OPEN ACCESS

University of BRISTOL

Citation (Version of Record):
Kline, J., Stevens, S., Maalouf, G., Rondeau Saint-Jean, C., Nguyen Ngoc, D., Mirmehdi, M., Guerin, D., Burghardt, T., Pastucha, E., Costelloe, B. R., Watson, I. M., Richardson, T. S., & Pagh Schultz Lundquist, U. (2025). *MMLA: Multi-Environment, Multi-Species, Low-Altitude Drone Dataset*. Paper presented at Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, Tennessee, United States.

Version:
Peer reviewed version

License (if available):
CC BY

Link to publication record on the Bristol Research Portal:
Access record including PDF (if available)

University of Bristol – Bristol Research Portal

This is the accepted author manuscript (AAM) of the article which has been made Open Access under the University of Bristol's Scholarly Works Policy. The final published version (Version of Record) can be found on the publisher's website. The copyright of any third-party content, such as images, remains with the copyright holder.

Citing this paper
Guidance on citing material from the Bristol Research Portal is available under 'Manuscripts and Citations' in the Terms of Use (linked below)

Terms of Use
Full Terms of Use for the Bristol Research Portal are available online: https://www.bristol.ac.uk/research-enterprise-innovation/research-policy/pure/brp-terms/

Has this research benefited you?
Let us know how reading this research here has been meaningful to you by sharing your story at https://forms.office.com/e/ffL5CZ0fGT

Bristol Research Portal Coversheet last updated 21-07-2026

MMLA: Multi-Environment, Multi-Species, Low-Altitude Drone Dataset

Jenna Kline[1], Samuel Stevens[1], Guy Maalouf[2], Camille Rondeau Saint-Jean[2], Dat Nguyen Ngoc[3], Majid Mirmehdi[3], David Guerin[4], Tilo Burghardt[3], Elzbieta Pastucha[2], Blair Costelloe[5], Matthew Watson[3], Thomas Richardson[3], Ulrik Pagh Schultz Lundquist[2]

[1]The Ohio State University [2]University of Southern Denmark [3] University of Bristol [4] WildDroneEU [5] Max Planck Institute of Animal Behavior

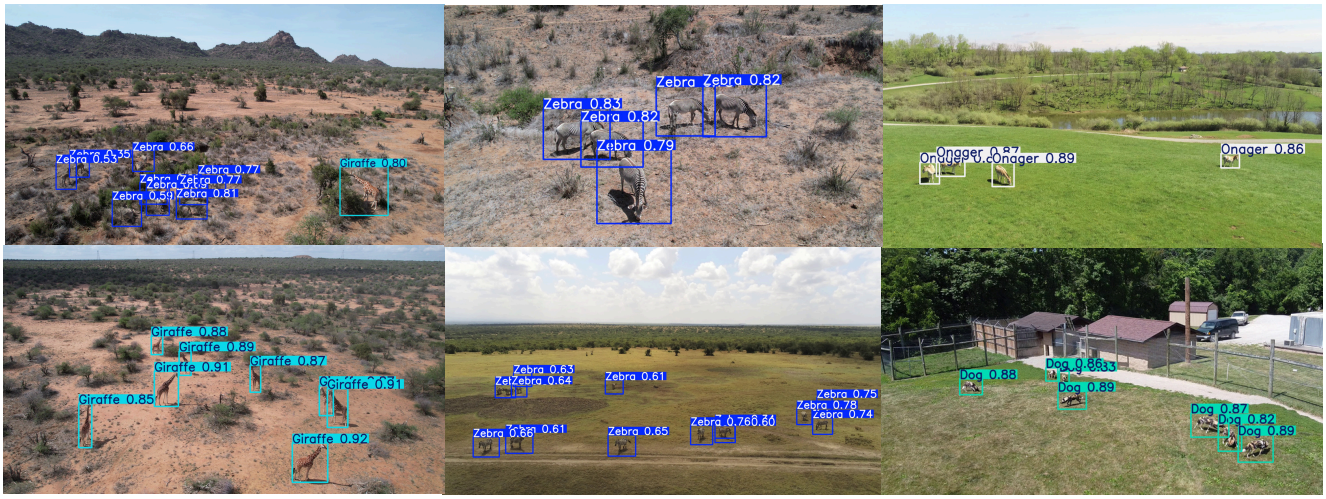

Figure 1. Examples of the MMLA dataset and example detection results from a YOLOv11m model fine-tuned on MMLA. The dataset spans multiple species and environments, enabling robust model training for low-altitude wildlife detection.

Abstract

Real-time wildlife detection in drone imagery supports critical ecological and conservation monitoring. However, standard detection models like YOLO often fail to generalize across locations and struggle with rare species, limiting their use in automated drone deployments. We present MMLA, a novel multi-environment, multi-species, low-altitude drone dataset collected across three sites (Ol Pejeta Conservancy and Mpala Research Centre in Kenya, and The Wilds in Ohio), featuring six species (zebras, giraffes, onagers, and African wild dogs). The dataset contains 811K annotations from 37 high-resolution videos. Baseline YOLO models show performance disparities across locations while fine-tuning YOLOv11m on MMLA improves mAP50 to 82%, a 52-point gain over baseline. Our results underscore the need for diverse training data to enable robust animal detection in autonomous drone systems.

