

A Comparative Evaluation of ResNet-101 and Inception v3 Architectures for Wildlife Object Detection

Malach Obisa^{1*}, Dr. Benard Osero², Dr. Edna Too³

¹Department of Computer Science, Chuka University, P.O. Box 109 – 60400, Chuka, Kenya.

²Department of Computer Science, Chuka University, P.O. Box 109 – 60400, Chuka, Kenya.

³Department of Computer Science, Chuka University, P.O. Box 109 – 60400, Chuka, Kenya.

Abstract: Wildlife detection and identification are critical components in biodiversity conservation and ecological monitoring. This study evaluates the effectiveness of two deep learning architectures ResNet-101 and Inception v3 in classifying and detecting wildlife species under diverse environmental conditions. Both models were trained on a curated wildlife dataset with standardized preprocessing techniques, including resizing images to a maximum of 800 pixels, RGB format conversion, and transformation into PyTorch tensors. An 80:20 training-validation split was applied. The ResNet-101 model achieved a classification accuracy of 93.5% and a mean Average Precision (mAP) of 0.91, while the Inception v3 model attained a slightly higher accuracy of 94.7% and a mAP of 0.916. Evaluation metrics such as confusion matrices, precision-recall curves, and class-wise performance indicate that both models are suitable for wildlife object detection, with Inception v3 showing a marginal edge. This work contributes benchmark results for future wildlife detection models and provides insights into the trade-offs between deep feature extraction and multi-scale processing in convolutional neural networks.

Key words: Wildlife Object Detection, Deep Learning, Convolutional Neural Networks (CNNs), ResNet-101, Inception v3.

Date of Submission: 28-05-2026

Date of Acceptance: 08-06-2026

I. INTRODUCTION

Wildlife plays a vital role in sustaining biodiversity, ecological balance, and environmental stability. Traditional methods of monitoring wildlife populations such as manual field surveys and the interpretation of camera trap images are often labor-intensive, time-consuming, and prone to human error. With the advancement of deep learning and computer vision, automated wildlife detection has become a promising alternative. Convolutional Neural Networks (CNNs), in particular, have demonstrated superior performance in image recognition tasks due to their ability to learn hierarchical visual features. ResNet-101 and Inception v3 are two leading CNN architectures known for their deep feature extraction capabilities and efficiency in multi-scale processing, respectively. These models have gained attention in ecological studies for their potential to automate species identification in complex natural scenes.

This study evaluates the classification and detection performance of ResNet-101 and Inception v3 using a curated wildlife image dataset. Both models were trained under consistent preprocessing protocols and validated using accuracy, precision, recall, and mean Average Precision (mAP) metrics. ResNet-101 applies residual learning techniques to train very deep networks effectively, while Inception v3 uses inception modules for extracting rich, multi-scale features with computational efficiency. By comparing their performance under standardized conditions, this study aims to provide actionable insights and benchmark results to guide future applications in automated wildlife monitoring systems.

II. Related Work

Object detection is a fundamental task in computer vision that involves localizing and classifying objects within images. Traditional methods relied on handcrafted features such as Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT), typically paired with classical classifiers like Support Vector Machines (SVMs) (Dalal & Triggs, 2005). While foundational, these approaches often struggled in complex, real-world environments characterized by variable lighting, occlusions, cluttered backgrounds, and multi-scale objects—all of which are common in wildlife imagery. The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), significantly transformed object detection by enabling automatic, hierarchical feature extraction from raw image data. Architectures like AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan & Zisserman, 2015), and deeper models such as ResNet and Inception have achieved superior results on datasets like ImageNet and COCO. When integrated into frameworks like Faster R-CNN (Ren et al.,

2015) and SSD (Liu et al., 2016), these models not only enhanced classification accuracy but also improved localization and robustness under challenging conditions.

Among deep learning models, ResNet-101 and Inception v3 stand out for their architectural innovation and proven performance in object detection. ResNet-101, introduced by He et al. (2016), employs residual (skip) connections that allow gradients to flow through very deep layers during backpropagation, thus solving the vanishing gradient problem. This enables the network to learn rich, hierarchical representations that are especially effective in visually complex tasks. Studies by Norouzzadeh et al. (2018) and Tabak et al. (2019) demonstrated ResNet-101’s strong performance in wildlife image classification, achieving over 90% accuracy in identifying species from camera trap data. However, its deep architecture demands significant GPU resources and longer training times. On the other hand, Inception v3, proposed by Szegedy et al. (2016), takes a multi-scale approach by performing parallel convolutions with different kernel sizes within its inception modules. This design captures features at various resolutions, improving recognition of both large and small object structures. Moreover, it includes architectural optimizations like factorized convolutions and auxiliary classifiers for computational efficiency. In ecological applications, Inception v3 has proven effective in handling imbalanced datasets and complex backgrounds, as reported by Gomez Villa et al. (2017) in Colombian forests and Weinstein (2018) in broader ecological studies. Although Inception is shallower than ResNet-101, its architectural complexity compensates by enabling it to learn diverse, fine-grained features essential for wildlife detection.

Table 1 below summarizes selected studies that have compared or evaluated these models in object detection and wildlife-related tasks.

