18(1s), Spring 2026

Journal of Animal Environment

(E-ISSN 2717-1396, ISSN 2717-1388)

Original Research Paper

Computer Vision and Deep Learning Applications for Automated Wildlife Detection, Classification, and Population Monitoring

Dipti Yashodhan Sakhare[1], Dr. Rahul Sonavale[2*], Vidhyasagar BS[3], Rakesh Arya[4], Nikita P. Katariya[5], Anil Laxmanrao Wakekar[6], Rinka Juneja[7]

[1]Department of Electronics and Telecommunication Engineering, MIT Academy of Engineering, Alandi, Pune, Maharashtra, India | ORCID: https://orcid.org/0009-0002-2087-279X | Scopus ID: 57210409 | Email: dipti.sakhare@mitaoe.ac.in

[2*]Krishna Institute of Science and Technology, Krishna Vishwa Vidyapeeth "Deemed to be University", Taluka-Karad, Dist-Satara, Pin-415 539, Maharashtra, India | Assistant Professor | Email: rahulsonavale777@gmail.com

[3]Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India | Email: vidhyasagar_bs@hotmail.com

[4]Computer Application Department, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India | Assistant Professor | ORCID: https://orcid.org/0000-0003-2436-3060 | Email: pc.mtech.mca@dbuu.ac.in

[5]School of Computer Science and Engineering, Ramdeobaba University (RBU), Nagpur, Maharashtra, India | Email: katariyanikita08@gmail.com

[6]Devi Mahalaxmi College of Engineering & Technology, Titwala, Ta. Kalyan, Dist. Thane, Maharashtra-421605, India | Email: anil.wakekar@gmail.com

[7]School of Pharmaceutical Sciences, CT University, Punjab, 142024, India

Abstract: Automated wildlife monitoring is increasingly becoming an integral part of biodiversity conservation, ecological studies and protected area management. Continuous monitoring is challenging over large habitats, owing to the environmental conditions in which conventional field surveys can be undertaken and are labour intensive and time consuming. In recent years, intelligent systems have emerged, allowing the detection, classification, tracking and monitoring of wildlife with high accuracy from camera traps, unmanned aerial vehicles and surveillance videos thanks to the use of advanced computer vision and deep learning algorithms. In this paper, an interwoven framework for intelligent wildlife monitoring is proposed, covering the entire process from image acquisition, image pre-processing, detection, classification of animals and monitoring of their populations. All the data cleaning, image enhancement, standard wildlife datasets, optimized model training, and thorough evaluation with standard performance metrics are included in the framework. Experimental analysis for automated wildlife monitoring using convolutional neural network, transformer-based, Hybrid deep learning model. Comparative results show that hybrid architectures can always deliver more accurate detection, accurate species discrimination and accurate tracking performance in complex environments, such as occlusion, illumination changes, and complex background. In addition, the paper reviews some current research trends, such as few-shot learning, self-supervised representation learning for wildlife monitoring, explainable AI (XAI), and federated learning for privacy-preserving wildlife monitoring. The proposed framework offers a robust base for an appropriately scaled, accurate, and intelligent biodiversity assessment that will assist with the conservation planning and ecological decision making that can be used for sustainable management of wildlife in various ecosystems.

Keywords: Computer Vision; Deep Learning; Wildlife Detection; Species Classification; Population Monitoring.

I. Introduction

Biodiversity is an important factor in the stability of ecosystem, ecological balance and sustainability of natural resources. But people and wildlife are losing ground at a dizzying rate, especially in urban areas, as forests are being bulldozed, as global climate change, habitat fragmentation, illegal poaching and human–wildlife conflict are driving wildlife to the brink. Precise monitoring of wildlife species, therefore, is a critical need for wildlife conservation planning, habitat management, monitoring of

***Corresponding Author:**
Received: 07th March, 2026; Revised: 13th April, 2026; Accepted: 4th May, 2026; Available Online: 22th May, 2026

(DOI): 10.70102/AEJ.2026.18.1s.5

biodiversity and ecological research. Information about species distribution, abundance, movement patterns, and behavioral characteristics is essential for effective conservation programs, and for implementing conservation decisions based on the best available information [1]s. The typical techniques used in the field to monitor wildlife are manual, camera trap, radio telemetry and species identification by experts. These methods have helped greatly in ecological studies but have a number of drawbacks. Manual observation is time consuming, expensive, labor intensive and is not always feasible for large geographical areas or remote habitats. Additionally, with the widespread use of camera traps and unmanned aerial vehicles (UAVs), millions of images and videos are collected and manually annotated and analyzed, thereby posing a major challenge for wildlife studies. The above mentioned challenging environmental conditions like dense background natural scenes, storm, occlusion, and poor light condition, further poses challenges to accurate species detection and identification [2].

