Multi-animal behavioral tracking and environmental reconstruction using drones and computer vision in the wild

Benjamin Koger[1,2,3], Adwait Deshpande[1,2,3], Jeffrey T. Kerby[4,5,6], Jacob M. Graving[7,1,2,3], Blair R. Costelloe[1,2,3]#, Iain D. Couzin[1,2,3]#

1. Department of Collective Behavior, Max Planck Institute of Animal Behavior, Konstanz, Germany
2. Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany.
3. Department of Biology, University of Konstanz, Konstanz, Germany.
4. Aarhus Institute of Advanced Studies, Aarhus University, Aarhus, Denmark.
5. Neukom Institute for Computational Science, Dartmouth College, Hanover, USA.
6. Department of Bioscience, Aarhus University, Aarhus, Denmark.
7. Advanced Research Technology Unit, Max Planck Institute of Animal Behavior, Konstanz, Germany

Co-senior authors

Corresponding authors: B. Koger (benkoger@gmail.com) and I.D. Couzin (icouzin@ab.mpg.de)

Keywords: behavioral tracking, computer vision, drones, environmental reconstruction, gelada monkey, pose, posture, video analysis, wildlife, zebra,

Abstract

1. Methods for collecting animal behavior data in natural environments, such as direct observation and bio-logging, are typically limited in spatiotemporal resolution, the number of animals that can be observed, and information about animals' social and physical environments.
2. Video imagery can capture rich information about animals and their environments, but image-based approaches are often impractical due to the challenges of processing large and complex multi-image datasets and transforming resulting data, such as animals' locations, into geographic coordinates.
3. We demonstrate a new system for studying behavior in the wild that uses drone-recorded videos and computer vision approaches to automatically track the location and body posture of free-roaming animals in georeferenced coordinates with high spatiotemporal resolution embedded in contemporaneous 3D landscape models of the surrounding area.
4. We provide two worked examples in which we apply this approach to videos of gelada monkeys and multiple species of group-living African ungulates. We demonstrate how to track multiple animals simultaneously, classify individuals by species and age-sex class, estimate individuals' body postures (poses), and extract environmental features, including topography of the landscape and game trails.
5. By quantifying animal movement and posture, while simultaneously reconstructing a detailed 3D model of the landscape, our approach opens the door to studying the sensory ecology and decision-making of animals within their natural physical and social environments.

Introduction

Studying animals in the wild is essential for understanding how they behave within, and are shaped by, the environments in which they have evolved. Animal behavior impacts, and is impacted by, biological processes across vast scales, from the neural, genetic and endocrine interactions within organisms to the emergent functional complexity of groups, populations and ecosystems. Furthermore, we are in the midst of a biodiversity crisis (Ceballos et al., 2020), which has created an urgent need to understand how the decisions, movements, and, ultimately, the fitness of organisms are influenced by anthropogenic impacts and environmental change.

Historically, studies of animal behavioral ecology have relied on direct observation by humans (Altmann, 1974), an approach that remains foundational today, but also has inherent limitations. For example, humans' attention capacity constrains the spatial and temporal resolution of the data they can collect, the number of animals they can observe, and the duration of observations (Dell et al., 2014). Data collected via human observation is also prone to observer bias and subjective classification (Tuyttens et al., 2014), issues that cannot be resolved without an objective record of behavior. Video recording allows researchers to preserve a less biased record and extract more detailed data than is possible via direct observation. However, manually scoring behavior from video footage is also subjective and time-consuming, which introduces errors, and limits scalability (Dell et al., 2014).

Major advances in sensor, battery, and communications technologies over the past two decades have led to the rapid development of animal-borne data-logging devices, known as bio-loggers. Bio-logging has become a powerful approach for studying behavior, allowing scientists to collect data over expansive spatial and temporal scales and from inaccessible environments (Brown et al., 2013; Kays et al., 2015). This approach has generated important insights into behaviors that were previously difficult or impossible to study effectively, including migration (Jesmer et al., 2018), dispersal (Klarevas-Irby et al., 2021), energetics (Flack et al., 2020; Williams et al., 2014), sleep (Loftus et al., 2022; Rattenborg et al., 2016), and individual and collective decision-making (Flack et al., 2018; Strandburg-Peshkin et al., 2015, 2017).

A key challenge of bio-logging is understanding the behavioral states and contemporaneous social and environmental factors that give rise to animal movements. Secondary sensors affixed to tags can provide context for geolocation data (Williams et al., 2020), but are themselves limited. For example, accelerometers are commonly used to record stereotyped behavioral states such as resting, feeding, and locomotion, but generally cannot capture more complete behavioral repertoires (Brown et al., 2013). On-board cameras can record the social interactions, foraging decisions, and physical surroundings of instrumented individuals, but are restricted in their fields of view and prone to occlusion (Ehlers et al., 2021; Naganuma et al., 2021). Interpreting data from secondary sensors can also be challenging. Translation of accelerometer data to human-recognizable behavioral labels requires validation through synchronous acceleration logging and visual observation. This is time-consuming and often conducted using captive animals or relatively brief observations of wild animals, which is likely to exclude rare or context-dependent behaviors (Brown et al., 2013; Wang et al., 2015). Data from animal-borne cameras typically must be manually reviewed by experts, and variable image quality can prevent extraction of relevant data (Moll et al., 2007; Naganuma et al., 2021). Alternatively, social behavior can be studied by instrumenting most or all animals of interest (e.g. Strandburg-Peshkin et al., 2017), but this approach is costly, logistically challenging and not viable for species with unstable group composition (Hughey et al., 2018). To understand environmental context, movement data can be fused with external data streams, such as meteorological records or satellite and drone imagery (Benitez-Paez et al., 2021; Strandburg-Peshkin et al., 2017); however, these data are typically asynchronous with behavioral data or too coarse-grained for analysis at finer scales (Brum-Bastos et al., 2020).

Beyond scientific constraints, bio-logging presents other ethical, financial, and logistical limitations. These include the expense of the devices, costs of remote data transmission, logistics of capturing animals, and potential reduced fitness of instrumented individuals (Hughey et al., 2018; Murray & Fuller, 2000). Consequently, the costs of equipment and potential impact of instrumentation on animals means that, outside of well-funded research initiatives, bio-logging studies typically have small sample sizes and are rarely replicated over time or across populations.

Video-based observation, in combination with vision-based machine learning tools, is an emerging approach and an alternative to bio-logging for studying animals in the wild (Hughey et al., 2018; Tuia et al., 2022). These methods originated in laboratory settings, where physical on-animal markers and controlled recording conditions enabled early automated movement tracking solutions (reviewed in Dell et al., 2014). However, field settings pose significant challenges for conventional vision-based tracking solutions due to the variable lighting conditions and complex visual scenes inherent to natural environments. Recent increases in computational processing power and efficiency, and advances in deep learning techniques, have transformed automated vision-based tools in field settings (Zhao et al., 2019). Still, these methods are technically complex and require expert knowledge for successful implementation.

Similar to the overhead imaging common in many laboratory-based animal tracking paradigms (Buhl et al., 2006; Katz et al., 2011), drones mounted with high resolution cameras allow for efficient top down filming that enables precise quantification of animal positions and reduces occlusion by environmental features or other group members (e.g. Inoue et al., 2019; Torney et al., 2018). Modern consumer imaging drones are affordable and easy to pilot. Drones' mobility makes them a particularly powerful platform for field observations because they can be used to film free-roaming animals as they move around the landscape. However, this camera motion becomes a major source of error that must be accounted for (Haalck et al., 2020). The topography of natural landscapes also introduces challenges in accurately projecting 2D image coordinates into 3D space in standard units. While multi-camera 3D imaging can surmount these issues, they are logistically and computationally difficult to deploy (Francisco et al., 2020). Therefore, previous drone-based behavioral studies have typically relied on approaches like manually processing videos (Sprogis et al., 2020), analyzing a subset of still frames (Inoue et al., 2019), approximating animal movement paths using the drone's position (Raoult et al., 2018), or recording trajectories in relation to the video frame dimensions rather than geographic coordinate systems (Ringhofer et al., 2020; Torney et al., 2018).

Here we describe a method for using aerial video and computer vision to collect high-resolution georeferenced locational and behavioral data on free-ranging animals without capturing or tagging them. In our approach, we use drones to record overhead video of focal animals and subsequently use a deep learning-based pipeline to automatically locate and track all individuals in the video. We use Structure-from-Motion (SfM) techniques to reconstruct the 3D topography of the surrounding habitat (D'Urban Jackson et al., 2020) which, combined with further image processing, can accurately transform animal movement data into geographic coordinate systems independent of camera movement. We simultaneously generate additional behavioral and environmental information, including estimates of each animal's body posture (pose) and landscape features such as game trails. This method, illustrated in Fig. 1, is applicable to a wide range of species, generates behavioral data at sub-second and sub-meter resolution, and produces synchronous information about the surrounding physical and biotic environment. The use of inexpensive consumer drones, that can be flexibly redeployed, greatly reduces the cost of data collection relative to bio-logging, and promotes larger sample sizes and replication of studies. Although not without its own limitations (see *Limitations and Considerations*), this approach has many advantages that make it a powerful alternative or complement to current methodologies. Our approach thus has the potential to broaden the scope of behavioral ecology to encompass questions and systems that are unsuitable for direct observation or bio-logging approaches in isolation.

