DeepForestVisionV2: Ecology-Driven Taxonomy Expansion for Camera-Trap Monitoring in African Tropical Forests

Hugo Magaldi[1,2], Théau d'Audiffret[1,2], Etienne François Akomo-Okoue[3,4], Bala Amarasekaran[5], Naomi Anderson[5], Claire Auger[1,2], Noémie Cappelle[6], Daniel Cornélis[7], Raphaël Cornette[8], Tobias Deschner[9,10], Gabriel Dubus[1,2], Davy Fonteyn[7], Rosa M. Garriga[5], Jennifer Hatlauf[11], Innocent Kasekendi[2], Raymond Katumba[2], Aram Kazandjian[5], Alfred Ngomanda[3], Stéphan Ntie[12,13], Simone Pika[10], Xavier Rufray[6], Harold Rugonge[2], John Justice Tibesigwa[14], Peter van Lunteren[15], Hadrien Vanthomme[7], Joeri A. Zwerts[16], and Sabrina Krief[1,2]

[1] UMR7206 Eco-Anthropologie, MNHN, Paris, France; One Forest Vision initiative
[2] Sebitoli Chimpanzee Project, Sebitoli, Kibale National Park, Uganda
[3] Centre National de la Recherche Scientifique et Technologique, Libreville, Gabon
[4] Institut de Recherche en Ecologie Tropicale, Libreville, Gabon
[5] Tacugama Chimpanzee Sanctuary, Freetown, Sierra Leone
[6] International Department, Biotope, Montpellier, France
[7] CIRAD, UPR Forêts et Sociétés, Montpellier, France
[8] Institut de Systématique, Evolution, Biodiversité (ISYEB), UMR7205 CNRS, MNHN, SU, EPHE-PSL, UA, Paris, France
[9] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[10] Comparative BioCognition, Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany
[11] BOKU University, Institute of Wildlife Biology and Game Management, Vienna, Austria
[12] Département de Biologie, Université des Sciences et Techniques de Masuku, Franceville, Gabon
[13] Agence Nationale des Parcs Nationaux du Gabon, Gabon
[14] Uganda Wildlife Authority, Kampala, Uganda
[15] Addax Data Science, Utrecht, The Netherlands
[16] Wildlife Ecology and Nature Restoration, Utrecht University, The Netherlands

Abstract. Camera-trap monitoring in African tropical forests increasingly extends beyond closed-canopy interiors to riverbanks, clearings, and park edges. Among available open tools for African forest camera-trap classification, DeepForestVision is the only one providing a matched offline workflow for both photographs and videos, and previous work showed that it outperformed other available baselines on a comparable video benchmark. However, it was originally designed for closed-canopy, ground-level forest interiors and uses a 35-class prediction space that becomes too coarse when deployments encounter arboreal primates, birds, semi-aquatic taxa, or human-associated confounders such as livestock. We present DeepForestVisionV2, an ecology-driven expansion from 35 to

64 prediction classes (61 animal classes plus *human*, *vehicle*, and *blank*) designed to address three recurrent deployment gradients: vertical stratification, scene openness, and anthropogenic interfaces. DeepForestVisionV2 retains the same offline workflow and is trained on 1,535,010 photographs and 243,354 videos from multi-country African tropical-forest projects. Evaluation combines a cross-country cropped-photo validation set, used to assess robustness across sites and camera-trap settings, with three held-out Uganda video benchmarks spanning the targeted gradients. On the validation set, DeepForestVisionV2 reaches 0.86 accuracy, 0.82 macro-F1, and 0.81 balanced accuracy. On the deployment benchmarks, it preserves or improves baseline accuracy despite its harder classification task, while increasing the number of identified taxa from 22 to 29 in forest-interior videos and from 4 to 9 at riverbanks. In the park-edge use case, it raises accuracy from 0.62 to 0.86 and reduces false alarms from 11 to 0. These results show that DeepForestVisionV2 materially improves field utility while preserving robustness across sites, habitats, and camera-trap settings.

Keywords: camera-traps · tropical forests · species identification · biodiversity monitoring · AI for conservation

1 Introduction

Camera-traps are now a standard instrument for wildlife monitoring because they support species inventories, activity analyses, and disturbance assessment at spatial and temporal scales that are difficult to achieve by direct observation [2]. As the volume of collected data grows, automated species identification becomes essential for practical workflows [18]. In practice, the ecological value of identification tools depends on more than their accuracy: a system must be deployable under field constraints, remain robust across the habitats where camera-traps are actually placed, and provide the taxonomic resolution required by downstream ecological questions [17,21].

