drones9070455.

Alisauskas, R. T., and P. B. Conn. 2019. "Effects of Distance on Detectability of Arctic Waterfowl Using Double-Observer Sampling During Helicopter Surveys." *Ecology and Evolution* 9, no. 2: 859–867.

Amini-Naieni, N., T. Han, and A. Zisserman. 2024. "CountGD: Multi-Modal Open-World Counting." arXiv. https://arxiv.org/abs/2407.04619.

Amstrup, S. C., T. L. McDonald, and B. F. Manly. 2010. *Handbook of Capture-Recapture Analysis*. Princeton University Press.

Barnas, A. F., D. Chabot, A. J. Hodgson, D. W. Johnston, D. M. Bird, and S. N. Ellis-Felege. 2020. "A Standardized Protocol for Reporting Methods When Using Drones for Wildlife Research." *Journal of Unmanned Vehicle Systems* 8, no. 2: 89–98.

Beaver, J. T., R. W. Baldwin, M. Messinger, C. H. Newbolt, S. S. Ditchkoff, and M. R. Silman. 2020. "Evaluating the Use of Drones Equipped With Thermal Sensors as an Effective Method for Estimating Wildlife." *Wildlife Society Bulletin* 44, no. 2: 434–443.

Beery, S., G. Van Horn, and P. Perona. 2018. "Recognition in Terra Incognita." In *Proceedings of the European Conference on Computer Vision (ECCV)*, 456–473.

Brack, I. V., A. Kindel, D. O. Berto, et al. 2023. "Spatial Variation on the Abundance of a Threatened South American Large Herbivore Using Spatiotemporally Replicated Drone Surveys." *Biodiversity and Conservation* 32, no. 4: 1291–1308.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2004. *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press.

Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. *Distance Sampling: Methods and Applications*. Springer.

Burke, C., M. Rashman, S. Wich, A. Symons, C. Theron, and S. Longmore. 2019. "Optimizing Observing Strategies for Monitoring Animals Using Drone-Mounted Thermal Infrared Cameras." *International Journal of Remote Sensing* 40, no. 2: 439–467.

Burke, C., M. F. Rashman, S. N. Longmore, et al. 2019. "Successful Observation of Orangutans in the Wild With Thermal-Equipped Drones." *Journal of Unmanned Vehicle Systems* 7, no. 3: 235–257. https://doi.org/10.1139/juvs-2018-0035.

Burton, A. C., E. Neilson, D. Moreira, et al. 2015. "Wildlife Camera Trapping: A Review and Recommendations for Linking Surveys to Ecological Processes." *Journal of Applied Ecology* 52, no. 3: 675–685. https://doi.org/10.1111/1365-2664.12432.

Campbell, C., P. Worth, D. Bratanov, and G. Hamilton. 2026. "A Single Drone as Two Observers: Increasing Wildlife Detection Availability in Complex Environments Using Repeated Drone Flights With Offset Paths." *Remote Sensing* 18: 327.

Cantu de Leija, A., R. E. Mirzadi, J. M. Randall, M. D. Portmann, E. J. Mueller, and D. E. Gawlik. 2023. "A Meta-Analysis of Disturbance Caused by Drones on Nesting Birds." *Journal of Field Ornithology* 94, no. 2: 3.

Chabot, D., and D. M. Bird. 2015. "Wildlife Research and Management Methods in the 21st Century: Where Do Unmanned Aircraft Fit in?" *Journal of Unmanned Vehicle Systems* 3, no. 4: 137–155.

Chang, B., B. Hwang, W. Lim, et al. 2025. "Enhancing Wildlife Detection Using Thermal Imaging Drones: Designing the Flight Path." *Drones* 9, no. 1: 52. https://doi.org/10.3390/drones9010052.

Christie, K. S., S. L. Gilbert, C. L. Brown, M. Hatfield, and L. Hanson. 2016. "Unmanned Aircraft Systems in Wildlife Research: Current and Future Applications of a Transformative Technology." *Frontiers in Ecology and the Environment* 14, no. 5: 241–251.

Clement, M. J., J. J. Hervert, and J. L. Bright. 2022. "Telemetry-Based Aerial Surveys for Estimating Abundance of Sparse Populations." *Wildlife Society Bulletin* 46, no. 1: e1253.

Cordier, C. P., D. A. Ehlers Smith, Y. Ehlers Smith, and C. T. Downs. 2022. "Camera Trap Research in Africa: A Systematic Review to Show Trends in Wildlife Monitoring and Its Value as a Research Tool." *Global Ecology and Conservation* 40: e02326.