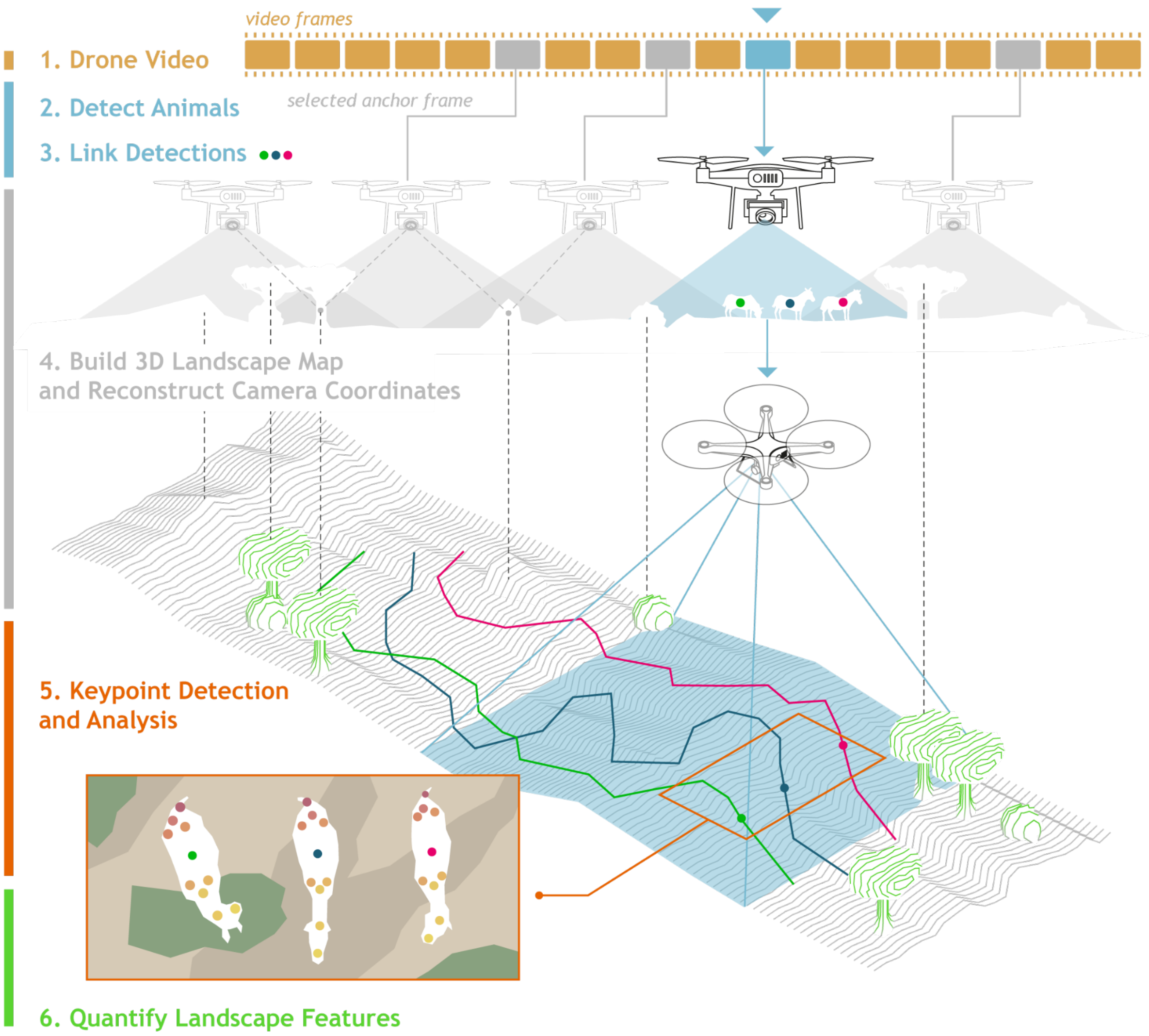


Figure 1. Overview of the processing pipeline for extracting movement, behavioral, and landscape data from aerial drone footage of wildlife. Numbered steps correspond to the numbered sections in the text. First, the animals of interest are video recorded from above (Step 1). Next, an object detection algorithm is used to localize each animal in every video frame (Step 2), and these locations are then linked across frames to generate movement trajectories in pixel coordinates (Step 3). In parallel, anchor frames are selected from the footage and used to build a 3D model of the landscape and estimate the locations of the drone across the observation. Camera locations at

anchor frames and local visual features are combined to estimate camera locations for all frames allowing the transformation of the animal trajectories from Step 3 into geographic coordinate space (Step 4). Optionally, further analyses can be performed to extract more detailed behavioral and landscape information, for example through keypoint detection (Step 5) and landscape feature detection (Step 6).

Methods

Below we outline the major steps of our methodological approach (Fig. 1). Full details of each step are given in the Supplement. To illustrate our processing pipeline, we provide two worked examples with complete code and data available on GitHub (https://github.com/benkoger/overhead-video-worked-examples). We encourage readers to explore these examples and to modify the code for their own datasets.

In the first example, we apply our method to multiple African ungulate species. We recorded ungulate groups at Ol Pejeta and Mpala Conservancies in Laikipia, Kenya over two field seasons, from November 2 to 16, 2017 and from March 30 to April 19, 2018. In total, we recorded thirteen species, but here we focus our analyses on the endangered Grevy's zebra (*Equus grevyi*) (Rubenstein et al., 2016). We used DJI Phantom 4 Pro drones [DJI, Shenzhen, China], and deployed two drones sequentially in overlapping relays to achieve continuous observations longer than a single battery duration. Our method is agnostic to the exact drone-type used, and is scalable to the higher-resolution video made possible by employing more recent models.

In the second worked example, we process video recordings of grassland-dwelling gelada monkeys (*Theropithecus gelada*). Aerial video recordings of gelada monkeys were provided by the Guassa Gelada Research Project. The recordings were collected between October 16, 2019 and February 28, 2020 at the Guassa Community Conservation Area in the Ethiopian highlands. Geladas were recorded with a DJI Mavic 2 Pro [DJI, Shenzhen, China].

In each worked example, we generate movement trajectories for each animal in the example videos and 3D models of the surrounding landscape. In the ungulate example, we also track body keypoints and analyze landscape imagery to detect game trails, which zebras tend to follow while moving across the landscape. For the geladas, we train our detection model (Step 2), to distinguish between adult males and other individuals.

Step 1. Video Recording

Videos should be recorded from above the animals of interest with the camera pointing directly down. The visibility of the animals is of vital importance and will be affected by the habitat type, video resolution and characteristics of the animals themselves. As a general rule, videos in which the animals are easily detectable by humans will significantly ease processing. Although any camera platform can be used to capture the footage, drone-mounted cameras will likely be the most common approach and subsequently we use "drone" and "camera" interchangeably. For specific points to consider when planning drone-based data collection, see *Limitations and Considerations*, below and *Step 1* in the Supplement.

Step 2. Detection

Animal detection and localization in each video frame is accomplished with deep convolutional neural networks (CNNs), which we build, train and deploy to predict localizing **bounding boxes** (see Table 1 for definitions of bolded terms) for all individuals in all frames of each video. We first manually annotate frames from the video footage to build image sets for training the model and evaluating its performance. **Annotation** can be tedious, and the content of the images will strongly affect model performance; therefore, it is important to carefully consider the best annotation strategy to achieve high information value for each annotation while minimizing human labor (see Supplement section 2.1). Users can improve the efficiency of the annotation process via **model-assisted labeling** and **active learning** techniques, which concentrate annotation effort on examples that are particularly challenging for the model to detect.

To train a deep learning-based object detection model, researchers must first choose an appropriate software framework and a specific model to train within that framework (see Supplement section 2.2). We use the Detectron2 API within the PyTorch framework (Paszke et al., 2019; Wu et al., 2019) in our worked examples, but the user may choose a different framework depending on their level of coding proficiency or prior experience with other programming libraries. For simple use cases with clearly-visible individuals, many common models can be readily re-configured for the researchers' data. Users with more challenging footage may need to make more considered choices and in some cases move towards incorporating more recent high-performance algorithms. Generally, users should choose models that have been **pretrained** on general image datasets and that incorporate **image augmentation**, as these steps increase training efficiency and require smaller annotated training sets. See Table S1 for annotation statistics and model performance metrics for the datasets and trained models used in the worked examples.

