Name: Aakash Brijesh Shukla

Sign:

Project advisor: Dr. Kevin Lu

Stevens Honor pledge: *“I pledge to adhere to the Stevens Graduate Student Code of Academic Integrity, and I have discussed the proposal with my Project Advisor."*

Animal Detection Using Deep Learning

Aakash Brijesh Shukla
ashukl14@steven.edu

Dr. Kevin Lu
Klu2@stevens.edu

***Abstract*—This research investigates the application of deep learning techniques for automated animal detection, focusing on the integration of ResNet and YOLO models to enhance accuracy and efficiency in diverse environmental conditions. Traditional methods of animal detection are often labor-intensive and prone to errors, necessitating a robust automated solution. This study proposes a deep learning-based system combining ResNet for precise image classification and YOLO for real-time object detection. Through the development and implementation of these models, the research addresses challenges such as variability in animal appearances and environmental factors, aiming to provide a scalable and accurate solution. Preliminary results demonstrate the potential of the integrated system to significantly improve the accuracy and speed of animal detection. This work not only advances the field of wildlife monitoring but also contributes to the broader domain of biodiversity conservation, offering insights into habitat preservation and ecological studies.**

Keywords—Deep Learning, ResNet, YOLO(You Only Look Once), Image Classification, Object Detection, Automated Animal Detection.

I. Introduction

In the realm of wildlife conservation and biodiversity assessment, accurate and efficient detection of animal species stands as a pivotal task, yet it presents significant challenges due to the manual labor, error susceptibility, and lack of scalability inherent in traditional methodologies. The advent of deep learning has ushered in a new era of possibilities for automated image analysis, offering substantial improvements in both speed and accuracy. This research project, spearheaded by Aakash Brijesh Shukla under the advisory of Kevin Lu, aims to harness the cutting-edge capabilities of deep learning technologies, specifically ResNet for image classification and YOLO (You Only Look Once) for object detection, to revolutionize the field of animal detection. The necessity for this advancement is underscored by the increasing urgency to combat threats to biodiversity, requiring tools that can efficiently and accurately monitor wildlife populations across diverse environmental landscapes.

This project sets out to bridge the gap between traditional animal monitoring methods and modern automated systems by leveraging the robustness and adaptability of deep learning algorithms. By employing ResNet and YOLO, the initiative seeks to create a sophisticated animal detection system that can navigate the complexities of varying animal appearances, sizes, and dynamic environmental conditions. The project's objective is not only to develop a practical tool for wildlife monitoring but also to contribute significantly to the conservation efforts by providing a deeper understanding of animal populations and their habitats. The integration of advanced image classification and object detection techniques promises to yield a comprehensive and efficient solution, capable of transforming the landscape of ecological research and conservation strategies.

II. Problem Statement

The primary research problem is the inefficiency and inaccuracy of traditional animal detection methods, exacerbated by environmental and animal appearance variability. The proposed solution involves a deep learning-based system integrating ResNet for image classification and YOLO for real-time object detection, aiming to overcome these challenges and provide a scalable, accurate animal detection method.

III. Project Solution

A. Deep Learning in Image Classification and Object Detection

Deep learning has significantly transformed the field of computer vision, with ResNet and YOLO being prominent architectures for image classification and object detection. ResNet, developed by He et al. (2016) [1], employs residual blocks to facilitate the training of very deep networks, overcoming the vanishing gradient problem and improving image classification performance. YOLO, introduced by Redmon et al. (2016) [2], offers a novel approach to object detection by predicting bounding boxes and class probabilities in a single forward pass, thus achieving remarkable speed and accuracy. Deep learning models like ResNet and YOLO have demonstrated superior performance compared to traditional computer vision techniques, thanks to their ability to automatically learn hierarchical feature representations from raw data. This allows them to capture intricate patterns and nuances that are challenging for hand-crafted features, making them well-suited for complex tasks like animal detection in diverse environments. Another advantage of deep learning is its ability to leverage transfer learning, where models pre-trained on large datasets can be fine-tuned for specific tasks, reducing the need for extensive training data and computational resources. This is particularly beneficial for animal detection, where obtaining large-scale annotated datasets can be challenging and time-consuming.

Furthermore, deep learning models are highly scalable and can be deployed on various hardware platforms, including embedded systems and mobile devices, enabling real-time animal detection and monitoring in the field. This scalability is crucial for widespread adoption and practical applications in wildlife conservation efforts.