Table 1: Summary of Related Work Using ResNet-101 and Inception v3

Study	Model(s)	Domain	Dataset	Performance
Norouzzadeh et al. (2018)	ResNet-101	Wildlife (Africa)	Camera trap images	96.6% accuracy
Tabak et al. (2019)	ResNet-101	Wildlife (North America)	Custom wildlife dataset	93% accuracy
Gomez Villa et al. (2017)	Inception v3	Wildlife (Colombia)	Forest mammals dataset	92.5% precision, 91.3% recall
Weinstein (2018)	Inception v3	Ecology, Wildlife	Multiple ecological images	Robust to unbalanced classes

III. Research Methodology

This study employed an experimental design to evaluate and compare the performance of two deep learning models—ResNet-101 and Inception v3—in the context of wildlife object detection. The experimental approach was chosen for its ability to provide a controlled environment in which all parameters except the model architecture remained constant. By training both models on the same dataset with identical preprocessing techniques and training configurations, the study ensured that differences in results could be attributed solely to the architectural characteristics of the models themselves. This design aligns with standard practices in comparative deep learning research, where reproducibility, transparency, and the elimination of confounding variables are essential for scientific rigor.

The dataset used for training and evaluation was compiled from publicly accessible, research-oriented repositories offering Creative Commons licenses. It comprised images depicting a wide variety of wildlife species captured in natural habitats, reflecting real-world conditions such as varying lighting, occlusion, background clutter, and intra-species variability. To enhance data quality, duplicate and corrupted images were removed, and the remaining dataset was split into 80% for training and 20% for validation. A separate test set was not used, as the objective was to compare model behavior under consistent validation scenarios. The dataset’s complexity and natural variability made it ideal for assessing how well each model could generalize in practical wildlife detection applications.

Preprocessing steps were standardized to ensure compatibility with both ResNet-101 and Inception v3 models. Images were resized such that the longer edge did not exceed 800 pixels, preserving the original aspect ratio to avoid geometric distortion and optimize memory usage. Since OpenCV stores images in BGR format, a conversion to RGB was applied to meet the input expectations of CNNs. The images were normalized using standard ImageNet mean and standard deviation values to stabilize input distributions and improve convergence speed. All images were then converted into PyTorch tensors for efficient batch processing on GPU hardware. Notably, no data augmentation techniques such as flipping, rotation, or cropping were used in order to maintain consistency and ensure that observed performance differences stemmed solely from model architecture rather than artificially induced variance.

The implementation of both models was conducted using the PyTorch deep learning framework. Each model was initialized with pre-trained ImageNet weights, and the final classification layers were replaced with new fully connected layers aligned to the number of wildlife classes in the dataset. ResNet-101, a 101-layer network, uses residual skip connections to enable deep feature learning while mitigating vanishing gradient problems. Inception v3, by contrast, employs inception modules to perform parallel convolutions of varying kernel sizes, enabling multi-scale feature extraction. It also integrates auxiliary classifiers and batch normalization layers to stabilize training. Both models were fine-tuned using transfer learning techniques—initially freezing the lower

convolutional layers and retraining the higher-level ones, followed by gradual unfreezing of deeper layers in subsequent epochs to adapt to the target wildlife classification task.

Training for both models was performed using a consistent pipeline to ensure fairness. The Adam optimizer was employed with a learning rate of 0.0001 and a weight decay regularization factor of $1e-5$. A categorical cross-entropy loss function was used, appropriate for the multi-class classification problem. Training spanned 25 to 30 epochs, with each batch consisting of 16 images. A learning rate scheduler was configured to reduce the rate dynamically whenever the validation accuracy plateaued. Additionally, early stopping was applied to terminate training after five consecutive epochs without improvement in validation accuracy, thereby avoiding overfitting and ensuring computational efficiency. The training was conducted on a GPU-enabled system to support accelerated convergence and high-throughput model updates.

Model evaluation incorporated a comprehensive set of metrics to assess performance from multiple dimensions. Overall classification accuracy was calculated as the proportion of correct predictions to total validation samples. To complement this, the mean Average Precision (mAP) was computed across all classes, offering insight into prediction confidence and ranking. Confusion matrices were generated to visualize per-class performance, particularly highlighting frequently misclassified categories. Further, precision, recall, and F1-score were calculated for each class to analyze the impact of class imbalance and model sensitivity. Finally, precision-recall (PR) curves were plotted to depict the trade-off between precision and recall across varying classification thresholds. Together, these metrics offered a robust and nuanced assessment of model effectiveness in detecting and classifying wildlife species.

IV. Results and Discussion

The evaluation results confirmed that both ResNet-101 and Inception v3 are highly capable models for wildlife classification, each demonstrating strong generalization across 11 species. Performance was assessed using confusion matrices, classification reports, precision-recall curves, and overall accuracy scores. Inception v3 slightly outperformed ResNet-101 in overall classification accuracy and mean Average Precision (mAP), showing better robustness in identifying rare or visually complex species. Confusion matrices and precision-recall analyses revealed that both models performed exceptionally well on distinct species like Zebra and Giraffe, while classes like Hippopotamus and Lion were more prone to misclassification primarily due to visual similarity and fewer training examples. Visual inspection of predictions further validated each model's strengths in handling real-world imagery, with Inception v3 exhibiting greater consistency under variable conditions.