1. Introduction

Animal ecology studies increasingly rely on vision-based technologies for efficient, noninvasive wildlife monitoring. Camera traps, high-altitude and low-altitude drones each offer distinct advantages for different research goals. Machine learning models tailored for animal ecology data can rapidly analyze these datasets to produce valuable insights [31], but their effectiveness varies significantly across monitoring approaches. While camera traps and high-altitude drone surveys have established methodologies and datasets [33,34], low-altitude drone studies represent an untapped opportunity. These deployments offer more detailed imagery than high-altitude approaches, covering broader areas than stationary camera traps and providing valuable insights into animal interactions and decision-making [21]. Critically, low-altitude drones are increasingly deployed as autonomous systems that can actively track wildlife [18,20], requiring models with high precision and localization.

While effective on benchmark datasets, standard computer vision models like YOLO [14] perform poorly when applied to wildlife detection in low-altitude drone imagery across varied natural environments. This performance gap stems from the specific requirements of autonomous wildlife-tracking drones: high precision detection with minimal false positives and accurate object localization (measured by IoU) are essential, as false detections or imprecise bounding boxes can cause autonomous systems to make incorrect flight decisions [20]. Additionally, environmental variability across field sites introduces substantial domain shifts, causing models trained in one location to fail when deployed elsewhere.

To address this gap, we present a novel **multi-environment, multi-species, low-altitude drone dataset (MMLA)**, the first low-altitude dataset explicitly designed for developing robust computer vision models for autonomous wildlife-tracking drones. Our dataset spans three distinct locations (Ol Pejeta Conservancy and Mpala Research Centre in Kenya and The Wilds Conservation Center in Ohio, USA) and six species: plains zebras, Grevy's zebras, reticulated and Masai giraffes, Persian onagers, and African wild dogs. MMLA contains 155K frames (1 hr, 26 min), with manually annotated bounding boxes around each animal optimized for training high-precision detection models.

The creation of MMLA represents a significant investment in advancing wildlife conservation technology, involving over 150 person-hours of fieldwork across two continents. By releasing this dataset and fine-tuned model weights, we invite the computer vision community to tackle these challenging real-world ecological problems and advance the state of the art in autonomous wildlife monitoring technology.

2. Background & Related Work

2.1. Monitoring Approaches for Animal Ecology

Vision-based wildlife monitoring approaches differ significantly in their technical specifications, applications, and computer vision requirements (Tab. 1 and Fig. 2). Camera traps capture medium to large-sized objects (128-256 pixels) [12] and excel at species identification but are stationary. High-altitude drones capture very small objects (0-16 pixels) [34] and are optimized for population counts across large areas but lack detail for individual ID [29] or behavioral analysis [15]. Low-altitude drone studies capture medium-sized objects (100 pixels) [15], offering a balance that enables behavioral analysis, individual ID, and tracking studies. The optimal resolution for these ecological studies is approximately 100-300 pixels per animal [20], which low-altitude drones provide consistently. Unlike camera traps and high-altitude drones, which typically process data in post-hoc batches, low-altitude drone applications often require real-time processing to support autonomous flight decisions [18, 20]. These fundamental differences in monitoring approaches have driven the development of specialized computer vision models for camera traps (e.g., MegaDetector [3]) and high-altitude imagery (e.g., HerdNet [9], POLO [24]). However, no pre-trained models specifically address the unique challenges of low-altitude drone imagery.

Feature	**Camera Trap**	**High-alt. Drone**	**Low-alt. Drone**
Size[1]	med-large (128-256)	v. small (0-16)	med (64-128)
Processing	Post-hoc batches	Post-hoc batches	Real-time video
Metric	Low false negative	F1-score	High prec. & mAP50
Vision task	Detect, Species ID	Counting, Species ID	Localization
Models	MegaDetector [3]	POLO [24], HerdNet [9]	**None**

Table 1. Comparison of vision-based animal ecology monitoring approaches. Low-altitude drone studies differ from camera trap and high-altitude drone studies for processing, relevant metrics, and computer vision tasks. **There are no pre-trained models aimed at low-altitude drone studies.**
[1]Typical animal size in pixels [12].

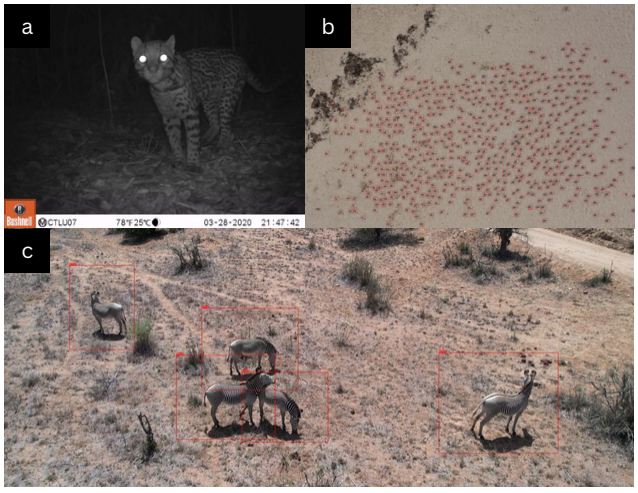

Figure 2. Animal imagery collected via (a) camera traps [33], (b) high-altitude drone [13], and (c) low-altitude drone missions [15].

2.2. Existing Wildlife Drone Datasets

A critical gap exists in wildlife monitoring datasets: no multi-species, multi-location aerial footage dataset pro-

Dataset	Categories	Image Count	Size[1]	Multi-Location[2]
Population estimates [10]	elephant, zebra, giraffe	561 RGB	50	✗
Bird detection [35]	bird	24K RGB	35	✓
Seabird colonies [11]	albatross, penguin	4K RGB	300	✗
Conservation Drones [5]	human, elephant, lion, dog	48 thermal videos	35	✓
Group behavior [21]	zebra, gazelle, waterbuck, buffalo, gelada, human	2K RGB	56	✓
KABR [15]	giraffe, zebra	130K videos	100	✗
WAID [25]	sheep, cattle, seal, camel, kiang, zebra	14K RGB	166	✗
WildLive [8]	zebra, giraffe, elephant	19K RGB	100	✗
WildlifeMapper [22]	zebra, hartebeest, cattle, shoats, zebra	11K RGB	—	✗
BuckTales [27]	antelope	22.5K RGB	1000	✗
MMLA (ours)	**zebra, onager, giraffe, African wild dog**	**155K RGB**	**100**	✓

[1]Size indicates the typical animal size in pixels. [2]Multi-Location indicates datasets capturing the same species across multiple geographic locations.