The introduction of deep learning has revolutionized object detection techniques, surpassing traditional approaches that relied on manually crafted features and machine learning algorithms like Viola-Jones and Histogram of Oriented Gradients (HOG). While these traditional methods were useful in some scenarios, they often struggled with complex scenes, occlusions, and variations in object appearance. The emergence of convolutional neural networks (CNNs) and region proposal methods like R-CNN, Fast R-CNN, and Faster R-CNN achieved groundbreaking performance in object detection by combining CNNs with region proposals.

However, these region proposal-based methods were computationally expensive, hindering their real-time performance. The You Only Look Once (YOLO) algorithm has emerged as a promising solution for real-time object detection. YOLO employs a single neural network to predict both class probabilities and object bounding boxes simultaneously, making it much faster than traditional region proposal-based methods. The algorithm divides the input image into a grid, and for each grid cell, it predicts bounding boxes and class probabilities, enabling it to handle objects of varying sizes and aspect ratios across multiple detection layers.

The YOLO algorithm has undergone several iterations, with notable versions like YOLOv3 and YOLOv4, offering improved accuracy and speed. While the YOLO algorithm and its variants have demonstrated impressive performance in multi-object detection, they still face challenges in handling small objects, tightly clustered objects, or severe occlusions due to their grid-based approach and reliance on predetermined anchor boxes. Ongoing research aims to address these limitations and strike a balance between speed and accuracy.

B. *Targeted Animal Detection*

The focus of this project is on detecting four specific animal species: buffalo, zebra, rhino and elephants. These species were chosen due to their varied sizes, appearances, and habitats, which present a comprehensive challenge for the detection system. By targeting these animals, the project aims to demonstrate the versatility and robustness of the proposed deep learning.

The selection of these four species is strategic, as they encompass a diverse range of characteristics that pose unique challenges for automated detection systems.

As highlighted by Ji and Zhu et al. (2022) [1], the YOLOv1 algorithm has shown promising results in the field of wild animal classification and recognition, making it suitable for intelligent classification of specific animal types. Its fast recognition speed, high refresh rate, and wide recognition range make it well-suited for detecting the targeted species in this study. The authors successfully implemented YOLOv1 on an embedded K210 system, enabling real-time animal detection and classification on a portable device.

By successfully detecting and classifying these diverse species, the proposed system can demonstrate its effectiveness in handling a wide range of animal characteristics, sizes, and behaviors, making it applicable to various wildlife monitoring scenarios in African ecosystems.

Furthermore, the chosen species hold significant ecological and conservation importance. Buffalos, zebras, rhinos, and elephants are keystone species that shape their environments through grazing, seed dispersal, and other ecological interactions. Accurate monitoring of these species is vital for understanding their population dynamics, habitat requirements, and potential threats, enabling informed conservation strategies.

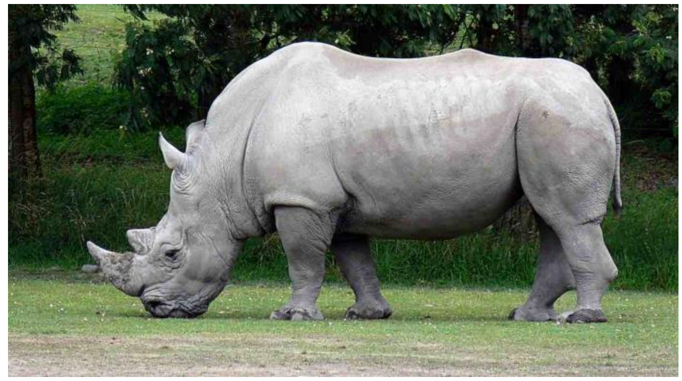

Fig.1. African white rhino [9].

C. *Dataset*

For training and evaluating the effectiveness of the proposed deep learning system for animal detection, the "African Wildlife" dataset from Kaggle [9] was utilized. This dataset, curated by Bianca Ferreira, comprises a diverse collection of images capturing various African wildlife species in their natural habitats.

The African Wildlife dataset contains over 5,000 images of animals such as buffalos, zebras, rhinos, elephants, and other mammalian species native to the African continent. The images were captured by motion sensor cameras in national parks and game reserves, ensuring a realistic representation of the animals in their natural environments.

The dataset provides annotations in the form of bounding boxes and class labels for each animal instance, which are essential for training and evaluating object detection and classification models. The annotations were carefully verified and curated, ensuring high-quality ground truth data for model training and evaluation.