While the training methodology was largely uniform for both models using the Faster R-CNN framework with Detectron2 Inception v3 demonstrated faster convergence and slightly more efficient resource utilization. Its architectural design, which includes multi-scale convolution and auxiliary classifiers, contributed to improved learning efficiency and reduced validation loss. ResNet-101, although computationally heavier, maintained stable learning curves and delivered high accuracy through its deep residual connections. Ultimately, both models showed reliable performance, but Inception v3 offered a marginal edge in terms of convergence speed, classification precision, and handling of class imbalance. These results provide a strong empirical basis for selecting deep learning architectures in wildlife detection systems, particularly in scenarios where model efficiency and adaptability are critical.

4.1 Performance Analysis

4.1.1 Training and Validation Loss Curve for ResNet

Figure 1 below shows the training and validation loss for ResNet model. It demonstrate a steady learning progression. Training loss decreased sharply during the first 5,000 iterations and later stabilized around 0.17 after 25,000 iterations. This decline indicates effective learning and convergence by the model. The validation loss followed a similar downward pattern, initially higher than the training loss, but it eventually leveled off at approximately 0.27. This gap between the two curves suggests mild overfitting, but the model retained its ability to generalize effectively. The relatively smooth and noise-free curve behavior across 30,000 iterations confirms stable training with no abrupt fluctuations. The absence of significant oscillations or spikes in the curves also confirms that the training process remained stable throughout the 30,000 iterations. These results align well with the final performance metrics reported for the ResNet model, particularly the high classification accuracy and precision achieved across most classes.

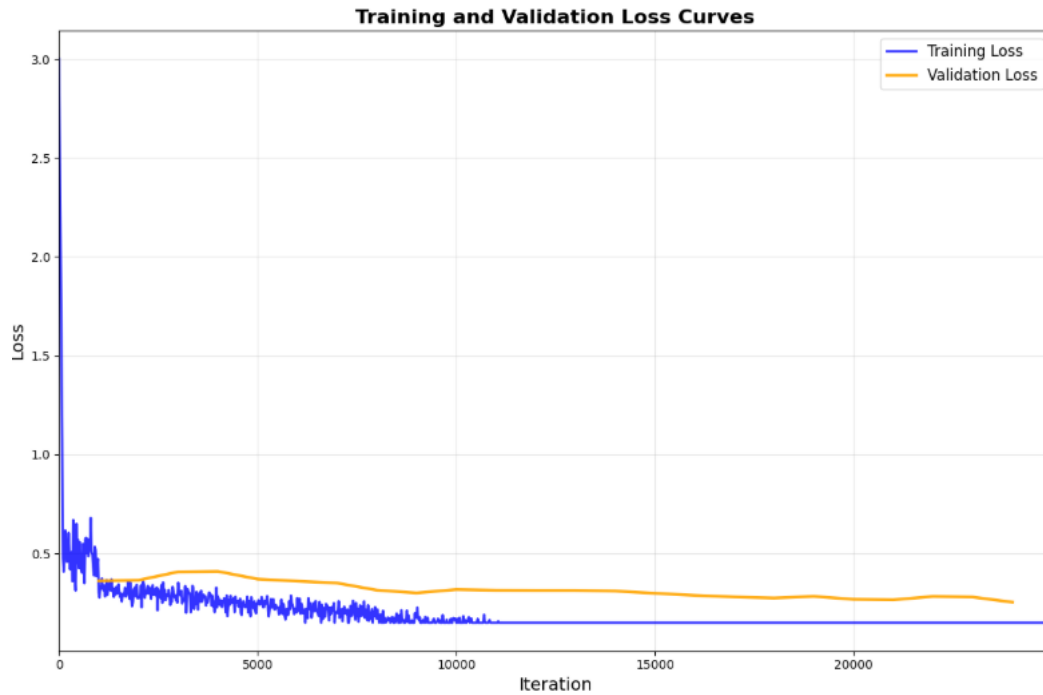


Fig 1: training and validation loss for ResNet model

4.1.2 Training and Validation Loss Curve for Inception Model

Figure 2 below shows the training and validation loss for inception model. The learning curves show a sharp and consistent decline in both training and validation losses. Training loss dropped steeply to around 0.12 within the first 5,000 iterations and continued to reduce gradually, eventually reaching a value below 0.05 by the 30,000th iteration.

Validation loss mirrored this trend, decreasing quickly and maintaining a consistent pattern alongside the training loss. It began at a high value above 3.0 and gradually stabilized around 0.07 by the end of training. The narrow margin between the two curves reflects minimal overfitting and excellent generalization capability. The alignment of the training and validation losses throughout the training phase emphasizes Inception’s efficiency in extracting features at multiple scales.