For African tropical forests, several open tools are available, including Mbaza [21], Zamba [6], SpeciesNet [7], and DeepForestVision [10]. These systems differ in taxonomy, but also in deployment constraints: Mbaza and SpeciesNet process photographs only, whereas Zamba requires online or code-based local execution. DeepForestVision is the only tool for non-programmers that provides offline processing of both photographs and videos through AddaxAI, a no-code interface for local model deployment [9,10]. Previous work already compared DeepForestVision with Mbaza, Zamba, and SpeciesNet on a video benchmark from Kibale National Park, Uganda, by adding the engineering needed to run photo-native methods on videos, and found that DeepForestVision outperformed all three in accuracy (Mbaza by 45.0%, Zamba by 13.1% and SpeciesNet by 37.7%) [10]. Accordingly, this paper does not aim to establish another cross-tool benchmark. Instead, it shows that an ecology-driven expansion of the DeepForestVision label space improves field utility within the same offline workflow for photographs and videos.

DeepForestVision was designed around closed-canopy ground-level deployments. However, ongoing projects increasingly deploy camera-traps along ecological gradients, understood here as recurrent changes in deployment conditions that shift the observed community composition: (i) a *vertical gradient*, where arboreal primates become visible near clearings, mineral licks, or canopy openings [12,14]; (ii) an *openness gradient*, where riverbanks and open views increase detections of birds [13] and expose cameras to riparian and semi-aquatic taxa; and (iii) an *anthropogenic gradient*, where park-edge cameras must distinguish wildlife from livestock [11,19,20]. In these regimes, DeepForestVision's coarse labels such as *bird*, *monkey*, or *civet/genet* are often insufficient for ecological interpretation or management action. DeepForestVisionV2 addresses this mismatch by redesigning the label space to better capture taxa that become relevant along these deployment gradients, while preserving the robustness of a field-ready workflow.

Contributions. This paper makes three contributions:

1. an ecology-driven expansion from 35 to 64 prediction classes (61 animal classes, plus *human*, *vehicle*, and *blank*) aligned with three ecologically motivated deployment gradients;
2. a geographically diverse, Africa-wide cropped-photo validation benchmark used to assess robustness of the expanded classifier across sites and camera-trap settings;
3. a deployment-centric evaluation on three held-out Uganda video benchmarks, comparing DeepForestVisionV2 with the original DeepForestVision under a matched offline workflow.

2 Material and methods

2.1 Study system and datasets

We assembled a multi-country corpus of labeled camera-trap data from African tropical forests intended to support geographic robustness, by combining partner projects and publicly available resources (Table 1). In total, the training and validation pool contains 1,535,010 photographs and 243,354 videos acquired across different sites, habitats, camera-trap protocols, and imaging conditions. The data were collected non-invasively and required no contact with animals. All necessary authorizations and permits were obtained prior to data collection and are available upon request.

Evaluation is split into two complementary parts. First, the data sources listed in Table 1 were transformed into cropped animal detections using MegaDetector v5 [3] and split into training and validation sets using a random partition at camera level prior to crop extraction, so as to reduce spurious correlations between the two sets. To limit long-tail imbalance in evaluation while preserving rare taxa for training, each class contributes at most 1,000 validation crops and no more than 10% of the available crops for that class. This cropped-detection

Table 1: Training and validation data sources.

Dataset	**Region**	**Sites**	**Type**	**Items**
Sebitoli Chimpanzee Project	Uganda	1	Video	165,942
CIRAD-SWM [4]	Gabon	1	Video	24,430
Biotope Baseline Biodiversité	Gabon	1	Video	17,046
Tacugama Chimpanzee Sanctuary	Sierra Leone	1	Video	15,194
Ozouga Chimpanzee Project	Gabon	1	Video	14,703
J. A. Zwerts [22]	Congo, Gabon	14	Photo	643,660
Wildlife Insights [1]	Global	10	Photo	617,882
LLA BC	Global	3	Photo	267,778
Pan African Programme	Africa	29	Video	6,039
InterMuc Project	Global	1	Photo	5,094
Golden Jackal Project [8,16]	Global	1	Photo	596
Total photographs				**1,535,010**
Total videos				**243,354**

Table 2: Ecological gradients motivating the taxonomy redesign and the corresponding held-out deployment benchmarks.

Gradient	**Ecological motivation**	**Benchmark**
Vertical stratification	Arboreal primates become visible near clearings, mineral licks, and canopy openings.	Forest interior, 17,899 videos (Feb.-Jul. 2025), 29 animal classes including 9 primates.
Openness	Riverbanks and open views expose birds and semi-aquatic taxa that rarely appear in closed-canopy interiors.	Riverbanks, 1,016 videos (Oct.-Nov. 2025), 9 animal classes including 6 birds or semi-aquatic taxa.
Anthropogenic interface	Park-edge monitoring must distinguish genuine wildlife intrusions from domestic passages.	Park-edge, 146 videos (Jul.-Sep. 2025), 3 animal classes including one domestic animal.

validation set is designed to measure the robustness of the classification step across countries, sites, habitats, and camera-trap settings and covers the full 61-animal target label space.

Second, to assess operational value in realistic field settings, we evaluate the whole DeepForestVisionV2 pipeline on three video benchmarks corresponding to the three targeted ecological gradients (Table 2, Figure 1). These benchmarks were collected and annotated by the Sebitoli Chimpanzee Project in Kibale National Park, Uganda, and were not included in the training or validation pool.