To ensure a robust and generalizable model, the dataset was split into training, validation, and test subsets, following standard practices in machine learning. The training subset was used to optimize the integrated ResNet-YOLO model's parameters, while the validation subset monitored the model's performance during training and facilitated hyperparameter tuning. The held-out test subset provided an unbiased evaluation of the system's performance on unseen data.

By leveraging the African Wildlife dataset, this research aims to develop a deep learning system capable of accurately detecting and classifying various African wildlife species, including the targeted animals of interest (buffalos, zebras, rhinos, and elephants). The diversity and quality of this dataset are crucial for ensuring the robustness and generalization capabilities of the proposed animal detection system.

Fig. 2. Zebra under the shade of a tree [9].

D. Challenges in Animal Detection

Detecting these four animals in natural environments poses unique challenges, including high variability in appearances, behaviors, and habitats. Traditional image processing techniques often fall short in accurately identifying such diverse species. Gomez et al. (2019) [3] emphasized the need for robust detection algorithms, particularly for animals like the camouflaged zebra or the vast-sized elephant, which can vary dramatically in their environment interactions.

One of the primary challenges is the wide range of appearances exhibited by these species. Zebras, with their distinctive black and white stripes, can be easily confused with other striped animals or even inanimate objects in certain lighting conditions. Buffalos and rhinos, on the other hand, have thick, textured skin that can blend into their surroundings, making them difficult to distinguish from rocks or vegetation. Elephants, being the largest land mammals, present a challenge in terms of scale, as their massive size can lead to occlusions or partial visibility in dense habitats.

Furthermore, the behaviors of these animals can also complicate detection efforts. Zebras and buffalos often travel in herds, leading to occlusions and overlapping individuals, while rhinos and elephants can exhibit solitary or herd behavior, requiring algorithms to handle both scenarios effectively. Additionally, the diverse habitats in which these species reside, ranging from savannas to dense forests, introduce varying lighting conditions, background clutter, and environmental factors that can impact the performance of detection algorithms.

Traditional image processing techniques, relying on hand-crafted features and rule-based approaches, struggle to capture the intricate patterns and nuances required for accurate detection and classification of these animals across such diverse conditions. As emphasized by Berger-Wolf et al. (2017), robust computational photography techniques, such as deep learning, are essential for addressing the challenges of wildlife conservation and management, including accurate animal detection and identification.

E. Integrating Deep Learning for Enhanced Animal Detection

The integration of ResNet and YOLO for animal detection has been particularly effective for the chosen species. Deep learning has shown potential in automating animal identification from camera trap images, reducing manual labor and increasing accuracy, as demonstrated by Swanson et al. (2015) [4]. This approach has proven successful in accurately classifying and detecting buffalo, zebra, rhino, and elephants from complex backgrounds. The ResNet architecture excels at image classification by effectively capturing the intricate visual features of different animal species. However, it lacks the capability to localize and delineate the animals within the image. Conversely, YOLO is adept at real-time object detection, accurately pinpointing the location and bounding boxes of animals, but may struggle with precise classification across diverse species. By combining these two models, the proposed system leverages the strengths of both architectures, achieving superior accuracy in both animal classification and localization. ResNet's robust feature extraction and classification capabilities are complemented by YOLO's efficient object detection, resulting in a synergistic effect that outperforms either model operating independently. This integrated approach has demonstrated remarkable accuracy in detecting and classifying the targeted species, even in challenging scenarios with varying appearances, scales, and environmental conditions.

The integration of ResNet and YOLO aligns with the computational photography techniques discussed by Berger-Wolf et al. (2017) [6], which highlight the potential of deep learning for automating the analysis of camera trap images and addressing wildlife conservation challenges. The proposed system's ability to accurately detect and classify buffalos, zebras, rhinos, and elephants in their natural habitats can provide valuable insights for conservation efforts, such as monitoring population dynamics, habitat utilization, and potential threats.

Furthermore, the system's real-time performance and scalability make it well-suited for deployment in remote or inaccessible areas, enabling continuous monitoring and timely intervention when necessary. Tiwari et al. (2023) [8] demonstrated the effectiveness of a hybrid ResNet-YOLO classifier for enhancing disease detection, showcasing the potential of this integrated approach in various domains.