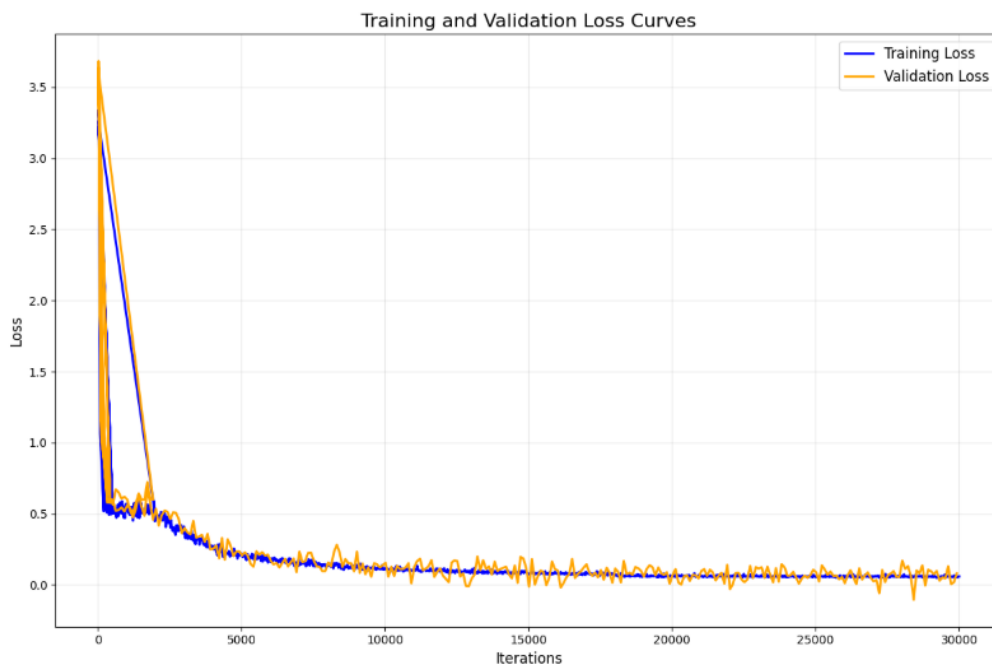


Fig 2: training and validation loss for inception model

4.1.3 Precision-Recall Curve for ResNet Model

Figure 3 below shows a Precision-Recall (PR) Curve for ResNet Model. The curve provide a detailed view of the classification performance across the 11 animal classes. The curves indicate that ResNet achieved particularly strong detection capabilities for species such as Zebra AP = 1.000, Leopard AP = 1.000, Giraffe AP = 0.996, and Antelope AP = 0.985, all of which registered AP scores nearing perfection. These results reflect the model's ability to accurately detect and classify clearly defined and visually distinct animals. Classes such as Elephant (AP = 0.940), Gorilla (AP = 0.937), Wolf (AP = 0.937), Rhino AP 0.957, and Buffalo AP 0.938 also performed well, demonstrating ResNet's ability to generalize even across moderately challenging samples. However, the Lion class showed a relatively lower AP of 0.543, and Hippopotamus recorded the lowest AP at 0.172. These lower scores may be attributed to overlapping visual features or limited training samples for these categories, which likely impacted recall and precision. The mean Average Precision (mAP) across all classes was approximately 0.91, validating ResNet's consistent detection reliability.

