ENVIRONMENTAL RESEARCH
COMMUNICATIONS

PAPER • OPEN ACCESS

Evaluating machine learning-based elephant recognition in complex African landscapes using drone imagery

To cite this article: Chris McCarthy *et al* 2024 *Environ. Res. Commun.* **6** 115035

IOP Publishing Environ. Res. Commun. 6 (2024) 115035 https://doi.org/10.1088/2515-7620/ad9380

Environmental Research Communications

CrossMark

OPEN ACCESS

RECEIVED
20 July 2024

REVISED
27 September 2024

ACCEPTED FOR PUBLICATION
15 November 2024

PUBLISHED
28 November 2024

PAPER

Evaluating machine learning-based elephant recognition in complex African landscapes using drone imagery

Chris McCarthy[1,2,*], **Lumbani Benedicto Banda**[3,4], **Daud Jones Kachamba**[2], **Zuza Emmanuel Junior**[5], **Cornelius Chisambi**[6], **Ndaona Kumanga**[3], **Luciano Lawrence**[3] **and Troy Sternberg**[7,8]

1 Zanvyl Krieger School of Arts & Sciences, Johns Hopkins University, Baltimore, MA, 21218, United States of America
2 Lilongwe University of Agriculture and Natural Resources (LUANAR), Bunda College of Agriculture Campus, Department of Forestry, PO Box 219, Lilongwe, Malawi
3 Lilongwe University of Agriculture and Natural Resources (LUANAR), Bunda College of Agriculture Campus, Department of Environment and Natural Resources, PO Box 219, Lilongwe, Malawi
4 Africa Centre of Excellence for Climate Smart Agriculture and Biodiversity Conservation, Haramaya University, PO Box 138, Dire Dawa, Ethiopia
5 School of Agricultural Science and Practice, Royal Agricultural University, GL7 6JS, Cirencester, Gloucestershire, United Kingdom
6 Graduate School of Global Environmental Studies, Kyoto University, Kyoto, 606-8501, Japan
7 CEI Centre for International Studies ISCTE—University Institute Lisbon, Avenida das Forças Armadas, 1649, Lisbon, Portugal
8 School of Geography, University of Oxford, Oxford OX1 3QY, United Kingdom
* Author to whom any correspondence should be addressed.

E-mail: cmccar27@jh.edu, lumbabanda@gmail.com, dkachamba@gmail.com, Emmanuel.Zuza@rau.ac.uk, corneliuschisambi79@gmail.com, nkumanga@gmail.com, lawrenceluciano40@gmail.com, tsgyr@iscte-iul.pt and Troy.sternberg@geog.ox.ac.uk

Keywords: African elephants, artificial intelligence, AWS, drone imagery, machine learning, Malawi, wildlife identification

Abstract

This paper evaluates a machine learning-based approach for identifying and analyzing African bush elephants within complex terrains using high-resolution drone imagery. With human-wildlife conflict posing a significant threat to elephants worldwide, accurate and efficient monitoring techniques are crucial, yet challenging in diverse landscapes. Our study utilizes approximately 3,180 drone-captured images from Kasungu National Park in Malawi, encompassing various terrains including dense forests and open bushlands. These images were systematically preprocessed and analyzed using three distinct ML algorithms: Faster R-CNN, RetinaNet, and Mask R-CNN, each fine-tuned for identification of elephants across different age groups. Comparative performance metrics revealed nuanced strengths and limitations: Faster R-CNN showed notable proficiency in detecting adult elephants, particularly in dense foliage. Mask R-CNN, while less precise overall, demonstrated increased effectiveness in identifying juveniles and infants. RetinaNet, optimized for larger images, showed particular adeptness with adult elephants but less so with younger ones. Despite these promising results, overall recognition rates were lower than ideal, highlighting the complexities of wildlife identification in natural settings. This study not only facilitates the identification and counting of individual elephants but also provides insights into the challenges of applying ML in complex ecological contexts. The derived insights can assist conservationists and park officials in making informed decisions related to wildlife protection and habitat preservation. Furthermore, the study offers a valuable blueprint for integrating AI and machine learning technology into wildlife conservation strategies, presenting a scalable model with potential applications for different species and geographic regions, while acknowledging the need for further refinement to enhance accuracy and reliability in diverse ecological settings.