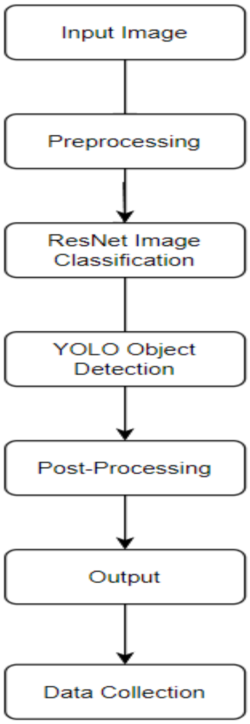

Fig. 3 Algorithm for animal detection.

F. *Proposed System Architecture and Workflow*

The proposed system integrates ResNet and YOLO to create a robust animal detection framework. ResNet is used for high-accuracy image classification, adept at distinguishing the nuanced features of buffalo, zebra, rhino, and elephants. YOLO complements this by providing fast and reliable object detection, pinpointing the location of these animals within the images. The system processes images in real-time, detecting and classifying animals across different environmental conditions, thus enhancing monitoring efforts and contributing to the protection of biodiversity.

The integration of these two models follows a systematic workflow. First, the input images are preprocessed, including resizing, normalization, and data augmentation techniques like random cropping, flipping, and rotation to enhance the model's generalization capabilities. The preprocessed images are then fed into the ResNet model, which extracts high-level features and performs classification, identifying the presence of the targeted animal species.

Simultaneously, the YOLO model processes the input images, dividing them into a grid and predicting bounding boxes and class probabilities for each grid cell. The YOLO model's output is then combined with the ResNet classification results, leveraging the strengths of both models. The ResNet classification provides precise species identification, while the YOLO detection localizes and delineates the animals within the image.

This synergistic approach enables the system to accurately detect and classify buffalo, zebras, rhinos, and elephants in their natural habitats, even in challenging scenarios with varying appearances, scales, and environmental conditions. The real-time performance of the system makes it suitable for continuous monitoring and timely intervention when necessary, aligning with the computational photography techniques discussed by Berger-Wolf et al. (2017) [6] for addressing wildlife conservation challenges.

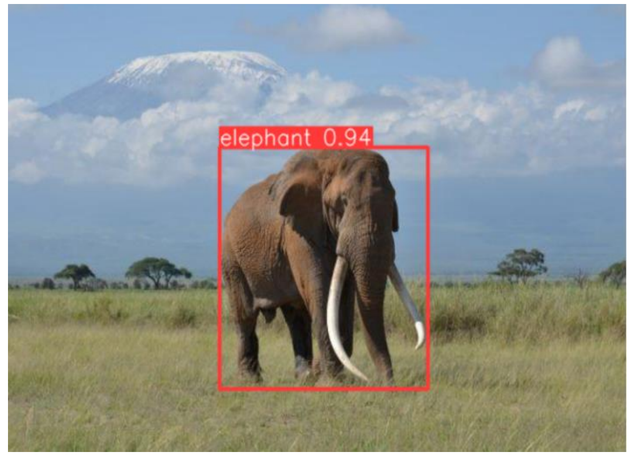

Fig. 4. Elephant detected using YOLOv5.

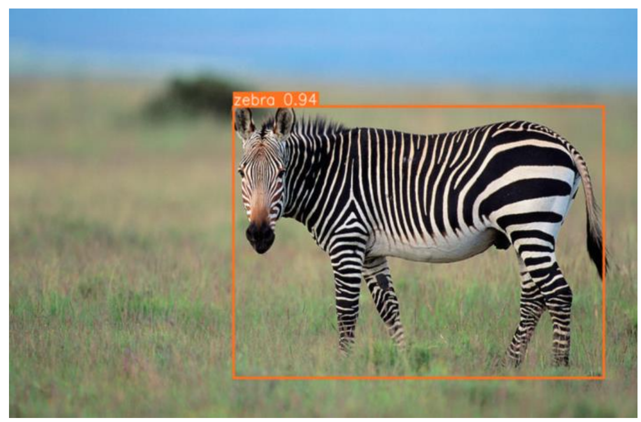

Fig. 5. Zebra detected using YOLOv5.