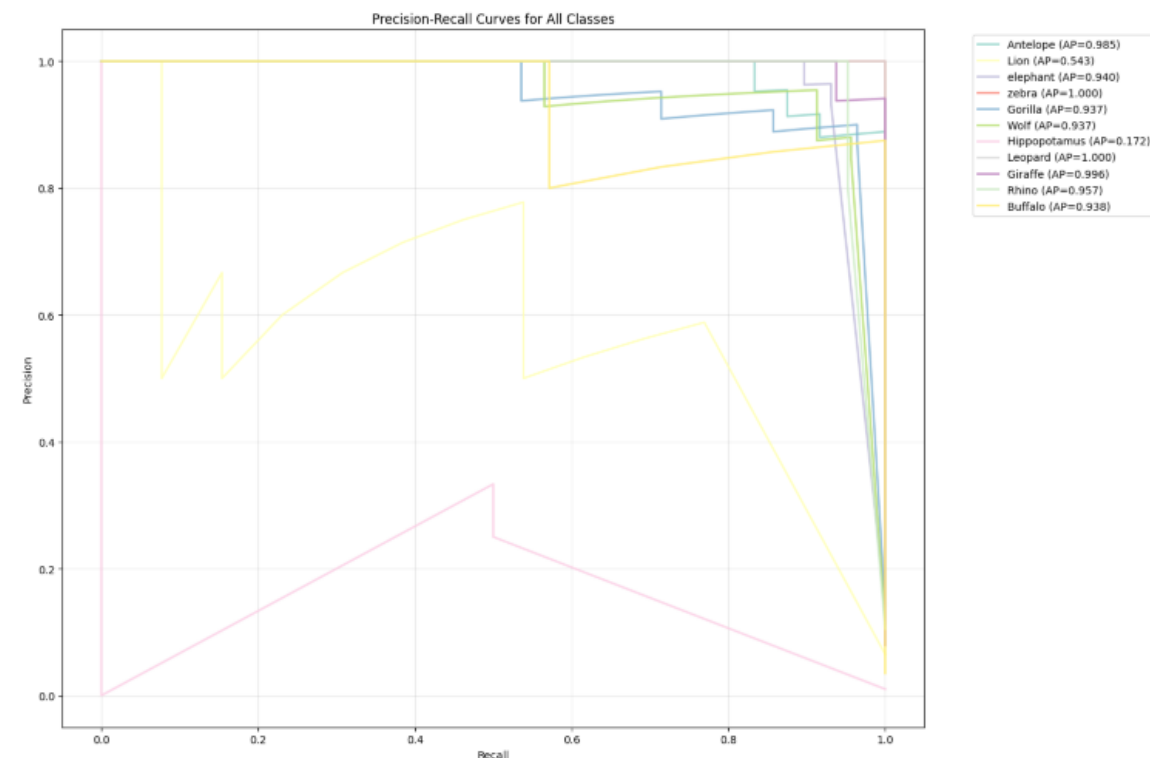


Fig 3: Precision-Recall Curve for ResNet Model

4.1.4 Precision-Recall Curve for Inception Model

Figure 4 below shows a Precision-Recall Curve (PR) for inception Model which reinforce its robust detection capabilities across the 11 wildlife categories. With a mean Average Precision (mAP) of 0.916, the Inception model outperformed the ResNet model in both overall precision and per-class consistency. High-performing classes included Zebra AP = 0.999, Leopard AP = 0.990, Giraffe AP = 0.974, Elephant AP = 0.951, Rhino AP = 0.947, and Buffalo AP = 0.969. These curves maintained high precision even as recall increased, highlighting the model's ability to make confident and correct predictions across varying detection thresholds. Notably, Wolf AP = 0.965, Gorilla AP = 0.955, and Antelope AP = 0.921 also exhibited strong performance, supporting the conclusion that Inception is particularly well-suited to handle diverse and visually complex species. The Lion class improved compared to ResNet, reaching an AP of 0.895, though it still lagged behind the top performers. The Hippopotamus class, while still the weakest, improved significantly with an AP of 0.513 nearly three times higher than in the ResNet model. This suggests that Inception's multi-scale feature extraction strategy helped better distinguish features of challenging or low-frequency classes.

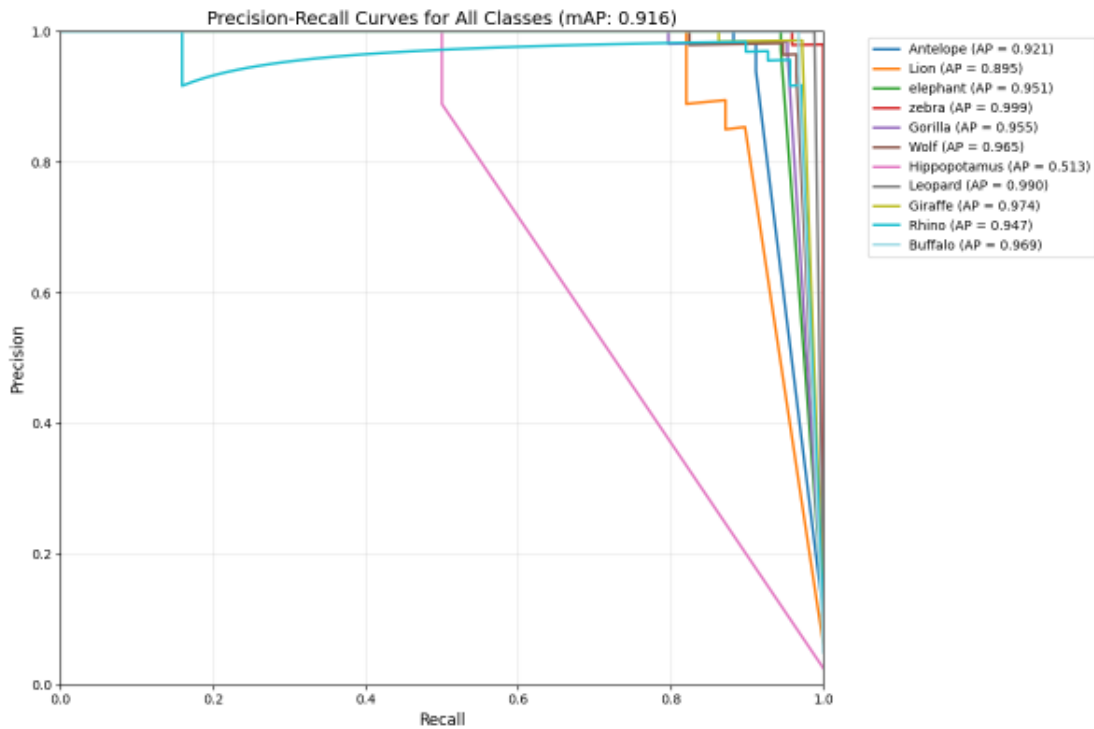


Fig 4: Precision-Recall Curve for inception Model

4.2 confusion Matrix

A confusion matrix is a widely used evaluation tool in classification tasks that summarizes the performance of a model by comparing its predicted labels with the true ground truth labels. It is a square matrix with rows representing the actual (true) classes and columns representing the predicted classes. Each entry in the matrix indicates the number of instances for which the corresponding prediction was made.