1. Introduction

Wildlife conservation has gained substantial attention due to escalating threats like habitat destruction, climate change, and poaching, resulting in significant biodiversity losses worldwide [1–3]. In particular, African

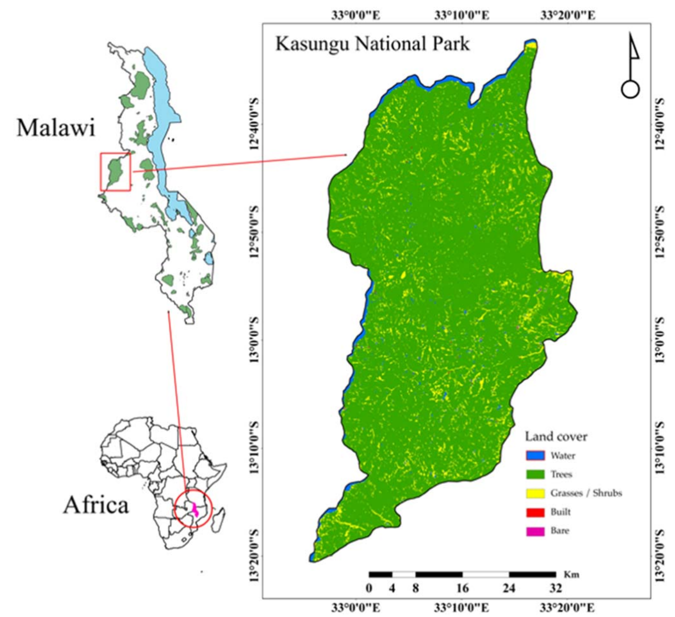


Figure 1. Map showing the location of Kasungu National Park in Malawi.

elephants (Loxodonta africana), essential megafauna for ecological balance and biodiversity, have seen a significant population decline. A continent-wide aerial survey reported a 30% decrease from 2007 to 2014 [4], and recent studies continue to highlight the enduring issue of poaching [5].

These elephants are vital for environmental equilibrium and significantly contribute to the tourism industry, a critical economic driver in many African nations [6–8]. Thus, understanding and monitoring their populations address conservation goals and support broader sustainable development objectives in regions dependent on wildlife tourism [9].

Our study focuses on Kasungu National Park (KNP), located 175 km north of Malawi's capital, Lilongwe, sharing its boundaries with Zambia (figure 1). Covering an area of over 2,300 km^2, it is the second-largest national park in Malawi. The park's terrain is a mosaic of diverse ecosystems, with predominant vegetation consisting of miombo woodlands, interspersed with grasslands and occasional wetlands, providing an ideal habitat for its varied fauna.

A significant inhabitant of KNP is the African bush elephant (Loxodonta africana). As of 2021, this species has been listed as 'Endangered' on the IUCN Red List, with the primary threats being habitat destruction and poaching for ivory and bush meat. Given this status, the park plays a crucial role in the conservation of the species. In July 2022, 250 elephants, along with 250 other animals, were translocated from Liwonde National Park to KNP. This translocation was one of the largest in Malawi's history, emphasizing the importance of KNP in the country's wildlife conservation efforts [10].

To further protect the elephants and mitigate human-elephant conflict, 60 km of fencing were added to the eastern border of KNP's existing boundary. This extension plays a crucial role in safeguarding both the park's elephants and the surrounding communities. However, despite these efforts, problems persist. Gaps in the fence still allow for the possibility of elephants exiting the park boundaries, posing ongoing challenges for conservation and community safety.

KNP faces significant challenges. The proximity to the Zambian border has made it vulnerable to cross-border poaching activities. This problem is compounded by a limited staff of rangers and a lack of adequate resources to effectively patrol and safeguard the vast expanse of the park. The poaching issue, coupled with insufficient manpower and resources, highlights an urgent need for innovative monitoring solutions.

Traditional monitoring techniques, including ground-based surveys conducted on foot or by vehicle, and aerial counts using helicopters, often fall short due to limitations in resources, accessibility, and their capability to effectively survey large, difficult terrains [11, 12]. These methods struggle to cover extensive areas and detect illegal activities effectively, emphasizing an urgent need for innovative monitoring solutions.

Given these challenges faced by Kasungu National Park and the limitations of traditional monitoring techniques, innovative technological solutions are urgently needed to enhance conservation efforts.

Technological advancements, specifically machine learning (ML) and drones, present a promising conservation opportunity [13, 14]. The increased use of drones, capable of capturing high-resolution imagery across large and often inaccessible areas, has proven invaluable in wildlife conservation [15]. This trend is driven by increased versatility, decreasing costs, and the growing sophistication of available models and technologies [16–18]. Drones, when used at appropriate altitudes, can provide a high vantage point for observing wildlife with minimal disturbance, enabling the collection of valuable data [19].

