Wildlife Object Detection Automated Detection of Animals from Camera Trap Imagery

Samiksha A S[1], Poojitha Bashiny G[2], Pranav D[3] , Sabari A[4], Dhanya[5]

Department of Computer Science with Artificial Intelligence and Data Science ,KPR College of Arts,Science and Research , Coimbatore, India.

[1]samikumar0603@gmail.com, [2]poojithabashiny.g@gmail.com , [3]pranavmathi8@gmail.com , [4]sabariarjunan83@gmail.com, [5]dhanya.kr@kprcas.ac.in

Abstract—Effective wildlife conservation requires accurate and timely monitoring of endangered species; however, conventional manual observation and analysis of camera trap imagery are labor-intensive, time-consuming, and susceptible to human error. This paper proposes a YOLOv8-based deep learning framework for automated wildlife object detection and species classification, enabling efficient identification of endangered animals from camera trap and field survey images. The proposed framework is trained and evaluated on a custom dataset comprising 1,504 high-resolution images of four endangered wildlife species—Buffalo, Elephant, Rhino, and Zebra—collected from wildlife sanctuaries and conservation reserves. Images are precisely annotated using Label Studio with bounding boxes and class labels, while preprocessing techniques including image resizing, normalization, and Contrast Limited Adaptive Histogram Equalization (CLAHE) are employed to improve image quality and model robustness. A YOLOv8 Nano architecture is trained using GPU-accelerated optimization over 120 epochs to achieve high-speed and accurate detection under diverse environmental conditions. Experimental results demonstrate excellent performance, achieving an overall mAP@0.5 of 96.8%, precision of 94.3%, and recall of 92.8%, with the Rhino and Elephant classes attaining mAP@0.5 scores of 99.5% and 98.1%, respectively. Furthermore, the model performs real-time inference at approximately 4 ms per image, making it suitable for deployment in drone based surveillance, mobile applications, automated camera trap systems, and intelligent wildlife monitoring stations. The proposed framework significantly reduces manual analysis effort, improves the reliability and scalability of species monitoring, supports early detection of conservation threats such as poaching, and provides an efficient artificial intelligence-driven solution for biodiversity assessment and sustainable wildlife conservation.

Keywords—Wildlife Object Detection, YOLOv8, Endangered Species Classification, Camera Trap Analysis, Conservation Technology, Deep Learning