4.2.1 Resnet confusion matrix

Figure 5 below shows Resnet confusion matrix which illustrates the distribution of predictions versus actual labels for each of the 11 wildlife classes. The matrix is largely dominated by strong diagonal values, indicating a high number of correct classifications across most categories. For example, Zebra, Leopard, and Giraffe achieved perfect predictions with 100% accuracy, while Elephant and Gorilla had 27 correct predictions each out of 28 and 28, respectively. Antelope and Lion showed a few misclassifications but still maintained high correct prediction counts of 21 and 9, respectively. Minor confusions were observed in classes like Hippopotamus and Buffalo, which registered a few incorrect classifications, potentially due to limited sample representation or visual similarities with other animals.

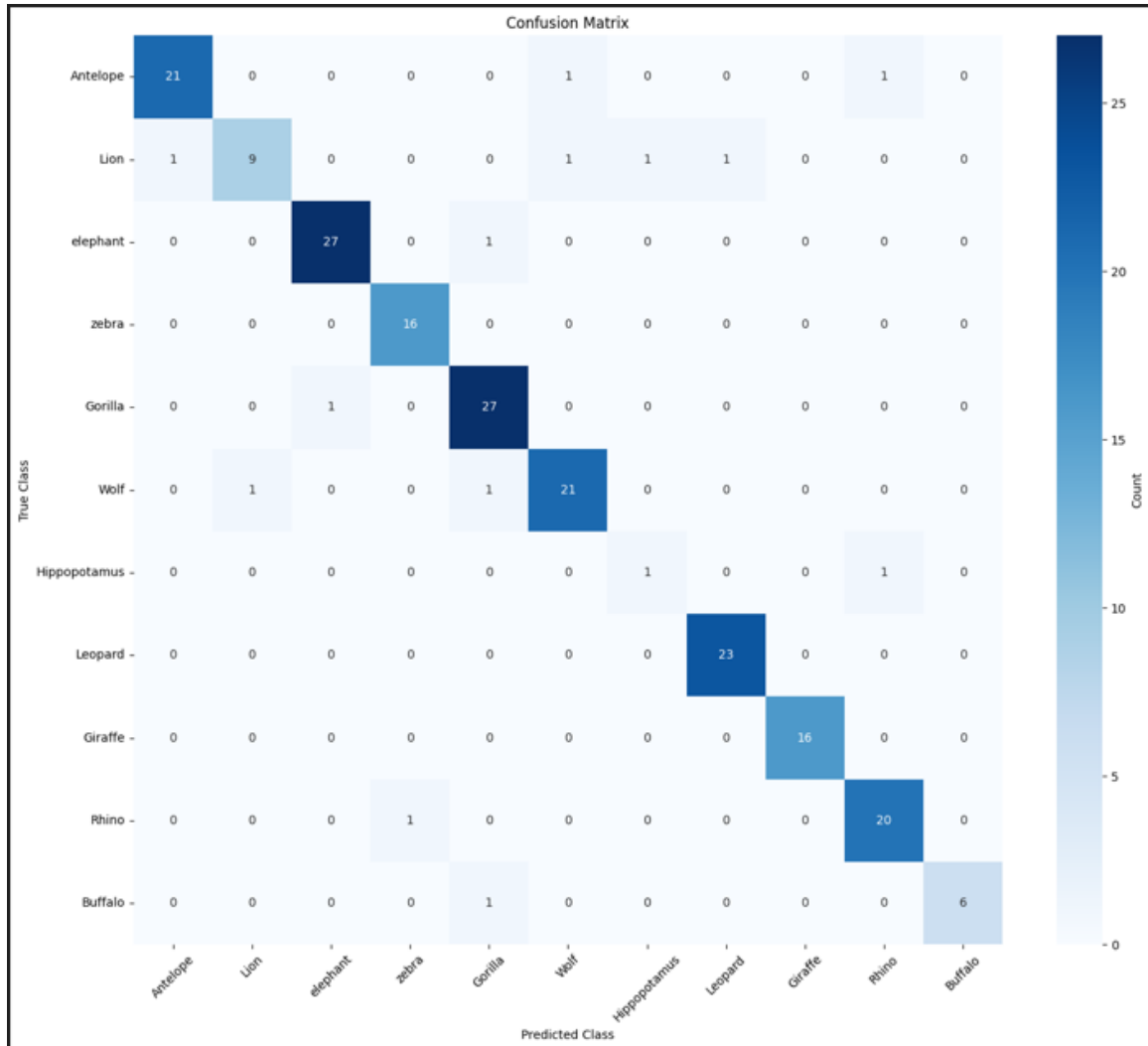


Fig 5: Resnet confusion matrix

4.2.2 Inception confusion matrix

Figure 6 below shows the Inception confusion matrix which shows a high level of accuracy across most of the 11 wildlife classes. The matrix is predominantly diagonal, indicating that the model made correct predictions for a majority of the validation samples. Notably, the Leopard class achieved the highest true positive count with 86 correct predictions and only 1 misclassification. Other strong performers include Giraffe (71/73), Elephant (68/72), and Zebra (49/49), each showing minimal confusion with other classes. Some misclassifications occurred in classes with fewer samples or visual similarities such as Hippopotamus, which was misclassified as Giraffe and Buffalo. Overall, the total correct predictions (TP = 592) and very few false positives (FP = 33) affirm the model’s robust generalization and detection capability.