Previous studies have applied various machine learning methods for wildlife detection and identification [13, 14, 20, 21]. Convolutional neural networks have been used to identify, count, and describe animals in camera-trap images, achieving 96.6% accuracy across 48 species [13]. In complex terrains, UAVs and crowd-sourced annotations have been utilized to detect animals in African savannas, highlighting the difficulties of identifying wildlife in varied landscapes [21]. However, key challenges remain in using drone images to quickly and efficiently distinguish wildlife from their intricate environments, such as dense forests or sprawling bushlands. Our study builds upon this previous work by specifically focusing on elephant identification and age classification in the diverse and challenging terrain of Kasungu National Park.

Our research aims to evaluate the effectiveness of widely used object identification machine learning models, specifically Faster R-CNN, RetinaNet, and Mask R-CNN, in precisely identifying elephants of different age groups across the challenging landscapes of African terrains. The study seeks to offer insights and a framework for technology-assisted wildlife conservation applicable in various geographic and ecological contexts. It aims to answer the following questions:

1. How effective are these machine learning models in identifying and monitoring elephants, including distinguishing between adult, juvenile, and infant elephants, from high-resolution drone imagery within intricate and complex environments like forests and bushlands?
2. What challenges and limitations exist in integrating drone technology and machine learning for wildlife conservation, especially when focusing on complex terrains?

By addressing these questions, our study aims to contribute to the development of more effective and efficient methods for elephant monitoring and conservation, potentially offering solutions to the challenges faced in Kasungu National Park and similar environments across Africa.

While this study focuses on elephant conservation in Kasungu National Park, the methodologies and insights gained have the potential to be adapted for other species and conservation areas worldwide, contributing to broader wildlife monitoring and protection efforts in various ecosystems.

2. Methodology

2.1. Image collection

Our study utilized a DJI Mavic 2 Pro drone equipped with a high-resolution 20 MP camera featuring a 1-inch CMOS sensor, oriented in a nadir position for consistent imagery across Kasungu National Park (KNP). Between October 2022 and September 2023, we collected 5,237 high-resolution images covering diverse landscapes at a flight altitude of 100 m to balance area coverage and minimize wildlife disturbance. Images were initially labeled 'DJI_####', averaging 15 MB in size, and uploaded to an AWS S3 bucket, then renamed using a Python script for efficient dataset management. Figure 2 provides a representative image from the collection.

The workflow of the study, encompassing the planning, execution, and data management stages, is illustrated in figure 3.

2.2. Preprocessing

Preprocessing involved normalization and resizing of images using a Python script within an AWS Sagemaker Jupyter Notebook environment. After analyzing various resolutions (256 × 256, 512 × 512, 1024 × 1024, and 2048 × 2048 pixels), we determined that 2048 × 2048 pixels provided optimal balance between computational load and image detail, as shown in figures 4(a)–(d). This resolution was key for maintaining the high-quality imagery needed to accurately identify different age groups of elephants (adults, juveniles, and infants), especially in challenging environments like dense forests. Moreover, we preserved the aspect ratio of each image during resizing.

We used AWS SageMaker Ground Truth for labeling, collaborating with experienced park rangers to ensure accurate identification of elephant age groups. We labeled 3,180 images, encompassing 3 classes including elephants categorized as adult (12 years and above), juvenile (4 to 12 years), and infant (under 4 years). The dataset was divided into training (70%), validation (15%), and test (15%) subsets, resulting in a class imbalance reflecting natural population distributions, with adult elephants being more represented than juveniles and infants.

Figure 2. Representative image from the dataset captured by a drone camera at an altitude of 100 m, showcasing the diverse terrain of Kasungu National Park.