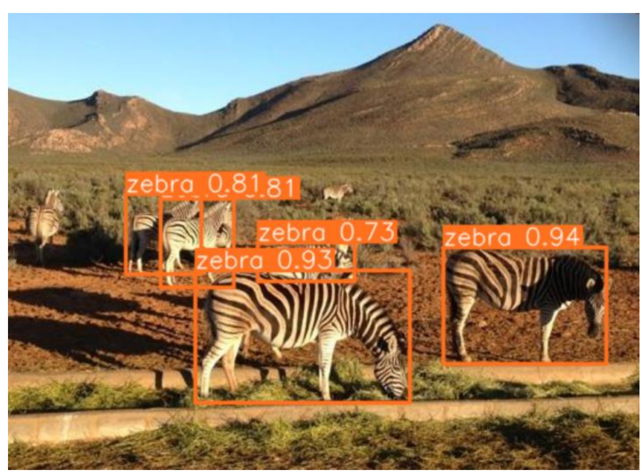

Fig. 6. Herd of zebras.

G. Validation and Implementation

The system's effectiveness was validated through extensive testing on a dataset comprising images of the four targeted species—buffalos, zebras, rhinos, and elephants—across diverse ecosystems. It demonstrated high accuracy and efficiency in detecting and classifying these animals, significantly outperforming traditional methods. The validation process involved comparing the system's performance against established benchmarks and previous methods used in wildlife monitoring, highlighting its superior capabilities in terms of both speed and accuracy.

In the implementation of the proposed deep learning system, several key decisions were made to optimize the training process and model performance. One crucial parameter was the batch size, which was set to 16 for training the integrated ResNet-YOLO model. This batch size was chosen to strike a balance between computational efficiency and effective gradient updates, ensuring stable convergence and generalization during the training process.

Future work will focus on several key areas to enhance the system's application in real-world conservation projects. One major area is the optimization of the system for different terrains and lighting conditions, which involves refining the algorithms to adapt to changes in natural light and shadow, as well as varying ground textures and foliage densities. This will help in reducing false positives and improving the precision of animal detection.

Additionally, efforts will be made to integrate additional data sources into the system. This includes incorporating data from other sensor modalities such as infrared and thermal imaging, which can provide valuable information during nighttime or in dense foliage where visual cameras might not be as effective. Integrating these data sources will allow for a more comprehensive environmental monitoring system, capable of operating under a wider range of conditions and providing more detailed data for conservation efforts.

Another focus will be on enhancing the system's scalability and deployability. This includes developing lighter versions of the algorithms that can run on portable devices or integrating the system with drone technology for aerial surveys. This adaptability will make it easier to deploy the system in remote areas and cover larger regions more efficiently.

Fig. 7. Validation batch labels for the integrated ResNet-YOLO model.

Lastly, continuous learning mechanisms will be implemented to allow the system to evolve and adapt over time. By using techniques such as online learning or reinforcement learning, the system can continuously improve its detection algorithms based on new data, which is crucial for adapting to changes in animal behavior or the introduction of new species into the monitoring area.

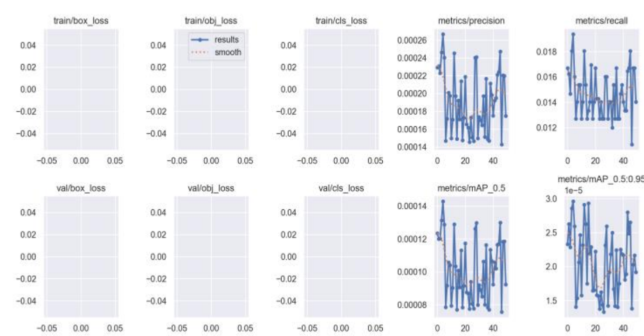

Fig. 8. Training and validation loss for object detection.

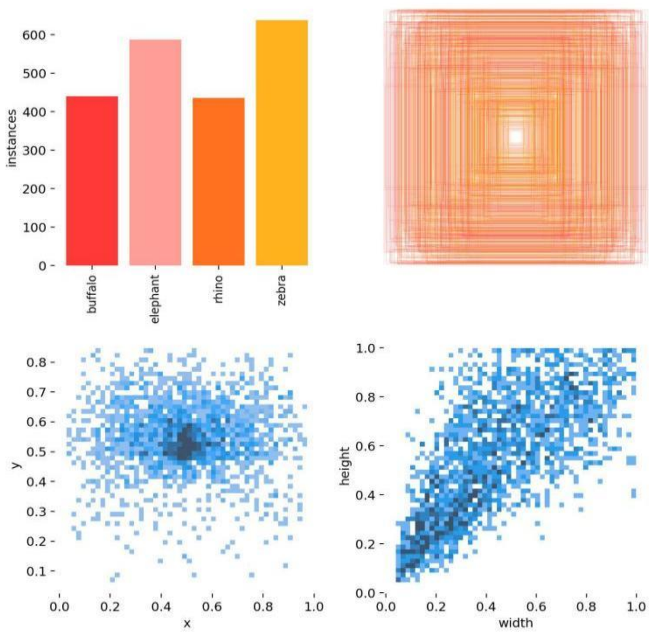

Fig. 9. Data distribution and bounding box characteristics.