Table 2. Summary of multi-species aerial datasets. Note that MMLA has large animals (100px) and multiple locations, unlike prior work.

vides sufficient resolution with the same species captured across multiple environments (Tab. 2). While the global bird detection dataset [35] contains multi-location imagery across 13 ecosystems, the animals are very small (35 pixels) and limited to avian species. Existing mammal datasets [8, 10, 15, 21, 22, 25] contain thousands of annotated samples but are geographically limited to single locations for each species, creating detection models that may not transfer well across environments [6, 32]. Specialized datasets like BuckTales [27] and WildLive [8] focus on single locations or specific computer vision tasks. BuckTales provides 22.5K RGB frames of antelopes for multi-object tracking and re-identification. The WildLive dataset includes multiple species (zebras, giraffes, and elephants) but all were captured at a single location (Ol Pejeta Conservancy) [8]. These datasets highlight the importance of high-quality, ecologically relevant datasets tailored to specific computer vision problems but do not address the critical need for cross-environment generalization.

2.3. Animal Detection Models

Existing wildlife detection approaches are tailored for camera traps and high-altitude imagery, making them ill-suited for the demands of low-altitude drone studies. Camera trap pipelines like MegaDetector [3] focus primarily on reducing false negatives to eliminate blank images [33] before specialized species classification models process them. Models for high-altitude imagery use point-based approaches optimized for accurate counting from minimal pixel information [9, 24]. These methods fail to meet the high precision and accurate localization requirements of automated low-altitude drone studies. Low-altitude drones often need real-time processing with high precision metrics to guide autonomous flight decisions [18], where false positives can significantly impair performance by causing drones to track non-animal objects. Our MMLA dataset aims to bridge this gap by enabling the development of robust models that can generalize across environments and species with autonomous drones.

3. MMLA Dataset

The MMLA dataset[1] contains annotated drone collected for wildlife monitoring and conservation research. The data was collected at three locations: Ol Pejeta Conservancy and Mpala Research Centre, both in Laikipia, Kenya, and The Wilds Conservation Center, in Ohio, USA. The dataset contains six species: Plains zebra (*Equus quagga*), Grevy's zebra (*Equus grevyi*), Persian onager (*Equus hemionus onager*), Reticulated giraffe (*Giraffa reticulata*), Masai giraffe (*Giraffa camelopardalis tippelskirchi*), and African Wild Dog (*Lycaon pictus*). The variety of species and environments in the MMLA dataset enables the evaluation of model generalizability in detecting animals across different settings. The dataset includes 155,074 video frames, captured at 30 frames per second, and contains a variety of habitats (details in App. 9 Tab. 4). Each video frame has COCO format annotations (bounding boxes with normalized coordinates). The dataset contains 811K total annotations, which were completed manually by a team of experts using CVAT [7] and `kabr-tools` [19].

The **Ol Pejeta dataset** contains 29,268 frames of Plains zebras at Ol Pejeta Conservancy in Kenya (Jan-Feb 2025), collected using the autonomous WildWing system [18] across seven video sessions during the 2025 WildDrone Hackathon [36]. The **Mpala dataset** contains 104,062 frames of reticulated giraffes, Grevy's and Plains zebras at Mpala Research Center in Kenya (Jan 2023), collected manually using a DJI Air 2S drone across five sessions [15]. The **Wilds dataset** contains 21,744 frames of African wild dogs, Persian onagers, reticulated and Masai giraffes and

[1]https://huggingface.co/collections/imageomics/mmla-project

Grevy's zebras at The Wilds Conservation Center in Ohio (Apr-Jul 2024), collected using both DJI Mini and Parrot Anafi drones, with with a combination of autonomous WildWing system [16] and manual collection.

4. Evaluation

We evaluate the performance of three popular YOLO model (YOLOv5m, YOLOv8m, and YOLOv11m) for animal detection in the MMLA dataset. The YOLO (You Only Look Once) [14] family of models represents a popular choice for animal ecology field studies due to their speed, ease of use, and accuracy [17]. We selected medium-sized variants to balance detection performance with computational efficiency, as field studies often utilize edge devices, such as laptops and Raspberry Pis, with limited memory and compute [4, 38]. YOLO models perform detection in a single forward pass, making them suitable for real-time applications on resource-constrained devices, such as those used for autonomous drone deployments [8, 18].

4.1. Pretrained YOLO

Our initial evaluations were conducted using the pretrained models without additional fine-tuning to assess their out-of-the-box performance on our dataset. We extracted approximately 500 frames from each of the ten sampling sessions to ensure even coverage across the MMLA dataset for all the locations and species. Note: we combined the two sessions collected at the Ol Pejeta Conservancy into one session for the testing, as all videos captured Plains zebras in the same park area on consecutive days. The evaluation subset consists of 5,073 frames, of which 1,050 from Ol Pejeta Conservancy, 1,923 from Mpala Research Center, and 2,100 from The Wilds Conservation Center. See App. 10 for detailed assessment of the baseline models' performance for each location and selected species present at most locations (zebra and giraffe).