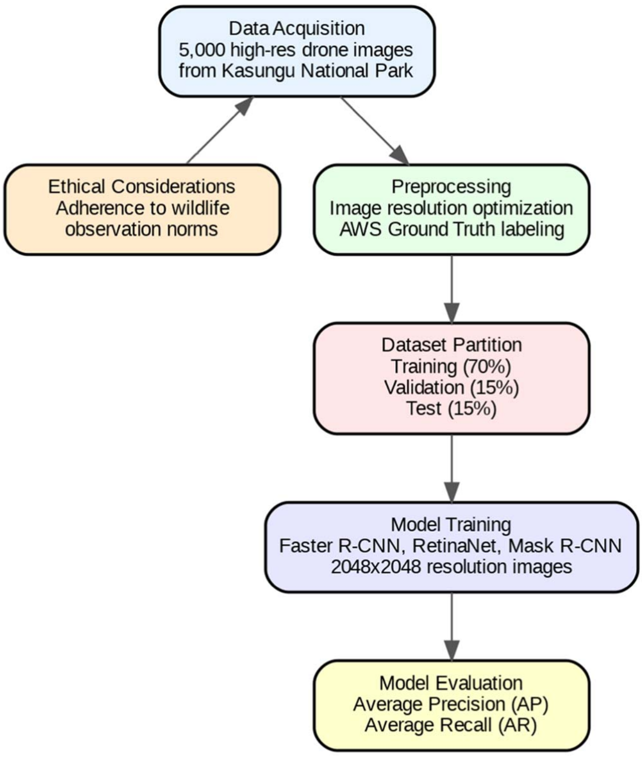


Figure 3. Flowchart illustrating the systematic methodology employed in the study, from drone image acquisition at Kasungu National Park to the model evaluation phase.

2.3. Analysis & model training

We evaluated three machine learning algorithms: Faster R-CNN, RetinaNet, and Mask R-CNN, chosen for their effectiveness in object detection and potential for adaptation to wildlife identification. The rationale for selecting each model is outlined in table 1.

To enhance the robustness of our models and address the challenges of limited data in wildlife contexts, we implemented several data augmentation techniques. During the training process, we applied augmentations with a 50% probability for each image. These augmentations included random brightness and contrast adjustments (0.9 to 1.1 factor), horizontal flips, rotations between −10 and 10 degrees, and random cropping to 80% of the original size. These augmentation techniques expanded our dataset's diversity without physically creating new images, which improved the models' ability to generalize across various conditions.

To address the class imbalance inherent in our dataset, where adult elephants were more represented than juveniles and infants, we implemented a balanced sampling approach using the RepeatFactorTrainingSampler provided by Detectron2. This sampler oversamples the minority classes (juvenile and infant elephants) during

(a)

(b)

(c)

(d)

Figure 4. Comparative analysis of image resolutions for elephant detection. (a) 256 × 256 pixel resolution. (b) 512 × 512 pixel resolution. (c) 1024 × 1024 pixel resolution. (d) 2048 × 2048 pixel resolution.

Table 1. Comparative justifications for selecting machine learning models in elephant detection.

Feature/Model	Convolutional neural network (faster R-CNN)	Single shot multibox detector (RetinaNet)	Mask R-CNN
Purpose of Choice	High accuracy in object detection, particularly effective in complex scenes	Modified for large image size (2048 × 2048 pixels)	Adapted for bounding box detection, facilitating direct comparison with other models
Strengths	Excellent in detecting subtle distinctions, crucial for differentiating between elephant age groups	Superior at detecting small objects, ideal for identifying smaller elephants in dense foliage	Exceptional at precise object delineation, beneficial for identifying individual animals and small features using bounding boxes
Image Processing Capability	Effective in diverse environments, including forests and bush	Enhanced capability for high-resolution images, crucial for complex terrains	Advanced capabilities in handling varied textures and patterns in wildlife imagery

training, effectively balancing the classes without physically duplicating the data. The repeat factor was calculated based on the frequency of each category, with a threshold set at 0.005. This approach ensured that the models received a more balanced representation of all classes during training, potentially improving their performance on underrepresented classes.

We standardized key parameters across all models to ensure a fair comparison. The batch size (IMS_PER_BATCH) was set to 4, accommodating the high-resolution images while working within memory constraints. We used a base learning rate (BASE_LR) of 0.001, which was decreased by a factor of 0.1 (GAMMA) at 13000 and 14500 iterations (STEPS). The total number of training iterations (MAX_ITER) was set to 15000 to allow for comprehensive learning given the complexity of our dataset.

For all models, we maintained the input image size at 2048 × 2048 pixels, allowing for detailed analysis of the complex scenes in our dataset. Uniform preprocessing techniques were applied to all training data, ensuring consistency across models. This included resizing all images to 2048 × 2048 pixels while maintaining aspect ratios, and normalizing pixel values.

Model assessment focused on Average Precision (AP) and Average Recall (AR) metrics, which provided insight into each model's classification accuracy across the different elephant age categories and overall performance. We used a custom COCO evaluator to compute per-category recall, which gave us a more detailed understanding of each model's performance on the different elephant age groups.

Table 2. Average precision (AP) for elephant detection models.