2.2 Taxonomy of DeepForestVisionV2

DeepForestVisionV2 outputs one label among 64 prediction classes: 61 animal classes spanning mammals and birds, plus *human*, *vehicle*, and *blank*. To sup-

(a) Riverbank view with a Grey crowned crane (*Balearica regulorum*).

(b) Chimpanzees (*Pan troglodytes*) crop-feeding in maize fields.

Fig. 1: Two deployment regimes targeted by the taxonomy expansion: openness on riverbanks and anthropogenic interfaces at park edges. © SCP

Table 3: DeepForestVisionV2 animal classes not represented as such in DeepForestVision, grouped by the main ecological contexts. Boldface indicates newly introduced classes, non boldface refinements of previously coarser classes.

Ecological grouping	**DeepForestVisionV2 classes**
Primates	agile mangabey; grey-cheeked mangabey; moustached monkey; red-capped mangabey; red-tailed monkey; spot-nosed monkey; red colobus; galago; potto
Birds	bird of prey; black guineafowl; casqued hornbill; crane; dove; duck; francolin; hornbill; nkulengu rail; passeriform; pelecaniform; rockfowl; turaco
Semi-riparian or aquatic taxa	**hippopotamus**; **otter**; **sitatunga**
Other mammals	black-legged mongoose; civet; genet; warthog; **cattle**; **dog**; **goat**

port finer-grained downstream ecological analyses, this redesign increases DeepForestVision's native 35-class resolution for arboreal primates, birds, riparian taxa, and domestic species (Table 3). The full taxonomy, including scientific names, taxonomic levels, IUCN status, training image counts, and mappings to the original DeepForestVision classes, is reported in Appendix A.

2.3 Photo-and-video pipeline and training

Pipeline. DeepForestVisionV2 retains the original DeepForestVision architecture and produces one label per camera-trap item (defined as one photograph or one video). Videos are important because they are a standard output of many camera-trap deployments and field teams typically review full clips rather than isolated frames.

For each photograph, or for each sampled frame extracted from a video, MegaDetector v5 [3] is first applied to detect *animal*, *human*, and *vehicle* instances. Human and vehicle detections are assigned directly from detector outputs, while animal detections are cropped and sent to a DINOv3 ViT-B/16 classifier [15]. Class scores are computed on retained detections across sampled frames and then averaged to produce one item-level prediction. If no detection remains after thresholding, the item is labeled *blank*.

Training. The detector stage was used without additional training. For classification, we fine-tuned a pretrained DINOv3 ViT-B/16 on MegaDetector crops (threshold 0.5) in PyTorch. Training augmentation comprised random resized crop to 224×224 (scale $[0.9, 1.0]$, aspect ratio $[0.9, 1.1]$), random horizontal flip ($p = 0.5$), color jitter ($p = 0.2$; brightness ± 0.15, contrast ± 0.20, saturation ± 0.10), random grayscale ($p = 0.05$), random autocontrast ($p = 0.2$), random sharpness adjustment (factor 1.5, $p = 0.2$), and Gaussian blur (kernel size 9, $\sigma \in [0.1, 1.0]$, $p = 0.05$). Images were converted to tensors and normalized with ImageNet statistics [5]; validation preprocessing used resize to 256, center crop to 224×224, and the same normalization. Optimization used AdamW with separate learning rates for backbone (10^{-5}) and head (10^{-4}), weight decay 0.05, label smoothing 0.05, cosine scheduling with 5% warm-up, automatic mixed precision, and gradient clipping at 1.0. Training ran for 20 epochs, with checkpoint selection by validation balanced accuracy.

2.4 Baseline, comparison protocol, and metrics

Baseline. The original DeepForestVision is the baseline on the deployment benchmarks because it is the only available open tool sharing with DeepForestVisionV2 a matched offline workflow for processing both photographs and videos.

Protocol. We evaluate the original DeepForestVision in its native 35-class label space using the mapping given in Appendix A, which defines an easier classification task than the 64-class label space used to evaluate DeepForestVisionV2. The six DeepForestVisionV2 classes with no corresponding DeepForestVision class -*cattle*, *dog*, *goat*, *hippopotamus*, *otter*, and *sitatunga* (Table 7)- are retained, because the goal is to quantify the ecological utility gained from the expanded DeepForestVisionV2 label space.

Metrics. We report accuracy, balanced accuracy, and macro-averaged F1-score as global metrics, along with per-class precision, recall, F1-score, and support. Since the DeepForestVisionV2 class distribution is strongly long-tailed (Appendix A), balanced accuracy and macro-F1 are given to reflect performance beyond the most frequent taxa. Blank and human items are excluded so that the reported scores reflect differences in animal classification rather than shared detector outputs, since both labels are assigned by MegaDetector and are therefore identical for DeepForestVision and DeepForestVisionV2.