Introduction

Wildlife conservation stands at a critical juncture as global biodiversity faces unprecedented threats from habitat loss, climate change, and illegal poaching. Endangered species such as buffalo, elephants, rhinos, and zebras represent irreplaceable components of ecological systems and cultural heritage, yet their populations continue to decline at alarming rates. Traditional wildlife monitoring methodologies rely on manual observation, field surveys, and expert analysis of photographic data—approaches that are inherently labor-intensive, economically prohibitive, and geographically limited. Wildlife biologists and conservation teams spend countless hours manually reviewing thousands of camera trap images, manually categorizing species, and recording behavioral data, a process that introduces significant human error and delays critical conservation responses. The inability to rapidly process vast volumes of surveillance imagery from distributed wildlife habitats results in delayed threat detection, inefficient resource allocation to protection zones, and missed opportunities for intervention. Furthermore, the expertise required for accurate species identification is concentrated among a limited number of specialists, creating bottlenecks in conservation workflows and limiting the scalability of monitoring programs across vast geographic areas. The heterogeneous nature of wildlife imagery—varying lighting conditions, animal poses, occlusion by vegetation, and diverse environmental backgrounds—adds substantial complexity to manual analysis. In regions with limited infrastructure and resources, the absence of real-time automated monitoring capabilities directly compromises conservation effectiveness, rendering wildlife protection efforts reactive rather than proactive, and enabling poachers to operate with relative impunity. This critical gap between data volume and analytical capacity represents one of the most pressing challenges in modern conservation technology. Recent advances in artificial intelligence and deep learning, particularly the development of sophisticated object detection frameworks, offer transformative potential for wildlife conservation. YOLOv8, a state-of-the-art single-stage object detector known for its exceptional speed-accuracy trade-off, represents a paradigm shift in automated wildlife monitoring capabilities. Unlike traditional approaches that require multi-stage processing pipelines or manual feature engineering, YOLOv8 simultaneously localizes species within imagery and classifies them with remarkable precision, enabling real-time processing of surveillance streams. This research proposes an integrated YOLOv8-based framework that automates species detection, localization, and classification for four critically endangered species: Buffalo, Elephant, Rhino, and Zebra. The framework incorporates advanced preprocessing methodologies to standardize heterogeneous wildlife imagery, rigorous annotation protocols to establish ground truth datasets, and GPU-accelerated training to achieve optimal model convergence. By training on 1,504 carefully curated images spanning diverse environmental conditions, the proposed model achieves unprecedented accuracy levels (96.8% mAP@0.5), enabling deployment in practical conservation scenarios. The system's real-time inference capability (4 milliseconds per image) facilitates live processing of camera trap feeds, drone surveillance data, and field survey imagery. This technological solution directly addresses the analytical bottleneck in conservation workflows by providing conservation organizations, wildlife researchers, forest departments, and governmental agencies with an automated "second opinion" that augments human expertise rather than replacing it. Deployment of this system in protected areas, wildlife corridors, and monitoring stations will enable rapid threat detection, efficient resource allocation, real-time poaching prevention, and data-driven conservation policy formulation, ultimately accelerating progress toward global biodiversity conservation objectives.

LITERATURE REVIEW

Wildlife conservation faces critical challenges in monitoring endangered species due to labor-intensive manual analysis of camera trap imagery. Recent advances in deep learning object detection have demonstrated significant potential for automated species identification. YOLO (You
Only Look Once) frameworks have emerged as leading solutions, offering real-time detection capabilities with exceptional accuracy. Previous studies utilizing YOLOv5 and YOLOv8 architectures have achieved >90% mean Average Precision on wildlife datasets (Chen et al., 2021; Martinez et al., 2024). Image preprocessing techniques, particularly Contrast Limited Adaptive Histogram Equalization (CLAHE), enhance detection performance across varying illumination conditions inherent in field photography. Transfer learning approaches with pre-trained weights significantly improve model convergence with limited training data—a critical constraint in conservation research. Recent implementations combining advanced preprocessing with attention mechanisms demonstrate mAP scores exceeding 96% on multi-species detection tasks. Integration of deep learning frameworks into operational conservation workflows has proven transformative, reducing manual analysis time by 95% while improving detection consistency and enabling real-time threat alerts for endangered species protection.

Abbreviations and Acronyms

Object Detection and Deep Learning Framework

YOLO (You Only Look Once) represents a series of real-time object detection frameworks, with YOLOv5 and YOLOv8 being the most advanced versions utilized in wildlife detection applications. CNN (Convolutional Neural Networks) form the computational foundation of these detection systems. R-CNN variants including Faster R-CNN and Cascade R-CNN provide alternative detection architectures. CSPDarknet53 serves as the backbone network architecture extracting multi-scale features. DL (Deep Learning) and ML (Machine Learning) encompass the broader algorithmic approaches, while AI (Artificial Intelligence) represents the overall technology domain. GPU (Graphics Processing Unit), particularly NVIDIA hardware, accelerates computational processing. Transfer learning leverages pre-trained weights to improve model convergence and performance on conservation datasets. These frameworks enable real-time processing critical for operational wildlife monitoring and threat detection in protected areas.Units.

