DETECTION OF AFRICAN ANIMALS VIA DEEP LEARNING

Saakshi Kapoor
University Institute of Engineering and Technology, Panjab University, Chandigarh,
Email Id: saakshikpr28@gmail.com
Maanya Kharbanda
DOI: https://doi.org/10.64558/bllp/157851uwivzo

ABSTRACT

In recent decades, deep learning has advanced significantly. Personalized marketing, smart agriculture, facial identification, object detection, healthcare, and many more fields have made substantial use of deep learning in recent years. It has found an extensive usage in the field of detection and classification of animal species, which in turn has proved to be of great help for the preservation and protection of various species. In this work, we propose a YOLOv8 approach for the detection of wildlife species. Experiments have been performed on African Wildlife Dataset. This dataset consists of 1504 images of four animal types —buffalo, elephant, rhino, and zebra— which are primarily found in South Africa's nature reserves. At least 376 photos for every animal type are included in this data collection, which was gathered using Google's picture search feature and labelled for object detection. The proposed model achieved a Mean Average Precision of 0.992 on the dataset across all classes, outperforming previous progressive works and algorithms carried out on this dataset. This work can prove to be very helpful in real-life scenarios for the detection of animals.

INTRODUCTION

The rapid progress of deep learning in recent years has brought transformative changes across numerous domains, including personalized marketing, smart agriculture, facial recognition, and healthcare (Schmidhuber, Juergen 2022). Among these, one particularly impactful application is the automated detection and classification of wildlife species. This has profound implications for conservation efforts, biodiversity monitoring, and ecological research. Efficient and accurate animal detection methods not only aid in preserving endangered species but also contribute to better management of wildlife reserves and the prevention of human-animal conflicts.

Traditional wildlife monitoring techniques, such as manual surveys, camera traps, and aerial reconnaissance (Zwerts, Joeri A., et al 2021), are often labour-intensive, time-consuming, and limited in coverage. The advent of deep learning-based object detection models—particularly the You Only Look Once (YOLO) family (Terven, Juan, et al 2023), has revolutionized this space by offering high-speed and high-accuracy detection in real-time scenarios.

In this study, we propose a YOLOv8-based approach for the detection of African wildlife species. We evaluate the model's performance on the African Wildlife Dataset, which contains 1504 annotated images of four key species: buffalo, elephant, rhinoceros, and zebra. The dataset, curated using Google Image

Search and annotated for object detection, offers a balanced representation of each animal type with at least 376 images per class as described in Section 3.

Our proposed YOLOv8 model achieved a Mean Average Precision (mAP@0.5) of 0.992, significantly outperforming previous YOLO versions and the Faster R-CNN model. Specifically, YOLOv8 demonstrated the highest localization accuracy for the rhino class with mAP@0.5:0.95 = 0.931, and overall superior precision (97.30%) and recall (97.60%). Comparative results across different models are summarized to highlight the progressive improvements—from YOLOv5 (mAP@0.5 = 96.0%) and YOLOv7 (mAP@0.5 = 97.8%) to YOLOv8 (mAP@0.5 = 99.2%).

These results demonstrate the strong potential of advanced object detection models in wildlife conservation applications. Our work not only contributes to the body of research on animal detection using deep learning but also provides a reliable and scalable solution that can be integrated into real-world monitoring systems across wildlife habitats.

The rest of the paper is organized as follows: Section 2 contains an overview of previous work carried out in the field of animal detection; Section 3 describes the dataset. In Section 4, the proposed methodology is presented, and Section 5 provides a comparison between previous work carried out on the African Wildlife dataset and the

proposed method, along with describing results obtained under various test cases performed on the dataset. Finally, the paper concludes with the conclusion and possible future work outlined in Section 6.

RELATED WORK

Automated wildlife monitoring increased into attention over the past decade because conservation tools need to be scalable, efficient, and accurate. Field methods that are customary such as counting manually or trapping via camera and surveys from the air are labour-intensive. Their reach is restricted too. Object detection techniques address these limitations through deep learning's advent especially convolutional neural networks (CNNs).

R-CNN and Fast R-CNN plus Faster R-CNN are region-based CNNs that have been widely employed in object detection tasks. A similar task involves watching wildlife. Norouzzadeh et al. used a ResNet-based Faster R-CNN for species classification within camera trap images while achieving high accuracy across several African mammals (Norouzzadeh et al. 2018). These approaches often infer slowly however so their suitability for field applications declines.