4.2. Fine-tuned YOLOv11m

YOLOv11m achieved the best performance on the MMLA dataset, with an 18.6% F1 score and 32.8% mAP50 score, so this model was selected for fine-tuning on the dataset. We implemented a stratified data partitioning strategy to ensure robust model generalization across our multi-location wildlife imagery dataset. We divided the data into training (58%) and test (42%) splits, carefully maintaining proportional representation across all three ecological zones and five target species (reticulated and Masai giraffe, Plains zebra, Grevy's zebra, Persian onager, and African wild dog. We assigned entire video sequences to a single partition to prevent data leakage, ensuring that temporally adjacent frames never appeared in different splits when possible. Furthermore, we regularly sampled non-consecutive frames (every 10) to reduce temporal redundancy while preserving diversity in environmental conditions, animal behaviors, and camera perspectives. This approach maintained statistical consistency across partitions while ensuring independence between the training and evaluation data, with each species and location represented in at least two partitions to facilitate a thorough assessment of the model's cross-habitat performance. Please refer to our GitHub repository [2] for instructions on generating the train/test split. This division of training/test data by session and location ensures that the model is evaluated on unseen environments, individuals, and contexts, making it a robust benchmark for testing generalization across ecological and geographic domains.

We trained the model for 50 epochs (approximately 2 hours) on the Ohio Supercomputer Cluster, using two Tesla V100-PCIE-16GB GPUs. The hyperparameters and optimizers were left as the default for YOLO fine-tuning using scripts provided in the `ultralytics` library [14]. We provide scripts to reproduce our results in our GitHub repository and provide the fine-tuned weights[3]. Additional training details may be found in App. 11.

5. Results

We report each model's IoU, precision, recall, F-1, and mAP50 scores in Tab. 3. Comparing the baseline YOLOv11m performance against the fine-tuned model reveals dramatic improvements across all metrics. Fine-tuning significantly improved model performance across locations and species. At Mpala, F1 increased by 34.3% and mAP50 by 33.8%, while The Wilds showed the most dramatic gains with F1 and mAP50 rising by 90.8% and 83.5%, respectively, indicating strong generalization to this previously underperforming site. These location-specific variations highlight distributional differences between fine-tuning data and habitat features. Species-level improvements were substantial, particularly for African wild dogs (+94.9% F1), onagers (+70.2%), and giraffes (+30.4%). Even zebras, which had the highest baseline performance, improved by 22.0% F1.

The normalized confusion matrix (Fig. 3) for the fine-tuned YOLOv11m model reveals strong species discrimination capabilities with minimal cross-species confusion. Zebras show minimal confusion with other classes due to their distinctive stripe patterns. Giraffes occasionally register false positives with background elements, likely due to their complex body patterns and tall stature blending with vegetation. Onagers demonstrate strong classification accuracy, while African wild dogs exhibit near-perfect classification due to their distinctive coat patterns and body shapes. Background elements occasionally trigger false positives

[2] https://github.com/Imageomics/mmla

[3] https://huggingface.co/imageomics/mmla

Model	IoU (%)	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)
YOLOv5m	20.88	34.04	24.94	27.09	28.79
YOLOv8m	20.23	32.20	24.78	26.61	28.00
YOLOv11m	21.77	35.34	25.83	28.14	29.85
YOLOv11m Finetuned on MMLA	**74.39**	**83.43**	**81.39**	**81.76**	**82.40**

Table 3. Performance of YOLO variants on the MMLA dataset. Fine-tuning YOLOv11m yields the highest performance across all metrics.

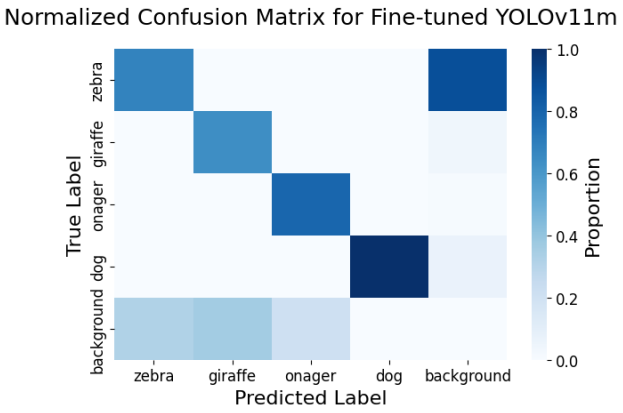

Figure 3. Prediction accuracy for fine-tuned YOLOv11m across the four classes (zebra, giraffe, onager, and dog) and background. Results suggest that the model could be improved to more accurately detect the background and might be overfitted for African wild dog detection.

across species, with slightly higher rates for giraffes and onagers, indicating successful learning of discriminative features while maintaining robustness for autonomous wildlife tracking applications.

6. Limitations

Our work has three main limitations that may be addressed in future work: the dataset itself, model architecture, and computational efficiency. One, our dataset has species-specific location limitations, particularly for African wild dogs, Masai giraffes, and Persian onagers, which were only present at The Wilds. Future work could combine others' datasets (Tab. 2) with MMLA to include these species across multiple environments. Further, we combine the two zebra species and giraffe species into one class. Two, we used a baseline YOLO model without any architectural improvements specific to animal detection from drone imagery. Other works on small-object animal detection and models tailored to drone imagery (WILD-YOLO [26], WildARe-YOLO [2], and Drone-YOLO [37]), could further improve performance on the MMLA dataset. Three, our current approach does not address deployment on resource-constrained drone platforms. Future work could incorporate energy-aware algorithms that explicitly consider edge constraints, such as dynamic pruning for edge inference [28].

