Impact Factor (SJIF): 6.126 E-ISSN: 2583-1615

International Journal of Research in Academic World

Received: 10/February/2025 **IJRAW:** 2025; 4(3):90-93 **Accepted:** 18/March/2025

Detection of Wild Animal Activity Using Deep Learning Techniques: A Review

[*1]Dr. T Suvarna Kumari, [2]Sree Pratiksha Sanjivalla and [3]Jayendra Medoju

[*1]Assistant Professor, Department of Computer Science and Engineering (CSE), Chaitanya Bharathi Institute of Technology (CBIT), Telangana, India.

[2, 3]Student, Department of Computer Science and Engineering (CSE), Chaitanya Bharathi Institute of Technology (CBIT), Telangana, India.

Abstract

Human-wildlife conflicts and the need for conservation have made efficient wildlife monitoring systems increasingly important. Traditional animal detection methods often rely on high-resolution, multi-frame imaging, which demands substantial computational power and is often expensive, particularly in remote and resource-constrained environments. Addressing these challenges, our project introduces a deep learning model designed to detect wild animals in images with both accuracy and efficiency. By combining the VGG-19 Convolutional Neural Network (CNN) for spatial feature extraction and a Bidirectional Gated Recurrent Unit (Bi-GRU) for temporal pattern recognition. Additionally, we employ YOLOv5 for precise localization, helping track animal positions within each frame. This combined model enhances detection capabilities, reduces computational demands, and lowers costs, making it ideal for deployment in remote areas with limited resources. Our solution is designed to contribute to real-time wildlife monitoring, aiding in conservation efforts and minimizing human-wildlife conflicts by providing a reliable, low-cost detection system that serves both environmental and community needs.

Keywords: VGG-19 Convolutional Neural Network (CNN), Bidirectional Gated Recurrent Unit (Bi-GRU), YOLOv5.

1. Introduction

The growing demand for effective monitoring of wildlife has brought out the limitations of traditional detection methods, which tend to be expensive and resource-heavy. With increasing human encroachment into natural habitats, the need for effective monitoring systems of animal activity to mitigate conflicts with humans, assist in conservation efforts, and better biodiversity research is becoming increasingly urgent. Traditional methods usually require high-resolution imaging equipment and substantial computational resources, thus making them inapplicable in remote or resource-scarce environments. Tackling such issues, the exact approach of our project is to utilize a deep learning model toward better and accurate detection of wild animals in images with efficiency that helps bring in advancements in computer vision and sequence modeling. Our model blends together the VGG-19 Convolutional Neural Network (CNN) with Bidirectional Gated Recurrent Unit (Bi-GRU). Overall, this will bring both spatial and temporal data together for comprehensive detection of wild animals. VGG-19 captures strong spatial features, including shape, patterns, and textures, that will distinguish the different animal species. Bi-GRU adds temporal context, analyzing sequences across frames for movement patterns and the odd behaviors of wild animals. This technology further utilizes YOLOv5 for real-time localization of the animal through correct bounding boxes to detect the actual position of the animal in every frame. That is why this hybrid model, through accuracy increases and computation cost reductions, suits deployment even in remote or under-resourced areas. This system promises great benefits to wildlife conservation, hence supporting initiatives needing low-cost monitoring tools but high accuracy. Such information by researchers and conservationists would be absolutely essential for protecting ecosystems and promoting sustainable coexistence between humans and wildlife. This contributed to the development of a scalable and efficient detection model on the energetic streams of ecological monitoring, research, and conservation.

2. Related Works

In recent years, much focus has been directed toward developing advanced methods for the detection of wildlife activity using deep learning. Applying neural networks of vast architecture including hybrid models and CNNs should enhance the robustness to variance in the environment and the real-time processing capacity while enhancing detection accuracy. Here, some relevant contributions and results obtained by various approaches on studies regarding animal activity detection are discussed. Each of these studies addresses different challenges in optimal performance and

***Corresponding Author:** Dr. T Suvarna Kumari

robustness concerning deep learning models applied to wildlife applications.