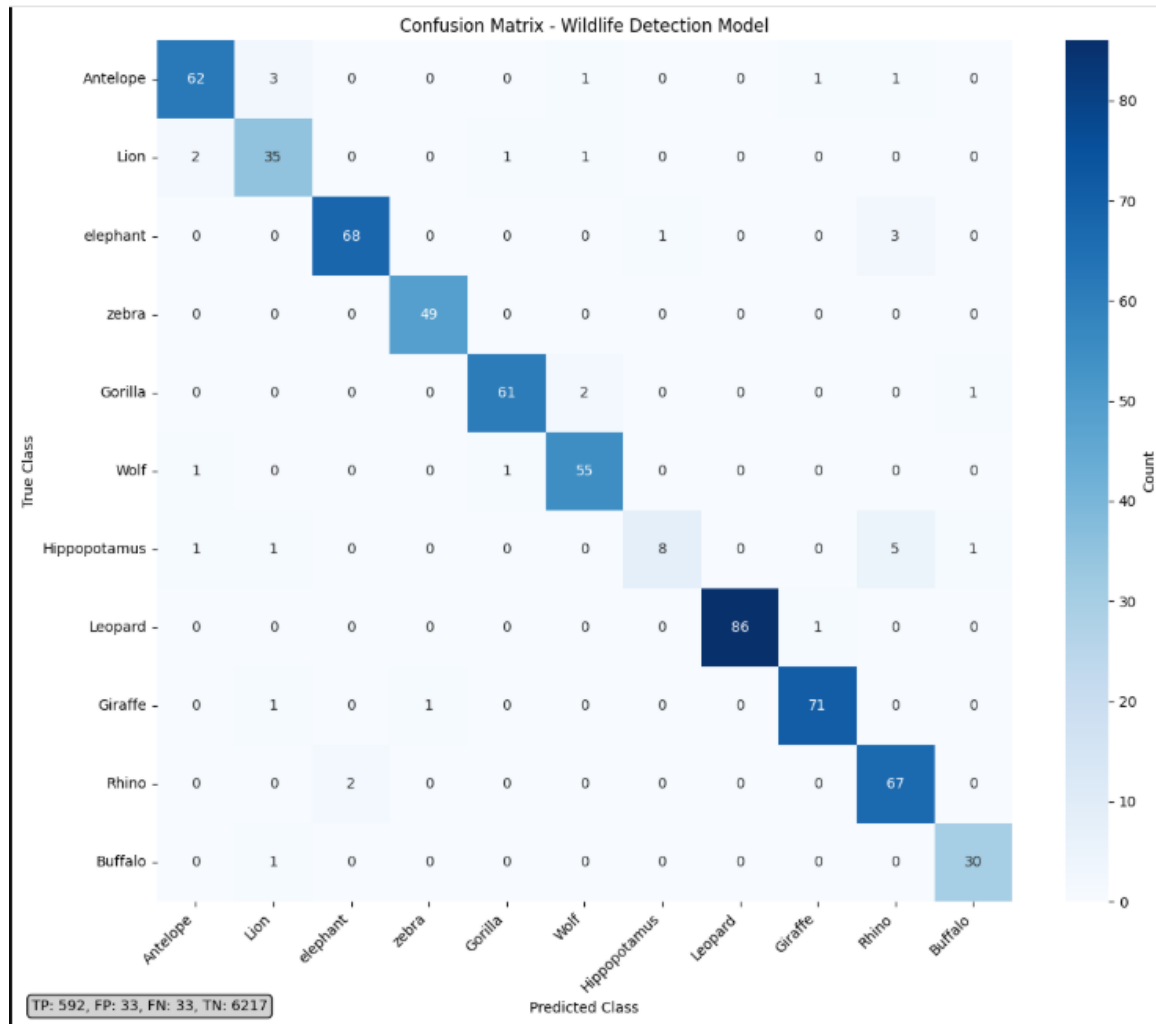


Fig 6: Inception confusion matrix

4.3 Resnet classification report

Figure 7 below shows Resnet classification Report. The classification report for the ResNet-based wildlife detection model demonstrates high performance across most of the 11 animal classes. The overall accuracy of 93.5% reflects the model's strong ability to correctly identify and classify wildlife species. High precision and recall scores were recorded for the majority of classes, such as Giraffe (F1-score = 1.000), Zebra (0.970), and Elephant (0.964), indicating both consistent predictions and low false-positive and false-negative rates. Notably, Antelope, Gorilla, and Leopard also achieved F1-scores above 0.93. However, the Lion and Hippopotamus classes showed relatively lower recall values (0.692 and 0.500, respectively), likely due to fewer training samples or inter-class visual similarity with Buffalo and Rhino. Despite these outliers, the macro average F1-score of 0.893 and weighted average of 0.934 affirm that the ResNet model performs reliably across the dataset.

