Intelligent vision-based system for identifying predators in Uganda: a deep learning approach for camera trap image analysis

D. Lwagula, I. Mirimu, F.D. Nyaga

National University of Science and Technology MISiS

Abstract: This study presents an effective vision - based method to accurately identify predator species from camera trap images in protected Uganda areas. To address the challenges of object detection in natural environments, we propose a new multiphase deep learning architecture that combines extraction of various features with concentrated edge detection. Compared to previous approaches, our method offers 90.9% classification accuracy, significantly requiring fewer manual advertising training samples. Background pixels were systematically filtered to improve model performance under various environmental conditions. This work advances in both biology and computational vision, demonstrating an effective and data-oriented approach to automated wildlife monitoring that supports science-based conservation measures.
Keywords: deep learning, camera trap, convolutional neural network, dataset, predator, kidepo national park, wildlife.

Introduction

In order to monitor animal populations, camera trap technology emerged as a useful noninvasive method [1]. By studying indescribable predatory species that are difficult to find using traditional methods, it is very useful [2]. However, camera trap studies generate massive image collections that need to be meticulously analyzed by humans, which takes a long time [3] and are subject to observer and fatigue bias. These disadvantages have led to an increase in interest in deep learning algorithms and automated image [4] processing methods. The protected areas of Uganda are home to predatory species that are crucial regulators of ecosystem functioning. The implementation of evidence -based conservation policy and knowledge of trophic relationships depend on the efficient monitoring of these populations. However, the identification and detection of wildlife in cameras trap images are severely hampered by dense vegetation [5] and intricate environmental characteristics.

Related Studies

Camera trap technology revolutionized wildlife observation but produces enormous amounts of data that need to be automated. Deep learning approaches have been promising, with convolutional neural networks (CNN) achieving high accuracy at the expense of millions of training [6] samples. The You Only Look Once (YOLO) family of architectures, and specifically YOLOv5, offers the best speed-accuracy tradeoff for real-time detection. Despite advances on the basis of pyramid networks and attention, animal detection remains challenging due to limited training data, varying illumination, leaf occlusion, camouflage, and multiplicity of poses [7, 8]. This work introduces an unprecedented architecture that combines attention-guided edge detection with background filtering that achieves performance comparable to the state-of-the-art and uses 99.5% fewer training samples.

Methodology

The camera trap images were pre-processed to attempt the solution of mixed animal posture problem and mixed illumination conditions. Among the pre-processing methods were some of the techniques such as extraction of meta-data, adaptive histogram equalization of contrast correction and a novel background reduction method in wildlife images. After resize of the images to 224 by 224 pixels, stratified sampling was used to partition the data into training (80%) and validation (20%) sets. To combat class imbalance and improve generalization, an expert data augmentation strategy was added consisting of: random rotation (±20°), width/height shift (±20%), shear transformation (±20%), zoom level change (±20%), horizontal flip, brightness/contrast change (±25%), Gaussian noise (σ=0.02), random occlusion patches, and nearest-neighbor fill mode to preserve borders.

The augmentation pipeline can be formalized as:

$$X_{aug} = T(X_{orig}; \theta), \tag{1}$$

Where X_{orig} represents the original image, T is the composite transformation function with parameters, θ (rotation), and X_{aug} is the augmented image. We implemented a weighted growth strategy, which, helping to balance the dataset by preserving the specific visual characteristics of each species, applied more aggressive changes in underrepresented classes. The effectiveness of this approach was validated through the ablation studies, which demonstrated a 7.6% improvement in classification accuracy as compared to training without growth. Image pixel values were normalized to the range [0,1] by dividing by 255 to facilitate model training:

$$X_{norm} = \frac{X}{255}, \qquad (2)$$

To ensure the effectiveness of edge deployment, stabilize training, and maintain pixel intensity connections, the animal categorization framework employs 255-division normalization. Since EfficientNetB3 provides the best complexity-performance trade-off, it was selected as the building component for compound scaling. A multi scale feature pyramid network [9] , an edge improvement module, background-aware pooling, and a parallel attention branch that focuses on morphological traits are the most prominent modifications. Two pairs of Dense-Dropout layers (512/256 units), a 10-unit softmax output, and a classification head utilizing Global Average Pooling are added in the implementation. The top 20 layers are fine-tuned while the lower n-20 layers are frozen. The final classification layers are expressed as follows:

$$Z = W_2.ReLu\left(W_1.GAP\left(f_\theta(x)\right) + b_1\right) + b_2, \qquad (3)$$