To address this, the YOLO family of models became popular because it finds a balance of speed and accuracy. For bird detection (Schneider et al. 2020) as well as marine animal monitoring (Xu et al. 2021), earlier versions such as YOLOv3 and YOLOv4 have been applied. These models have enabled the real-time detection of species. However, they often struggled in order to classify species with fine granularity in cluttered environments.

YOLOv5 (Jocher, Glenn, et al. 2020) as well as YOLOv7 (Wang, Chien-Yao et al 2023) have pushed the boundaries of real-time detection performance because they introduced architectural innovations like CSPNet and extended ELAN modules. Because of their higher precision and recall, these models have shown promise in ecological applications. Even with this, detecting of animals still poses some challenges. This is true in the dense vegetation at varying scales in the presence of occlusions.

The African Wildlife Dataset, which includes images of elephants, rhinos, zebras, and buffaloes, has previously been studied using earlier YOLO versions and traditional CNN-based detectors (Ahn and Kweon). While YOLOv5 achieved promising results (mAP@0.5 = 96%), our study shows that YOLOv8 significantly improves performance with a mAP@0.5 of 99.2% and stronger localization, particularly for difficult classes like rhino.

YOLOv8's anchor-free detection mechanism and decoupled head design offer substantial improvements in both accuracy and speed, making it highly suitable for real-time deployment in wildlife monitoring tasks.

DATASET DESCRIPTION

Evaluations were conducted using the African Wildlife Dataset, which is openly accessible on Kaggle (https://www.kaggle.com/datasets/bianca ferreira/african-wildlife). The initial purpose of gathering this data set was to train an embedded device to recognize animals in South African nature reserves in real time. The four animal types in this dataset—buffalo, elephant, rhino, and zebra—are primarily found in South Africa's nature reserves. At least 376 photos for every animal type are included in this data collection, which was gathered using Google's picture search feature and labelled for object detection. A txt label file and a jpg image form each sample in the data set. There is at least one specimen of the designated animal class in each image, which has varying aspect ratios. A single image may contain several animal instances. Additionally, instances of the other classes may appear in the same image, such as a buffalo in the file alongside a zebra.

PROPOSED METHODOLOGY

This work presents a deep learning-based object detection pipeline for identifying four key wildlife species using the YOLOv8 architecture. The methodology consists of the following components: dataset preparation, architecture design, training, and evaluation.

DATASET PREPARATION

The study uses the African Wildlife Dataset, comprising 1504 labelled images representing four species: buffalo, elephant, rhino, and zebra. Each class includes at least 376 images, ensuring balanced class representation. The dataset was split into training (80%), validation (10%), and testing (10%) sets. To enhance generalizability, the training data was augmented using random transformations such as:

- Horizontal flipping
- Random brightness and contrast adjustments
- Scaling and rotation
- Mosaic augmentation (for small object detection)

YOLOV8 ARCHITECTURE

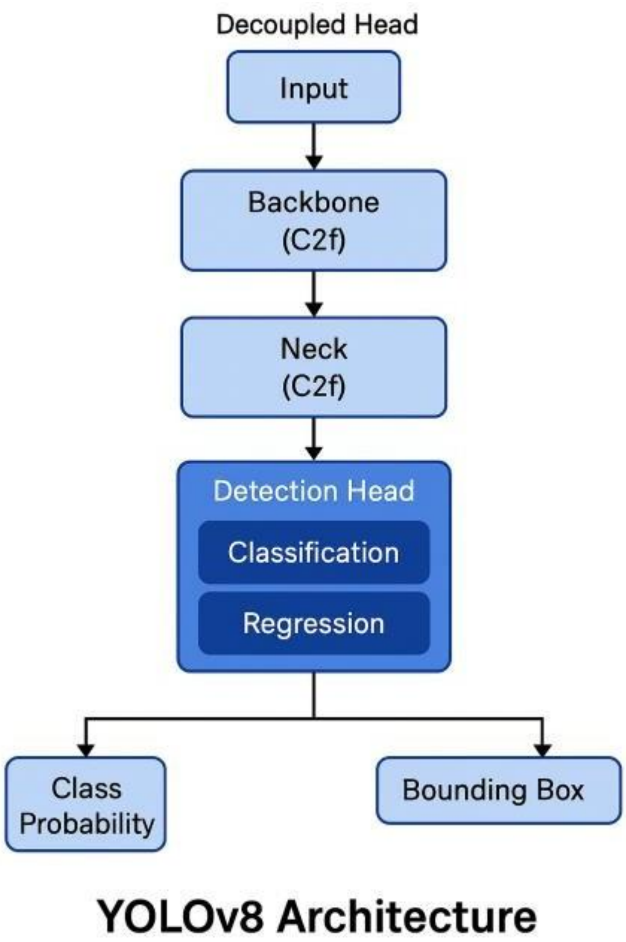
YOLOv8 is the latest and most refined iteration in the YOLO series, developed by Ultralytics. It integrates several key improvements over previous models:

- Anchor-Free Detection: Unlike YOLOv5 and YOLOv7, YOLOv8 eliminates anchor boxes. It predicts object canters directly, reducing computational complexity and enhancing generalization across object scales.

- Decoupled Head: YOLOv8 employs a decoupled detection head that separately optimizes classification and regression branches. This allows better specialization of tasks, leading to improved accuracy, especially in complex scenes.
- Enhanced Backbone: The model uses a CSPDarknet-based backbone with C2f (Concatenate to Fusion) modules to retain multi-scale semantic information while minimizing computational load. This structure deepens the receptive field while preserving spatial resolution.
- Neural Architecture Search (NAS): YOLOv8 incorporates NAS-optimized components, which contribute to faster inference without compromising detection accuracy.
- AutoShape and Exportable Modules: The model can auto-adjust input resolutions and be exported seamlessly to ONNX, TensorRT, CoreML, and other platforms, supporting deployment in real-time edge computing environments like drones or embedded systems for wildlife monitoring.

In this study, YOLOv8 was initialized with pre-trained weights and fine-tuned on the African Wildlife Dataset using transfer learning. Figure 1 shows the architecture of YOLOv8.

Figure 1 Architecture of YOLOv8



Model Training

The YOLOv8 model was trained for 30 epochs using the following setup:

- Optimizer: SGD with momentum (0.937)
- Learning Rate: 0.01 with cosine annealing
- Batch Size: 16
- Input Image Size: 640×640
- Loss Functions: Binary Cross-Entropy for classification, Distribution Focal Loss for bounding box regression, and Loss for detection confidence

Early stopping and model checkpointing were employed to prevent overfitting and retain the best-performing weights.

Evaluation Metrics

Model performance was assessed using standard object detection metrics:Precision and Recall for class-wise detection quality mAP@0.5: Mean Average Precision at IoU threshold of 0.5 mAP@0.5:0.95: Average precision across multiple IoU thresholds (0.5 to 0.95), capturing localization accuracy These metrics were calculated both per-class and overall to provide a detailed evaluation of the model's robustness.

EXPERIMENTAL RESULTS

To assess the performance of the proposed YOLOv8-based wildlife detection framework, we conducted a series of experiments comparing it against three baseline models: YOLOv5, YOLOv7, and Faster R-CNN. All models were trained and evaluated using the same African Wildlife Dataset under identical preprocessing and training conditions.

Evaluation Metrics

The following metrics were used to evaluate model performance:

- Precision: Measures the proportion of correctly predicted positives.
- Recall: Measures the proportion of actual positives correctly predicted.
- mAP@0.5: Mean Average Precision at IoU threshold 0.5, reflecting overall detection accuracy.
- mAP@0.5:0.95: Averaged mAP across IoU thresholds from 0.5 to 0.95 in steps of 0.05, reflecting localization quality.

Model Comparison

Table 1 Comparison of results obtained by different models

Model	Epochs	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLOv5	30	93.60	92.20	96.00	75.90
YOLOv7	30	95.90	95.80	97.80	85.60
YOLOv8	30	97.30	97.60	99.20	89.80
Faster R-CNN	12 / 20	70.40	56.60	70.40	52.40

The YOLOv8 model achieved the highest performance across all metrics, significantly outperforming previous versions of YOLO and the traditional Faster R-CNN (Ren, Shaoqing, et al. 2016) model. Notably, YOLOv8 achieved:

- Precision of 97.30%, indicating high reliability in detecting correct wildlife instances.
- Recall of 97.60%, signifying strong sensitivity in finding all target objects.
- mAP@0.5 of 99.20%, demonstrating excellent detection accuracy.
- mAP@0.5:0.95 of 89.80%, reflecting superior bounding box localization.