After the model is trained, all frames from all videos can be automatically processed. For each video frame the model generates bounding box coordinates, predicted object classes and confidence scores for every detected object (Fig. 2). We initially take the mean of the coordinates of the bounding box corners as an individual's location in the frame, which we subsequently use for tracking.

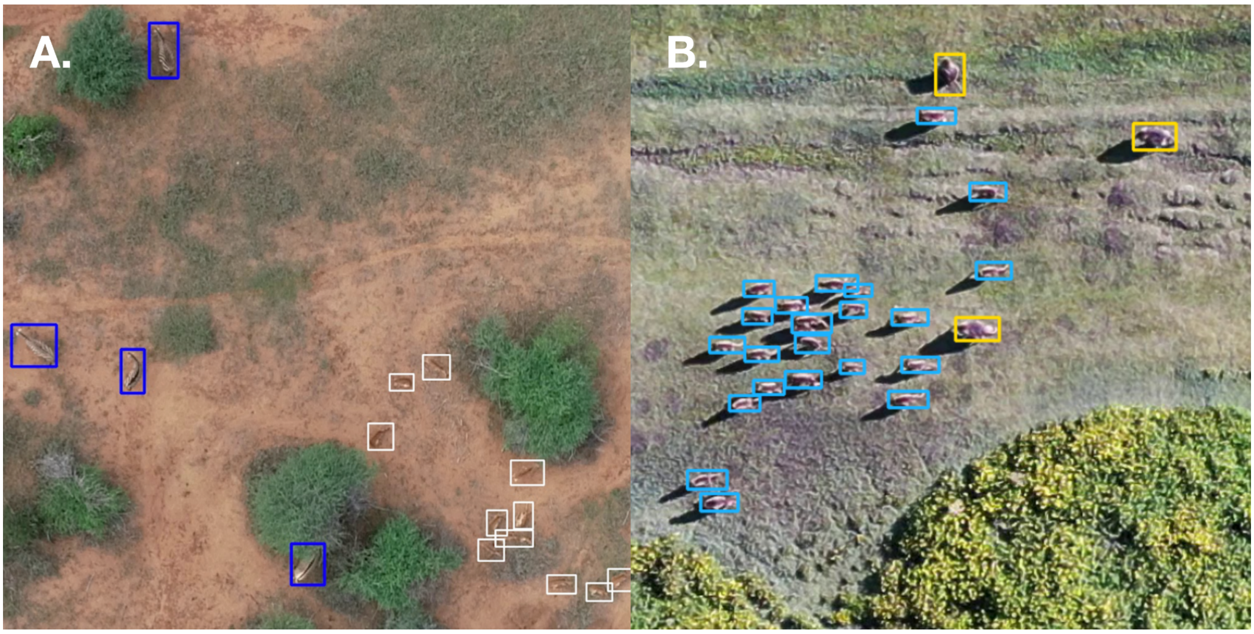


Figure 2. Predicted bounding boxes for two example video frames using the trained models from our worked examples. A) In the ungulates example, the model distinguishes between five classes (zebra, impala, buffalo, waterbuck, and other). The animals in dark blue bounding boxes are Grevy's (*Equus grevyi)* and plains zebras (*E. burchelli*) and animals in white bounding boxes are impala (*Aepyceros melampus*). B) In the gelada monkey example, we distinguish between species and age-sex class (human-observer, adult-male gelada, and other-gelada). Yellow boxes are predicted adult-males while light blue boxes are a mix of females and juveniles.

Step 3. Tracking

Tracking, or linking positions across video frames, allows us to generate trajectories for all detected individuals in the pixel coordinate system of the video. To match individual locations across consecutive frames, we use a modified version of the Hungarian algorithm (Kuhn, 1955), which finds the pairing of trajectories and new positions that minimizes the total distance between all pairs. We have incorporated additional distance- and time-based rules for connecting detections to tracks that make the algorithm more robust to missing or false detections by the detection model or from individuals entering or leaving the camera's field of view during the observation (see Step 3, in the Supplement).

The initial process of linking positions is automated but can result in multiple partial trajectories for a single individual in the case of environmental occlusions and other detection issues. For these instances, we provide a **graphical user interface (GUI)** for easy track validation and error correction. This allows the user to obtain, with limited manual effort, human-verified continuous trajectories of all individuals in each video within the pixel-based coordinate system of the video frame, which can then be transformed to a geographic coordinate system.

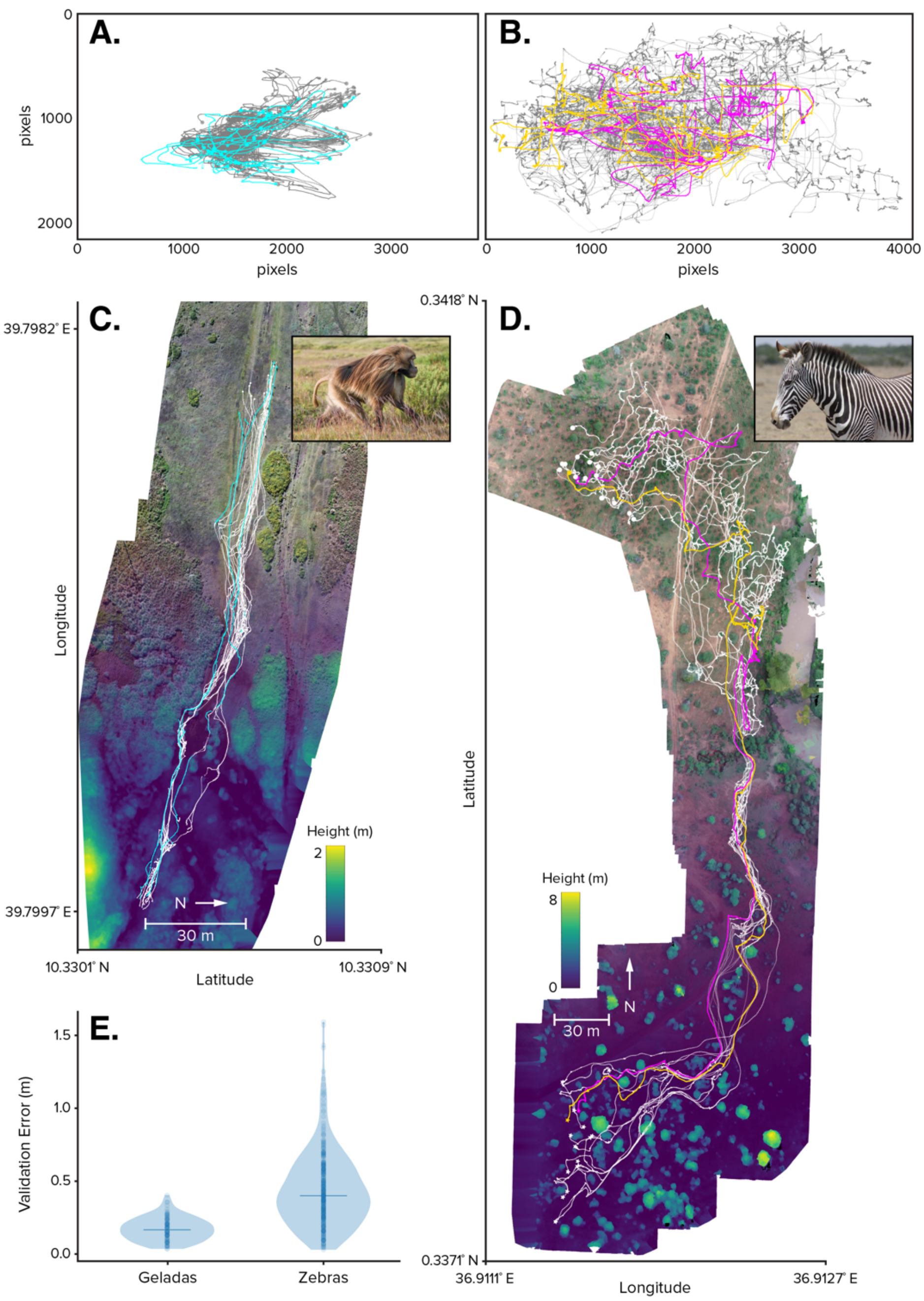


Figure 3. Trajectories extracted from drone videos. Trajectories of A) a gelada monkey group over a period of 5 minutes and B) a Grevy's zebra herd over a period of 50 minutes, 24 seconds plotted in video coordinates. Drone and animal movement are entangled and the position of the animals on the earth is unknown. In C) and D), the same trajectories are plotted in geographic coordinates and embedded in a reconstructed landscape. The landscape reconstruction includes visual information (top) as well as topographic information (bottom). E) The trajectories have submeter human-validated error between observed locations in video frames and embedded map locations (0.17m median, interquartile range 0.11-0.23m from 100 validated gelada locations; 0.40m median, interquartile range 0.26m-0.58m from 350 validated zebra locations). In A and C, light blue tracks denote adult male geladas. In B and D, two tracks are highlighted in yellow and pink for easier comparison. In E, the horizontal line shows the median.