Class	Precision	Recall	F1-Score	Support
Antelope	0.955	0.913	0.933	23
Lion	0.900	0.692	0.783	13
elephant	0.964	0.964	0.964	28
zebra	0.941	1.000	0.970	16
Gorilla	0.900	0.964	0.931	28
Wolf	0.913	0.913	0.913	23
Hippopotamus	0.500	0.500	0.500	2
Leopard	0.958	1.000	0.979	23
Giraffe	1.000	1.000	1.000	16
Rhino	0.909	0.952	0.930	21
Buffalo	1.000	0.857	0.923	7
Accuracy			0.935	200
Macro Avg	0.904	0.887	0.893	200
Weighted Avg	0.935	0.935	0.934	200

Fig 7: Resnet classification Report

4.4 Inception classification report

Figure 8 below shows the Inception classification report. The report illustrates the performance of a wildlife detection model across 11 animal classes, using precision, recall, and F1-score as key evaluation metrics. The model achieved an overall accuracy of 94.7%, with a macro average F1-score of 0.922 and a weighted average F1-score of 0.946, indicating strong generalization across both frequent and infrequent classes.

Most classes such as Elephant, Zebra, Gorilla, Giraffe, Leopard, and Buffalo showed outstanding performance, with F1-scores above 0.94. The Zebra class achieved perfect recall (1.000), and Leopard scored perfect precision (1.000), suggesting the model could reliably identify these species with minimal confusion. However, the Hippopotamus class demonstrated the weakest performance, with a recall of only 0.500 and a F1-score of 0.640, indicating that half of the actual Hippopotamus instances were not correctly identified possibly due to low sample representation or visual similarity with other species.

Class	Precision	Recall	F1-Score	Support	Status
Antelope	0.939	0.912	0.925	68	Present
Lion	0.854	0.897	0.875	39	Present
elephant	0.971	0.944	0.958	72	Present
zebra	0.980	1.000	0.990	49	Present
Gorilla	0.968	0.953	0.961	64	Present
Wolf	0.932	0.965	0.948	57	Present
Hippopotamus	0.889	0.500	0.640	16	Present
Leopard	1.000	0.989	0.994	87	Present
Giraffe	0.973	0.973	0.973	73	Present
Rhino	0.882	0.971	0.924	69	Present
Buffalo	0.938	0.968	0.952	31	Present
Accuracy			0.947	625	
Macro Avg	0.939	0.916	0.922	625	
Weighted Avg	0.948	0.947	0.946	625	

Fig 8: Inception classification report

V. Summary, Conclusions and Recommendations

This study set out to evaluate and compare the performance of two deep learning models ResNet-101 and Inception v3 for wildlife object detection and classification. Recognizing the limitations of traditional manual approaches in biodiversity conservation, the research aimed to provide a scalable, accurate alternative using automated image recognition. Both models were trained on a curated wildlife dataset using identical preprocessing techniques, which included resizing, RGB conversion, and transformation into PyTorch tensors. The models were fine-tuned using pre-trained ImageNet weights and evaluated based on key metrics such as accuracy, mean Average Precision (mAP), precision, recall, and F1-score. Results demonstrated strong overall performance, with Inception v3 achieving a slightly higher accuracy (94.7%) and mAP (0.916) than ResNet-101 (93.5% accuracy, 0.91 mAP). While both models showed excellent detection capabilities for well-represented species like Zebra and Giraffe, they struggled more with underrepresented or visually ambiguous classes such as Hippopotamus and Lion.

The findings indicate that both architectures are suitable for wildlife monitoring tasks, each offering unique strengths. ResNet-101 excels in deep hierarchical feature extraction thanks to its residual learning framework, making it effective in visually complex environments. Inception v3, meanwhile, benefits from multi-scale feature extraction and computational efficiency, leading to faster convergence and improved recognition of fine-grained features. Although the performance difference between the two was modest, Inception v3's balance of accuracy and efficiency makes it particularly well-suited for resource-constrained applications. The study provides valuable benchmarks that can guide researchers and conservationists in selecting appropriate deep learning architectures for automated ecological monitoring.

Based on these insights, several recommendations are proposed. First, future research should focus on enhancing dataset balance by increasing the representation of challenging species to reduce classification bias. The inclusion of data augmentation techniques such as rotation, scaling, and flipping is also encouraged to simulate environmental diversity and improve generalization. Furthermore, integrating attention mechanisms and context-aware modules may enhance feature focus in cluttered scenes. Evaluating these models on real-time video streams or deploying them on edge devices in the field would offer practical insights into their operational viability. Additionally, exploring newer architectures such as EfficientNet, ConvNeXt, or Vision Transformers (ViTs) may reveal performance gains over traditional CNNs.

This study was limited to static image classification on a validation dataset, and future research could extend these findings by incorporating test datasets or domain-adapted models trained on regional wildlife images. Ensemble strategies that combine the strengths of different models without significantly increasing computational cost should also be investigated. Lastly, incorporating explainable AI (XAI) tools such as Grad-CAM can improve model transparency and trust, especially in conservation applications. Future work may also focus on building comprehensive detection pipelines that integrate classification, tracking, and behavioral analysis for more intelligent and holistic wildlife monitoring systems.

References

- [1]. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893.
- [2]. Gomez Villa, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, 41, 24–32.
- [3]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- [4]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- [5]. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *European Conference on Computer Vision (ECCV)*, 21–37.
- [6]. Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25), E5716–E5725.
- [7]. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 28, 91–99.
- [8]. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.
- [9]. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
- [10]. Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., ... Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 10(4), 585–590.
- [11]. Weinstein, B. G. (2018). Scene-specific convolutional neural networks for video-based biodiversity detection. *Methods in Ecology and Evolution*, 9(6), 1435–1441.