Model	Overall	Adults	Juveniles	Infants
Faster R-CNN	27.3	34.2	29.1	27.7
RetinaNet	29.8	35.5	31.4	28.8
Mask R-CNN (bbox)	25.8	33.8	30.9	29.6

Table 3. Average recall (AR) for elephant detection models.

Model	Overall	Adults	Juveniles	Infants
Faster R-CNN	37.7	44.3	40.8	39.5
RetinaNet	40.1	46.2	43.6	41.8
Mask R-CNN (bbox)	36.5	44.7	42.9	42.2

3. Results

Our study evaluated three machine learning algorithms—Faster R-CNN, RetinaNet, and Mask R-CNN—for identifying and classifying adult, juvenile, and infant elephants in the complex terrains of Kasungu National Park. Tables 2 and 3 provide a quantitative comparison of the models' performance in terms of Average Precision (AP) and Average Recall (AR), respectively.

Faster R-CNN showed moderate performance in elephant monitoring within complex environments. As shown in table 2, it achieved an AP of 34.2% for adult elephants, 29.1% for juveniles, and 27.7% for infants, with an overall AP of 27.3%. Table 3 indicates an overall AR of 37.7% for Faster R-CNN, with 44.3% for adults, 40.8% for juveniles, and 39.5% for infants. These results indicate some ability to distinguish between different elephant age groups, but with significant room for improvement.

RetinaNet demonstrated slightly better overall performance, with the highest overall AP of 29.8% and AR of 40.1% among the three models, as evidenced in tables 2 and 3. Table 2 shows that it achieved an AP of 35.5% for adult elephants, 31.4% for juveniles, and 28.8% for infants. The AR scores in table 3 further support RetinaNet's performance, with 46.2% for adults, 43.6% for juveniles, and 41.8% for infants. While these scores are the highest among the three models, they still indicate substantial challenges in accurately detecting and classifying elephants.

Mask R-CNN, with its bounding box configuration, showed comparable but slightly lower performance across all age categories. Table 2 indicates it had the lowest overall AP of 25.8%, with 33.8% for adults, 30.9% for juveniles, and the highest AP for infant elephants at 29.6%. As shown in table 3, Mask R-CNN's overall AR was 36.5%, with 44.7% for adults, 42.9% for juveniles, and 42.2% for infants.

The overall AP and AR scores across all models, as presented in tables 2 and 3, reflect the significant challenges posed by the complexity of the natural environments in our dataset and the task of distinguishing between three age categories of elephants. The results underscore the difficulty of applying object detection algorithms to wildlife monitoring in complex, real-world environments.

To illustrate the models' performance in different scenarios, we provide visual examples:

Figures 5(a) and (b) demonstrates Faster R-CNN's capability in processing high-resolution drone images. These images showcase the model's effectiveness in accurately identifying and classifying elephants, adeptly differentiating between various elephant age groups and individual elephants within the natural bushland setting.

Figures 6(a) and (b) showcases RetinaNet's performance in natural settings. It demonstrates the model's effectiveness in detecting and classifying elephants within a diverse forested environment. The zoomed-in view in figure 6(b) highlights the model's precision in differentiating individual elephants from anthropogenic factors, emphasizing its capacity to discern objects amidst dense vegetation and multifaceted environments.

Figures 7(a) and (b) provides insight into Mask R-CNN's proficiency in detailed segmentation. These images demonstrate the model's effectiveness in detecting and classifying elephants within environments of dense foliage, particularly its precision in identifying elephants of different age groups in areas where overlapping tree canopies add complexity to the visual scene.

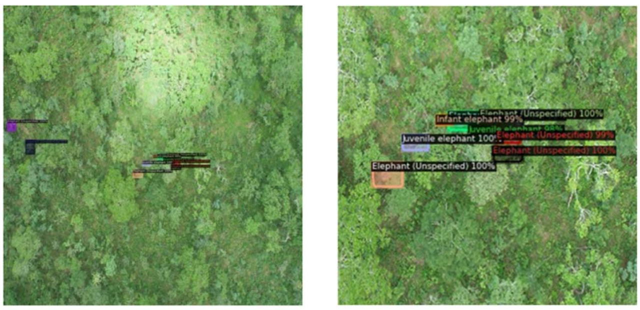
While each model demonstrates unique strengths, all showed room for improvement, especially in enhancing accuracy and reducing misclassifications within challenging terrains. These visual examples underscore both the potential and the current limitations of using machine learning models for elephant identification in complex African terrains.