Ecological utility. To quantify ecological gain, we also report the number and identity of taxa surfaced by DeepForestVisionV2 but not represented as such in DeepForestVision and benchmark performance across the three motivating ecological gradients. For the park-edge benchmark containing primates and goats, we design a simple alarm task in which the alarm group contains all primates, and goat passages are nuisance confounders because they trigger unnecessary interventions if the system cannot separate them from wildlife. We report true alarms, false alarms, and missed alarms for primate intrusions, as well as precision and recall computed from the alarm counts.

Calibration. Confidence calibration is assessed with a reliability diagram and the Expected Calibration Error (ECE), which measures the average absolute difference between predicted confidence and empirical accuracy across confidence bins. Since model scores may be used for review prioritization, filtering, or semi-automated alerting in field deployment, calibration is operationally important.

3 Results

3.1 Validation set.

On the cross-country validation set, used here to assess robustness of the expanded classifier over the full animal label space, DeepForestVisionV2 reaches 0.86 accuracy, 0.81 balanced accuracy, and 0.82 macro-F1. All per-class validation supports and metrics are reported in Appendix B.

3.2 Deployment benchmarks and surfaced taxa

Global performance metrics on the deployment benchmarks are summarized in Table 4. On the forest-interior benchmark, DeepForestVisionV2 matches the original DeepForestVision in accuracy (0.89 for both models) while increasing the number of surfaced taxa from 22 to 29. On riverbanks, DeepForestVisionV2 remains close in accuracy (0.72 vs. 0.73) and raises the number of surfaced taxa from 4 to 9. On the park-edge benchmark, DeepForestVisionV2 improves accuracy from 0.62 to 0.86.

Table 5 details the additional taxa recovered by the expanded label space on the video benchmarks. On the forest-interior benchmark, these added detections include several arboreal primates and bird groups that were previously merged into coarser labels. On riverbanks, the additional detections are concentrated in bird and semi-aquatic taxa associated with open riparian views.

3.3 Operational use case: filtering crop-raiding alarms

On the park-edge benchmark, DeepForestVisionV2 removes all false alarms while keeping high recall, and increases missed alarms from 3 to 5 (Table 6). Overall, it raises benchmark accuracy from 0.62 to 0.86 (Table 4).

Table 4: Deployment-benchmark metrics for DeepForestVisionV2 (64 classes) and DeepForestVision (35 classes).

Benchmark	Model	Acc.	Bal. acc.	Macro-F1
Forest interior	DeepForestVisionV2	**0.89**	**0.77**	**0.73**
	DeepForestVision	**0.89**	0.70	0.70
Riverbanks	DeepForestVisionV2	0.72	**0.71**	**0.59**
	DeepForestVision	**0.73**	0.63	0.55
Park-edge	DeepForestVisionV2	**0.86**	**0.86**	**0.92**
	DeepForestVision	0.62	0.62	0.62

Table 5: Taxa surfaced by DeepForestVisionV2 that are absent as such from the DeepForestVision label space, with weighted F1 per ecological group.

Group	Taxa	Weighted F1	Support
Interior - primates	galago, grey-cheeked mangabey, red colobus, red-tailed monkey	0.78	265
Interior - birds	bird of prey, dove, francolin	0.79	100
Interior - other	civet, genet, hippopotamus, dog	0.78	105
Riverbanks - birds	bird of prey, crane, duck, pelecaniform	0.53	20
Riverbanks - other	hippopotamus, otter	0.44	15
Park-edge - domestic	goat	0.86	50

3.4 Calibration

On pooled video-benchmark predictions, DeepForestVisionV2 has an Expected Calibration Error of 0.17. The reliability diagram in Figure 2 shows that, over most confidence bins, empirical accuracy is higher than predicted confidence, indicating under-confidence rather than over-confidence. The apparent drop in the highest-confidence bin is unstable because it contains only six predictions.

4 Discussion

On the geographically diverse validation set, DeepForestVisionV2 remains strong in the full animal target space, which indicates that the expanded taxonomy classification is robust across sites, habitats and camera-trap settings. However, classes with low training support are more likely to exhibit low F1 scores (worst performers are *honey badger* 0.42, *potto* 0.49 and *red-tailed monkey* 0.58) and would benefit from further training on additional data. Performance on the video benchmarks is lower than cropped-image classification performance on the validation set, showing that validation metrics are informative but optimistic with respect to end-to-end deployment. This difference is expected in part because the video pipeline also depends on detection quality, and therefore includes errors

Table 6: Alarm-filtering performance on the park-edge benchmark. Alarm precision and recall are derived from the true, false, and missed alarm counts.