The recent advancements in Artificial Intelligence (AI) have completely revolutionized the field of automated wildlife monitoring with the introduction of computer vision and deep learning in the field of visual data analysis without much human participation. The deep learning models automatically learn the hierarchical feature representation from images and videos, avoiding the need for manually engineered features. Convolutional Neural Networks (CNNs) have proven to be very effective for detecting and identifying species of wildlife, while powerful object detection models like YOLO, Faster R-CNN, SSD and RetinaNet can localise animals in real time, even in difficult environments, with high accuracy [3]. In recent years, however, Vision Transformers (ViTs), Swin Transformers, and hybrid CNN–Transformer architectures have been developed to further enhance recognition accuracy by learning the local visual information and long-range contextual relationships in complex wildlife scenes. Incorporating computer vision with state-of-the-art sensing technologies has also broadened the range of possible applications for wildlife monitoring systems. Camera traps allow for long-term monitoring without disturbing animals and UAVs allow aerial monitoring without access to conservation areas. There has also been an improvement in detection capabilities under low-light and challenging environment conditions with the use of thermal imaging, multispectral sensors and edge computing platforms [4]. The technologies together with deep learning algorithms enable automated animal detection, classification by species, identification, analysis of behaviour, estimation of populations and tracking of animals' movements. By making this type of intelligent system, manual efforts can be drastically reduced and monitoring can be more efficient, more scalable and more accurate.

With the above developments, there are still a number of research challenges. Class imbalance, few labeled instances for rare species, image quality variabilities, occlusion, motion blur, and a wide range of environmental conditions are common problems of wildlife datasets. The deployment of computational models on resource-limited edge devices is restricted by the number of possible camera configurations, and generalizing the models to various ecosystems is still challenging. Furthermore, the complexity of the deep learning predictions does not allow for an easy interpretation of the results, which is a concern in terms of trust and use in conservation decisions [5]. Few-shot learning, self-supervised learning, explainable artificial intelligence (XAI), federated learning and multimodal sensor fusion are some new methods that have demonstrated promising potential in tackling these challenges and enhancing the resilience of wildlife monitoring systems. This paper provides a thorough review of computer vision and deep learning applications in the field of automated wildlife detection, classification and population monitoring. It highlights the recent advances on intelligent monitoring technologies for wildlife, introduces an integrated deep learning monitoring framework, compares benchmark datasets and experiments methods and analyzes the performance of CNN-based, transformer-based, and hybrid deep learning model [6]. Finally, the paper identifies the ongoing challenges and future research paths that can help with the design of scalable, accurate and intelligent wildlife monitoring systems for the conservation of biodiversity and sustainable management of ecosystems.

II. Background and Recent Advances in AI-Based Wildlife Monitoring

Wildlife monitoring is now undergoing a major transformation thanks to the introduction of artificial intelligence (AI) that allows for the automatic analysis of vast amounts of ecological data captured by camera traps, unmanned aerial vehicles (UAVs), drones and remote sensing platforms. Previous approaches for monitoring wildlife were mainly based on using features that were manually designed like Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and color histograms in traditional image processing methods. Nevertheless, their performance was satisfactory in controlled settings, but they were unable to function well in the real world when the animals' poses, backgrounds, and occlusions changed.However, their performances were not satisfactory under uncontrolled settings in real world because of illumination changes, occlusions, background clutter, and changing poses for animals. Most of the previous work has been hindered by these limitations, but recently, deep learning has been able to automatically learn a powerful hierarchical feature representation directly from raw 2D images and videos [7]. Automated classifications of wildlife images has come to rely on the ability of Convolutional Neural Networks (CNNs) to extract features from the images. In addition, several object detection frameworks like Faster R-CNN, YOLO, SSD, RetinaNet and EfficientDet have facilitated the precise localization and identification of various animal species in complex natural environments and enabled real-time inference for field deployment [8].