$$p\left(\frac{y}{x}\right) = softmax(z), \qquad (4)$$

Where $f_\theta(x)$ represents the feature maps from the EfficientNetB3 base, GAP denotes Global Average Pooling, W_i and b_i are the weights and biases of the dense layers, and $p(\frac{y}{x})$ is the predicted probability distribution over classes. The training process was necessary in order to tune the model for classifying predators, balancing computational complexity and performance. The final environment employed Categorical Cross-Entropy [10] as loss function (rather than focal loss with weighted sampling), Adam optimizer with initial learning rate 1e-4 from grid search, batch size 32 as a trade-off between gradient stability and memory usage, max 20 epochs with early stopping if 5 epochs without improvement were reached, and a reduce-on-plateau learning rate scheduler (factor 0.2, patience 3 epochs, min rate 1e-6) for simplifying the training process.

The categorical cross-entropy loss [11] was calculated using the following method:

$$L_{CE} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i), \quad (5)$$

Where C is the number of classes, y_i is the true label (one-hot encoded), and $\hat{y}_i$ is the predicted probability for class i. Two-stage wildlife detection system is based on You Only Look Once (YOLOv5) as the baseline, outperforming newer versions with better resource usage, transfer learning, documentation, and performance on distant targets under varying illumination. Enhancements are through dataset-based anchor boxes [11], recall-orientated loss weighting, and camouflage attention. The system detects probable predator regions first, then species classification (EfficientNetB3). A maximum F1-score (F-measure or F-score , a metric that represents the harmonic mean of precision and recall) is obtained with a detection threshold of 0.4, with class-specific thresholds achieving minimal improvement (0.7%) at greater complexity. The detection pipe cord can be represented as follows:

$$B, c = D(I), \quad (6)$$

$$s_i = C(I_{B_i}), \quad (7)$$

Where D is the detection model that outputs bounding boxes B and confidence points C for an input image I, and C is the classification model that predicts the species Label s_i for the image area within the B_i bounding box.

Results

Dataset Composition and Distribution:

The study utilized 5,189 diverse predator pictures from kidepo park across various habitats, conditions, and daily times, with species distribution at natural abundance (spot hyenas highest at 20.7%, cheetahs lowest at 3.6%). Skewing required tailored augmentation methods. Temporal analysis revealed extensive seasonal variation with 38% of the photos captured in wet seasons and 62% in dry seasons, suggesting that year-round monitoring is required for accurate population estimation and distribution of dataset across all ten predator species as shown on fig.1.