Model	True alarms	False alarms	Missed alarms	Precision	Recall
DeepForestVisionV2	91	0	5	1.00	0.95
DeepForestVision	93	11	3	0.89	0.97

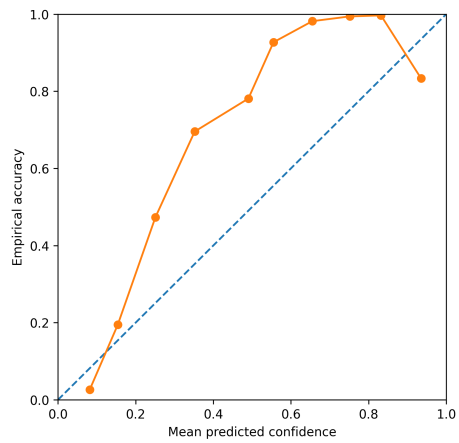

Fig. 2: Reliability diagram for pooled benchmark predictions: empirical accuracy by confidence bin (orange) versus perfect calibration (dashed blue).

from missed or incorrect detections in addition to classification errors. It may be further amplified by limited validation support for some taxa, with nine classes represented by fewer than 100 validation images, and by calibration choices made on the validation set. Validation performance is therefore best interpreted as a controlled comparison signal rather than a direct estimate of global field performance.

Our video benchmarks ask an operational workflow question under field conditions in Uganda. Across the three deployment gradients, DeepForestVisionV2 preserves or increases the accuracy of DeepForestVision despite its harder classification task, and provides different kinds of ecological gains. In the forest interior, DeepForestVisionV2 surfaces additional arboreal primates and birds. On riverbanks, the benefit is primarily taxonomic coverage in an open-scene regime where the original taxonomy was not designed to operate. However, this benchmark is relatively small and should be expanded to additional cameras, seasons, and riparian settings. At the park edge, the benefit is directly operational: domestic classes suppress nuisance alarms and make the system more suitable for future edge-AI alerting pipelines in which review capacity is limited and false positives are costly [11,20]. DeepForestVisionV2 removes all 11 false alarms triggered by DeepForestVision, but adds 2 more missed alarms. This trade-off suggests that calibration and threshold tuning are practical steps needed before fully automated alerting.

The reliability diagram and Expected Calibration Error show that DeepForestVisionV2 is underconfident on the deployment benchmarks, and suggest that confidence scores should not be interpreted as calibrated probabilities. Post-hoc calibration, such as temperature scaling on held-out data, would be advisable before using confidence thresholds for automated triage or alarming. Finally, the pipeline as we present it here emits a single label per clip, which is acceptable for most present deployments but will miss the occasional multi-species event. This can be changed in the AddaxAI interface to output all predictions above a set confidence threshold.

5 Conclusion

DeepForestVisionV2 expands DeepForestVision from a 35-class forest-interior camera-trap model to a 64-class system better aligned with field realities in African tropical forests. This expansion is motivated by three recurrent ecological gradients - vertical stratification, scene openness, and anthropogenic interfaces - and evaluated on benchmarks matched to each of them. On a cross-country validation set, DeepForestVisionV2 reaches 0.86 accuracy, 0.81 balanced accuracy, and 0.82 macro-F1 over the full animal target space. On held-out video benchmarks, it preserves or improves performance despite the harder classification task, while surfacing more taxa in forest-interior and riverbank settings and providing a direct operational benefit for park-edge alarm filtering. Together, these results indicate that ecology-driven taxonomy expansion increases the practical value of a field-ready offline camera-trap workflow and makes it more informative for conservation use across contrasting habitats.

Data availability. Camera-trap imagery cannot be released publicly because of data-sharing agreements and permitting restrictions. Derived, non-identifying outputs such as label mappings and evaluation scripts are available from the authors on reasonable request. DeepForestVisionV2 is openly available through the offline no-code interface AddaxAI `https://addaxdatascience.com/addaxai`. Model weights and inference code are provided through the OneForestVision initiative GitHub repository: `https://github.com/MNHN-OFVI/DeepForestVisionV2`.

Acknowledgements. The authors thank the Fondation pour la Nature et l'Homme, Fondation Prince Albert 2 de Monaco, Fonds Français pour l'Environnement Mondial, and the One Forest Vision initiative, funded by the French Ministry of Higher Education and Scientific Research and the Ministry of Europe and Foreign Affairs, for supporting this work. Tacugama Chimpanzee Sanctuary was supported by USFS, USFWS, and USAID, with the collaboration of the National Protected Area Authority of Sierra Leone. In Gabon, the authors thank ANPN, CENAREST, PWG-CEB, Station d'Etudes des Gorilles et Chimpanzés, and Lopé National Park for authorizations and access to the study sites, and acknowledge support from the European Union through the Sustainable Wildlife Management Programme, the Wild Chimpanzee Foundation, the Sievert Foundation, the Commonwealth Scholarship Commission UK, the University of Oxford's Hertford College Mortimer May Senior Scholarship in Geography, D. Lehmann, J. Edzang Ndong, E. Dimoto, and the Zooniverse team and citizen scientists.