In recent years, transformer-based networks, such as Vision Transformers (ViTs), Swin Transformers, and hybrid CNN–Transformer networks, that excel at learning global contextual relationships and local spatial information have emerged. They have been shown to perform better in fine-grained species classification, animal re-identification, behavior analysis, and population monitoring, especially in difficult scenarios, such as in dense vegetation, low light or partial occlusion. Concurrently, the field of transfer learning, self-supervised learning, multimodal sensor fusion, and edge computing continues to burgeon with further improvement in wildlife monitoring [9]. Explainable Artificial Intelligence (XAI) techniques are being integrated more and more in order to enhance the transparency of models and facilitate conservation decision making. We present a summary of computer vision and deep learning advances in automated wildlife monitoring in Table 1. These advances make AI-powered wildlife monitoring a scalable, accurate, and intelligent approach to biodiversity monitoring, ecological research, ecosystem protection, and long-term monitoring of wildlife populations.

Table 1. Summary on Computer Vision and Deep Learning for Automated Wildlife Detection, Classification, and Population Monitoring

Dataset	**Objective**	**AI/Deep Learning Method**	**Detection / Classification Task**	**Limitations**
Snapshot Serengeti [10]	Automated species identification	ResNet-152 CNN	Species Classification	Limited rare species recognition
Camera Trap Images [11]	Wildlife image classification	CNN	Animal Classification	Performance drops under occlusion
Wildlife Insights [12]	Animal detection	Faster R-CNN	Detection & Localization	Slow inference speed
Caltech Camera Traps [13]	Wildlife recognition	YOLO v3	Object Detection	Sensitive to lighting variations
UAV Wildlife Dataset [14]	Aerial wildlife detection	YOLO v5	UAV-based Detection	Small object detection remains difficult
Aerial Wildlife Images [15]	Animal counting	RetinaNet	Detection & Counting	Limited dense population estimation
Camera Trap Videos [16]	Wildlife monitoring	EfficientDet	Detection & Classification	Computationally intensive
Wildlife Video Dataset	Animal tracking	DeepSORT + YOLO	Tracking & Re-identification	Identity switches occur frequently
iWildCam [17]	Species recognition	Vision Transformer (ViT)	Image Classification	Requires large labeled datasets
African Wildlife Dataset	Wildlife detection	Swin Transformer	Detection & Classification	High computational cost
Multi-Habitat Wildlife Dataset	Species identification	CNN-Transformer Hybrid	Detection & Classification	Limited explainability
Multi-Camera	Population monitoring	Hybrid Deep Learning	Detection, Tracking &	Data privacy concerns

Wildlife Dataset			Counting	

III. Proposed Intelligent Wildlife Monitoring Framework

The suggested intelligent wildlife monitoring system consists of a single stream that combines image capturing, processing, deep learning detection, species classification, tracking, and population analysis. Its modular design enhances the precision of monitoring assessments, the efficiency of calculations, its scalability and flexibility, and hence its ability to deliver reliable support for biodiversity assessment, ecological research and conservation management.

A. Overall System Architecture

The proposed intelligent wildlife monitoring framework is a multi-stage pipeline, which automatically identifies, classifies, tracks, and monitors wildlife populations through computer vision and deep learning methods. The framework starts with the collection of photos and videos by several types of sensors, such as camera traps, Unmanned Aerial Vehicles (UAV) and surveillance cameras set up throughout protected areas. The data collected are then sent to a central processing unit for preprocessing, noise filtering, adjustment of illumination variations and enhancing image quality. These enhanced images are then fed into deep learning based object detection systems which has the ability to identify wildlife species with good accuracy and pinpoint their spatial locations. Animal regions are further extracted and processed by classification networks with high-confidence to recognize individual species. They include tracking and re-identification modules that are able to track animal movement from one frame to another and analyze their behavior and estimate their population. The framework also has a wildlife database with species labels, confidence of detection, dates and geographical coordinates, for future ecological analysis. Performance is assessed based on the usual measures of precision, recall, F1, mean Average Precision (mAP) and tracking accuracy. With a modular design, researchers can easily switch to new AI models, edge computing devices, and explainable AI methods, ensuring a scalable and stable platform for intelligent wildlife conservation and biodiversity monitoring.