Correia, C. A. M., F. A. A. Andrade, A. Sivertsen, et al. 2022. "Comprehensive Direct Georeferencing of Aerial Images for Unmanned Aerial Systems Applications." *Sensors* 22, no. 2: 604. https://doi.org/10.3390/s22020604.

Dave, B., M. Mori, A. Bathani, and P. Goel. 2023. "Wild Animal Detection Using YOLOv8." *Procedia Computer Science* 230: 100–111.

Davis, A. J., D. A. Keiter, E. M. Kierepka, et al. 2020. "A Comparison of Cost and Quality of Three Methods for Estimating Density for Wild Pig (*Sus scrofa*)." *Scientific Reports* 10: 2047. https://doi.org/10.1038/s41598-020-58937-0.

Delaney, D. M., T. M. Harms, and S. J. Dinsmore. 2025. "Method Matters: Use of Thermal-Imaging Drones to Assess the Assumptions of Density Estimation Techniques." *Ecological Applications* 35, no. 8: e70164.

Delisle, Z. J., P. G. McGovern, B. G. Dillman, and R. K. Swihart. 2022. "Imperfect Detection and Wildlife Density Estimation Using Aerial Surveys With Infrared and Visible Sensors." *Remote Sensing in Ecology and Conservation* 9: 222–234. https://doi.org/10.1002/rse2.305.

Du, Z., D. Wu, Q. Wen, et al. 2025. "An Improved Lightweight Model for Protected Wildlife Detection in Camera Trap Images." *Sensors* 25, no. 23: 7331. https://doi.org/10.3390/s25237331.

Duporge, I., M. Kholiavchenko, R. Harel, et al. 2025. "BaboonLand Dataset: Tracking Primates in the Wild and Automating Behaviour Recognition From Drone Videos." *International Journal of Computer Vision* 133, no. 9: 6578–6589.

Elchik, C. C., F. Karimi Nejadasl, S. S. Mohammadi Ziabari, and A. M. M. Alsahag. 2025. "A Framework for Multi-View Multiple Object Tracking Using Single-View Multi-Object Trackers on Fish Data." arXiv. https://arxiv.org/abs/2505.17201.

Elchik, C. C., S. Wich, and A. Burger. 2025. "A Framework for Detecting and Tracking Elephants in Drone Videos." *Drone Systems and Applications* 13: 1–16. https://doi.org/10.1139/dsa-2025-0032.

Esmukov, K., and GeoPy Contributors. 2023. "GeoPy: Geocoding Library for Python (Version 2.4.1)." Computer Software.

Espíndola, W. D., A. Cruz-Mendoza, A. Garrastazú, M. A. Nieves, F. F. Rivera-Milán, and T. A. Carlo. 2023. "Estimating Population Size of Red-Footed Boobies Using Distance Sampling and Drone Photography." *Wildlife Society Bulletin* 47, no. 1: e1406.

Fang, C., C. Li, P. Yang, et al. 2024. "Enhancing Livestock Detection: An Efficient Model Based on YOLOv8." *Applied Sciences* 14, no. 11: 4809.

Fonod, R., H. Cho, H. Yeo, and N. Geroliminis. 2025. "Advanced Computer Vision for Extracting Georeferenced Vehicle Trajectories From Drone Imagery." *Transportation Research Part C: Emerging Technologies* 178: 105205. https://doi.org/10.1016/j.trc.2025.105205.

Franklin, J., and J. A. Miller. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.

Gabrlik, P. 2015. "The Use of Direct Georeferencing in Aerial Photogrammetry With Micro UAV." *IFAC-PapersOnLine* 48, no. 4: 380–385. https://doi.org/10.1016/j.ifacol.2015.07.064.

Garcia-Garin, O., A. Aguilar, A. Borrell, et al. 2020. "Who's Better at Spotting? A Comparison Between Aerial Photography and Observer-Based Methods to Monitor Floating Marine Litter and Marine Mega-Fauna." *Environmental Pollution* 258: 113680.

Google LLC. 2024. "GNSS Logger (Version 4.0.3.0)." Android Mobile Application. Google Play Store.

Hansen, J., and R. Figueiredo. 2024. "Active Object Detection and Tracking Using Gimbal Mechanisms for Autonomous Drone Applications." *Drones* 8: 55. https://doi.org/10.3390/drones8020055.

Harvey, P. 2025. "ExifTool: Read, Write and Edit Meta Information." https://exiftool.org/.