(a) (b)

Figure 5. Elephant detection using Faster R-CNN model. (a) An example demonstrating the model's ability to accurately identify and classify elephants in a complex environment, using a standard 2048 × 2048 resolution image captured from a drone height of 100 m. (b) Zoomed-in view of (a) highlighting the model's precision in distinguishing between different age groups and individual elephants in an environment with typical African bushland.

(a) (b)

Figure 6. Elephant detection using RetinaNet model. (a) This image displays the model's ability to identify and classify elephants within a complex habitat, using a 2048 × 2048 resolution captured from a drone height of 100 m. (b) A zoomed-in view of (a), highlighting the RetinaNet model's precision in distinguishing between elephants in complex environments.



(a) (b)

Figure 7. Elephant detection using Mask R-CNN model. (a) This image illustrates the model's capability to accurately identify and classify elephants in diverse environments, using a standard 2048 × 2048 resolution captured from a drone height of 100 m. (b) A close-up from (a), this image exhibits the Mask R-CNN model's proficiency in detailed segmentation, particularly in distinguishing different elephant age groups amidst dense foliage.

The performance metrics and visual examples highlight the unique challenges posed by our specific task and dataset. Factors contributing to the observed performance include the complexity of the terrain, the inclusion of multiple classes beyond just elephants, and the difficulty in distinguishing between elephant age groups, especially in challenging visual conditions.

These findings emphasize the need for further refinement of the models, potentially through more extensive and diverse training data, advanced data augmentation techniques, and possibly the development of ensemble methods that combine the strengths of different models.

(a) (b)

Figure 8. Challenges in elephant detection using RetinaNet model. (a) An instance where the model produced a false positive, showcasing the challenges faced in complex environmental conditions. (b) A closer look at the same scene, emphasizing the difficulties the model encounters in distinguishing elephants from their surroundings due to the dense foliage, highlighting areas for model refinement.

4. Discussion

Our study explored the application of three machine learning models - Faster R-CNN, RetinaNet, and Mask R-CNN - for identifying and classifying elephants of various age groups in the complex terrains of Kasungu National Park. We implemented several enhancements to our methodology, including data augmentation techniques and a balanced sampling approach, to address the challenges of limited data and class imbalance inherent in wildlife conservation contexts. The results reveal both the potential and significant challenges of applying these technologies to wildlife conservation in intricate natural environments.

Our results, while promising, show lower accuracy compared to some previous wildlife detection studies. For instance, one study achieved 96.6% accuracy in identifying animals in camera-trap images across 48 species [14]. However, this study used stationary camera traps in more controlled environments, unlike our drone-based approach in complex, varied terrains. Another study, using UAVs for animal detection in African savannas, faced similar challenges to our study in terms of environmental complexity [21]. Their use of crowd-sourced annotations to improve detection accuracy could be a valuable approach for future iterations of our work. The performance of our models, particularly in detecting adult elephants (AP: 33.8%–35.5%), is encouraging given the added complexities of our study environment and the challenge of age group classification.

Each model demonstrated unique strengths in elephant detection but also faced considerable challenges. Faster R-CNN showed balanced performance across age categories, with an overall AP of 27.3% and an AR of 37.7%. It achieved an AP of 34.2% for adults, 29.1% for juveniles, and 27.7% for infants. RetinaNet achieved the highest overall AP (29.8%) and AR (40.1%), performing slightly better with adult elephants (AP: 35.5%) compared to juveniles (31.4%) and infants (28.8%). Mask R-CNN, while having the lowest overall AP (25.8%) and AR (36.5%), showed balanced performance across age categories and was particularly effective in identifying infant elephants (AP: 29.6%).

The overall performance metrics were lower than those typically seen in standard object detection tasks. This discrepancy highlights the unique challenges posed by our specific application. The diverse landscapes of Kasungu National Park, ranging from dense forests to open grasslands, present a more challenging environment for object detection than typical benchmark datasets. Our task of classifying elephants into three age categories (adult, juvenile, infant), while more representative of real-world scenarios, increased the complexity of the classification task. The subtle differences between elephant age groups, especially in challenging visual conditions, added an extra layer of difficulty not present in many object detection tasks.