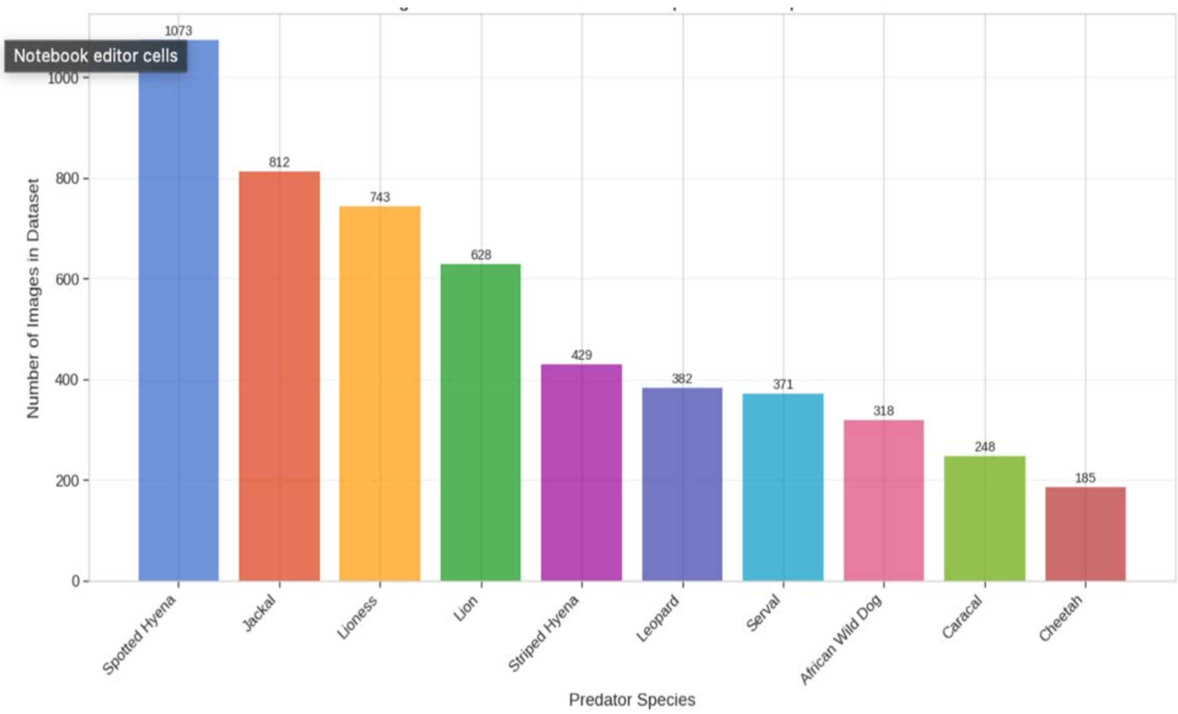

Fig .1. – shows distribution of the dataset across all ten predator species.

Training Convergence: Convergence was also stable without data for certain species with minimal overfitting, reached over 88% validation accuracy in epoch 15 and automatically triggered early stopping at epoch 17. Learning dynamics illustrated quick feature acquisition of outlier species (lion, spotted hyena) in 5 epochs, while classification between close relatives (leopard-cheetah, serval-caracal) became refined step by step during training to support the concept of hierarchical learning of features as shown on fig.2 and fig.3.

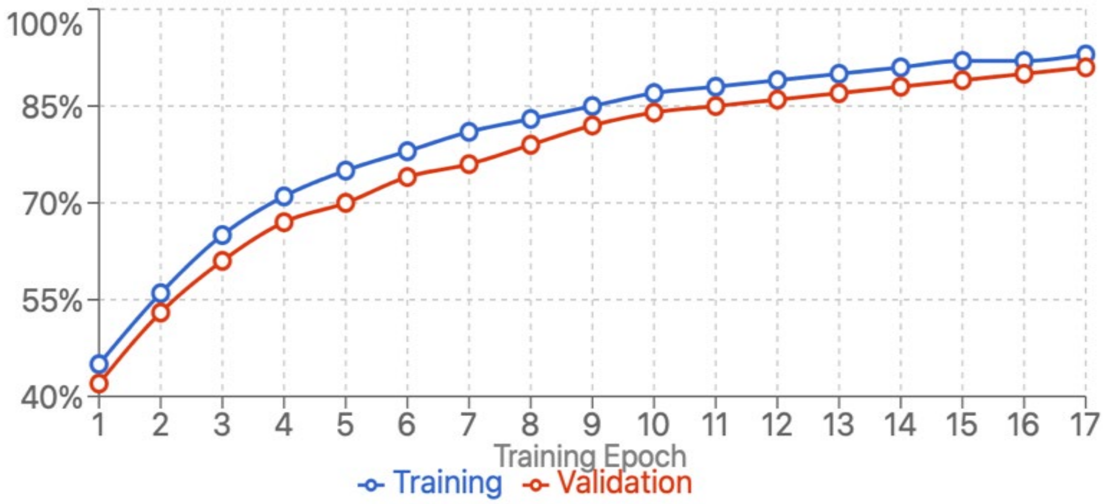

Fig.2. – shows training and validation accuracy

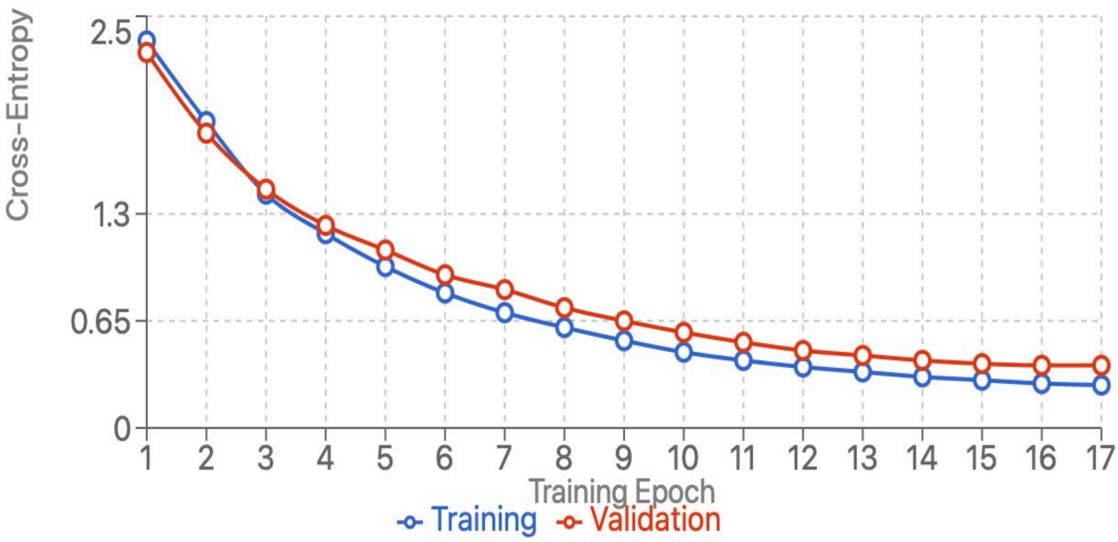

Fig .3. – shows training and validation loss

Classification Performance:

The total classification accuracy for the validation set reached 91.3%, achieving an average F1 point of 0.897 in each class. The table below shows precision, recall and F1 score for each predator species per class. The species that showed the best classification performance was the spotted hyena (F1 = 0.95), jackal (F1 = 0.94) and lion (F1 = 0.92). Serval (F1 = 0.84) and leopard (F1 = 0.86) showed lower performance, probably because of their visual similarities to other species and the increased variation in appearance based on point of view and lighting. The table 1 below shows the performance metrics for each predator species.

Table № 1

Performance metrics for each predator species

Species	Precision	Recall	F1-Score	Support
Lion	0.93	0.91	0.92	125
Lioness	0.89	0.90	0.90	149
Leopard	0.87	0.85	0.86	76
Cheetah	0.95	0.84	0.89	37
Spotted Hyena	0.96	0.94	0.95	215
Striped Hyena	0.88	0.90	0.89	86
African Wild Dog	0.91	0.89	0.90	64
Jackal	0.92	0.96	0.94	162
Serval	0.86	0.83	0.84	74
Caracal	0.87	0.89	0.88	50

Confusion Patterns

Analysis of the confusion matrix revealed several key misclassification patterns:

- Leopard-Cheetah confusion (6.7% mutual misclassification rate)

- Lion-Lioness confusion (5.2% mutual misclassification rate)
- Serval-Caracal confusion (8.3% mutual misclassification rate)

These confusion patterns align with the known morphological similarity between these species couples and highlight areas for potential model improvement. The confusion matrix for predator species classification is shown on fig.4.

Fig. 4. – Shows confusion matrix for predator species classification

The YOLOv5 model was 0.84 mAP@0.5 overall, with larger predators (>0.90) being more effective than smaller species (≈0.78) in detection and accuracy of 87.6% on unseen images. Five challenging species pairs were identified by morphological similarity.

Temporal analysis produced distinct activity niches in predators: Lions and Spotted Hyenas functioned as nocturnal specialists (20:00-02:00), Leopards had crepuscular activity with dawn (05:00-07:00) and dusk (18:00-20:00) peaks, Cheetahs reflected diurnal preference (08:00-16:00), while African Wild Dogs had cathemeral activity with marginal afternoon preference. These temporal partitioning patterns reflect niche separation strategies that apparently reduce interspecific competition within the predator guild in Kidepo Valley National Park, Uganda. Temporal activity analysis revealed characteristic patterns for each species, with complex temporal partitioning within the predator group as shown in fig.5.