7. Conclusion

Low-altitude drone studies offer a valuable middle ground between the detailed imagery of camera traps and the broad coverage of high-altitude surveys, enabling new behavioral and movement studies. Autonomous drones combined with robust detection models can transform wildlife monitoring from a static, pre-planned process to an adaptive system that responds to animal movements in real-time [18]. Deploying such systems involves important tradeoffs for ecologists. Lower flight altitudes provide better resolution but may increase animal disturbance, potentially altering natural behaviors [1, 30]. The balance between human and automated analysis also presents challenges [17]. Manual analysis by trained biologists remains the reliability standard but limits study scale, while automated detection can dramatically increase processing speed but requires domain-specific training to achieve necessary reliability.

The dramatic improvements achieved through fine-tuning suggest that with appropriate dataset development, AI-assisted wildlife monitoring can achieve the reliability needed for scientific applications while significantly increasing wildlife studies' temporal and spatial coverage. By contributing both the MMLA dataset and fine-tuned model weights to the research community, we hope to accelerate the development of robust computer vision solutions for wildlife conservation and establish a foundation for the next generation of intelligent, autonomous wildlife monitoring systems.

8. Acknowledgments

We want to thank Samuel Mutisya and William Njoroge at Ol Pejeta Conservancy for their support in facilitating the WildDrone 2025 Hackathon. This work was supported by WildDroneEU (EU Horizon Europe MSCA grant 101071224), the Imageomics Institute (NSF HDR Award 2118240), and ICICLE (NSF grant OAC-2112606).

References

[1] Saadia Afridi, Lucie Laporte-Devylder, Guy Maalouf, Jenna Kline, Samuel G Penny, Kasper Hlebowicz, Dylan Cawthorne, and Ulrik Pagh Schultz Lundquist. Impact of drone disturbances on wildlife: A review. *Drones*, 9(4):311, 2025.

[2] Sibusiso Reuben Bakana, Yongfei Zhang, and Bhekisipho Twala. Wildare-yolo: A lightweight and efficient wild animal recognition model. *Ecological Informatics*, 80:102541, 2024.

[3] Sara Beery, Dan Morris, and Siyu Yang. Efficient pipeline for camera trap image review. *arXiv preprint arXiv:1907.06772*, 2019.

[4] Marc Besson, Jamie Alison, Kim Bjerge, Thomas E. Gorochowski, Toke T. Høye, Tommaso Jucker, Hjalte M. R. Mann, and Christopher F. Clements. Towards the fully automated monitoring of ecological communities. *Ecology Letters*, 25(12):2753–2775, 2022.

[5] Elizabeth Bondi, Raghav Jain, Palash Aggrawal, Saket Anand, Robert Hannaford, Ashish Kapoor, Jim Piavis, Shital Shah, Lucas Joppa, Bistra Dilkina, and Milind Tambe. Birdsai: A dataset for detection and tracking in aerial thermal infrared videos. In *WACV*, 2020.

[6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018.

[7] CVAT.ai Corporation. Computer Vision Annotation Tool (CVAT), Nov. 2023.

[8] Nguyen Ngoc Dat, Tom Richardson, Matthew Watson, Kilian Meier, Jenna Kline, Sid Reid, Guy Maalouf, Duncan Hine, Majid Mirmehdi, and Tilo Burghardt. Wildlive: Near real-time visual wildlife tracking onboard uavs. *arXiv preprint arXiv:2504.10165*, 2025.

[9] Alexandre Delplanque, Samuel Foucher, Jérôme Théau, Elsa Bussière, Cédric Vermeulen, and Philippe Lejeune. From crowd to herd counting: How to precisely detect and count african mammals using aerial imagery and deep learning? *ISPRS Journal of Photogrammetry and Remote Sensing*, 197:167–180, 2023.

[10] Jasper A.J. Eikelboom, Jelle Wind, Evert van de Ven, Lizzie M. Kenana, Bennet Schroder, Henrik J. de Knegt, Frank van Langevelde, and Herbert H.T. Prins. Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, 10(11):1875–1887, 2019.

[11] Madeline C Hayes, Patrick C Gray, Guillermo Harris, Wade C Sedgwick, Vivon D Crawford, Natalie Chazal, Sarah Crofts, and David W Johnston. Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. *Ornithological Applications*, 123(3):duab022, Aug. 2021.

[12] Wentong He, Ze Luo, Xinyu Tong, Xiaoyi Hu, Can Chen, and Zufei Shu. Long-tailed metrics and object detection in camera trap datasets. *Applied Sciences*, 13(10), 2023.

[13] Jarrod C. Hodgson, Rowan Mott, Shane M. Baylis, Trung T. Pham, Simon Wotherspoon, Adam D. Kilpatrick, Ramesh Raja Segaran, Ian Reid, Aleks Terauds, and Lian Pin Koh. Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, 9(5):1160–1167, 2018.

[14] Glenn Jocher, Jing Qiu, and Ayush Chaurasia. Ultralytics YOLO.

[15] Maksim Kholiavchenko, Jenna Kline, Michelle Ramirez, Sam Stevens, Alec Sheets, Reshma Babu, Namrata Banerji, Elizabeth Campolongo, Matthew Thompson, Nina Van Tiel, et al. Kabr: In-situ dataset for kenyan animal behavior recognition from drone videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 31–40, 2024.

[16] Jenna Kline. Wildwing system deployment data. https://zenodo.org/records/14838101, 2025.

[17] Jenna Kline, Austin O'Quinn, Tanya Berger-Wolf, and Christopher Stewart. Characterizing and modeling ai-driven animal ecology studies at the edge. In *2024 IEEE/ACM Symposium on Edge Computing (SEC)*, pages 220–233. IEEE, 2024.