These challenges are visually represented in figures 8(a) and (b), which illustrate instances where the models struggled with accurate identification and false positives:

Furthermore, as shown in figure 9, the models faced difficulties in detecting elephants from images captured at various camera angles, a challenge particularly relevant to drone-based studies:

Despite these challenges, our study represents a significant step forward in applying machine learning to wildlife conservation, particularly for elephant monitoring. The ability of these models to distinguish between different age groups of elephants, even with moderate accuracy, could provide valuable demographic data for conservation efforts. Currently, traditional elephant monitoring methods often rely on manual counts from aerial surveys or ground observations, which can be time-consuming, expensive, and potentially disruptive to

Figure 9. A scenario where the Faster R-CNN model failed to detect elephants in an image captured at a non-vertical angle. Moreover, the model also faced issues of misclassifying other classes, such as human settlements, and generating false positives by identifying objects that were not elephants, highlighting the need for a more diverse and dynamic training dataset.

the animals. Our approach, once refined, could offer a less invasive and more efficient method for regular population monitoring.

To address the challenges identified in this study and handle the intricacies of real-world scenarios, several avenues for future research and improvements are apparent. Advanced data augmentation techniques could artificially increase the representation of underrepresented classes, particularly juvenile and infant elephants. Incorporating more diverse camera angles in the training data would improve the models' robustness to varied perspectives, better mimicking real-world drone operations. Expanding the training dataset to include images from different seasons, times of day, and weather conditions would help the models better handle the variability encountered in real-world monitoring scenarios.

Transfer learning techniques, where models pre-trained on large datasets are fine-tuned for our specific task, could potentially improve performance, especially given our relatively small dataset. Exploring ensemble methods that combine the strengths of multiple models could also enhance overall performance. Developing specialized age-group classifiers as a second-stage process after initial elephant detection might improve age classification accuracy.

Incorporating temporal analysis by using sequences of images rather than individual frames could potentially improve detection accuracy and provide insights into elephant behavior and movement patterns. Developing context-aware detection algorithms that take into account the surrounding environment and typical elephant behavior patterns could improve detection accuracy and reduce false positives.

It's important to acknowledge the limitations of our current approach. While our implementation of data augmentation and balanced sampling techniques represents a step forward, our dataset may still not fully capture the full range of conditions and scenarios encountered in elephant monitoring. The high-resolution images necessary for accurate detection, combined with our augmentation techniques, require significant computational resources, which may limit real-time application in resource-constrained environments. The performance of our models is heavily dependent on image quality, which can be affected by factors like atmospheric conditions, drone stability, and camera specifications.

Current models lack the ability to understand broader contextual clues that human observers might use, such as animal behavior or group dynamics. While our models can detect and classify elephants by age group, they cannot yet reliably identify specific individuals, which is important for some conservation applications. The use of drones for wildlife monitoring, while less invasive than some traditional methods, still raises ethical questions about wildlife disturbance that need ongoing consideration.

In conclusion, while our study demonstrates the potential of integrating drone technology and machine learning for elephant monitoring, it also underscores the complexities involved in applying these technologies in challenging natural environments. The relatively low performance metrics should not be seen as a failure, but rather as a baseline for future improvements and a realistic assessment of the current state of this technology in complex, real-world scenarios.

This research contributes to the field by providing a detailed analysis of the challenges involved in elephant detection and classification in their natural habitats. It sets the stage for further refinements that could eventually lead to powerful new tools for wildlife conservation. As these technologies continue to evolve, they have the potential to revolutionize our approach to wildlife monitoring, offering new solutions to longstanding challenges in protecting elephants and other species in their natural habitats.

5. Conclusion

This research demonstrates both the potential and challenges of using advanced machine learning models—Faster R-CNN, RetinaNet, and Mask R-CNN—with drone technology for elephant monitoring in complex African habitats. The models showed moderate success in identifying elephants in complex environments, with overall AP scores ranging from 25.8% to 29.8%. RetinaNet demonstrated the highest overall performance (AP: 29.8%, AR: 40.1%), while Faster R-CNN and Mask R-CNN showed more balanced performance across age groups. However, all models faced significant challenges with accuracy, particularly in dense foliage and in classifying younger elephants, highlighting the complexity of real-world conservation applications.

Key challenges in integrating drone technology and machine learning for wildlife conservation in complex terrains include varied lighting conditions, diverse vegetation cover, and the need for high-resolution imagery to distinguish between age groups. The subtle differences between elephant age categories, especially in challenging visual conditions, add an extra layer of difficulty. Class imbalance in natural populations and computational demands of processing large, high-resolution datasets pose additional difficulties.

These findings provide a baseline for future improvements in wildlife monitoring technology. While the performance metrics are lower than those typically seen in standard object detection tasks, they represent a promising start given the complexity of the task. Refinements in training datasets, including more diverse environmental conditions and expanded representation of underrepresented classes, could potentially enhance wildlife monitoring significantly. This study contributes to elephant identification techniques and opens avenues for broader conservation applications across various species and habitats.