An adaptive feature selection approach with manifold learning is proposed to improve image classification accuracy. This method dynamically selects features from image data to filter out redundant or irrelevant information, focusing instead on key points critical for classification. This approach enhances feature selection in wild animal activity detection, improving the performance of classification systems by emphasizing discriminative features and reducing background noise. Manifold learning enables the model to adapt to data structure variations, especially valuable in high-dimensional wildlife imagery with diverse backgrounds. Experiments on large-scale datasets demonstrate improved classification performance, especially in complex, variable scenes, showing this approach's effectiveness in wildlife monitoring where environmental factors frequently alter image features [1].

A study addressing animal detection in cluttered natural scenes utilizes spatiotemporal object region proposals and patch verification to handle dense foliage and complex backgrounds common in wildlife monitoring. This model differentiates moving animals from stationary objects by analyzing both spatial and temporal dimensions, effectively identifying animals in noisy environments. It generates region proposals to identify potential animal locations and verifies each region through patch validation, reducing false positives. Incorporating spatiotemporal features allows the model to detect movement patterns and contextual information across frames, making it suitable for continuous wildlife surveillance where animals are in motion and backgrounds are challenging [2].

Another research utilizes deep convolutional neural networks (CNNs) for animal species recognition from camera trap images, a key tool in ecological research. This model achieves high accuracy in classifying different species by identifying fine-grained, species-specific features such as fur patterns and body shapes. Tested on a large, diverse dataset with images from various ecological conditions, CNNs proved robust, handling noise and variability from different angles, lighting, and distances. This approach holds practical value for automating biodiversity assessment and monitoring, providing accurate species identification across diverse natural environments [3].

A hybrid deep neural network is proposed to enhance real-time alert systems based on detected animal activity. This system combines object detection and classification layers, allowing it to identify animal types and movement patterns. The hybrid approach improves the system's responsiveness, enabling real-time classifications and behavior predictions essential for applications like early warning systems in wildlife conservation and human-wildlife conflict prevention. By distinguishing passive from potentially hazardous behaviors, the model supports real-time alerts, aiding in efforts to minimize conflicts between wildlife and local communities [4].

In an analysis of deep learning models for remote sensing image classification, researchers evaluate models like ResNet, Inception, and DenseNet in classifying diverse, complex landscapes. This study is highly relevant to wildlife habitat monitoring, where remote sensing helps identify habitats and migration patterns over large areas. The comparative analysis helps practitioners select appropriate models for project-specific needs—whether prioritizing accuracy, speed, or environmental adaptability. These insights are beneficial for large-scale wildlife monitoring and conservation efforts requiring image classification over varied terrains [5].

A fuzzy discriminative block representation learning method is introduced for image feature extraction, which emphasizes key areas of an image while minimizing background noise. This approach improves deep learning models' focus on animal-specific features, beneficial in wildlife environments where foliage noise can hinder detection accuracy. By enhancing animal-associated image regions and de-emphasizing irrelevant sections, this fuzzy logic-based method improves detection outcomes, supporting wildlife monitoring in settings with challenging visual elements like shadows and camouflage [6].

A novel animal detection system featuring a cascaded YOLOv8 model incorporates adaptive preprocessing and feature extraction techniques. This system adapts preprocessing steps to optimize for lighting and noise in varied environments, enhancing performance. The cascaded architecture enables high-speed detections without compromising accuracy, making it suitable for real-time wildlife monitoring. Adaptive feature extraction helps the model capture class-specific features for species differentiation, even in cluttered landscapes or partial occlusions, establishing the system's robustness for real-world animal tracking and behavioral analysis [7].

In another study, T-YOLO, a lightweight model based on YOLO, is designed for detecting small objects, initially aimed at vehicles but applicable to wildlife detection for small animals. Using multi-scale convolutional layers to capture detailed features, T-YOLO performs well in scenarios requiring high sensitivity for small or distant objects. This efficient model, suited for deployment in resource-limited environments, balances speed and accuracy, proving useful in large-scale, continuous monitoring applications where detail is crucial [8].