[18] Jenna Kline, Alison Zhong, Kevyn Irizarry, Charles V Stewart, Christopher Stewart, Daniel I Rubenstein, and Tanya Berger-Wolf. Wildwing: An open-source, autonomous and affordable uas for animal behaviour video monitoring. *Methods in Ecology and Evolution*, 2025.

[19] Jenna Kline, Alison Zhong, Maksim Kholiavchenko, and Elizabeth Campolongo. kabr-tools: Tools for working with data for annotating animal behavior, 2024.

[20] Jenna M Kline, Maksim Kholiavchenko, Otto Brookes, Tanya Berger-Wolf, Charles V Stewart, and Christopher Stewart. Integrating biological data into autonomous remote sensing systems for in situ imageomics: A case study for kenyan animal behavior sensing with unmanned aerial vehicles (uavs). *arXiv preprint arXiv:2407.16864*, 2024.

[21] Benjamin Koger, Adwait Deshpande, Jeffrey T. Kerby, Jacob M. Graving, Blair R. Costelloe, and Iain D. Couzin. Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. *Journal of Animal Ecology*, n/a(n/a), 2022.

[22] Satish Kumar, Bowen Zhang, Chandrakanth Gudavalli, Connor Levenson, Lacey Hughey, Jared A Stabach, Irene Amoke, Gordon Ojwang, Joseph Mukeka, Stephen Mwiu, et al. Wildlifemapper: Aerial image analysis for multi-species detection and identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12594–12604, 2024.

[23] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.

[24] Giacomo May, Emanuele Dalsasso, Benjamin Kellenberger, and Devis Tuia. Polo–point-based, multi-class animal detection. *arXiv preprint arXiv:2410.11741*, 2024.

[25] Chao Mou, Tengfei Liu, Chengcheng Zhu, and Xiaohui Cui. Waid: A large-scale dataset for wildlife detection with drones. *Applied Sciences*, 13(1818):10397, Jan. 2023.

[26] Chao Mou, Chengcheng Zhu, Tengfei Liu, and Xiaohui Cui. A novel efficient wildlife detecting method with lightweight deployment on uavs based on yolov7. *IET Image Processing*, 18(5):1296–1314, 2024.

[27] Hemal Naik, Junran Yang, Dipin Das, Margaret Crofoot, Akanksha Rathore, and Vivek Hari Sridhar. Bucktales: A multi-uav dataset for multi-object tracking and re-identification of wild antelopes. *Advances in Neural Information Processing Systems*, 37:81992–82009, 2024.

[28] Austin O'Quinn, Conor Snedeker, Siyuan Zhang, and Jenna Kline. Environment-aware dynamic pruning for pipelined edge inference. *arXiv preprint arXiv:2503.03070*, 2025.

[29] Edouard Rolland, Kasper Grøntved, Lucie Laporte-Devylder, Jenna Kline, Ulrik Lundquist, and Anders Christensen. Drone swarms for animal monitoring: A method for collecting high-quality multi-perspective data. In T. Richardson, editor, *15^{th} annual International Micro Air Vehicle Conference and Competition*, pages 316–323, Bristol, United Kingdom, Sep 2024. Paper no. IMAV2024-38.

[30] Lukas Schad and Julia Fischer. Opportunities and risks in the use of drones for studying animal behaviour. *Methods in Ecology and Evolution*, 14(8):1864–1872, 2023.

[31] Devis Tuia, Benjamin Kellenberger, Sara Beery, Blair R Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W Mathis, Frank Van Langevelde, Tilo Burghardt, et al. Perspectives in machine learning for wildlife conservation. *Nature communications*, 13(1):792, 2022.

[32] Vidit Vidit, Martin Engilberge, and Mathieu Salzmann. Clip the gap: A single domain generalization approach for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3219–3229, 2023.

[33] Juliana Vélez, William McShea, Hila Shamon, Paula J. Castiblanco-Camacho, Michael A. Tabak, Carl Chalmers, Paul Fergus, and John Fieberg. An evaluation of platforms for processing camera-trap data using artificial intelligence. *Methods in Ecology and Evolution*, 14(2):459–477, 2023.

[34] Dongliang Wang, Quanqin Shao, and Huanyin Yue. Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (uass): A review. *Remote Sensing*, 11(1111):1308, Jan. 2019.

[35] Ben Weinstein, Lindsey Garner, Vienna R. Saccomanno, Ashley Steinkraus, Andrew Ortega, Kristen Brush, Glenda Yenni, Ann E. McKellar, Rowan Converse, Christopher D. Lipitt, Alex Wegmann, Nick D. Holmes, Alice J. Edney, Tom Hart, Mark J. Jessopp, Rohan Clarke, Dominik Markowski, Henry Senyondo, Ryan Dotson, Ethan P. White, Peter Frederick, and S.K Morgan Ernest. A global model of bird detection in high resolution airborne images using computer vision, June 2021.

[36] WildDrone EU Consortium. Wilddrone eu project, 2025. Accessed: 03-Apr-2025.

[37] Zhengxin Zhang. Drone-yolo: An efficient neural network method for target detection in drone images. *Drones*, 7(8):526, 2023.

[38] Imran Zualkernan, Salam Dhou, Jacky Judas, Ali Reza Sajun, Brylle Ryan Gomez, and Lana Alhaj Hussain. An iot system using deep learning to classify camera trap images on the edge. *Computers*, 11(11):13, Jan. 2022.