Step 4. Landscape Reconstruction and Geographic Coordinate Transformation

Transforming trajectories into precise geographic coordinates allows us to disentangle the movement of the tracked animals over variable terrain and the movement of the drone (Fig. 3), and also allows the user to analyze the resulting movement data in standard units and in relation to external georeferenced data sources, such as satellite imagery. To achieve an accurate transformation, we must estimate the topography of the landscape over which the animals are moving and the location of the camera relative to that landscape (Fig. 4). We obtain this information by feeding a subset of frames ("**anchor frames**") from each video into **Structure-from-Motion (SfM)** software (Supplement section 4.1). This builds a detailed 3D model of the landscape surrounding each observation and also calculates the location of the camera when each anchor frame was captured. To accurately locate and scale the model in geographic space, we georeference it using either **ground control points** collected in the field or information from the drone's onboard GPS sensor (Supplement section 4.2). We track visual features in each frame to estimate local camera movement between anchor frames allowing us to calculate the camera position for every frame in the video (Supplement section 4.3). We then transform animal positions from the pixel coordinates of the frames into the geographic coordinates of the 3D model by projecting rays from the estimated camera location to the surface of the landscape model (Supplement section 4.4). In our examples, this method yields movement trajectories with sub-meter mean and median error (Fig. 3E); we provide a GUI that allows the user to efficiently assess the accuracy of their own track locations (Supplement section 4.5).

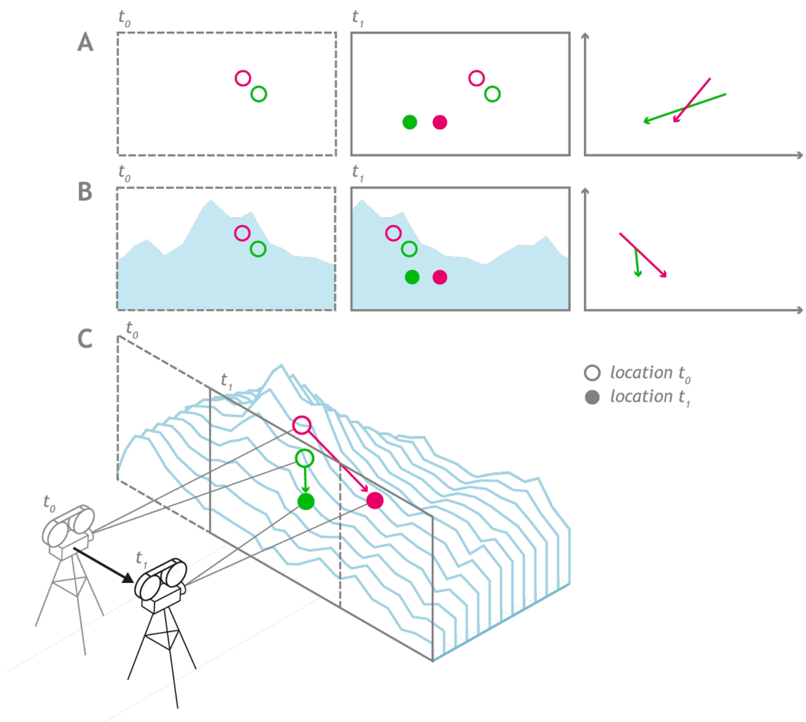
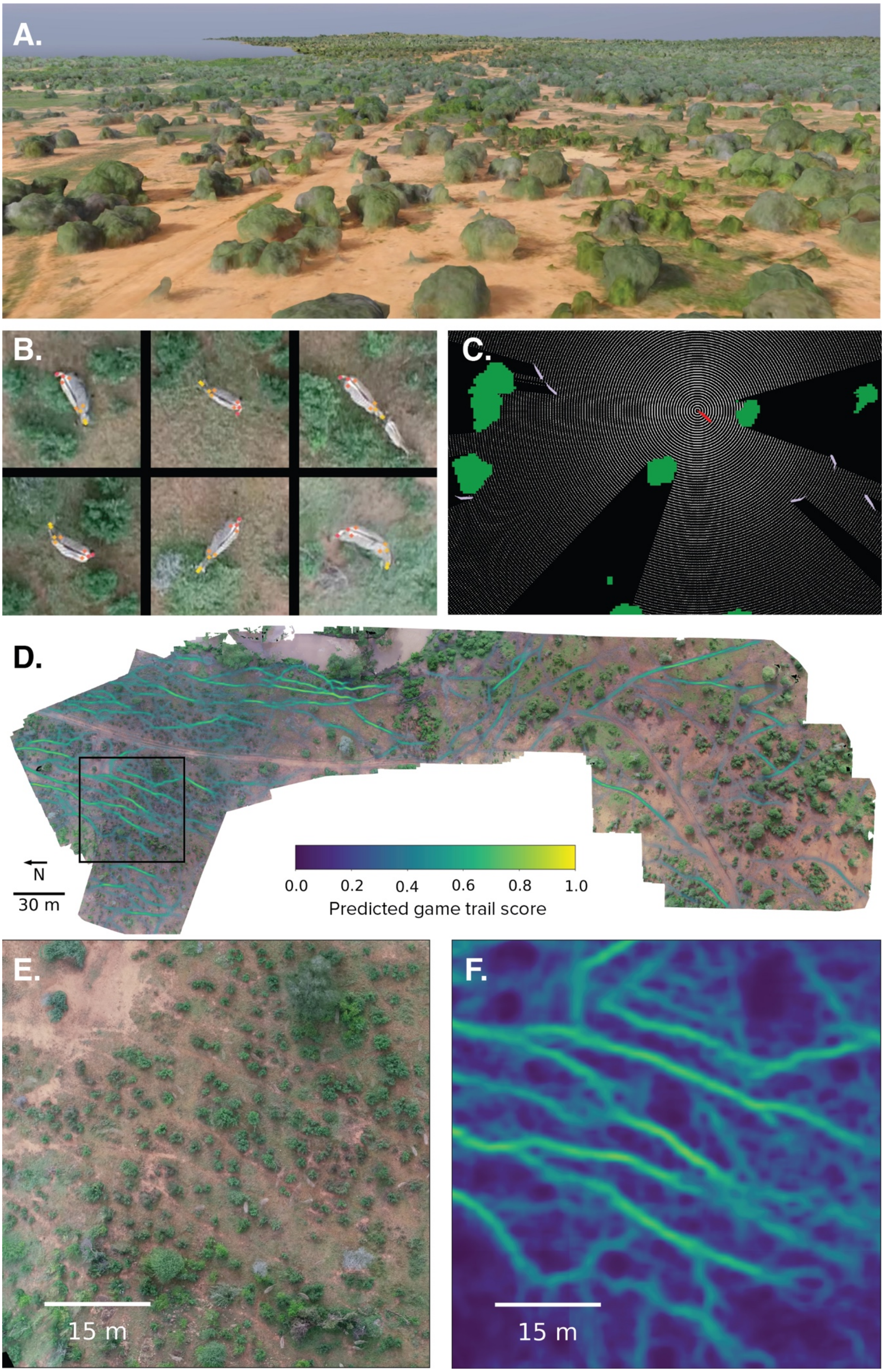


Figure 4. Both camera position and landscape topography are necessary to generate accurate trajectories from locations extracted from 2D images. A) At the end of Step 3 in the data extraction pipeline, we have linked detections across frames to generate trajectories for individual animals in the pixel-based coordinate space of the video frames. These trajectories do not distinguish between camera movement and animal movement, and thus changes in animal positions across frames (here from t0 to t1) do not accurately represent animal movements in the real world. B) By incorporating landscape features in the background of the video frames we calculate each frame's relative camera position and so disentangle camera movement from animal movement. C) By combining camera location information with the 3D topography of the environment, we can reconstruct accurate movement trajectories in 3D space. Without a model of the landscape structure, the user must assume a flat earth, which introduces error into trajectories.

Step 5. Body-part Keypoint Detection

A major advantage of image-based techniques is that each image contains much more behaviorally-relevant information than does position alone. If the video resolution is sufficient, the user can, for example, use existing tools for markerless pose estimation (e.g. SLEAP (Pereira et al., 2020), DeepLabCut (Mathis et al., 2018; Nath et al., 2019), DeepPoseKit (Graving et al., 2019)) to extract positions of user-defined body parts on each detected animal, generating time-varying postural information for tracked individuals (Step 5 in the Supplement; Fig. 5B). Such postural data are particularly amenable to automated behavioral annotation, and other downstream tasks, because the user can pair the (relatively) low dimensional keypoint information with human verified annotations from the raw video.