A study on wild animal detection using deep learning evaluates various neural network architectures in detecting animals in complex forest and open landscapes. Training models on diverse datasets with images of animals in various postures and backgrounds, this study examines how architectures like VGG and ResNet handle environmental noise and background diversity. Results indicate that deeper networks, while computationally intensive, achieve higher accuracy. Data augmentation improves robustness, suggesting that diverse habitat data enhances detection accuracy. The study highlights deep learning's potential in automating wildlife surveillance, a significant advancement for conservation efforts [9].

Lastly, a survey of deep learning techniques for underwater image classification provides insights into methods that could also apply to challenging terrestrial environments, such as dense woods or low-light wildlife settings. The study examines CNNs and RNNs for handling complex image characteristics, including noise, poor visibility, and occlusions, similar to issues in wildlife monitoring. Techniques for adapting to image quality variance and lighting conditions are emphasized, offering valuable guidance for selecting architectures suited for animal detection in visually challenging conditions, relevant to both underwater and terrestrial wildlife ecosystems [10].

The iWildCam dataset is a vast-sized wildlife camera trap image dataset created to assist in the monitoring and conservation of wildlife. It falls under the Wildlife Conservation Society's WCS, in collaboration with Microsoft's AI for Earth initiative. iWildCam has been hosted through the WILDS benchmark of Wildlife Imaging and

Learning DataSet. It comprises millions of images from hundreds of camera traps set up in natural habitats of different geographic locations. These images contain a wide diversity of species; when making use of species type, activity, and location metadata annotations, they may be used to create and test computer vision models targeting the automatic detection, classification, and behavioral analysis of animals.

The iWildCam dataset is unique in its focus on capturing real-world complexity including variability in lighting and weather conditions, partial occlusions, and camouflage over animals. This variability makes it difficult to be picked up by models and hence renders iWildCam an essential source of benchmarking the robustness and flexibility of a model. Other benchmark tasks apart from species classification and location-specific generalization tests focus on the testability of a model on accuracy and generalizability. iWildCam advances machine learning for conservation technology and thus improves biodiversity monitoring while supporting research, which will help inform a more precise environmental policy and a more effective species preservation strategy.

3. Background

Notable improvement in detection accuracy and robustness has been noted using techniques related to deep learning in the field of wildlife monitoring. Recent work by Ashraf *et al.* [1] introduced adaptive feature selection techniques along with manifold learning that promotes improved classification of images while emphasizing the relevance of dynamic feature selection within complex visual environments on the choice made for VGG-19 for extracting rich spatial features in wild animal activity detection.

Zhang *et al.* [2] addressed the challenge of detecting animals in extremely cluttered natural scenes using spatiotemporal region proposals. The application requires capturing both spatial and temporal information elements, thus motivating our use of Bi-LSTM to model temporal dependencies for consistent tracking of animals across frames. In the same direction, Islam *et al.* [3] deployed deep CNNs for species recognition in ecological images, as CNNs can readily address complicated habitat conditions. This serves to further confirm the reasons why we chose to use VGG-19 in the extraction of high spatial features from natural scenes.

Natarajan *et al.* [4] further extended the detection system by using hybrid neural networks for generating alerts based on activities. Their hybrid methodology for the processing of spatial and temporal data is similar to our effort to integrate VGG-19 with Bi-LSTM as we are targeting both location and movement cues simultaneously to enhance the detection. Alem and Kumar [5] recommend testing a model in terms of efficiency on resource-constrained systems, therefore we compromise on the computation for both VGG-19 as well as Bi-LSTM.

Recently, Wang *et al.* introduced the fuzzy block representation for improving feature extraction-a concept we apply to fine-tune feature discrimination of VGG-19. Meanwhile, Chappidi and Sundaram made use of adaptive preprocessing techniques with YOLOv8 in order to demonstrate the importance of customization to specific environmental demands. It is precisely this adaptability that forms the foundation of our design: we customize preprocessing for diverse wildlife habitats.