Hodgson, J. C., S. M. Baylis, R. Mott, A. Herrod, and R. H. Clarke. 2016. "Precision Wildlife Monitoring Using Unmanned Aerial Vehicles." *Scientific Reports* 6: 22574. https://doi.org/10.1038/srep22574.

Hodgson, J. C., R. Mott, S. M. Baylis, et al. 2018. "Drones Count Wildlife More Accurately and Precisely Than Humans." *Methods in Ecology and Evolution* 9, no. 5: 1160–1167. https://doi.org/10.1111/2041-210X.12974.

Hoefer, S., D. T. McKnight, S. Allen-Ankins, E. J. Nordberg, and L. Schwarzkopf. 2023. "Passive Acoustic Monitoring in Terrestrial Vertebrates: A Review." *Bioacoustics* 32, no. 5: 506–531.

Hoffmann, A., J. Decher, F. Rovero, J. Schaer, C. Voigt, and G. Wibbelt. 2010. "Field Methods and Techniques for Monitoring Mammals." In *Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories*, edited by D. VandenSpiegel, Y. Samyn, J. C. Monje, C. Häuser, and J. Degreef, vol. 8 (part 2), 482–529. Pensoft Publishers.

Israel, M. 2011. "A UAV-Based Roe Deer Fawn Detection System." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38, no. 1/C22: 1–5. https://doi.org/10.5194/isprsarchives-XXXVIII-1-C22-1-2011.

Jachmann, H. 2002. "Comparison of Aerial Counts With Ground Counts for Large African Herbivores." *Journal of Applied Ecology* 39, no. 5: 841–852.

Jones, L. R., J. A. Elmore, B. S. Krishnan, et al. 2023. "Controllable Factors Affecting Accuracy and Precision of Human Identification of Animals From Drone Imagery." *Ecosphere* 14, no. 9: e4657.

Kays, R., J. Sheppard, K. McLean, et al. 2019. "Hot Monkey, Cold Reality: Surveying Rainforest Canopy Mammals Using Drone-Mounted Thermal Infrared Sensors." *International Journal of Remote Sensing* 40, no. 2: 407–419. https://doi.org/10.1080/01431161.2018.1523580.

Kellenberger, B., D. Marcos, and D. Tuia. 2018. "Detecting Mammals in UAV Images: Best Practices to Address a Substantially Imbalanced Dataset With Deep Learning." *Remote Sensing of Environment* 216: 139–153.

Khanam, R., and M. Hussain. 2024. "YOLOv11: An Overview of the Key Architectural Enhancements." arXiv. https://arxiv.org/abs/2410.17725.

Kholiavchenko, M., J. Kline, M. Kukushkin, et al. 2024. "Deep Dive Into KABR: A Dataset for Understanding Ungulate Behavior From In-Situ Drone Video." *Multimedia Tools and Applications* 84, no. 21: 1–20.

Laake, J. L., and D. L. Borchers. 2004. "Methods for Incomplete Detection at Distance Zero." In *Advanced Distance Sampling*, edited by S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, 108–189. Oxford University Press.

Lalgudi, C. K., M. E. Leone, J. V. Clark, S. Madrigal-Mora, and M. Espinoza. 2025. "Zero-Shot Shark Tracking and Biometrics From Aerial Imagery." *Methods in Ecology and Evolution* 16, no. 9: 2023–2035. https://doi.org/10.1111/2041-210X.70116.

Li, Z., N. Hovakimyan, V. Dobrokhodov, and I. Kaminer. 2011. "Vision-Based Target Tracking and Motion Estimation Using a Small UAV." In *49th IEEE Conference on Decision and Control (CDC)*, 2505–2510. IEEE.

Linchant, J., J. Lisein, J. Semeki, P. Lejeune, and C. Vermeulen. 2015. "Are Unmanned Aircraft Systems the Future of Wildlife Monitoring? A Review of Accomplishments and Challenges." *Mammal Review* 45, no. 4: 239–252. https://doi.org/10.1111/mam.12046.

Liu, S., Z. Zeng, T. Ren, et al. 2023. "Grounding DINO: Marrying DINO With Grounded Pre-Training for Open-Set Object Detection." arXiv. https://arxiv.org/abs/2303.05499.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. "Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One." *Ecology* 83, no. 8: 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. Bailey, and J. E. Hines. 2017. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. 2nd ed. Academic Press.

Marques, T. A., L. Thomas, S. W. Martin, et al. 2013. "Estimating Animal Population Density Using Passive Acoustics." *Biological Reviews* 88, no. 2: 287–309.