References

1. Ahumada, J.A., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T.G., Palmer, J., Schuttler, S., Zhao, J.Y., Jetz, W., et al.: Wildlife insights: A platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. Environmental Conservation **47**(1), 1–6 (2020). https://doi.org/10.1017/S0376892919000298
2. Ahumada, J.A., Hurtado, J., Lizcano, D.: Monitoring the status and trends of tropical forest terrestrial vertebrate communities from camera trap data: A tool for conservation. PLOS ONE **8**(9), 1–10 (09 2013). https://doi.org/10.1371/journal.pone.0073707, https://doi.org/10.1371/journal.pone.0073707
3. Beery, S., Morris, D., Yang, S.: Efficient pipeline for camera trap image review. arXiv (2019)
4. Cornelis, D.: Camera trap survey data of the project "Sustainable Wildlife Management" in Gabon (Mulundu Department). CIRAD Dataverse (2023). https://doi.org/10.18167/DVN1/TNBXCN
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
6. Dorne, E., et al.: Zamba: Computer vision for wildlife conservation. Proceedings of the Python in Science Conference (SciPy) (2025), accessed 2026-01-04
7. Gadot, T., Istrate, S., Kim, H., Morris, D., Beery, S., Birch, T., Ahumada, J.: To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. IET Computer Vision **18**(8), 1193–1208 (2024). https://doi.org/https://doi.org/10.1049/cvi2.12318, https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/cvi2.12318
8. Hatlauf, J., Hackländer, K.: Preliminary results of golden jackal (canis aureus) survey in austria. Beiträge zur Jagd- und Wildforschung **41**, 295–306 (2016)
9. van Lunteren, P.: Addaxai: A no-code platform to train and deploy custom yolov5 object detection models. Journal of Open Source Software **8**(88), 5581 (2023). https://doi.org/10.21105/joss.05581, https://doi.org/10.21105/joss.05581
10. Magaldi, H., Cornette, R., Tibesigwa, J.J., Katumba, R., Rugonge, H., Amarasekaran, B., Anderson, N., Cappelle, N., Cardoso, A.W., Cornélis, D., Deschner, T., Fonteyn, D., Garriga, R.M., van Lunteren, P., Rufray, X., Vanthomme, H., Zwerts, J.A., Krief, S.: Deepforestvision: Automated wildlife identification for camera traps of african tropical forests. Ecological Solutions and Evidence **6**(4), e70167 (2025). https://doi.org/https://doi.org/10.1002/2688-8319.70167, https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1002/2688-8319.70167, e70167 ESO-25-07-156.R1

11. McKaughan, J.E.T., Stephens, P.A., Hill, R.A.: Estimating abundance of crop-foraging primates in anthropogenic landscapes using camera traps. American Journal of Primatology **87**, e70087 (2025). https://doi.org/10.1002/ajp.70087
12. Moore, J.F., Pine, W.E., Mulindahabi, F., Niyigaba, P., Gatorano, G., Masozera, M.K., Beaudrot, L.: Comparison of species richness and detection between line transects, ground camera traps, and arboreal camera traps. Animal Conservation **23**(5), 561–572 (2020). https://doi.org/https://doi.org/10.1111/acv.12569, https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1111/acv.12569
13. O'Brien, T.G., Kinnaird, M.F.: A picture is worth a thousand words: the application of camera trapping to the study of birds. Bird Conservation International **18(S1)** (2008). https://doi.org/10.1017/S0959270908000348
14. Pebsworth, P.A., Gruber, T., Miller, J.D., Zuberb"uhler, K., Young, S.L.: Selecting between iron-rich and clay-rich soils: a geophagy field experiment with black-and-white colobus monkeys in the budongo forest reserve, uganda. Primates **62**(1), 133–142 (2021). https://doi.org/10.1007/s10329-020-00845-y
15. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P.: Dinov3 (2025), https://arxiv.org/abs/2508.10104
16. Suss, L., Hatlauf, J.: Focus on carnivore communities: photo traps and data analysis in biodiversity research. Acta Zoobot **160** (2024)
17. 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.: Perspectives in machine learning for wildlife conservation. Nature Communications **13**(1), 792 (2022). https://doi.org/10.1038/s41467-022-27980-y
18. Vélez, J., McShea, W., Shamon, H., Castiblanco-Camacho, P.J., Tabak, M.A., Chalmers, C., Fergus, P., Fieberg, J.: An evaluation of platforms for processing camera-trap data using artificial intelligence. Methods in Ecology and Evolution **14**(2), 459–477 (2023). https://doi.org/https://doi.org/10.1111/2041-210X.14044, https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14044
19. Walton, B.J., Findlay, L.J., Hill, R.A.: Camera traps and guard observations as an alternative to researcher observation for studying anthropogenic foraging. Ecology and Evolution **12**(4), e8808 (2022). https://doi.org/10.1002/ece3.8808
20. Whytock, R.C., Suijten, T., van Deursen, T., Świeżewski, J., et al.: Real-time alerts from ai-enabled camera traps using the iridium satellite network: A case-study in gabon, central africa. Methods in Ecology and Evolution **14**(3), 867–874 (2023). https://doi.org/10.1111/2041-210X.14036
21. Whytock, R.C., Świeżewski, J., Zwerts, J.A., et al.: Robust ecological analysis of camera trap data labelled by a machine learning model. Methods in Ecology and Evolution **12**, 1080–1092 (2021). https://doi.org/10.1111/2041-210X.13576
22. Zwerts, J.A., Sterck, E.H.M., Verweij, P.A., Maisels, F., van der Waarde, J., Geelen, E.A.M., Tchoumba, G.B., Donfouet Zebaze, H.F., van Kuijk, M.: FSC-certified forest management benefits large mammals compared to non-FSC. Nature **628**(8008), 563–568 (2024). https://doi.org/10.1038/s41586-024-07257-8