The authors' lightweight application of the multi-scale CNN, as realized in T-YOLO proposed by Carrasco *et al.* [8], encourages us to further work on fine-tuning feature extraction in our models to detect animals under changing environmental conditions. Edison and S. K. L. [9] discussed deep learning for wild animal detection and highlights data augmentation to handle natural variability, which we make use of in our study to handle the changeability of the environment.

Finally, Mittal *et al.* [10] presented techniques to classify underwater, which thus commented upon poor visibility and occlusion. These results are applicable to our project since it is dense forests which usually obscure sight; hence we must consider data diversity to find more resilient wildlife detection.

4. Methodology

The wildlife detection project can be divided into several stages of methodology, which are essential, starting from preprocessing the data, feature extraction, temporal modeling, object detection, and training of the model. Then, every step is part of constructing a highly efficient deep learning framework for real-time recognition of wildlife activities.

i). **Data Preprocessing:** Pre-processing data ensures uniformity and optimizes input images for model learning. All images are resized to 224 x 224 pixels to balance between spatial detail and computation efficiency in calculations. A set of techniques used for data augmentation include: random rotation, zoom, horizontal flipping, and scaling-to enrich the view and distance diversity, so the animal orientation across different views is enabled. Pixel values are normalized between 0 and 1 to optimize numerical stability and speed convergence.

ii). **Feature Extraction Using VGG-19:** Feature extraction by VGG-19 is a 19-layer convolutional network. This model is specifically famous for extracting intricate spatial features. The lower layers of VGG-19 focus on low-level features such as edges and textures, whereas the deeper layers are dedicated to more complex details that involve shapes and markings of animals. Max-pooling layers down-sample the feature map maintaining the important information while reducing noise and computational requirements. The output of the convolutional layer is flattened into a feature vector for the final output of this layer. It now forms a spatial representation of each frame as input to further modeling along the temporal dimension.

iii). **Bi-GRU Model for Temporal Modeling:** For the sequential dependencies between frames, the project utilizes a Bi-GRU-based model. In the Bi-GRU architecture, it considers both forward and backward directions in the sequence processing, which satisfactorily and wholly formulates movement patterns. Multiple Bi-GRU layers look over the features extracted by VGG-19 in time to capture continuous behaviors and interactions across a few frames. Bidirectional approach helps the model predict movement well as both past and future context is included. This makes it especially effective for the detection of many complex activities of animals within the sequence of a video.

iv). **Object Detection with YOLOv5:** YOLOv5 is used in real-time, accurate object detection. This single-shot detector is efficient in object localization and classification, identifying and localizing animals in each frame. YOLOv5's architecture is very specifically optimized for real-time detection tasks with low latency and the ability to detect objects at different scales. It generates bounding boxes around detected animals, which allows the model to focus on regions of interest, thus making it important for monitoring movement and animal presence over time.

v). **Model Training and Evaluation:** Therefore, in the dataset, the training, validation, and testing sets split separate it. In training a model, it permits effective training while ensuring that when tested on new data, the performance of the model is not biased. Fine-tuning the hyperparameters, such as learning rate and batch size, is often done through grid or random search in order to achieve an optimal balance between computational efficiency and accuracy. The Adam optimizer is used to achieve faster convergence, and categorical cross-entropy just so happens to be the loss function, suitable for multi-class classification. Dropout layers are applied post the Bi-GRU to prevent overfitting; thus, this model will generalize well when subjected to diverse datasets.

This methodology integrates spatial and temporal analysis features with real-time detection capabilities in a wildlife monitoring system. This method brings about the creation of a stable, robust, accurate, adaptable, and computationally efficient system.

