Can biodiversity monitoring benefit from deep learning? Evidence from Taï National Park

Andrew Bennett [1] Emmett Sexton [1] Luis Francisco Alvarez Poli [1]

[1]Barcelona School of Economics

Abstract

This poster examines the potential of deep learning to aid biodiversity monitoring efforts in Taï National Park, Côte d'Ivoire. We develop a convolutional neural network (CNN) to classify animal species in trail camera images. Using a dataset of 16,489 labeled images from the park, the model achieved 80% accuracy in classifying eight categories (seven animal types and blank images). The study demonstrates that deep learning can effectively automate species classification in camera trap data, potentially reducing costs and labor for conservation efforts. However, the model's performance varied across species, with some categories achieving high reliability (F1-scores >0.80) while others, particularly antelopes and blank images, showed lower accuracy.

Background: Biodiversity in Côte d'Ivoire

Balanced biodiversity is key to a well functioning ecosystem. However, recent biodiversity loss is associated with major disruptions across many areas of life, including food security and economic well-being. Nearly $43 trillion of economic output each year relies on healthy and robust biodiversity (Kurth et al.). A large share of this output will be under threat if biodiversity continues to decline. As part of this global decline, Côte d'Ivoire is making way for more agricultural land for cocoa, their largest cash crop, primarily through deforestation. Coupled with climate change, this deforestation is exacerbating biodiversity losses throughout the country which will exert material economic damage on the country's prospects (UNEP). However, Côte d'Ivoire has taken steps to stop the degradation of its biodiversity. One of its major initiatives has been protecting approximately 23% of all its land area from destructive human activities (World Bank).

Tāi National Park is one such protected area home to one of West Africa's largest rain-forests and a large share of the country's biodiversity. Our goal is to determine the feasibility of utilizing deep learning to assist the national park in its efforts to better understand the state of biodiversity within its park by classifying animal species across trail camera images. Trail camera data is an invaluable resource to allow researchers and conservationists to take motion activated photos at many distant locations at the same time. However, this monitoring has historically required individuals to sift through tens of thousands or more trail photos and manually classify the animals seen in each photo. These labor hours have imposed a material cost on the often small budgets of conservation groups doing this work. If our deep learning approach obtains significant reliability, there is potential for these groups to reduce their costs on species classification and direct their funds towards other pressing problems.

Previous trail camera use cases

Using trail cameras in conjunction with Deep Learning techniques is not a new area of research, and has been proven successful at reducing costs in the past. Companies and wildlife divisions have been employing machine learning techniques to help optimize planning, maintenance, and investment operations. The earliest examples of this research target counting techniques. These counting techniques use simple bounding boxes to classify images, which allows the a wildlife department to optimize resources, minimize ecological damage and increase ranger support (Staab).

Another more nuanced people counting method allows non-profit companies in Rwanda to observe the impact of bridges that they installed. These bridges are installed in rural parts of Rwanda so that treacherous terrain can be scaled more easily. The observation of these bridges allows researches to observe the impact of these bridges and in conjunction with satellite imagery allow observation of ecological impacts of the bridge (E. Thomas et al.).

An example highly related to our own use case, was counting species in a Texas, US wildlife park. Researches first divided images into night and day images tracking major differences in pixel color. Then created a layer which categorized each image as a bird or an animal. After classifying the birds and animals, they trained a model to classify species of birds. With their approach, they were able to count species of birds with more than 92% specificity and sensitivity (Golnaz et al.). In our research, we aim to use similar techniques to classify a more extensive set of animals.

Data: Trail Camera Trap Images

Our data is provided by the Wild Chimpanzee Foundation and the Max Planck Institute for Evolutionary Anthropology, who collected images from camera traps located in different sites around Taï National Park. They provide 16,489 images that are labelled with one of seven animal types: birds, civets, duikers, hogs, leopards, prosimians, and rodents. There are also images that contain no animals. The data was accessed through an on-going Driven Data image classification competition page (Driven Data). To construct, fine-tune, and then test our model we split our images using a 70%-15%-15% split for train-validation-test sets. Originally, these images were supposed to be the train dataset for the competition, but we create the train-val-test partitions to evaluate the out of sample predictive power of our trained model.