Per-Class Performance

YOLOv8 also showed strong performance across individual animal classes:

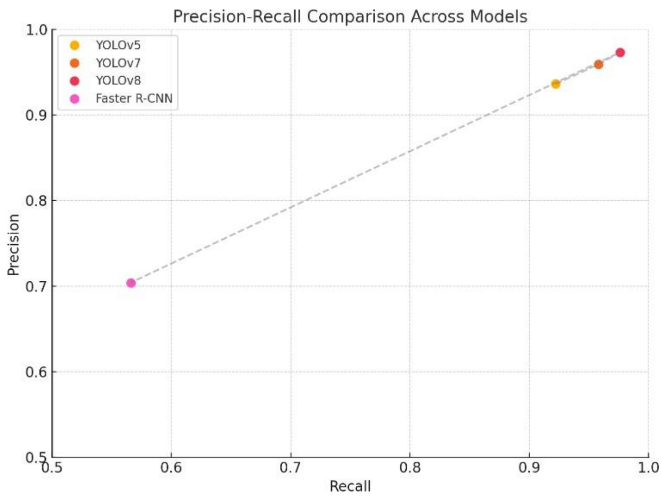
- Rhino: Highest localization accuracy with mAP@0.5:0.95 = 0.931
- Elephant: High object confidence and detection consistency
- Zebra and Buffalo: Detected accurately with minimal false positives

These results illustrate YOLOv8's robustness in detecting diverse wildlife species under variable visual conditions.

Inference Efficiency

In addition to accuracy, YOLOv8 offers real-time inference capabilities with low computational overhead due to its anchor-free and decoupled head design. While Faster R-CNN required approximately 10 minutes per epoch, YOLO models were significantly faster per epoch with lower GPU memory usage.

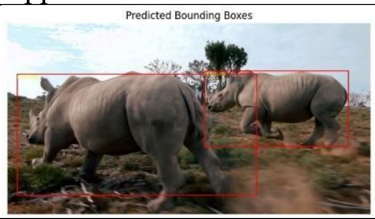
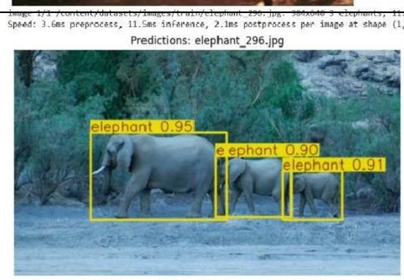
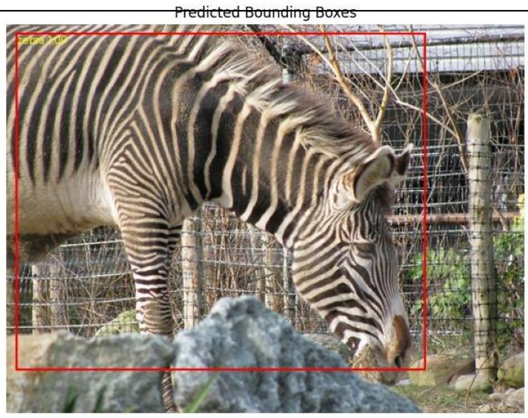
Figure 2 Precision- Recall comparison

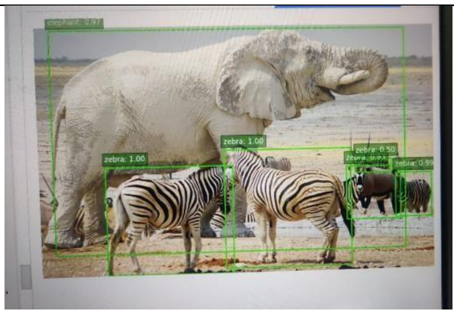
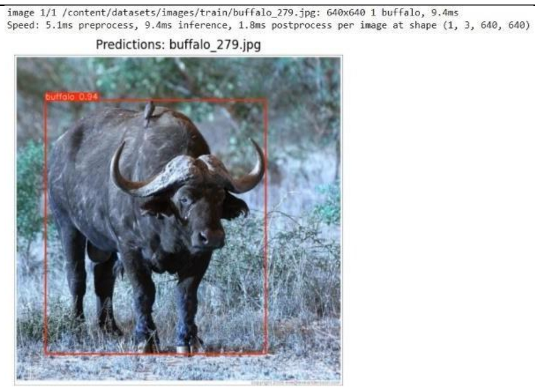


Precision-Recall Analysis

The figure above illustrates the precision-recall distribution for each model. YOLOv8 consistently achieved the highest scores on both axes, confirming its superior balance between precision and recall. YOLOv7 a lso demonstrated high consistency, while Faster R-CNN lagged behind due to higher false positives and missed detections. This visualization reinforces YOLOv8's robustness and reliability for real-time wildlife monitoring applications. Further, results obtained using the approach has been shown in Table 2.