9. MMLA Dataset Details

Location	Session	Date	Total Frames	Species
Ol Pejeta Conservancy, Laikipia, Kenya	1	1/31/25	16,726	Plains zebra
	2	2/1/25	12,542	Plains zebra
		Subtotal:	*29,268*	
Mpala Research Center, Laikipia, Kenya	1	1/12/23	16,891	Reticulated giraffe
	2	1/17/23	11,165	Plains zebra
	3	1/18/23	17,940	Grevy's zebra
	4	1/20/23	33,960	Grevy's zebra
	5	1/21/23	24,106	Giraffe, Plains and Grevy's zebras
		Subtotal:	*104,062*	
The Wilds Conservation Center, Ohio, USA	1	6/14/24	13,749	African wild dog
	2	7/31/24	3,436	Masai and reticulated giraffe
	3	4/18/24	4,053	Persian onager
	4	7/31/24	506	Grevy's zebra
		Subtotal:	*21,744*	
		Total:	**155,074**	

Table 4. Frame-level summary of the MMLA dataset by location, session, date, and species.

10. Baseline YOLO Evaluation

Our evaluation of baseline YOLO models (v5m, v8m, and v11m) reveals significant performance challenges when applied to low-altitude drone imagery across different environments (Table 5 and Table 6). Despite achieving 49-51% mAP50 on the standard MSCOCO dataset, these models show substantial degradation when applied to our MMLA dataset, with overall F1 scores dropping dramatically to just 16-19Performance varied significantly across locations and species. For instance, at the Mpala location, models achieved relatively strong mAP50 scores (64-66%) but struggled considerably at The Wilds (19-25%) and Ol Pejeta Conservancy (26-30%). This location-based performance disparity highlights the domain shift problem in cross-environment generalization. Species-specific performance also showed notable patterns. Giraffe detection generally outperformed zebra detection across all models and locations, with precision scores for giraffes reaching 76-80

The newest model, YOLOv11m, achieved the best overall F1 score (18.6%) and mAP50 (32.8%), demonstrating incremental improvements over earlier versions. However, even this best-performing model exhibits a nearly 20-point drop in mAP50 from MSCOCO to our wildlife dataset, underscoring the significant gap between standard object detection benchmarks and the challenges of real-world low-altitude wildlife detection. This performance deficit clearly demonstrates the need for specialized models and cross-environment training approaches for autonomous wildlife drone applications.

Model	Overall F1				Giraffe F1				Zebra F1			
	All	Mpala	OPC	Wilds	All	Mpala	OPC	Wilds	All	Mpala	OPC	Wilds
YOLOv5m	16.31	**22.60**	**16.34**	12.02	49.33	52.62	–	39.98	10.73	7.10	**18.42**	4.03
YOLOv8m	17.01	20.20	16.08	15.33	46.05	49.29	–	38.70	10.16	5.51	17.97	5.08
YOLO11m	**18.55**	21.86	15.52	**17.59**	**50.89**	**54.59**	–	**41.88**	**16.08**	**4.60**	16.36	**42.55**

Table 5. Base YOLO models F1 Scores (%) across species and locations in MMLA.

Model	MSCOCO (mAP50)	Overall mAP50				Giraffe Precision				Zebra Precision			
		All	Mpala	OPC	Wilds	All	Mpala	OPC	Wilds	All	Mpala	OPC	Wilds
YOLOv5m	49.0	31.58	64.88	**29.78**	22.17	**76.05**	77.91	–	**70.64**	31.17	35.44	**30.37**	22.40
YOLOv8m	50.2	31.77	63.88	26.33	**24.99**	66.85	77.90	–	47.47	29.67	**36.74**	27.05	37.30
YOLOv11m	**51.2**	**32.83**	**65.71**	27.06	19.51	71.53	**80.28**	–	53.30	**34.18**	28.84	21.59	**68.24**

Table 6. Precision Metrics (%) across MMLA Species and Locations. We report the mAP50 achieved by each YOLO variant on the MSCOCO dataset [23] as a reference baseline for object detection performance [14].

11. YOLO Finetune Training Details

11.1. Data Splitting Strategy for Model Fine-tuning and Evaluation

Please refer to our GitHub repository [4] for instructions on generating the train/test split.

Location	Training Set	Test Set
Mpala	Multiple full sessions for reticulated giraffes, Plains zebras, and Grevy's zebras, including mixed-species scenes	Separate zebra and mixed-species sessions (includes reticulated giraffes, Plains and Grevy's zebras) not used during training
Ol Pejeta	Full sessions of Plains zebras	Separate zebra and mixed-species sessions not used during training
The Wilds	70% of sessions for African wild dogs, giraffes (Masai and reticulated), and Persian onagers	Remaining 30% of sessions, including additional Grevy's zebra sessions used exclusively for testing

Table 7. MMLA Dataset Train/Test Split by Location and Species

11.2. Training and Optimization Details

Training was conducted for 50 epochs on two NVIDIA Tesla V100 GPUs (16 GB each) using Ultralytics YOLO v8.3.117, Python 3.10.16, and PyTorch 2.6.0 with CUDA 12.4. The total training time was 2 hour, 11 minutes. We used the Ohio Supercomputer to run our experiments. We used the YOLO framework's `auto` optimizer setting, which defaults to **stochastic gradient descent (SGD)** with momentum. The initial learning rate was set to 0.01 and decayed using a cosine schedule with a final learning rate fraction (`lrf`) of 0.01. Momentum was set to 0.937 and weight decay to 0.0005. A warmup phase of 3 epochs was applied, with a warmup momentum of 0.8 and a warmup bias learning rate of 0.0.

Loss weighting terms were: box loss = 7.5, class loss = 0.5, and Distribution Focal Loss (DFL) = 1.5. Mixed precision training was enabled via automatic mixed precision (AMP), and deterministic behavior was enforced using `seed = 0` and `deterministic = True`.

