BaboonLand Dataset: Tracking Primates in the Wild and Automating Behaviour Recognition from Drone Videos

Isla Duporge*, Maksim Kholiavchenko*, Roi Harel, Scott Wolf, Daniel Rubenstein, Meg Crofoot, Tanya Berger-Wolf, Stephen Lee, Julie Barreau, Jenna Kline, Michelle Ramirez, Charles Stewart

*** Contributed equally.**

Abstract

Using unmanned aerial vehicles (UAVs) to track multiple individuals simultaneously in their natural environment is a powerful approach for better understanding group primate behavior. Previous studies have demonstrated that it is possible to automate the classification of primate behavior from video data, but these studies have been carried out in captivity or from ground-based cameras. However, to understand group behavior and the self-organization of a collective, the whole troop needs to be seen at a scale where behavior can be seen in relation to the natural environment in which ecological decisions are made. This study presents a novel dataset for baboon detection, tracking, and behavior recognition from drone videos. The foundation of our dataset is videos from drones flying over the Mpala Research Centre in Kenya. The baboon detection dataset was created by manually annotating all baboons in drone videos with bounding boxes. A tiling method was subsequently applied to create a pyramid of images at various scales from the original 5.3K resolution images, resulting in approximately 30K images used for baboon detection. The baboon tracking dataset is derived from the baboon detection dataset, where all bounding boxes are consistently assigned the same ID throughout the video. This process resulted in half an hour of very dense tracking data. The baboon behavior recognition dataset was generated by converting tracks into mini-scenes, a video subregion centered on each animal, each mini-scene was manually annotated with 12 distinct behavior types, and one additional category for occlusion, resulting in over 20 hours of data. Benchmark results show mean average precision (mAP) of 92.62% for the YOLOv8-X detection model, multiple object tracking precision (MOTA) of 63.81% for the BotSort tracking algorithm, and micro top-1 accuracy of 63.97% for the X3D behavior recognition model. Using deep learning to rapidly and accurately classify wildlife behavior from drone footage facilitates non-invasive data collection on behavior enabling the behavior of a whole group to be systematically and accurately recorded.

The dataset can be accessed at https://baboonland.xyz.

Dataset

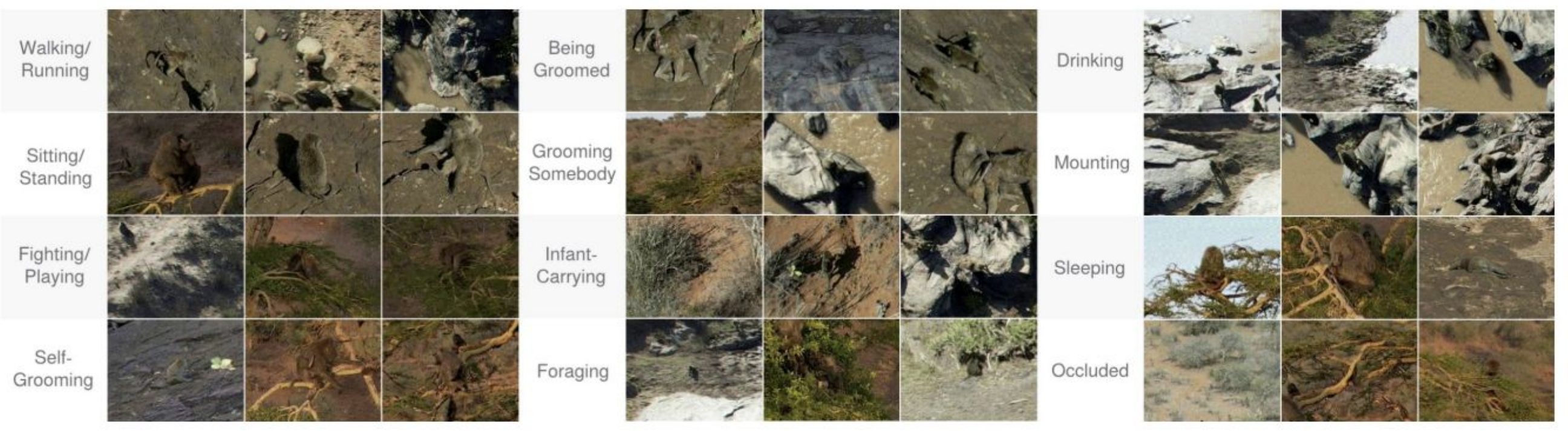
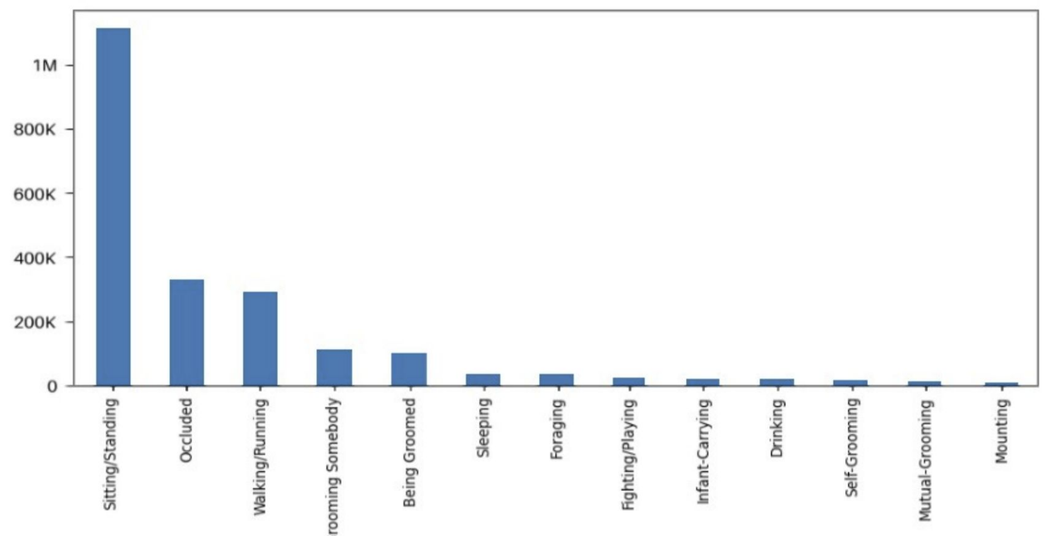