Mou, C., T. Liu, C. Zhu, and X. Cui. 2023. "WAID: A Large-Scale Dataset for Wildlife Detection With Drones." *Applied Sciences* 13, no. 18: 10397. https://doi.org/10.3390/app131810397.

Mulero-Pázmány, M., S. Jenni-Eiermann, N. Strebel, T. Sattler, J. J. Negro, and Z. Tablado. 2017. "Unmanned Aircraft Systems as a New Source of Disturbance for Wildlife: A Systematic Review." *PLoS One* 12, no. 6: e0178448. https://doi.org/10.1371/journal.pone.0178448.

Mulero-Pázmány, M., R. Stolper, L. D. van Essen, J. J. Negro, and T. Sassen. 2014. "Remotely Piloted Aircraft Systems as a Rhinoceros Anti-Poaching Tool in Africa." *PLoS One* 9, no. 1: e83873. https://doi.org/10.1371/journal.pone.0083873.

Nichols, J. D., and B. K. Williams. 2006. "Monitoring for Conservation." *Trends in Ecology & Evolution* 21, no. 12: 668–673. https://doi.org/10.1016/j.tree.2006.08.007.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. *Statistical Inference From Capture Data on Closed Animal Populations. Wildlife Monographs, No. 62*. Wildlife Society.

Pagacz, S., and J. Witczuk. 2023. "Estimating Ground Surface Visibility on Thermal Images From Drone Wildlife Surveys in Forests." *Ecological Informatics* 78: 102379.

Piel, A. K., A. Crunchant, I. E. Knot, et al. 2022. "Noninvasive Technologies for Primate Conservation in the 21st Century." *International Journal of Primatology* 43, no. 1: 133–167. https://doi.org/10.1007/s10764-021-00245-z.

Pinel-Ramos, E. J., F. Aureli, S. Wich, et al. 2025. "An Assessment of the Effectiveness of RGB-Camera Drones to Monitor Arboreal Mammals in Tropical Forests." *Drones* 9, no. 9: 622.

Plumptre, A. J. 2000. "Monitoring Mammal Populations With Line Transect Techniques in African Forests." *Journal of Applied Ecology* 37, no. 2: 356–368.

Povlsen, P., D. Bruhn, P. Durdevic, D. O. Arroyo, and C. Pertoldi. 2024. "Using YOLO Object Detection to Identify Hare and Roe Deer in Thermal Aerial Video Footage—Possible Future Applications in Real-Time Automatic Drone Surveillance and Wildlife Monitoring." *Drones* 8, no. 1: 2. https://doi.org/10.3390/drones8010002.

Rahman, D. A., R. Herliansyah, B. Subhan, et al. 2023. "The First Use of a Photogrammetry Drone to Estimate Population Abundance and Predict Age Structure of Threatened Sumatran Elephants." *Scientific Reports* 13, no. 1: 21311.

Ransom, J. I. 2012. "Detection Probability in Aerial Surveys of Feral Horses." *Journal of Wildlife Management* 76, no. 2: 299–307. https://doi.org/10.1002/jwmg.204.

Roboflow. 2025. "Roboflow Documentation: Build Vision Models With Roboflow." https://docs.roboflow.com/.

Romairone, J., J. Jiménez, J. J. Luque-Larena, and F. Mougeot. 2018. "Spatial Capture-Recapture Design and Modelling for the Study of Small Mammals." *PLoS One* 13, no. 6: e0198766.

Ryan, S. A., D. M. Southwell, C. T. Beranek, J. Clulow, N. R. Jordan, and R. R. Witt. 2025. "Estimating the Landscape-Scale Abundance of an Arboreal Folivore Using Thermal Imaging Drones and Binomial n-Mixture Modelling." *Biological Conservation* 309: 111207.

Schlossberg, S., M. J. Chase, and C. R. Griffin. 2016. "Testing the Accuracy of Aerial Surveys for Large Mammals: An Experiment With

African Savanna Elephants (*Loxodonta africana*).” *PLoS One* 11, no. 10: e0164904.

Schnupp, M. J., F. Hernández, E. J. Redeker, J. P. Sands, and R. M. Perez. 2013. “An Electronic System to Collect Distance-Sampling Data During Helicopter Surveys of Northern Bobwhite.” *Wildlife Society Bulletin* 37, no. 1: 236–245.

Seymour, A. C., J. Dale, M. Hammill, P. N. Halpin, and D. W. Johnston. 2017. “Automated Detection and Enumeration of Marine Wildlife Using Unmanned Aircraft Systems (UAS) and Thermal Imagery.” *Scientific Reports* 7, no. 1: 45127.