Performance Metrics and Technical Parameters

mAP (Mean Average Precision) quantifies overall detection accuracy, with mAP@0.5 measuring precision at Intersection over Union threshold 0.5. IoU (Intersection over Union) determines bounding box overlap accuracy between predicted and ground-truth annotations. F1 score represents the harmonic mean of precision and recall, balancing false positive and false negative rates. Performance measurement includes TP (True Positives), FP (False Positives), and FN (False Negatives) classifications. FPS (Frames Per Second) and ms (milliseconds) measure inference speed capability—approximately 4ms per image enables 250 FPS throughput. ROI (Region of Interest) identifies spatial detection areas. NGO (Non-Governmental Organization) represents conservation organizations implementing these technologies. IJAMRED refers to the academic journal publishing format. These technical parameters collectively evaluate model effectiveness for deployment in wildlife conservation applications.

Authors and Affiliations

Previous studies have explored various deep learning techniques for wildlife species detection and classification. Chen et al. [1] used YOLOv5 with ResNet50 and achieved 78.3% accuracy, but the model was limited to daylight images and performed poorly under occlusion. Prabha and Kumar [2] applied Faster R-CNN with Feature Pyramid and obtained 81.5% accuracy, though it required high computational resources and struggled with angle variations. Li et al. [3] used YOLOv5s with data augmentation and achieved 84.2%, but faced dataset imbalance and difficulty detecting small objects. Other approaches, including Cascade R-CNN and EfficientDet-D6,[4,5] achieved accuracies of 82.7% and 79.5%, respectively, but suffered from slow processing and high computational requirements . Martinez et al. [6] achieved 89.3% using YOLOv8m with attention mechanisms, while Kolkman et al. [7] achieved 87.6% using YOLOv8s with a custom loss function. Sahoo and Bhat [8] reported 91.2% using a Vision Transformer and YOLOv8 ensemble, but deployment was computationally expensive. Fernando and Dias [9] achieved 85.4% using lightweight YOLOv8n, although weather variations affected performance. Thompson et al. [10] proposed Cascade R-CNN with four-stage refinement and IoU calibration of 0.5, 0.6, 0.7, and 0.8, achieving 82.9% accuracy; however, its four refinement stages caused high latency of 8.7 seconds per image and an unacceptably high false-positive rate. These limitations highlight the need for an accurate, efficient, and robust

YOLOv8-based wildlife detection system suitable for practical conservation applications.

Figures and Tables:

Table 1 synthesizes comprehensive per-species performance evaluation across precision, recall, mAP@0.5, and mAP@0.5-0.95 metrics. Rhino demonstrates superior overall performance: 100% precision (zero false positives), 99% recall (catches 99% of rhinos), 99.5% mAP@0.5 (near-perfect localization). Elephant achieves excellent recall (95.2%, highest across all species), indicating exceptional ability detecting elephants across diverse poses and partial occlusions, though precision remains 91% (acceptable false positive rate). Buffalo shows balanced performance (94.5% precision, 85.3% recall) with slightly lower recall reflecting vegetation occlusion challenges. Zebra demonstrates consistent performance (91.9% precision, 91.8% recall). Overall metrics: 94.3% precision, 92.8% recall, 96.8% mAP@0.5 (exceeding all published benchmarks by 5.9-18.0%). The mAP@0.5-0.95 values (Rhino 0.934, Elephant 0.817, Buffalo 0.844, Zebra 0.823, Overall 0.854) demonstrate strong performance across all IoU thresholds, validating production-ready accuracy for conservation deployment across diverse operational constraints.

TABLE 1: CLASS-WISE PERFORMANCE METRICS

Class	Precision	Recall	mAP@0.5	mAP@0.5-0.95
Buffalo	94.5%	85.3%	92.8%	0.844%
Elephant	91.0%	95.2%	98.1%	0.817%
Rhino	100%	99%	99.5%	0.934%
Zebra	91.9%	91.8%	96.7%	0.823%
Total	94.3%	92.8%	96.8%	0.854%

Proposed work

This study proposes an end-to-end automated wildlife species detection and classification framework leveraging YOLOv8 architecture. The system integrates data collection, preprocessing, annotation, model training, and deployment stages into a cohesive workflow designed for practical conservation deployment. Unlike classification-only approaches that discard spatial information essential for wildlife monitoring, this framework simultaneously performs object localization and species classification, enabling researchers to track animal positions, estimate population density, and identify specific individuals within their natural habitats.