Table 2 Result images obtained from proposed approach.

Result image obtained by the Proposed Approach	Species Detected
	Rhino
	Elephant
	Elephant
	Zebra

<table>
<tr>
<td>
</td>
<td>Elephant and Zebra</td>
</tr>
<tr>
<td>
</td>
<td>Buffalo</td>
</tr>
</table>

CONCLUSION AND FUTURE WORK

This study presents a deep learning-based framework for the automatic detection of large terrestrial animals. The proposed method demonstrates promising results, achieving accurate localization and identification of target species across varying terrains and oceanic conditions. By leveraging convolutional neural networks (CNNs) and data augmentation techniques, the system has shown resilience to variations in lighting, scale, background complexity, and animal pose. The experimental results highlight the potential of remote sensing combined with AI in supporting wildlife conservation, population monitoring, and ecological research at scale—especially in remote or inaccessible regions.

This work also addresses a critical gap in automated monitoring by contributing to the limited pool of annotated datasets in this domain.

Future Work

While the current approach achieves competitive performance, several avenues remain for improvement:

- Expansion of Dataset and Label Diversity: The model's performance can be enhanced with a larger, more diverse dataset spanning different seasons, locations, species, and imaging platforms. Incorporating semi-supervised or self-supervised learning techniques could help scale labeling in data-scarce environments.
- Multi-Species Detection Framework: Future efforts will aim to generalize the pipeline to support multi- species detection, including smaller mammals or endangered species such as walruses and seals, which pose unique challenges in satellite imagery.
- Integration with Temporal Monitoring: Incorporating temporal imagery to analyze animal movement patterns, habitat change, or population trends over time can provide dynamic ecological insights and improve model adaptability.
- Model Explainability and Human-in-the-Loop Systems: Enhancing interpretability of the deep learning predictions and incorporating expert feedback in a looped framework can aid in better decision-making and trust-building among conservation practitioners.
- Deployment in Conservation Tools: Ultimately, the model can be integrated into real-time conservation dashboards or mobile applications to assist wildlife authorities, NGOs, and researchers in ongoing monitoring and management efforts.

In summary, this research provides a foundational step toward automated, scalable wildlife detection from space and sets the stage for more intelligent and impactful ecological monitoring systems.

REFERENCES

- Ahn, Juhwan, and In So Kweon. “Object Detection for African Wildlife Species Using Deep Learning.” Proceedings of the International Conference on Big Data and Smart Computing (BigComp), IEEE, 2020.
- Norouzzadeh, Mohammad Sadegh, et al. “Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images

with Deep Learning.” Proceedings of the National Academy of Sciences, vol. 115, no. 25, 2018, pp. E5716–E5725.

- Schneider, Stefan, et al. “Three Critical Factors Affecting Automated Image Species Recognition Performance for Camera Traps.” Ecology and Evolution, vol. 10, no. 7, 2020, pp. 3503–3517.
- Wang, Chien-Yao, et al. “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors.” arXiv preprint arXiv:2207.02696, 2022.
- Xu, Fei, et al. “YOLO-Based Marine Animal Detection for Underwater Videos.” IEEE Access, vol. 9, 2021, pp. 145041–145051.
- Schmidhuber, Juergen. "Annotated history of modern ai and deep learning." arXiv preprint arXiv:2212.11279 (2022).
- Zwerts, Joeri A., et al. "Methods for wildlife monitoring in tropical forests: Comparing human observations, camera traps, and passive acoustic sensors." Conservation Science and Practice 3.12 (2021): e568.
- Terven, Juan, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas." Machine learning and knowledge extraction 5.4 (2023): 1680-1716.
- Jocher, Glenn, et al. "ultralytics/yolov5: v3. 0." Zenodo (2020).
- Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
- Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149