A Full class taxonomy and mapping

Table 7: DeepForestVisionV2 class taxonomy with IUCN status, training support, and mapping to DeepForestVision. Classes with fewer than 1,000 training images are marked with †. IUCN codes: LC, NT, VU, EN, CR; NA = domestic taxa not assessed by IUCN. For higher taxa, $\geqslant X$ denotes the least threatened member species ranging in African tropical forests.

Class DeepForestVisionV2	Scientific name	Taxonomic level	IUCN	Support	Class DeepForestVision	Mapping
aardvark	*Orycteropus afer*	species	LC	1,328	aardvark	
african buffalo	*Syncerus caffer*	species	NT	207,162	african buffalo	
african elephant	*Loxodonta* sp.	genus	$\geqslant$ EN	642,161	african elephant	
african golden cat	*Caracal aurata*	species	VU	9,816	african golden cat	
agile mangabey	*Cercocebus agilis*	species	LC	22,951	monkey	refinement
baboon	*Papio* sp.	genus	$\geqslant$ LC	434,969	baboon	
bird of prey	*Accipitriformes*	order	$\geqslant$ LC	2,448	bird	refinement
black guineafowl	*Agelastes niger*	species	LC	35,626	guineafowl	refinement
black-and-white colobus	*Colobus* sp.	genus	$\geqslant$ LC	7,220	black-and-white colobus	
black-legged mongoose	*Bdeogale nigripes*	species	LC	7,211	mongoose	refinement
blue duiker	*Philantomba* sp.	genus	$\geqslant$ LC	489,020	blue duiker	
blue monkey	*Cercopithecus mitis*	species	LC	3,590	blue monkey	
bushbuck	*Tragelaphus scriptus*	species	LC	204,197	bushbuck	
bushpig	*Potamochoerus* sp.	genus	$\geqslant$ LC	196,209	bushpig	
casqued hornbill	*Ceratogymna atrata*	species	NT	1,156	bird	refinement
cattle	*Bos taurus*	species (domestic)	NA	86,688		new class
chimpanzee	*Pan troglodytes*	species	EN	198,350	chimpanzee	
civet	*Civettictis civetta*	species	LC	2,784	civet_genet	refinement
crane†	*Gruidae*	family	$\geqslant$ LC	921	bird	refinement
dog	*Canis lupus familiaris*	subspecies (domestic)	NA	16,624		new class
dove	*Columbidae*	family	$\geqslant$ LC	8,012	bird	refinement
duck	*Anatidae*	family	$\geqslant$ LC	3,205	bird	refinement
francolin	*Pternistis*	genus	$\geqslant$ LC	9,379	bird	refinement
galago	*Galagidae*	family	$\geqslant$ LC	1,512	galago_potto	refinement
genet	*Genetta*	genus	$\geqslant$ LC	16,014	civet_genet	refinement
goat	*Capra hircus*	species (domestic)	NA	21,358		new class
gorilla	*Gorilla* sp.	genus	$\geqslant$ CR	24,099	gorilla	
grey-cheeked mangabey	*Lophocebus albigena*	species	LC	5,698	monkey	refinement
guineafowl	*Numididae*	family	$\geqslant$ LC	181,440	guineafowl	
hippopotamus	*Hippopotamus amphibius*	species	VU	6,817		new class
honey badger	*Mellivora capensis*	species	LC	1,959	honey badger	

Table 7 (continued)