A.
B.
C.
D.
N
30 m
0.0
0.2
0.4
0.6
0.8
1.0
Predicted game trail score
E.
15 m
F.
15 m

308

Figure 5. Additional landscape and behavioral data generated beyond animal location. A) An SfM-generated 3D triangular mesh landscape model. B) Examples of keypoints detected on zebras. Nine user-defined keypoints (snout, head, base of the neck, left and right shoulder, left and right hindquarters, tail base and tip of the tail) are tracked using DeepPoseKit (Graving et al. 2019). C) Combining landscape data and animal body and head location allows the visualization of animal's visual fields in their landscapes. Here, the red object is the focal zebra, the white rays show its visual field, and the green and white polygons represent bushes and conspecifics, respectively. D) A SfM-generated 2D orthomosaic image showing predicted game trail presence score. Predictions are generated by a CNN trained on separate orthomosaics that were annotated by three independent human annotators. See worked examples for more detail. The black box indicates the area shown in E and F, which provide detailed views of the orthomosaic and the game trail predictions, respectively. The color bar applies to both D and F.

Step 6. Landscape Quantification

The SfM software used in Step 4 generates multiple forms of landscape data that provide valuable environmental context for the behavior of the observed individuals, including 2D rasters encoding elevation data, **orthomosaic images**, and 3D **point clouds** and **triangular meshes** (Fig. 5A, E). These outputs can be used directly or further processed to extract information regarding local landscape features and topography, classify habitat types, or even estimate the visual fields of tracked animals. In the ungulate worked example we apply pixel-wise classification algorithms, known as **semantic segmentation**, to orthomosaic images to automatically detect game trails, which we expect will influence animal movement (Fig. 5D, F). Further possibilities are explored in the section *Analysis, Extensions, & Applications*.

Limitations and Considerations

Here we discuss the logistics of using drones to capture behavioral data, the suitability of different research questions to this approach, and the coding skills necessary to implement this method. For required computing resources, please see the Supplement.

Ethical, logistical and legal considerations

It is important to consider the potential impact of the drone on the focal animals when planning future research. Behavioral and physiological responses to drone flights can negatively impact wildlife (Ditmer et al., 2015; Weimerskirch et al., 2018), and may lead to biased or misleading results in behavioral studies (Duporge et al., 2021). Most species that respond to drones seem to be primarily affected by the sound of the drone (Duporge et al., 2021; McEvoy et al., 2016). Pilots may be able to reduce disturbance by choosing quieter drone models, using low-noise propellers, launching the drone far away from target animals, approaching from downwind, and flying at higher altitudes. While still an emerging field of study, Duporge et al. (2021) and others (Bevan et al., 2018; Christiansen et al., 2016; Mulero-Pázmány et al., 2017) offer guidelines for flying drones near a variety of species. It is typically prudent to perform preliminary flights prior to data collection in order to assess the animals' response to the drone and establish appropriate protocols.

When choosing a study environment, factors such as extreme temperatures, dust, wind, precipitation and fog can reduce visibility, equipment longevity and flight performance. Furthermore, landscapes (such as open water and snow fields) without abundant and persistent visible landmarks are not suitable for SfM methods, and may preclude transformation of animal trajectories into geographic coordinates (Step 4). SfM height mapping accuracy is also sensitive to vegetation structure and motion (such as tall grass in the wind), which may introduce errors in those parts of the landscape relative to static regions (D'Urban Jackson et al., 2020). Additionally, the real-world geospatial accuracy of the final data is ultimately determined by the quality of the 3D landscape models. For sub-meter accuracy one must either be able to place ground control points in the landscape of interest or have access to existing georeferenced landscape models or imagery.

Finally, researchers must be aware of and abide by legal regulations regarding drone operations in any prospective study area both in terms of formal permissions and licenses required to deploy drones as well as limitations on in-flight maneuvers. Particular rules vary by location and are continually evolving but can include limits on flight altitudes, distances, or locations (i.e. airports, national parks, or certain government areas) and requirements for maintaining visual contact with drones and keeping safe distances from people and structures. Beyond legal requirements, researchers should also ensure that projects do not negatively impact local communities. For a thorough discussion of ethical drone use, see Duffy et al. (2018).

Research question suitability

In determining whether image-based data collection is appropriate for a research question, researchers should consider the required data resolution and spatial and temporal scales of all targeted study behaviors. Animal groups that are spread over large areas may be impossible to fully capture at an appropriate resolution, especially if automated tracking or individual posture data are required. Deploying multiple drones simultaneously can increase spatial coverage, but complicates flight protocols and data processing. Additionally, spatially-expansive behaviors like long-distance hunts may draw the drone beyond the legal operation distance from stationary pilots (Creel & Creel, 1995). Nocturnal or crepuscular animals would require imaging via thermal or high-sensitivity cameras, which imposes further constraints to the spatial, and temporal, resolution of video that can be obtained.

The timescale over which behaviors occur is also important. Drone battery efficiency has rapidly improved over the years, but flight times are still relatively short (currently 30-45 minutes for widely-used models). While observation time can be extended with multiple sequential flights, the duration of observations are still limited by battery supply, pilot fatigue, and the movement of the focal animals away from the operational site. Similar to direct observation, recording rare or unpredictable behaviors, such as predation events, requires deploying drones in the right place at the right time. Incorporating external observational or biologging data could help predict such behaviors and inform the location and timing of deployments. Relatedly, deploying a small number of geolocators on individuals in the target population could allow researchers to more reliably locate rare species or repeatedly target focal individuals and associated conspecifics for drone-based observation.

Programming proficiency

All code provided in the worked examples is written in the Python programming language. We expect limited additional programming will be required to apply this code to new datasets that are similar in scope to the examples. However, we expect many researchers will want to adapt our base code to better suit their needs, which will require some degree of programming skill, depending on the functionality required. Furthermore, since this method generates large volumes of high-resolution and high-dimensional data, certain programming capabilities will prove essential for effective visualization and analysis (see below for possible analysis directions).

Analysis, Extensions, & Applications

With recent advances in imaging technology and video analysis, drone-based behavioral observation is poised to become a widely-used approach in the study of wildlife ecology. Here we expand on the potential of this approach by outlining possible data analysis approaches as well as future extensions and applications of drone-based observation.

Data analysis

Our method provides a rich set of biologically-relevant features, but these data must be analyzed to address the research questions of interest. There exists a rich body of literature for exploring, visualizing, and analyzing animal trajectory data (Patterson et al., 2017; Seidel et al., 2018). The sub-second and sub-meter animal positions and detailed landscape data produced by this method allows researchers to use step-selection-type approaches to evaluate movement decisions across spatial and temporal scales, from actual individual steps to movement decisions at larger-scales (Fieberg et al., 2021). Tracking body keypoint locations (Step 5) can directly provide information about behaviorally-relevant body parts, such as head direction and body orientation. Depending on the position of the animal(s) within each frame, correctly interpreting animals' relative keypoint patterns may require taking lens geometry into account (see Supplement for details). When possible to estimate, this head and body information can be combined with the 3D landscape models (Step 4) to reconstruct estimates of each animal's visual field (Aben et al., 2018). These can provide a valuable estimate of the social and non-social information available to animals, and how they use this information to make decisions such as whom to follow (Strandburg-Peshkin et al., 2013) or how to respond to possible threats (Davies et al., 2016; Rosenthal et al., 2015; Sosna et al., 2019).

Future extensions

During detection (Step 2), if given sufficient image resolution, it may be possible to visually identify individual animals within and across observations. Individual recognition opens the door to studies of individual behavioral variation across time and contexts, and the role of individual behavior in driving processes at the population and community levels (Costa-Pereira et al., 2022). Beyond laboratory settings (Walter & Couzin, 2021) individual recognition in wild populations is increasingly feasible (Norman et al., 2017; Tuia et al., 2022). Alternatively, individuals could be identified using ground-level photographs or direct observation, and then these identities could be manually linked to individuals in the aerial recordings.

A further advance would be to use temporal keypoint data, or body posture trajectories (Fig. 1 - Step 5), to define finescale behavioral labels. While similar to the problem of accelerometer-based behavioral classification, a vision-based approach provides the added benefit of ground-truth videos for validation (Brown et al., 2013; Wang et al., 2015). Both supervised (Bohnslav et al., 2021) and unsupervised/self-supervised (Berman et al., 2014) approaches, or a combination of both (Sun et al., 2021), could be applied to achieve automated behavioral annotations describing the behavioral states of individuals and groups.