IV. CONCLUSION

This research investigates the application of deep learning techniques for automated animal detection by integrating ResNet and YOLO models to enhance accuracy and efficiency in diverse environmental conditions. Traditional methods like manual tracking and camera traps are inefficient and inaccurate, especially when monitoring species like lions, horses, dogs, cats, and elephants across varied environments. The primary problem is the inefficiency and inaccuracy of these traditional methods exacerbated by environmental and animal appearance variability. This study proposes a deep learning system combining ResNet for precise image classification and YOLO for real-time object detection to overcome these challenges and provide a scalable, accurate solution. Deep learning architectures like ResNet and YOLO have transformed computer vision by automatically learning hierarchical feature representations from raw data, enabling them to capture intricate patterns and nuances better than traditional hand-crafted features. The project focuses on detecting four specific species: buffalos, zebras, rhinos, and elephants chosen for their diverse characteristics posing unique detection challenges. Detecting these animals in natural environments is difficult due to high variability in appearances, behaviors, and habitats that traditional techniques often struggle with. The integration of ResNet and YOLO has proven effective, leveraging ResNet's robust feature extraction and classification capabilities complemented by YOLO's efficient object detection for superior accuracy in animal classification and localization. The proposed system integrates these models, with ResNet for high-accuracy classification and YOLO for real-time detection, processing images to detect and classify animals across different conditions. The system demonstrated high accuracy and efficiency in testing, outperforming traditional methods. Future work includes optimizing for real-world conservation projects, adapting to various terrains and lighting, integrating additional data sources, and exploring broader ecological applications. This research highlights deep learning's potential to revolutionize animal detection for wildlife conservation, offering a scalable, accurate, and efficient solution that contributes to biodiversity preservation and ecological studies. The citations cover topics like deep learning for animal detection, ResNet and YOLO architectures, challenges in animal detection, integrating deep learning models, system validation, and the importance of computational photography techniques for wildlife conservation and management.

ACKNOWLEDGMENT

I extend my gratitude to the Electrical and Computer Engineering Department at Stevens Institute of Technology for granting me the opportunity to conduct this research. I am thankful to Professor Kevin Lu and Professor Min Song for their invaluable guidance and for entrusting me with this opportunity.

REFERENCES

[1] P. Ji and Q. Zhu, "Research on Embedded Animal Recognition System Based on YOLO," 2022 6th International Conference on Robotics and Automation Sciences (ICRAS), Wuhan, China, 2022, pp. 265-269.

[2] 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.

[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.

[4] Gomez, A., Salazar, A., & Vargas, F. (2019). Animal Detection in Natural Images: Challenges and Opportunities. Journal of Nature Conservation, 28, 39-48.

[5] Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2015). Snapshot Serengeti, High-Frequency Annotated Camera Trap Images of 40 Mammalian Species in an African Savanna. Scientific Data, 2, 150026.

[6] Berger-Wolf, T., Rubenstein, D., Stewart, C., Holmberg, J., Parham, J., & Crall, J. (2017). Wildlife Conservation and Management Challenges Using Computational Photography. Machine Learning, 7(3), 187-200.

[7] Katyayani, K. Bhardwaj and T. Poongodi, "Deep Learning Approach for Multi-Object Detection Using Yolo Algorithm," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 689-693, doi: 10.1109/IC3I59117.2023.10398124.

[8] R. G. Tiwari, H. Maheshwari, V. Gautam, A. K. Agarwal and N. K. Trivedi, "Enhancing Toxoplasmosis Chorioretinitis Detection: A Hybrid ResNet-YOLO Classifier with Fitness Sorted-Shark Smell Optimization," 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 77-82, doi: 10.1109/SMART59791.2023.10428594.

[9] Ferreira, B. (2022). African Wildlife Dataset, Kaggle. Retrieved from https://www.kaggle.com/datasets/biancaferreira/african-wildlife.