Figure 1. Examples of the behavior categories in the BaboonLand dataset: "Walking/Running", "Sitting/Standing", "Fighting/Playing", "Self-Grooming", "Being Groomed", "Grooming Somebody", "Mutual Grooming", "Infant-Carrying", "Foraging", "Drinking", "Mounting", "Sleeping", and "Occluded".

Olive baboons (Papio anubis) were used as the test species for this study. The troops of baboons reside in Mpala, an open landscape in Laikipia county, central Kenya, comprising savannah and dry woodland habitat bordered by the Ewaso Ngiro and Ewaso Narok rivers. Mpala is not fenced; thus, these wild primates were observed in an unconstrained natural environment. The videos were collected in various background environments, including when sleeping on a tree, on a rock, during a river crossing, in an open savannah, and on a cliff. The baboon detection dataset was created by manually annotating all baboons in the videos with bounding boxes. A tiling method was then used to generate a pyramid of images at different scales from the original 5.3K resolution footage, resulting in approximately 30K images for baboon detection. The baboon tracking dataset was derived from the detection dataset, with each bounding box consistently assigned the same ID throughout the video, yielding half an hour of dense tracking data. The tracks were converted into mini-scenes for the baboon behavior recognition dataset, which are video subregions centered on each animal. Each mini-scene was manually annotated with 12 distinct behavior types and an additional category for occlusions, resulting in about 20 hours of data. Our behavior dataset exhibits a long-tailed distribution, indicating significant disparities in the number of samples across different categories. This is anticipated, as some behaviors are naturally more prevalent in baboons than others.



We provide the original drone videos along with the corresponding tracks and bounding boxes. Additionally, the dataset includes several evaluation sets: one for tracking algorithms (with 75% of each video used for training and 25% for testing), a YOLO-formatted dataset for training and evaluating detection (80% of images for training, 7% for validation, and 13% for testing), and a dataset in the Charades format for training and evaluating behavior recognition models (75% of videos for training and 25% for testing). We also provide all of the data processing scripts that we used to generate these datasets from original drone videos. Tracks and behavior annotations in the BaboonLand dataset are stored in the simplified CVAT for video 1.1 format. These can be uploaded to CVAT for exploratory data analysis and annotation adjustments. After making any annotation adjustments, the provided scripts can be used to regenerate all of the sub-datasets (detection, tracking, and behavior recognition).

Evaluation

5.1. Detection

Our dataset for object detection was evaluated using YOLOv8-X [56]. Given the small size of the baboons in the original footage, a specialized approach was necessary to train a detector effectively. We adopted a tiled methodology, generating a multi-scale pyramid (2x2, 3x3, 4x4) to encompass varying perspectives and dimensions of the target object.

Model	mAP@50	Precision	Recall
YOLOv8-X	**92.62**	**93.70**	**87.60**

Table 1. The results of YOLOv8-X model trained on our dataset. The model was trained for 64 epochs with 768x768 input resolution.

Examples of baboon detection from drone videos.

5.2. Tracking

We adopted ByteTrack [58] and BotSort [59] algorithms to evaluate the tracking performance. Both algorithms were built on the trained YOLOv8-X model described in the preceding section. The effectiveness of the tracking algorithm depends on several factors, including the quality of baboon detection and the precision of the detection association. In our evaluation, BotSort demonstrated superior results compared to ByteTrack.

Tracker	MOTA	MOTP	IDF1	P	R
ByteTrack	63.55	34.10	77.01	96.32	64.90
BotSort	**63.81**	**34.31**	**78.24**	**97.21**	**66.16**

Table 2. Evaluation results of ByteTrack and BotSort. The YOLOv8-X model was a backbone for both ByteTrack and BotSort algorithms.

Examples of tracking over a cliff, river, tree, and rock.