Beyond data processing, advancements with drones' on board automated visual tracking (Islam et al., 2019) and the ability to automatically coordinate flight among multiple drones (Zhou et al., 2022) could help to streamline complex operations, reduce the risk of human error, and also facilitate further observation techniques such as multiview 3D posture tracking (Tallamraju et al., 2019). Drones could thus be deployed to autonomously find and record individuals of a target species. While these are exciting future steps, current regulations in many countries would prevent the use of these methods without special permissions or certifications. Thus, regulatory rather than technological restrictions may be the most substantial barriers to large-scale automated observation.

Synergizing with other remote-sensing methods

There is great potential in combining this approach with existing remote sensing methods. In the context of biologging, drones could be deployed at key times or locations of interest to provide high-resolution behavioral snapshots of tagged individuals along with their social and environmental context. Image-based data could also be used to guide biologging study design choices. For example, one could downsample the high resolution image-based data to determine the optimal sampling frequency for bio-logging studies, using the videos to verify that the resulting data capture the target behaviors. Recording and quantifying the behavior of instrumented animals could also aid in the development of behaviorally activated "smart" sensors (Korpela et al., 2020; Yu, 2021).

The geolocated animal movement and environmental data generated by drone-based methods can be combined with multi-modal remote sensing data (which is typically acquired over a broad scale, but at substantially-lower resolution) to explore the interplay between animals and important environmental features. For example, SfM models can be combined with remote sensing data to enable calculation of microclimate variability (Maclean et al. 2019; Duffy et al. 2021) or water flow and saturation (Koci et al., 2020). Higher accuracy 3D landscape models, such as those generated by lidar imaging (D'Urban Jackson et al. 2020), could enable studies of herbivore impacts on vegetation via consumption or trampling effects, or allow higher precision visual field calculations in difficult-to-model environments like tall grass. Furthermore, integrating the 3D structure of landscapes with calibrated multi- or hyperspectral measures of landscape properties can produce maps of resource quality and accessibility for foraging studies (Jennewein et al. 2021).

Conclusion

We present a method that allows researchers to study animal behavior in its natural social and environmental context in a non-invasive and scalable way. Our approach is independent of specific species and can be deployed across a range of study systems making this a powerful and versatile tool for many researchers across fields. Importantly, as researchers work to understand the relationship between animals and their landscapes in a changing world, this method, which simultaneously records both at high resolution, is poised to be an important new way of observing the natural world.

500

Term	**Definition**
active learning	An approach related to model-assisted labeling wherein the researcher uses model-generated confidence scores to concentrate annotation effort on difficult examples.
anchor frames	The subset of video frames in our method input into the structure from motion software to generate the 3D landscape model and camera location information.
annotation	The process of labeling objects of interest in training imagery, for example by drawing bounding boxes around individual animals.
bounding box	A rectangle enclosing an object of interest in an image. Bounding boxes can be drawn by the user as a form of annotation, or can be generated by detection models to denote the predicted location of an object of interest.
graphical user interface (GUI)	A means of viewing or inputting data that relies on graphical elements (e.g. buttons) rather than coding inputs.
ground control point (GCP)	Locations or landmarks with known real-world geographic coordinates. These points are used to georeference the landscape model generated from the anchor frames.
image augmentation	A technique for increasing the effective size of the training set by modifying each training image each time it is shown to the model. Modifications, including blurring, rotating, and adjusting contrast, and are intended to mimic the diversity of images in the entire dataset.
model-assisted labeling	An iterative process where image annotation and model training are conducted in parallel. The user initially labels a small number of images and uses these to train an initial version of the model. This initial model is then used to generate annotations for the remaining images, which the user then confirms or corrects while also adding annotations for any animals that the model missed.
orthomosaic	Two-dimensional composite images generated by the SfM software that approximate the appearance of 3D structures from an overhead viewpoint.
point cloud	A set of 3D points often also containing a color value that can be used to represent landscapes or objects in space.
pretraining	A feature of many common deep learning models where the model has been initially trained on large datasets consisting of generic imagery of common scenes and objects. This allows the model to learn the basic, universal aspects of imagery before being fine-tuned on the user's specific dataset.
semantic segmentation	The process of labeling every pixel in an image as one of a set number of object classes.
structure from motion (SfM)	A technique for 3D reconstruction in which two-dimensional images from various overlapping viewpoints are used to define the geometry of

	the target structure (here, the landscape surrounding the observed animals).
triangular mesh	A 3D surface model created from a point cloud by connecting triads of points to create flat triangular surfaces.

Table 1. Glossary of terms. Defined terms are bolded at first appearance in the text.

Authors contributions

Benjamin Koger, Blair Costelloe and Iain Couzin developed the idea and goals of the method. Benjamin Koger, Blair Costelloe and Jeffrey Kerby coordinated and conducted data collection. Benjamin Koger, Blair Costelloe, Adwait Desphande, Jacob Graving and Jeffrey Kerby curated the data. Benjamin Koger designed the analytical pipeline; wrote the code in the worked examples and performed all aspects of model development, including training and validation. Jacob Graving advised on the design, organization, and development of earlier versions of the code. Blair Costelloe and Iain Couzin provided supervision and secured financial support. Blair Costelloe, Iain Couzin and Jeffrey Kerby administered and contributed material resources to the project. Benjamin Koger and Blair Costelloe wrote the initial manuscript draft and prepared the figures. All authors edited and revised the manuscript and gave final approval for publication.

Acknowledgements

Figures 1 and 4 were created by Mike Costelloe. We thank Mike Costelloe, Sven Lauke and Felicitas Oehler for assistance with collecting and annotating the ungulate data. We thank Guassa Gelada Research Project field managers Odin Bernardo and Nate Redon for collecting the gelada drone data, and project directors, camp staff members, and the local community for their support. We thank Anne Scharf for sharing code used to plan DJI mapping flights. We thank Jean Luo and David Basili for contributions to early versions of the game trail detection and track correction GUI. B.R.C. received support from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 748549. B.R.C. acknowledges support from the University of Konstanz's Investment Grant program. J.T.K acknowledges support from the Neukom Institute for Computational Science at Dartmouth College and the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 754513 and the Aarhus University Research Foundation. B.K., I.D.C, J.M.G., B.R.C., and A.D. acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – 'Centre for the Advanced Study of Collective Behaviour' EXC 2117-422037984. I.D.C gratefully acknowledges support from the Office of Naval Research (ONR) grant N00014-19-1-2556 and the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 860949. A.D. acknowledges the support by the Swiss National Science Foundation (Early PostDoc Fellowship: P2NEP3_200190). J.M.G. and B.R.C. acknowledge support from NVIDIA Corporation's Academic Hardware Grant Program.

Permissions and Ethics Statement

We imported and operated drones in Kenya with permission from the Kenya Civil Aviation Authority (import permits: KCAA/RPA/PERMIT-2017-0006, KCAA/RPA/PERMIT-2017-0007, KCAA/RPA/PERMIT-2017-0008, KCAA/ASSR/RPA/PERMIT-0016, KCAA/ASSR/RPA/PERMIT-0017, KCAA/ASSR/RPA/PERMIT-0018; authorization numbers: KCAA/OPS/2117/4 Vol. 2 (80), KCAA/OPS/2117/4 Vol. 2 (81), KCAA/OPS/2117/5 (86), KCAA/OPS/2117/5 (87); operator certificate numbers: RPA/TP/0005, RPA/TP/0000-0009). We conducted research in Kenya with permission from the Kenyan National Commission for Science, Technology and Innovation (research permits: NACOSTI/P/17/59088/15489, NACOSTI/P/59088/21567) and in affiliation with the Kenya Wildlife Service. All research activities pertaining to the ungulate data, including drone operation, were performed with the knowledge and support of management and security staff at our field sites, Ol Pejeta Conservancy and Mpala Research Centre. Data collection protocols for the ungulate data were reviewed and approved by Ethikrat, the independent ethics council of the Max Planck Society. All research activities pertaining to gelada monkeys, including drone operation, were undertaken with the knowledge and approval of the Guassa Community Conservation Area leadership and under the approval of a memorandum of understanding between the Ethiopian Wildlife Conservation Authority and the Guassa Gelada Research Project.

Data availability statement

Code for the worked examples is available on GitHub at https://github.com/benkoger/overhead-video-worked-examples. All data for the worked examples are archived on EDMOND at https://edmond.mpdl.mpg.de/privateurl.xhtml?token=9c1a978a-21f3-4843-bebe-40c296bffc73. DOIs and stable links to these resources will be generated upon final acceptance of the manuscript.