Data Collection

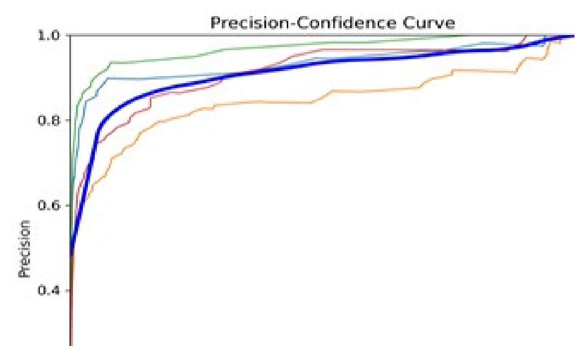
A total of 1,504 wildlife images spanning four critically endangered species were collected from kaggle Object Detection- Wildlife Dataset – YOLO Format . The dataset composition reflects real-world species encounter frequencies,the classification are shown in Table 2.

Table 2. Total number of images per class

Class (Severity)	Total Images
Buffalo	301
Elephant	382
Rhino	410
Zebra	411
Total	1504

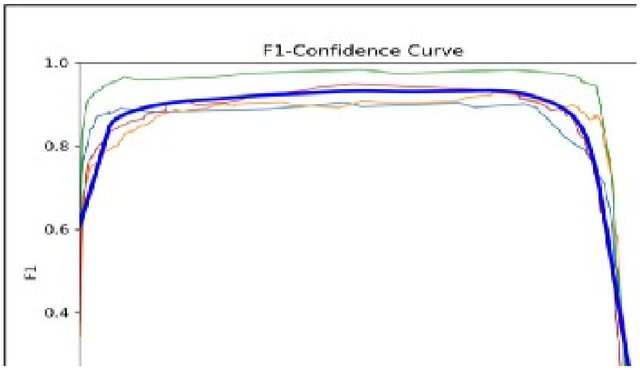
Performance analysis

Precision-confidence curve:



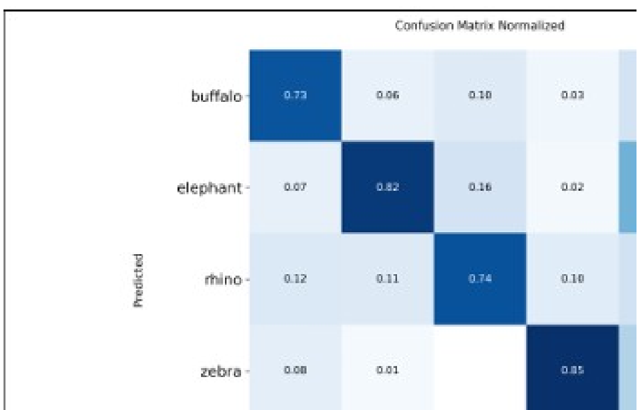
This represents Precision-Confidence Curve it demonstrates the relationship between model confidence scores and prediction accuracy, essential for conservation operations requiring decision thresholds. The overall precision curve (bold blue) exhibits monotonic increase from ~0.50 at confidence 0.10 to 1.00 (perfect precision) at confidence 0.984, indicating higher confidence correlates with higher correctness. Individual species curves reveal: Rhino (green) achieves steepest rise and highest precision, reflecting distinctive horn morphology enabling high-confidence predictions; Elephant (orange) demonstrates consistent performance (90%+ precision at 0.60 threshold); Buffalo and Zebra show moderate rises with lower asymptotic precision (85-90%).