Acknowledgments

The authors would like to thank Kasungu National Park and the Department of National Parks and Wildlife Malawi for their assistance in facilitating this study. We also extend our gratitude to the Explorers Club for providing funding for this research. Additionally, we are grateful to Dr Buho Hoshino for his valuable input on our study and Yamikani Nyoni for his assistance in collecting the drone imagery.

Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Chris McCarthy https://orcid.org/0000-0001-5563-3583
Zuza Emmanuel Junior https://orcid.org/0000-0001-5706-8637

References

[1] Jones K R *et al* 2018 One-third of global protected land is under intense human pressure *Science* **360** 788–91
[2] Morton O, Scheffers B R, Haugaasen T and Edwards D P 2021 Impacts of wildlife trade on terrestrial biodiversity *Nat. Ecol. Evol.* **5** 540–8
[3] Williams D R, Clark M, Buchanan G M, Ficetola G F, Rondinini C and Tilman D 2020 Proactive conservation to prevent habitat losses to agricultural expansion *Nat. Sustain.* **4** 314–22
[4] Chase M J *et al* 2016 Continent-wide survey reveals massive decline in African savannah elephants *Peer J.* **4** e2354
[5] Schlossberg S, Chase M J, Gobush K S, Wasser S K and Lindsay K 2020 State-space models reveal a continuing elephant poaching problem in most of Africa *Sci. Rep.* **10** 10166
[6] Lindsey P A *et al* 2013 The bushmeat trade in African savannas: impacts, drivers, and possible solutions *Biological Conservation* **160** 80–96

[7] Naidoo R, Weaver L C, Diggle R W, Matongo G, Stuart-Hill G and Thouless C 2016 Complementary benefits of tourism and hunting to communal conservancies in Namibia *Conserv. Biol.* **30** 628–38
[8] Naidoo R, Fisher B, Manica A and Balmford A 2016 Estimating economic losses to tourism in Africa from the illegal killing of elephants *Nat. Commun.* **7** 13379
[9] Towards Measuring the Economic Value of Wildlife Watching Tourism in Africa - Briefing paper. Madrid: Organización Mundial del Turismo 2016
[10] Mpaka C 2022 Jumbo task as Malawi moves 263 elephants to restock a degraded national park [Internet] cited 2023 Nov 10]. Available from: https://news.mongabay.com/2022/08/jumbo-task-as-malawi-moves-263-elephants-to-restock-a-degraded-national-park/
[11] Pimm S L *et al* 2015 Emerging technologies to conserve biodiversity *Trends Ecol. Evol.* **30** 685–96
[12] Schmeller D S *et al* 2017 Building capacity in biodiversity monitoring at the global scale *Biodivers. Conserv.* **26** 2765–90
[13] Chen A, Jacob M, Shoshani G and Charter M 2023 Using computer vision, image analysis and UAVs for the automatic recognition and counting of common cranes (Grus grus) *J. Environ. Manage.* **328** 116948
[14] Norouzzadeh M S *et al* 2018 Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning *Proc Natl Acad Sci USA [Internet]* **115** E5716–E5725
[15] Mulero-Pázmány M, Stolper R, Van Essen L D, Negro J J and Sassen T 2014 Remotely piloted aircraft systems as a rhinoceros anti-poaching tool in Africa ed H Suleman *PLoS One* **9** e83873
[16] Chabot D and Bird D M 2015 Wildlife research and management methods in the 21st century: where do unmanned aircraft fit in? *J. Unmanned Veh. Sys.* **3** 137–55
[17] Mulero-Pázmány M, Barasona J Á, Acevedo P, Vicente J and Negro J J 2015 Unmanned aircraft systems complement biologging in spatial ecology studies *Ecology and Evolution* **5** 4808–18
[18] Anderson K and Gaston K J 2013 Lightweight unmanned aerial vehicles will revolutionize spatial ecology *Frontiers in Ecology and the Environment* **11** 138–46
[19] Hodgson J C *et al* 2018 Drones count wildlife more accurately and precisely than humans ed N Yoccoz *Methods Ecol. Evol.* **9** 1160–7
[20] Tabak M A *et al* 2019 Machine learning to classify animal species in camera trap images: Applications in ecology ed T Photopoulou *Methods Ecol. Evol.* **10** 585–90
[21] Rey N, Volpi M, Joost S and Tuia D 2017 Detecting animals in African savanna with UAVs and the crowds *Remote Sens. Environ.* **200** 341–51