References

Aben, J., Pellikka, P., & Travis, J. M. J. (2018). A call for viewshed ecology: Advancing our understanding of the ecology of information through viewshed analysis. *Methods in Ecology and Evolution*, *9*(3), 624–633. https://doi.org/10.1111/2041-210X.12902

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3–4), 227–266. https://doi.org/10.1163/156853974X00534

Benitez-Paez, F., Brum-Bastos, V. da S., Beggan, C. D., Long, J. A., & Demšar, U. (2021). Fusion of wildlife tracking and satellite geomagnetic data for the study of animal migration. *Movement Ecology*, *9*(1), 31. https://doi.org/10.1186/s40462-021-00268-4

Berman, G. J., Choi, D. M., Bialek, W., & Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, *11*(99), 20140672. https://doi.org/10.1098/rsif.2014.0672

Bevan, E., Whiting, S., Tucker, T., Guinea, M., Raith, A., & Douglas, R. (2018). Measuring behavioral responses of sea turtles, saltwater crocodiles, and crested terns to drone disturbance to define ethical operating thresholds. *PLOS ONE*, *13*(3), e0194460. https://doi.org/10.1371/journal.pone.0194460

Bohnslav, J. P., Wimalasena, N. K., Clausing, K. J., Dai, Y. Y., Yarmolinsky, D. A., Cruz, T., Kashlan, A. D., Chiappe, M. E., Orefice, L. L., Woolf, C. J., & Harvey, C. D. (2021). DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. *ELife*, *10*, e63377. https://doi.org/10.7554/eLife.63377

Brown, D. D., Kays, R., Wikelski, M., Wilson, R., & Klimley, A. (2013). Observing the unwatchable through acceleration logging of animal behavior. *Animal Biotelemetry*, *1*(1), 20. https://doi.org/10.1186/2050-3385-1-20

Brum-Bastos, V., Long, J., Church, K., Robson, G., de Paula, R., & Demšar, U. (2020). Multi-source data fusion of optical satellite imagery to characterize habitat selection from wildlife tracking data. *Ecological Informatics*, *60*, 101149. https://doi.org/10.1016/j.ecoinf.2020.101149

Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., & Simpson, S. J. (2006). From Disorder to Order in Marching Locusts. *Science*, *312*(5778), 1402–1406. https://doi.org/10.1126/science.1125142

Ceballos, G., Ehrlich, P. R., & Raven, P. H. (2020). Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National*

Academy of Sciences, *117*(24), 13596–13602. https://doi.org/10.1073/pnas.1922686117

Christiansen, F., Rojano-Doñate, L., Madsen, P. T., & Bejder, L. (2016). Noise levels of multi-rotor unmanned aerial vehicles with implications for potential underwater impacts on marine mammals. *Frontiers in Marine Science*, *3*. https://doi.org/10.3389/fmars.2016.00277

Costa-Pereira, R., Moll, R. J., Jesmer, B. R., & Jetz, W. (2022). Animal tracking moves community ecology: Opportunities and challenges. *Journal of Animal Ecology*, 1365-2656.13698. https://doi.org/10.1111/1365-2656.13698

Creel, S., & Creel, N. M. (1995). Communal hunting and pack size in African wild dogs, Lycaon pictus. *Animal Behaviour*, *50*(5), 1325–1339. https://doi.org/10.1016/0003-3472(95)80048-4

Davies, A. B., Tambling, C. J., Kerley, G. I. H., & Asner, G. P. (2016). Limited spatial response to direct predation risk by African herbivores following predator reintroduction. *Ecology and Evolution*, *6*(16), 5728–5748. https://doi.org/10.1002/ece3.2312

Dell, A. I., Bender, J. A., Branson, K., Couzin, I. D., de Polavieja, G. G., Noldus, L. P. J. J., Pérez-Escudero, A., Perona, P., Straw, A. D., Wikelski, M., & Brose, U. (2014). Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution*, *29*(7), 417–428. https://doi.org/10.1016/j.tree.2014.05.004

Ditmer, M. A., Vincent, J. B., Werden, L. K., Tanner, J. C., Laske, T. G., Iaizzo, P. A., Garshelis, D. L., & Fieberg, J. R. (2015). Bears show a physiological but limited behavioral response to unmanned aerial vehicles. *Current Biology*, *25*(17), 2278–2283. https://doi.org/10.1016/j.cub.2015.07.024

Duffy, J. P., Cunliffe, A. M., DeBell, L., Sandbrook, C., Wich, S. A., Shutler, J. D., Myers-Smith, I. H., Varela, M. R., & Anderson, K. (2018). Location, location, location:

Considerations when using lightweight drones in challenging environments. *Remote Sensing in Ecology and Conservation*, *4*(1), 7–19. https://doi.org/10.1002/rse2.58

Duporge, I., Spiegel, M. P., Thomson, E. R., Chapman, T., Lamberth, C., Pond, C., Macdonald, D. W., Wang, T., & Klinck, H. (2021). Determination of optimal flight altitude to minimise acoustic drone disturbance to wildlife using species audiograms. *Methods in Ecology and Evolution*, *12*(11), 2196–2207. https://doi.org/10.1111/2041-210X.13691

D'Urban Jackson, T., Williams, G. J., Walker-Springett, G., & Davies, A. J. (2020). Three-dimensional digital mapping of ecosystems: A new era in spatial ecology. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1920), 20192383. https://doi.org/10.1098/rspb.2019.2383

Ehlers, L., Coulombe, G., Herriges, J., Bentzen, T., Suitor, M., Joly, K., & Hebblewhite, M. (2021). Critical summer foraging tradeoffs in a subarctic ungulate. *Ecology and Evolution*, *11*(24), 17835–17872. https://doi.org/10.1002/ece3.8349

Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A 'How to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, *90*(5), 1027–1043. https://doi.org/10.1111/1365-2656.13441

Flack, A., Nagy, M., Fiedler, W., Couzin, I. D., & Wikelski, M. (2018). From local collective behavior to global migratory patterns in white storks. *Science*, *360*(6391), 911–914. https://doi.org/10.1126/science.aap7781

Flack, A., Schaeffer, P. J., Taylor, J. R. E., Müller, I., Wikelski, M., & Fiedler, W. (2020). Daily energy expenditure in white storks is lower after fledging than in the nest. *Journal of Experimental Biology*, jeb.219337. https://doi.org/10.1242/jeb.219337

Francisco, F. A., Nührenberg, P., & Jordan, A. (2020). High-resolution, non-invasive animal tracking and reconstruction of local environment in aquatic ecosystems. *Movement Ecology*, *8*(1), 27. https://doi.org/10.1186/s40462-020-00214-w

Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., & Couzin, I. D. (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *ELife*, *8*, e47994. https://doi.org/10.7554/eLife.47994

Haalck, L., Mangan, M., Webb, B., & Risse, B. (2020). Towards image-based animal tracking in natural environments using a freely moving camera. *Journal of Neuroscience Methods*, *330*, 108455. https://doi.org/10.1016/j.jneumeth.2019.108455

Hughey, L. F., Hein, A. M., Strandburg-Peshkin, A., & Jensen, F. H. (2018). Challenges and solutions for studying collective animal behaviour in the wild. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1746), 20170005. https://doi.org/10.1098/rstb.2017.0005

Inoue, S., Yamamoto, S., Ringhofer, M., Mendonça, R. S., Pereira, C., & Hirata, S. (2019). Spatial positioning of individuals in a group of feral horses: A case study using drone technology. *Mammal Research*, *64*(2), 249–259. https://doi.org/10.1007/s13364-018-0400-2

Islam, M. J., Hong, J., & Sattar, J. (2019). Person-following by autonomous robots: A categorical overview. *The International Journal of Robotics Research*, *38*(14), 1581–1618. https://doi.org/10.1177/0278364919881683

Jesmer, B. R., Merkle, J. A., Goheen, J. R., Aikens, E. O., Beck, J. L., Courtemanch, A. B., Hurley, M. A., McWhirter, D. E., Miyasaki, H. M., Monteith, K. L., & Kauffman, Matthew. J. (2018). Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, *361*(6406), 1023–1025. https://doi.org/10.1126/science.aat0985

Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C., & Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, *108*(46), 18720–18725.

https://doi.org/10.1073/pnas.1107583108

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*(6240). https://doi.org/10.1126/science.aaa2478

Klarevas-Irby, J. A., Wikelski, M., & Farine, D. R. (2021). Efficient movement strategies mitigate the energetic cost of dispersal. *Ecology Letters*, *24*(7), 1432–1442. https://doi.org/10.1111/ele.13763

Koci, J., Sidle, R. C., Jarihani, B., & Cashman, M. J. (2020). Linking hydrological connectivity to gully erosion in savanna rangelands tributary to the Great Barrier Reef using structure-from-motion photogrammetry. *Land Degradation & Development*, *31*(1), 20–36. https://doi.org/10.1002/ldr.3421

Korpela, J., Suzuki, H., Matsumoto, S., Mizutani, Y., Samejima, M., Maekawa, T., Nakai, J., & Yoda, K. (2020). Machine learning enables improved runtime and precision for bio-loggers on seabirds. *Communications Biology*, *3*(1), 633. https://doi.org/10.1038/s42003-020-01356-8

Kuhn, H. W. (1955). The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly*, *2*, 83–97.