5.3. Behavior Recognition

To provide the baseline for behavior recognition, we trained I3D [60], SlowFast [61], and X3D [62] models on our dataset. We report micro (per instance) average and macro (per class) average accuracy. The confusion matrix depicted in Fig. 3 demonstrates the performance of the X3D model. We can see that the model performs quite well for common classes but rare behaviors are more challenging. The model tends to predict "Sitting/Standing for actions that encompass similar behaviors, such as "Drinking," "Foraging," and "Mounting". This indicates that while the model is effective at identifying frequent activities, it has difficulty distinguishing between less common behaviors that share visual similarities.

Average	Method	Top-1	Top-3	Top-5
Macro	I3D	26.53	54.51	65.47
	SlowFast	27.08	56.73	67.61
	X3D	**30.04**	**60.58**	**72.13**
Micro	I3D	61.29	89.38	92.34
	SlowFast	61.71	90.35	93.11
	X3D	**63.97**	**91.34**	**95.17**

Table 3. Results of I3D, SlowFast, and X3D models. I3D and X3D were trained with 16 input frames with a sampling rate of 5. For SlowFast, the Slow branch was trained with 16 input frames with a sampling rate of 5, and the Fast branch was trained with 4 input frames with a sampling rate of 5.

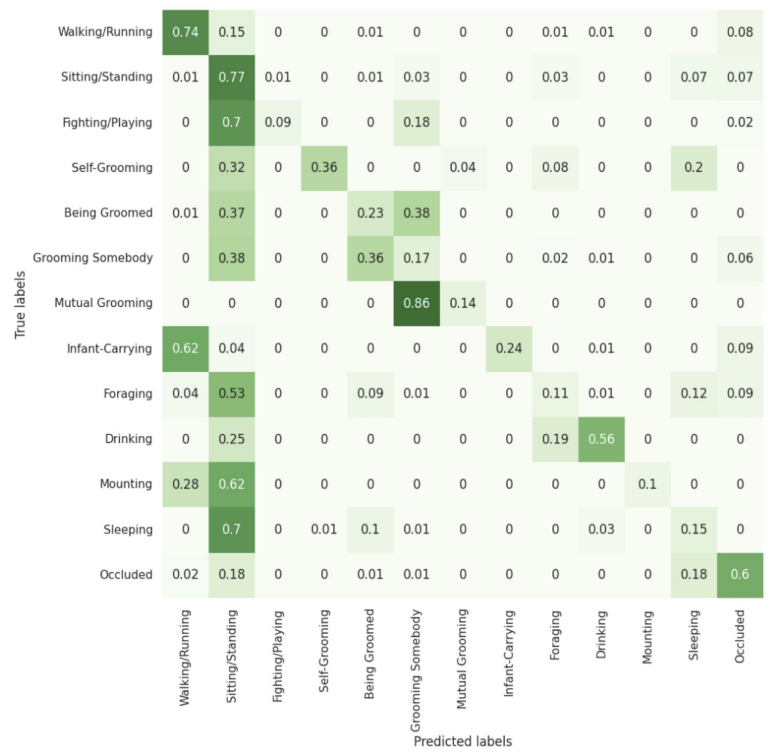


Figure 3. This confusion matrix showcases the performance of the X3D model, which is the top-performing model on the BaboonLand dataset based on our evaluation.

Conclusion

This paper presents a unique dataset encompassing a complex range of drone videos of Olive baboons moving between various background contexts, it can be used as a dataset to evaluate detection, tracking, and behavior recognition models. Our experiments demonstrate that the proposed dataset is a challenge for the evaluation of new state-of-the-art algorithms. The study of behavioral transitions of baboons provides a promising direction to explore how individuals are affected by their social environment, examining dyadic and ultra-dyadic interactions. The method we present here which automatically tracks collective behavior from UAV will be applicable to satellite video footage in the future. It is already possible to directly detect animal groups in satellite imagery and several companies are now developing satellites that can collect video. This will provide a more sophisticated method to monitor wildlife at grander spatial scales in the future.

Ethical Considerations

No humans can be distinguished in the videos, and the research was conducted under the authority of a Nacosti Research License. This license confirms adherence to the regulations enabling drone footage of animals to be collected in their natural habitats. We followed a protocol that strictly complies with the guidelines set forth by the Institutional Animal Care and Use Committee (IACUC). These guidelines are designed to ensure the ethical and humane treatment of animals involved in research activities. We flew at an altitude that did not disturb the baboons after calibrating this via several trial flights. We approached the animals from downwind, allowing drone noise to dissipate before reaching the animals.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Award No. 2118240 and Award No. 2112606. ID was supported by the National Academy of Sciences Research Associate Program and the United States Army Research Laboratory while conducting this study. ID collected all the UAV data on a Civil Aviation Authority Drone License CAA NQE Approval Number: 0216/1365 in conjunction with authorization from a KCAA operator under a Remote Pilot License. The data was gathered at the Mpala Research Centre in Kenya, in accordance with Research License No. NACOSTI/P/22/18214. The data collection protocol adhered strictly to the guidelines set forth by the Institutional Animal Care and Use Committee under permission No. IACUC 1835F.

Rensselaer
Princeton University
The Ohio State University
Max Planck Institute of Animal Behavior