Spaan, D., C. Burke, O. McAree, et al. 2019. “Thermal Infrared Imaging From Drones Offers a Major Advance for Spider Monkey Surveys.” *Drones* 3, no. 2: 34.

Stevenson, B. C., D. L. Borchers, R. Altwegg, R. J. Swift, D. M. Gillespie, and G. J. Measey. 2015. “A General Framework for Animal Density Estimation From Acoustic Detections Across a Fixed Microphone Array.” *Methods in Ecology and Evolution* 6, no. 1: 38–44.

Walker, A. 2020. “OpenTopoData (Version 1.5.0).” Computer Software. https://www.opentopodata.org/.

Wang, C.-Y., and H.-Y. M. Liao. 2024. “YOLOv1 to YOLOv10: The Fastest and Most Accurate Real-Time Object Detection Systems.” *APSIPA Transactions on Signal and Information Processing* 13, no. 1: 1–38. https://doi.org/10.1561/116.20240058.

Wang, X., J. Liu, and Q. Zhou. 2016. “Real-Time Multi-Target Localization From Unmanned Aerial Vehicles.” *Sensors* 17: 33. https://doi.org/10.3390/s17010033.

Watts, J., T. Holding, K. Anderson, et al. 2024. “Georectifying Drone Image Data Over Water Surfaces Without Fixed Ground Control: Methodology, Uncertainty Assessment and Application Over an Estuarine Environment.” *Estuarine, Coastal and Shelf Science* 305: 108853. https://doi.org/10.1016/j.ecss.2024.108853.

Wearn, O. R., and P. Glover-Kapfer. 2019. “Snap Happy: Camera Traps Are an Effective Sampling Tool When Compared With Alternative Methods.” *Royal Society Open Science* 6, no. 3: 181748. https://doi.org/10.1098/rsos.181748.

Whitworth, A., C. Pinto, J. Ortiz, E. Flatt, and M. Silman. 2022. “Flight Speed and Time of Day Heavily Influence Rainforest Canopy Wildlife Counts From Drone-Mounted Thermal Camera Surveys.” *Biodiversity and Conservation* 31, no. 13: 3179–3195.

Wich, S. A., N. Bonnin, A. Hutschenreiter, et al. 2023. “Using Drones to Determine Chimpanzee Absences at the Edge of Their Distribution in Western Tanzania.” *Remote Sensing* 15, no. 8: 2019. https://doi.org/10.3390/rs15082019.

Wich, S. A., D. Dellatore, M. Houghton, R. Ardi, and L. P. Koh. 2015. “A Preliminary Assessment of Using Conservation Drones for Sumatran Orang-Utan (*Pongo abelii*) Distribution and Density.” *Journal of Unmanned Vehicle Systems* 4, no. 1: 45–52.

Wich, S. A., and L. P. Koh. 2018. *Conservation Drones: Mapping and Monitoring Biodiversity*. Oxford University Press.

Witt, R. R., C. T. Beranek, L. G. Howell, et al. 2020. “Real-Time Drone Derived Thermal Imagery Outperforms Traditional Survey Methods for an Arboreal Forest Mammal.” *PLoS One* 15, no. 11: e0242204. https://doi.org/10.1371/journal.pone.0242204.

Wohlfahrt, S., C. Praschl, H. Leitner, et al. 2025. “Advancing Wildlife Monitoring: Drone-Based Sampling for Roe Deer Density Estimation.” arXiv. https://arxiv.org/abs/2508.03545.

Yang, L., B. Kang, Z. Huang, et al. 2024. “Depth Anything V2.” arXiv. https://arxiv.org/abs/2406.09414.

Zhao, X., F. Pu, Z. Wang, H. Chen, and Z. Xu. 2019. “Detection, Tracking, and Geolocation of Moving Vehicle From UAV Using Monocular Camera.” *IEEE Access* 7: 101160–101170. https://doi.org/10.1109/ACCESS.2019.2929760.

Zhou, X., W. Jia, R. He, and W. Sun. 2025. “High-Precision Localization Tracking and Motion State Estimation of Ground-Based Moving Target Utilizing Unmanned Aerial Vehicle High-Altitude Reconnaissance.” *Remote Sensing* 17, no. 5: 735. https://doi.org/10.3390/rs17050735.

Zhou, Y., D. Tang, H. Zhou, and X. Xiang. 2025. “Moving Target Geolocation and Trajectory Prediction Using a Fixed-Wing UAV in Cluttered Environments.” *Remote Sensing* 17, no. 6: 969. https://doi.org/10.3390/rs17060969.