Loftus, J. C., Harel, R., Núñez, C. L., & Crofoot, M. C. (2022). Ecological and social pressures interfere with homeostatic sleep regulation in the wild. *ELife*, *11*, e73695. https://doi.org/10.7554/eLife.73695

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

McEvoy, J. F., Hall, G. P., & McDonald, P. G. (2016). Evaluation of unmanned aerial vehicle shape, flight path and camera type for waterfowl surveys: Disturbance effects and species recognition. *PeerJ*, *4*. https://doi.org/10.7717/peerj.1831

Moll, R. J., Millspaugh, J. J., Beringer, J., Sartwell, J., & He, Z. (2007). A new 'view' of ecology and conservation through animal-borne video systems. *Trends in Ecology & Evolution*, *22*(12), 660–668. https://doi.org/10.1016/j.tree.2007.09.007

Mulero-Pázmány, M., Jenni-Eiermann, S., Strebel, N., Sattler, T., Negro, J. J., & Tablado, Z. (2017). Unmanned aircraft systems as a new source of disturbance for wildlife: A systematic review. *PLOS ONE*, *12*(6), e0178448. https://doi.org/10.1371/journal.pone.0178448

Murray, D. L., & Fuller, M. R. (2000). A critical review of the effects of marking on the biology of vertebrates. In L. Boitani & T. Fuller (Eds.), *Research Techniques in Animal Ecology: Controversies and Consequences* (pp. 15–64). Columbia University Press.

Naganuma, T., Tanaka, M., Tezuka, S., M.J.G. Steyaert, S., Tochigi, K., Inagaki, A., Myojo, H., Yamazaki, K., & Koike, S. (2021). Animal-borne video systems provide insight into the reproductive behavior of the Asian black bear. *Ecology and Evolution*, *11*(14), 9182–9190. https://doi.org/10.1002/ece3.7722

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Norman, B. M., Holmberg, J. A., Arzoumanian, Z., Reynolds, S. D., Wilson, R. P., Rob, D., Pierce, S. J., Gleiss, A. C., de la Parra, R., Galvan, B., Ramirez-Macias, D., Robinson, D., Fox, S., Graham, R., Rowat, D., Potenski, M., Levine, M., Mckinney, J. A., Hoffmayer, E., … Morgan, D. L. (2017). Undersea Constellations: The Global Biology of an Endangered Marine Megavertebrate Further Informed through Citizen Science. *BioScience*, *67*(12), 1029–1043. https://doi.org/10.1093/biosci/bix127

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch:

An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., & King, R. (2017). Statistical modelling of individual animal movement: An overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, *101*(4), 399–438. https://doi.org/10.1007/s10182-017-0302-7

Pereira, T. D., Tabris, N., Li, J., Ravindranath, S., Papadoyannis, E. S., Wang, Z. Y., Turner, D. M., McKenzie-Smith, G., Kocher, S. D., Falkner, A. L., Shaevitz, J. W., & Murthy, M. (2020). *SLEAP: Multi-animal pose tracking* (p. 2020.08.31.276246). https://doi.org/10.1101/2020.08.31.276246

Raoult, V., Tosetto, L., & Williamson, J. E. (2018). Drone-based high-resolution tracking of aquatic vertebrates. *Drones*, *2*(4), 37. https://doi.org/10.3390/drones2040037

Rattenborg, N. C., Voirin, B., Cruz, S. M., Tisdale, R., Dell'Omo, G., Lipp, H.-P., Wikelski, M., & Vyssotski, A. L. (2016). Evidence that birds sleep in mid-flight. *Nature Communications*, *7*(1), 12468. https://doi.org/10.1038/ncomms12468

Ringhofer, M., Go, C. K., Inoue, S., S. Mendonça, R., Hirata, S., Kubo, T., Ikeda, K., & Yamamoto, S. (2020). Herding mechanisms to maintain the cohesion of a harem group: Two interaction phases during herding. *Journal of Ethology*, *38*(1), 71–77. https://doi.org/10.1007/s10164-019-00622-5

Rosenthal, S. B., Twomey, C. R., Hartnett, A. T., Wu, H. S., & Couzin, I. D. (2015). Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, *112*(15), 4690–4695. https://doi.org/10.1073/pnas.1420068112

Seidel, D. P., Dougherty, E., Carlson, C., & Getz, W. M. (2018). Ecological metrics and methods for GPS movement data. *International Journal of Geographical Information Science*, *32*(11), 2272–2293. https://doi.org/10.1080/13658816.2018.1498097

Sosna, M. M. G., Twomey, C. R., Bak-Coleman, J., Poel, W., Daniels, B. C., Romanczuk, P., & Couzin, I. D. (2019). Individual and collective encoding of risk in animal groups. *Proceedings of the National Academy of Sciences*, *116*(41), 20556–20561. https://doi.org/10.1073/pnas.1905585116

Sprogis, K. R., Videsen, S., & Madsen, P. T. (2020). Vessel noise levels drive behavioural responses of humpback whales with implications for whale-watching. *ELife*, *9*, e56760. https://doi.org/10.7554/eLife.56760

Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., & Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science*, *348*(6241), 1358–1361. https://doi.org/10.1126/science.aaa5099

Strandburg-Peshkin, A., Farine, D. R., Crofoot, M. C., & Couzin, I. D. (2017). Habitat and social factors shape individual decisions and emergent group structure during baboon collective movement. *ELife*, *6*, e19505. https://doi.org/10.7554/eLife.19505

Strandburg-Peshkin, A., Twomey, C. R., Bode, N. W. F., Kao, A. B., Katz, Y., Ioannou, C. C., Rosenthal, S. B., Torney, C. J., Wu, H. S., Levin, S. A., & Couzin, I. D. (2013). Visual sensory networks and effective information transfer in animal groups. *Current Biology*, *23*(17), R709–R711. https://doi.org/10.1016/j.cub.2013.07.059

Sun, J. J., Kennedy, A., Zhan, E., Anderson, D. J., Yue, Y., & Perona, P. (2021). Task Programming: Learning Data Efficient Behavior Representations. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2875–2884. https://doi.org/10.1109/CVPR46437.2021.00290

Tallamraju, R., Price, E., Ludwig, R., Karlapalem, K., Bülthoff, H. H., Black, M. J., & Ahmad, A. (2019). Active Perception Based Formation Control for Multiple Aerial Vehicles.

IEEE Robotics and Automation Letters, *4*(4), 4491–4498. https://doi.org/10.1109/LRA.2019.2932570

Torney, C. J., Lamont, M., Debell, L., Angohiatok, R. J., Leclerc, L.-M., & Berdahl, A. M. (2018). Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1746), 20170385. https://doi.org/10.1098/rstb.2017.0385

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), 792. https://doi.org/10.1038/s41467-022-27980-y

Tuyttens, F. A. M., de Graaf, S., Heerkens, J. L. T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E., & Ampe, B. (2014). Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? *Animal Behaviour*, *90*, 273–280. https://doi.org/10.1016/j.anbehav.2014.02.007

Walter, T., & Couzin, I. D. (2021). TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *ELife*, *10*, e64000. https://doi.org/10.7554/eLife.64000

Wang, Y., Nickel, B., Rutishauser, M., Bryce, C. M., Williams, T. M., Elkaim, G., & Wilmers, C. C. (2015). Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. *Movement Ecology*, *3*(1), 2. https://doi.org/10.1186/s40462-015-0030-0

Weimerskirch, H., Prudor, A., & Schull, Q. (2018). Flights of drones over sub-Antarctic seabirds show species- and status-specific behavioural and physiological responses. *Polar Biology*, *41*(2), 259–266. https://doi.org/10.1007/s00300-017-2187-z

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., Grissac, S., Demšar, U., English, H. M., Franconi, N., Gómez-Laich, A., Griffiths, R. C., Kay, W. P., Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R. P., & Börger, L. (2020). Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology*, *89*(1), 186–206. https://doi.org/10.1111/1365-2656.13094

Williams, T. M., Wolfe, L., Davis, T., Kendall, T., Richter, B., Wang, Y., Bryce, C., Elkaim, G. H., & Wilmers, C. C. (2014). Instantaneous energetics of puma kills reveal advantage of felid sneak attacks. *Science*, *346*(6205), 81–85. https://doi.org/10.1126/science.1254885

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). *Detectron2*. https://github.com/facebookresearch/detectron2

Yu, H. (2021). *An evaluation of machine learning classifiers for next-generation, continuous-ethogram smart trackers*. 14.

Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zhou, X., Wen, X., Wang, Z., Gao, Y., Li, H., Wang, Q., Yang, T., Lu, H., Cao, Y., Xu, C., & Gao, F. (2022). Swarm of micro flying robots in the wild. *Science Robotics*, *7*(66), eabm5954. https://doi.org/10.1126/scirobotics.abm5954