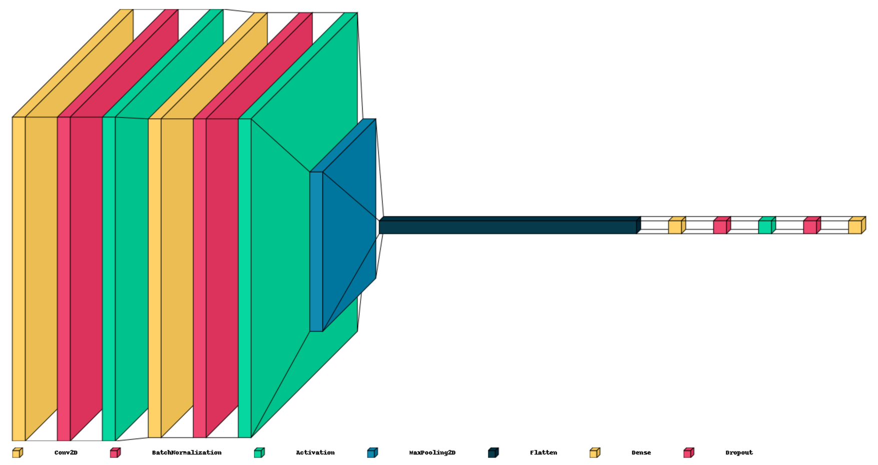
Figure 1. Example of a Hog

Figure 2. Example of a Prosimian

Model Architecture and Training

We developed a convolutional neural network (CNN) to classify our trail images into one of our eight categories. After training a wide range of model architectures (detailed in our Appendix), including more sophisticated CNN architectures, as well as transfer learning approaches using ResNet and EfficientNet, we quickly realized that the patterns in our trail camera images are better captured by a relatively simple architecture presented below in Figure 3:

Figure 3. Model architecture



- **Data Augmentation:** To enhance generalization, training data is augmented with rescaling, rotation, shear, and horizontal flipping, while validation data is only rescaled. We experimented with zoom and shifts, but these performed worse overall.
- **Model Architecture:** Our CNN processes 128x128 pixel RGB images. The architecture begins with a convolutional layer of 32 filters with a 3x3 kernel. This is followed by a second convolutional layer of 64 filters, along with a 2x2 max pooling layer to reduce spatial dimensions. The output is flattened and passed through a dense layer with 128 units, and a 50% dropout rate. Batch normalization and ReLU activations were used for each layer. Finally, the output layer uses softmax activation to classify the images into one of our eight categories: seven animal types or blank.
- **Training Configuration:** The model is compiled with the Adam optimizer and categorical cross-entropy loss function. Early stopping was implemented, monitoring validation loss to prevent overfitting. Using a batch size of 32, we obtain the best validation loss after 124 epochs.

Results

Our model accurately classifies 80% of all 2,474 test images. Macro or weighted averages for precision, recall, and F1-score across the eight classes were also tightly clustered, ranging from 0.80–0.82 depending on the statistic. Within individual classes precision and recall are typically high and well balanced. For six of the classes, precision, recall, and F1-score are all greater than 0.8. Antelope and the blank classes perform significantly worse with F1-scores of only 0.64 and 0.54 respectively.

Figure 4. Model performance on test images

	Precision	Recall	F1-score	Support
Macro average	0.82	0.81	0.81	2,474
Weighted average	0.80	0.80	0.80	2,474
Class				
Civet-genet	0.90	0.95	0.92	380
Leopard	0.94	0.89	0.92	336
Bird	0.90	0.89	0.90	227
Hog	0.95	0.85	0.90	142
Monkey-prosimian	0.83	0.83	0.83	401
Rodent	0.84	0.82	0.83	308
Antelope-duiker	0.59	0.69	0.64	357
Blank	0.58	0.50	0.54	323

For the test set classes with F1-scores of 0.80+, conservationists could implement our compact CNN and be confident that the image classification is reliable, saving material amounts of labor hours. However, future research should further dissect the driver behind the poorer performance for the Antelope and blank classes. This may take the form of alternative over-sampling techniques or the training of more specialized models that detect a fewer number of classes.

References

Driven Data. (2023). Conser-vision Practice Area: Image Classification. Link.

Kurth, T., Wübbels, G., Portafaix, A., Meyer zum Felde, A., and Zielcke, S. (2021). The Biodiversity Crisis Is a Business Crisis. *Boston Consulting Group*. Link.

Moallem, G., Pathirage, D. D., Reznick, J., Gallagher, J., and Sari-Sarraf, H. (2021). An explainable deep vision system for animal classification and detection in trail-camera images with automatic post-deployment retraining. *Knowledge-Based Systems*, vol. 216, pp. 106815. doi: Link.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423.

Staab, J., Udas, E., Mayer, M., Taubenböck, H., and Job, H. (2021). Comparing established visitor monitoring approaches with triggered trail camera images and machine learning based computer vision. *Journal of Outdoor Recreation and Tourism*, vol. 35, pp. 100387. doi: Link.

Thomas, E., Gerster, S., Mugabo, L., Jean, H., and Oates, T. (2020). Computer vision supported pedestrian tracking: A demonstration on trail bridges in rural Rwanda. *PLoS One*, vol. 15, no. 10, pp. e0241379. doi: Link.

United Nations Environment Programme (2024). Côte d'Ivoire - Biodiversity. Link.

World Bank (2022). Terrestrial protected areas - Cote d'Ivoire. Link.