Class DeepForestVisionV2	**Scientific name**	**Taxonomic level**	**IUCN**	**# Training Images**	**Class DeepForestVision**	**Mapping**
hornbill†	*Bucerotidae*	family	$\geqslant$ LC	963	bird	refinement
hyrax	*Procaviidae*	family	$\geqslant$ LC	8,090	hyrax	
leopard	*Panthera pardus*	species	VU	7,134	leopard	
lhoests monkey	*Allochrocebus lhoesti*	species	VU	30,919	lhoests monkey	
mandrill	*Mandrillus sphinx*	species	VU	82,529	mandrill	
mongoose	*Herpestidae*	family	$\geqslant$ LC	23,793	mongoose	
moustached monkey	*Cercopithecus cephus*	species	LC	1,487	monkey	refinement
nkulengu rail	*Himantornis haematopus*	species	LC	2,787	bird	refinement
otter	*Lutrinae*	subfamily	$\geqslant$ NT	1,257		new class
pangolin	*Manidae*	family	$\geqslant$ EN	6,832	pangolin	
passeriform	*Passeriformes*	order	$\geqslant$ LC	6,887	bird	refinement
pelecaniform	*Pelecaniformes*	order	$\geqslant$ LC	4,216	bird	refinement
porcupine	*Hystricidae*	family	$\geqslant$ LC	51,320	porcupine	
potto†	*Perodicticus potto*	species	NT	162	galago_potto	refinement
red colobus	*Piliocolobus* sp.	genus	$\geqslant$ VU	18,786	red colobus_red-capped mangabey	refinement
red duiker	*Cephalophus* sp.	genus	$\geqslant$ LC	633,121	red duiker	
red-capped mangabey	*Cercocebus torquatus*	species	EN	10,476	red colobus_red-capped mangabey	refinement
red-tailed monkey†	*Cercopithecus ascanius*	species	LC	619	monkey	refinement
rockfowl	*Picathartes*	genus	$\geqslant$ VU	1,921	bird	refinement
rodent	*Rodentia*	order	$\geqslant$ LC	70,027	rodent	
serval	*Leptailurus serval*	species	LC	3,664	serval	
side-striped jackal	*Canis adustus*	species	LC	6,560	side-striped jackal	
sitatunga	*Tragelaphus spekii*	species	LC	2,579		new class
spot-nosed monkey	*Cercopithecus nictitans*	species	LC	1,489	monkey	refinement
spotted hyena	*Crocuta crocuta*	species	LC	18,234	spotted hyena	
squirrel	*Sciuridae, Anomaluridae*	family	$\geqslant$ LC	91,131	squirrel	
turaco†	*Musophagidae*	family	$\geqslant$ LC	198	bird	refinement
warthog	*Phacochoerus* sp.	genus	$\geqslant$ LC	9,320	bushpig	refinement
water chevrotain	*Hyemoschus aquaticus*	species	LC	23,677	water chevrotain	
yellow-backed duiker	*Cephalophus silvicultor*	species	LC	92,400	yellow-backed duiker	

B Per-class validation metrics

Table 8: DeepForestVisionV2 per-class metrics and support on the validation set.

Class	Precision	Recall	F1	Support
red colobus	0.99	0.95	0.97	1654
genet	0.96	0.86	0.91	1649
gorilla	0.91	0.67	0.77	1319
mandrill	0.84	0.98	0.91	1158
red-capped mangabey	0.99	0.87	0.93	1155
blue duiker	0.82	0.95	0.88	1141
spotted hyena	0.88	0.88	0.88	1120
guineafowl	0.78	0.94	0.85	1117
rodent	0.84	0.91	0.87	1112
baboon	0.85	0.87	0.86	1091
water chevrotain	0.92	0.94	0.93	1085
porcupine	0.89	0.92	0.90	1076
bushpig	0.85	0.95	0.89	1060
goat	0.86	0.99	0.92	1045
black guineafowl	0.81	0.89	0.85	1013
red duiker	0.76	0.97	0.86	1002
african elephant	0.86	0.88	0.87	993
bushbuck	0.92	0.85	0.88	987
francolin	0.91	0.81	0.86	986
mongoose	0.73	0.82	0.77	970
agile mangabey	0.89	0.84	0.87	954
squirrel	0.75	0.95	0.84	949
cattle	0.96	0.85	0.90	934
lhoests monkey	0.86	0.87	0.87	927
african buffalo	0.76	0.92	0.83	888
yellow-backed duiker	0.78	0.91	0.84	876
chimpanzee	0.67	0.84	0.74	824
hyrax	0.97	0.86	0.91	801
black-and-white colobus	0.99	0.84	0.91	793
african golden cat	0.93	0.71	0.80	707
hippopotamus	0.93	0.83	0.88	706
warthog	0.98	0.88	0.93	676
side-striped jackal	0.98	0.62	0.76	676
black-legged mongoose	0.84	0.82	0.83	548
dog	0.82	0.71	0.76	514
leopard	0.92	0.91	0.92	492
dove	0.92	0.77	0.84	432
passeriform	0.91	0.61	0.73	417
pangolin	0.78	0.65	0.71	401
nkulengu rail	0.76	0.85	0.80	388
pelecaniform	0.99	0.93	0.96	372
serval	0.82	0.83	0.83	354
blue monkey	0.78	0.73	0.75	351
grey-cheeked mangabey	0.79	0.78	0.79	306
sitatunga	0.90	0.55	0.68	259
bird of prey	0.89	0.89	0.89	199
civet	0.82	0.93	0.87	195
moustached monkey	0.84	0.76	0.80	181
honey badger	0.58	0.33	0.42	160
casqued hornbill	0.96	0.89	0.93	139
rockfowl	0.76	0.88	0.81	124
aardvark	0.97	0.70	0.82	107
otter	0.89	0.68	0.78	98
spot-nosed monkey	0.52	0.85	0.64	80
red-tailed monkey	0.64	0.54	0.58	78
galago	0.81	0.52	0.64	65
hornbill	0.55	0.73	0.63	55
duck	1.00	0.71	0.83	31
turaco	0.86	0.57	0.69	21
potto	0.40	0.63	0.49	16
crane	1.00	1.00	1.00	5