We applied standard YOLO augmentations, including mosaic augmentation (probability = 1.0), random horizontal flipping (0.5), HSV color jittering (hue = 0.015, saturation = 0.7, value = 0.4), random translation (0.1), and scaling (0.5). We used **RandAugment** for policy-based image augmentation and random erasing with a probability of 0.4. MixUp and CopyPaste were disabled.

Training resumed from a pretrained checkpoint and saved model weights and training plots for all runs.

[4] https://github.com/Imageomics/mmla

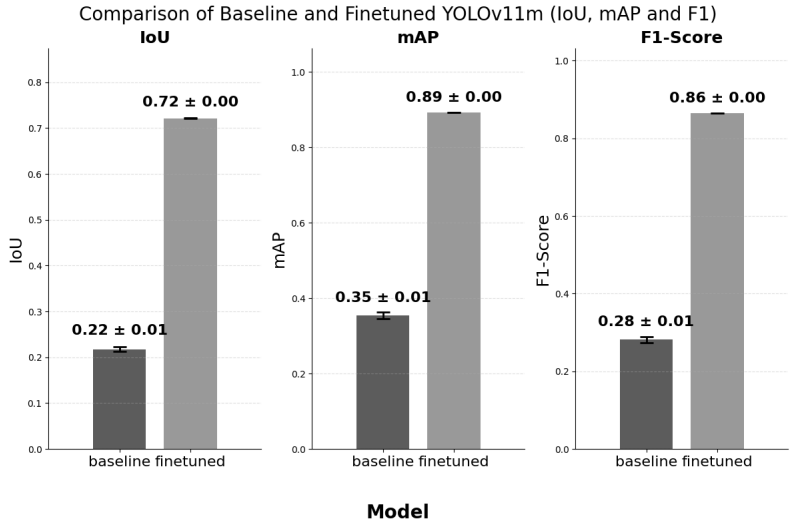

Figure 4. Bootstrapping performance of baseline and fine-tuned YOLOv11m models with confidence intervals for IoU, mAP, and F1. The finetuned model demonstrated statistically significant performance improvement across all metrics.

12. YOLO11m Finetuned Results

12.1. Bootstrapping

To ensure statistical significance, we performed bootstrapping with replacement on the test dataset for 1000 iterations on both the baseline and fine-tuned YOLOv11m model. We report the results with confidence intervals in Figure 4. The fine-tuned model shows statically significant performance gains across IoU (23% vs 64%) mAP (37% vs 89%), and F1 (29% vs 84%).

13. YOLOv11m Baseline vs Fine-tuned Model Comparison Performance by Location and Species

Table 8. Baseline Performance by Location

Location	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)
Mpala	50.5	47.4	48.2	31.2
Ol Pejeta	75.0	49.0	58.4	51.7
The Wilds	8.7	3.5	4.5	5.1
Overall	44.8	33.3	37.1	29.3

Table 9. Fine-Tuned Performance by Location

Location	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)
Mpala	83.0	82.8	82.6	64.9
Ol Pejeta	55.9	44.4	49.2	65.8
The Wilds	95.0	96.6	95.4	88.6
Overall	77.9	74.6	75.7	73.1

Table 10. Baseline Performance by Species

Species	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)
Dog	6.4	2.1	3.0	3.8
Giraffe	39.4	35.6	36.3	23.2
Onager	26.3	11.7	14.1	15.9
Zebra	57.0	47.7	50.7	36.5
Overall	32.3	24.3	26.0	19.9

Table 11. Fine-Tuned Performance by Species

Species	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)
Dog	96.6	99.6	97.9	90.8
Giraffe	100.0	50.0	66.7	71.2
Onager	92.9	81.5	84.3	79.2
Zebra	74.8	71.7	72.8	64.1
Overall	91.1	75.7	80.4	76.3

14. Performance Metrics for Wildlife Monitoring Approaches

For autonomous low-altitude drone systems like WildWing [18], precision metrics are crucial, more so than in traditional wildlife monitoring approaches. In these systems, false positives significantly impair performance by causing the drone to adjust its flight path or focus on non-animal objects. When an autonomous drone relies on detection algorithms to identify and track animals in real-time, each false positive potentially redirects valuable flight time and battery capacity away from actual wildlife subjects. For example, a drone following what it incorrectly identified as an animal could lose track of the actual subjects of interest, resulting in failed data collection missions and wasted resources. Additionally, Intersection over Union (IoU) metrics are particularly important for autonomous tracking, as precise bounding box predictions directly influence flight control decisions. A detection model with high IoU scores provides more accurate animal positioning data, allowing the drone to maintain optimal following distance and camera angles. Studies have shown that IoU values of at least 0.5 (mAP50) are necessary for stable autonomous tracking [20]. Autonomous drones also require faster inference times than post-hoc batch processing systems. While camera trap and high-altitude drone footage can be processed hours or days after collection, autonomous tracking requires near real-time analysis (typically ¡100ms per frame) to make timely flight control decisions. This creates a challenging optimization problem where precision, IoU accuracy, and inference speed must all be balanced.

The following metrics are critical for evaluating detection models for autonomous wildlife drones:

- **Precision**: The proportion of positive identifications that were actually correct. High precision minimizes wasteful flight path adjustments.
- **mAP50**: Mean Average Precision with IoU threshold of 0.5, measuring how accurately the model localizes animals within frames.
- **Inference Time**: Processing time per frame, which must be sufficiently fast for real-time flight decisions.
- **Cross-Environment Performance**: How well precision and mAP50 metrics transfer when the model is deployed in new environments not seen during training.

Our baseline model evaluations in this study prioritize these metrics, particularly focusing on precision and cross-environment performance as the most critical for autonomous wildlife tracking applications.