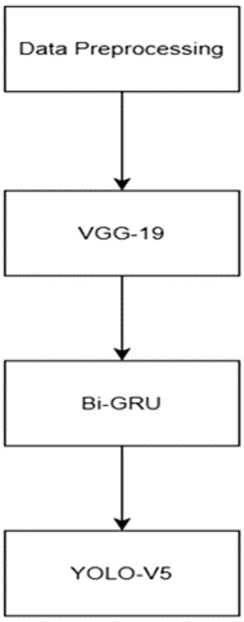

Fig 1: Flow of Hybrid (VGG-19 + Bi-GRU) Model

5. Conclusion

We designed a strong deep learning framework for wild animal activity detection based on the complementary strengths of VGG-19 and Bi-GRU. Our model, by strong feature extraction through the VGG-19 captures the detailed spatial information quite well, allowing it to differentiate between animals and very complex natural backgrounds. The Bi-GRU component enhances the temporal dimension and henceforth allows tracing and interpretation of sequential patterns, this is an important feature that requires the detection of movement of animals in one frame to another. The proposed hybrid approach is expected thus to be accurate and reliable while being efficient in terms of computation, as it addresses some critical defects of traditional approaches.

Our method is actually based on the integration of lessons from new developments concerning spatiotemporal detection, model adaptability, and lessons from new developments in image-classified models. Our model will therefore be capable of performing well under changing environmental conditions-from cluttered forest areas to poor-visibility situations, so tailored towards a more balanced architecture that emphasizes spatial and temporal dynamics. Thus, YOLOv5's incorporation into the system for real-time location identification further enhances our framework and provides localization at the high level as well as in species recognition capabilities.

This work realizes a promising solution for wildlife monitoring, combining deep convolutional and recurrent networks into a flexible, computationally manageable framework. Our model improves detection accuracy and opens the door for further development in conservation technology, encouraging real-time and scalable monitoring efforts that are crucial for the preservation of wildlife ecosystems.

References

1. Ashraf NM Nawi and M Aamir. "Adaptive Feature Selection and Image Classification Using Manifold Learning Techniques," *IEEE Access*, 2023; 12:40279-40289. DOI: 10.1109/ACCESS.2023.3322147.
2. Z Zhang, Z He, G Cao and W Cao, "Animal Detection From Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Patch Verification," *IEEE Transactions on Multimedia*. 2016; 18(10):2079-2091. DOI: 10.1109/TMM.2016.2594138.
3. SB Islam, D Valles, TJ Hibbitts, WA Ryberg, DK Walkup and MRJ Forstner. "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images," *Animals*, 2023, 13(9), Article 1526, May. DOI: 10.3390/ani13091526.
4. Natarajan, R Elakkiya, R Bhuvaneswari, K Saleem, D Chaudhary and SH Samsudeen, "Creating Alert Messages Based on Wild Animal Activity Detection Using Hybrid Deep Neural Networks," in *IEEE Access*. 2023; 11:67308-67321. DOI: 10.1109/ACCESS.2023.3289586.
5. A Alem and S Kumar. "Deep Learning Models Performance Evaluations for Remote Sensed Image Classification," in *IEEE Access*. 2022; 10:111784-111793. Doi: 10.1109/ACCESS.2022.3215264.
6. Y Wang, Z Li, F Li, Y Mi and J Yue. "Fuzzy Discriminative Block Representation Learning for Image Feature Extraction," in *IEEE Transactions on Image Processing*. 2022; 31:4994-5008. DOI: 10.1109/TIP.2022.3191846.
7. J Chappidi and DM Sundaram, "Novel Animal Detection System: Cascaded YOLOv8 with Adaptive Preprocessing and Feature Extraction," in *IEEE Access*. 2024; 12:110575-110587. DOI: 10.1109/ACCESS.2024.3439230.
8. Padilla Carrasco, HA Rashwan, MÁ García and D Puig, "T-YOLO: Tiny Vehicle Detection Based on YOLO and Multi-Scale Convolutional Neural Networks," in *IEEE Access*. 2023; 11:22430-22440. DOI: 10.1109/ACCESS.2021.3137638
9. SKL and A Edison. "Wild Animal Detection using Deep learning," *2022 IEEE 19th India Council International Conference (INDICON)*, Kochi, India, 2022, 1-5, DOI: 10.1109/INDICON56171.2022.10039799.
10. S Mittal, S Srivastava and JP Jayanth. "A Survey of Deep Learning Techniques for Underwater Image Classification," in *IEEE Transactions on Neural Networks and Learning Systems*. 2023; 34(10):6968-6982. DOI: 10.1109/TNNLS.2022.3143887.