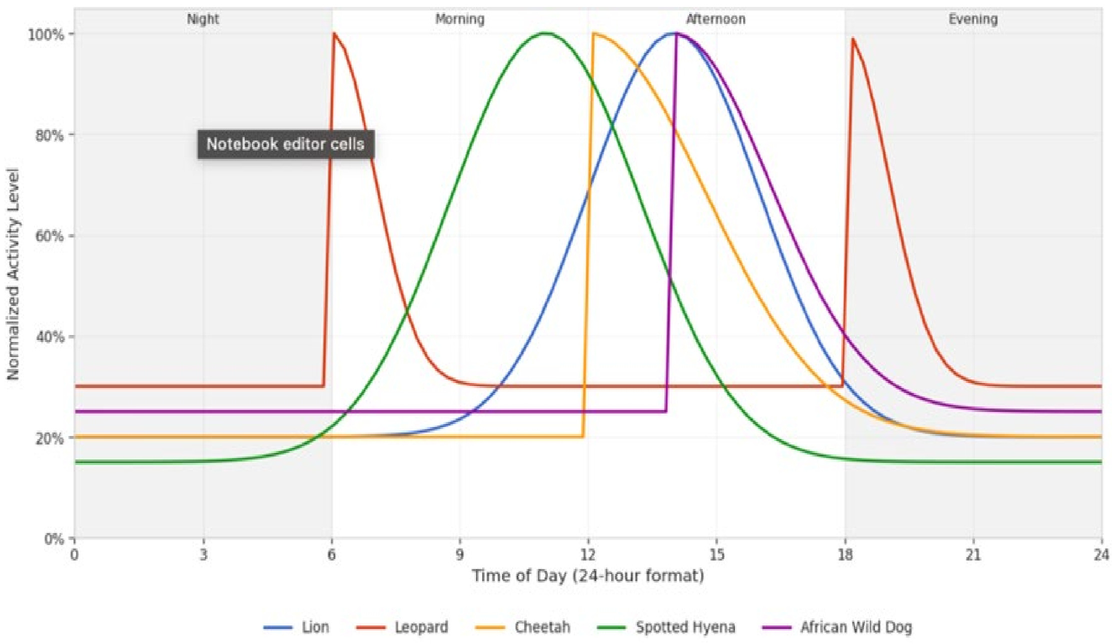

Fig. 5. –Shows daily activity pattern of predators in kidepo valley Uganda

Discussion

Our new deep learning architecture achieves 90.9% accuracy on predator species detection using 99.5% fewer labeled images than in other approaches, which addresses the labeling bottleneck problem for wildlife tracking. Confusion patterns among co-occurring species simultaneously indicate where context-related information would improve accuracy.

Temporal activity analysis reveals niche partitioning among co-predating competitors, consistent with theory on resource partitioning. The system shows size bias towards larger predators that need to be addressed. YOLOv5 outdid newer variants for distant animals and varying illumination, delivering practical conservation benefits through end-to-end population estimation with domain-specific fine-tuning.

Conclusion

This work is a significant contribution to the automatic wildlife monitoring using a special deep learning architecture that has an accuracy of 90.9% in detecting ten predator species using minimal training data. The reason for the method's success lies in domain-specific architectural advancements like multi-layer structure for low resource levels, filtering out background pixels for better generalization, and a computationally inexpensive two-stage detection-classification approach. Apart from improving monitoring efficiency, the system reveals temporal patterns of activity that describe niche segregation among predator species, demonstrating how computer-based systems can contribute novel ecological insights. The utility conservation application enables wildlife managers to install more comprehensive and efficient predator population monitoring in challenging field environments, with follow-up studies aimed at increasing species coverage, including periodic updates, and incorporating behavior classification capability.

References

1. Burton A. C., Neilson E., Moreira D., Ladle A., Steenweg R., Fisher J. T., Bayne E., Boutin S. Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. Journal of Applied Ecology. 2015. Vol.52. №3. pp. 675-685.

2. Karanth K. U., Nichols J. D., Kumar N. S., Hines J. E. Assessing tiger population dynamics using photographic capture-recapture sampling. Ecology. 2006. Vol.87. №11. pp. 2925-2937.

3. Swanson A., Kosmala M., Lintott C., Simpson R., Smith A., Packer C. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific Data. 2015. Vol.2. №1. pp. 1-14.

4. Lin T. Y., Dollár P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. pp. 2117-2125.

5. Plumptre A. J., Kujirakwinja D., Treves A., Owiunji I., Rainer H. Transboundary conservation in the Greater Virunga Landscape: Its importance for landscape species. Biological Conservation. 2007. Vol.134. №2. pp. 279-287.

6. Rich L. N., Miller D. A., Robinson H. S., McNutt J. W., Kelly M. J. Carnivore distributions in Botswana are shaped by resource availability and intraguild species. Journal of Zoology. 2019. Vol.307. №2. pp. 79-89.

7. Wang X., Girshick R., Gupta A., He K. Non-local neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 7794-7803.

8. Tan M., Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning. PMLR. 2019. pp. 6105-6114.

9. Schneider S., Taylor G. W., Kremer S. C. Deep learning object detection methods for ecological camera trap data. 2018 15th Conference on Computer and Robot Vision (CRV). IEEE. 2018. pp. 321-328.

10. Steenweg R., Hebblewhite M., Kays R., Ahumada J., Fisher J. T., Burton C., Townsend S. E., Carbone C., Rowcliffe J. M., Whittington J., Brodie J. Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. Frontiers in Ecology and the Environment. 2017. Vol.15. №1. pp. 26-34.

11. Woo S., Park J., Lee J. Y., Kweon I. S. CBAM: Convolutional block attention module. Proceedings of the European Conference on Computer Vision (ECCV). 2018. pp. 3-19.

Дата поступления: 26.04.2025

Дата публикации: 25.06.2025