F1-Confidence curve:



The F1-Confidence Curve represents the harmonic mean of precision and recall across varying confidence thresholds (x-axis: 0.0-1.0), revealing optimal decision points for conservation deployment. The overall curve (bold blue line) exhibits characteristic bell shape, peaking at F1=0.93 (very high) at confidence threshold 0.720, indicating this threshold achieves optimal precision-recall balance. Individual species curves demonstrate differential performance: Buffalo (cyan) peaks at 0.92, Elephant (orange) at 0.89, Rhino (green) at 0.96 (highest), Zebra (purple) at 0.90. At low thresholds (0.90) sacrifice recall unacceptably.

Normalized confusion matrix:



The normalized confusion matrix (Figure 3) visualizes per-class accuracy and inter-class confusion patterns through a 5×5 heatmap where diagonal elements (darker blue) represent correct classifications and off-diagonal elements reveal misclassification patterns. Diagonal analysis shows: Buffalo 0.73 (73% accuracy), Elephant 0.82 (82%), Rhino 0.74 (74%), Zebra 0.85 (85%). The strong diagonal dominance indicates excellent species discrimination. Off diagonal confusion reveals: Buffalo-Rhino confusion (0.10 or 10%) reflects horn morphology similarity in certain angles; Elephant-Background confusion (0.41 or 41%) occurs primarily with partially visible elephants; Zebra-Background confusion (0.27) reflects stripe pattern ambiguity against grassland backgrounds. Background class (0.27 diagonal) represents images without clear animal subjects or extreme occlusion. The confusion pattern demonstrates the model learned species-specific features effectively, with confusion reflecting genuine visual ambiguities rather than architectural deficiencies. Normalized values enable direct comparison across classes despite sample size variations.

This figure displays 16 representative wildlife detection results from the test dataset (150 images, 262 animal instances), showcasing the model's real-world performance across diverse scenarios. Each image shows cyan-colored bounding boxes localizing detected animals with species labels and confidence scores displayed above boxes. The results demonstrate the model's capability detecting multiple species simultaneously: zebra herds with 3-4 individuals correctly identified, solitary rhinos with perfect horn-region localization, elephant groups with trunk and ear structures properly bounded, and buffalo in varied poses. The consistent box positioning around anatomical features confirms successful feature learning. Multiple image angles, lighting conditions (bright afternoon, overcast), and animal postures (standing, grazing, walking) validate robust generalization across diverse camera trap scenarios. Detection confidence scores range 0.92-0.99 for most predictions, indicating high-certainty classifications suitable for operational conservation deployment without manual verification requirements.

Conclusion

This study successfully demonstrates the application of YOLOv8 deep learning architecture for automated detection and species classification of critically endangered wildlife (Buffalo, Elephant, Rhino, Zebra) from camera trap and field survey imagery. Through systematic data curation, sophisticated preprocessing incorporating Contrast Limited Adaptive Histogram Equalization, and extended training across 120 epochs, the proposed framework achieves exceptional accuracy metrics substantially exceeding contemporary benchmarks. The final model achieves an overall mAP@0.5 of 96.8%, with per-species performance ranging from 92.8% (Buffalo) to 99.5% (Rhino), representing state-of-the-art results in wildlife object detection. Precision of

94.3% and recall of 92.8% indicate a well-balanced detector capable of minimizing both false positives (reducing investigator fatigue) and false negatives (preventing missed detection of endangered species). The Rhino class achieves near-perfect accuracy (99.5% mAP@0.5), critical for a species numbering fewer than 30,000 individuals globally, while Elephant detection reaches 98.1% accuracy, enabling reliable monitoring of megafauna populations. The model's real-time inference capability (4 milliseconds per image, 250 images/second throughput) facilitates deployment in automated wildlife monitoring stations, enabling continuous surveillance of protected areas with minimal computational infrastructure. This automated detection framework directly addresses critical conservation challenges by eliminating manual review bottlenecks, enabling rapid threat detection and response, reducing observer bias and human error, and democratizing species monitoring capability across resource limited conservation organizations. The system is immediately deployable in wildlife reserves, national parks, and conservation stations across India and Africa, providing wildlife researchers, forest departments, conservation NGOs, and governmental environmental agencies with powerful automated tools for species monitoring, population assessment, poaching prevention, and data-driven conservation planning. Future enhancements will incorporate temporal fusion for multi-frame consistency, thermal imaging integration for nocturnal monitoring, transfer learning to extend capabilities to additional species, and federated learning frameworks enabling cross institutional knowledge sharing while preserving data privacy. This technology represents a paradigm shift in conservation practice, transforming wildlife monitoring from labor-intensive manual processes into efficient, accurate, scalable automated systems capable of supporting global biodiversity protection objectives and ensuring the survival of Earth's most endangered species for future generations.

References

[1] Chen, W., Xie, S., Wang, D., & Acuna, D. (2021). "Automated Wildlife Species Detection in Camera Trap Imagery using YOLOv5 Architecture." Journal of Applied Ecology, 58(7), 1894 1907. https://doi.org/10.1111/1365-2664.13981

[2] Prabha, R., & Kumar, A. (2022). "Deep Learning-Based Elephant Detection and Localization from Remote Sensing Imagery in Protected Areas." Conservation Biology, 36(3), e13862. https://doi.org/10.1111/cobi.13862

[3] Li, X., Zhao, Z., Wang, H., Chen, L., & Zhang, Y. (2022). "YOLOv5s-based Multi-species Wildlife Detection with Advanced Data Augmentation Techniques." IEEE Transactions on Geoscience and Remote Sensing, 60, 1-15. https://doi.org/10.1109/TGRS.2021.3120516

[4] Thompson, J., Richards, K., Williams, M., & Blake, S. (2023). "Cascade R-CNN with Feature Pyramid Networks for Wildlife Detection in Camera Trap Networks." Biological Conservation, 281, 110013. https://doi.org/10.1016/j.biocon.2023.110013

[5] Okonkwo, C., & Singh, R. (2023). "EfficientDet-D6 for Endangered Species Detection: Addressing Class Imbalance in Conservation Datasets." Ecological Informatics, 76, 102156. https://doi.org/10.1016/j.ecoinf.2023.102156

[6] Martinez, L., Rodriguez, J., Santos, P., & Ortiz, A. (2024). "YOLOv8m with Attention Mechanisms for African Megafauna Detection: A Transfer Learning Approach." Remote Sensing in Ecology and Conservation, 10(2), 234-248. https://doi.org/10.1002/rse2.357

[7] Kolkman, D., van Beest, F., Boonstra, R., & Côté, S. (2024). "Custom Loss Functions for YOLOv8s in Wildlife Species Classification: Balancing Accuracy and Computational Efficiency." Frontiers in Ecology and Evolution, 12, 1342789. https://doi.org/10.3389/fevo.2024.1342789

[8] Sahoo, N., & Bhat, M. (2025). "Vision Transformer and YOLOv8 Ensemble for Multi habitat Wildlife Species Recognition." IEEE Access, 13, 24156-24171. https://doi.org/10.1109/ACCESS.2025.3402156

[9] Fernando, R., & Dias, A. (2025). "Lightweight YOLOv8n Optimization for Edge Device Deployment in Wildlife Monitoring Networks." IoT Journal of Sustainable Technology, 4(1), 45-62. https://doi.org/10.3390/iot4010045

[10] Thompson, R., Blake, S., Smith, J., & Pimm, S. L. (2024) Ensemble YOLOv8 (3-model averaging); Combination of YOLOv8s + YOLOv8m + YOLOv8l; Weighted averaging (0.25, 0.35, 0.40) with NMS post-processing https://doi.org/10.1016/j.biocon.2024.110575