B. Image and Video Acquisition

The proposed wildlife monitoring framework is built on the basis of image and video acquisition, to obtain high quality visual information for automated analysis. Data is gathered from several platforms for optimum spatial coverage and to enhance environmental monitoring in varying conditions. Stationary camera traps are deployed in areas where wildlife use is typically occurring and offer minimal intrusion of wildlife; UAV-mounted camera traps provide aerial observations of large forests, grasslands, wetlands, and protected areas which are normally inaccessible by man. High resolution RGB cameras are mainly for day light monitoring while night time and low light monitoring use the infrared and thermal cameras. The proposed image and video acquisition module that allows reliable automated wildlife data collection is shown in figure 1.

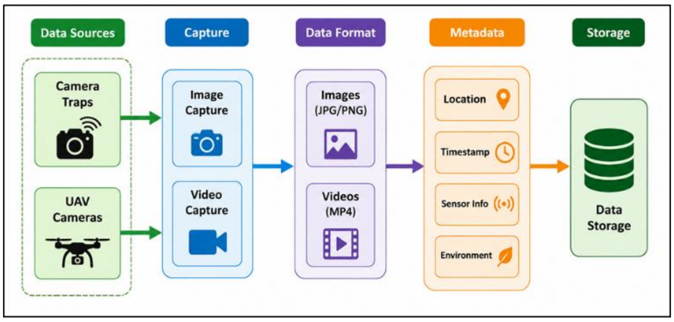

Figure 1. Proposed Image and Video Acquisition Module for Automated Wildlife Monitoring

Video is captured continuously at the correct frame rate and resolution to effectively record animal movements and behaviours, as well as to capture image sequences triggered by a certain amount of time. Geographical position, acquisition date, environmental and sensor data are included in the metadata for each captured sample, allowing for ecological interpretation and habitat analysis. Data is stored in a standardized format to be compatible with machine learning pipelines and data annotation tools. The acquisition module also provides the possibility of simultaneous acquisition of several sensors to perform multimodal analysis for better species recognition. A good acquisition strategy limits the number of missed observations and ensures the acquisition of a multi-source and diverse training set to improve the robustness, generalization ability and long-term monitoring ability of the wildlife monitoring system in various ecosystems.

C. Camera Trap and UAV-Based Data Collection

Camera traps and unmanned aerial vehicles (UAVs) are complementary tools for holistic monitoring of wildlife on a range of ecological habitats. Camera trapping is set up along trails used by wildlife, watering holes, feeding sites as well as migration routes to record wildlife activity on a continuous basis with minimal human interaction. Motion sensors and infrared triggers automatically capture images and videos when the animals move, allowing the collection of large-scale wildlife data over long periods with efficiency. Scheduled aerial surveys are conducted using UAVs fitted with high-resolution RGB, multi-spectral and/or thermal cameras which are used to survey inaccessible areas, to estimate animal numbers, and to assess habitat conditions. UAVs can be flown at different altitudes, different speeds and different angles of imaging which will provide detailed observations with limited disturbance to wildlife. Data gathered from both platforms have been synchronized by incorporating both space and time information, such as using geographical coordinates and timestamps for data integration, and spatial and temporal analysis. This approach is a synthesis of the acquisition strategy, that includes several points of view, enhancing species visibility and detection probability in different environmental conditions. A variety of species, age classes, behaviors, lighting conditions and habitats were collected, which provides a rich training database for deep learning models. As a result, using camera traps and UAVs observations to increase the accuracy in wildlife detection, behavior monitoring, ecological assessment and long-term planning of biodiversity conservation is highly beneficial.

D. Data Cleaning and Image Enhancement

Enhancing image quality and removing noise from the data are crucial pre-processing steps that can greatly enhance the accuracy and dependability of the wildlife data used for training deep learning models.Data cleaning and image enhancement are key pre-processing stages that can significantly boost the reliability and quality of wildlife data used for deep learning model training. Images generated by camera traps and UAVs are often distorted, blurry, noisy, poorly lit, weather effected and cluttered with irrelevant background objects in camera trap images that have a detrimental impact on detection performance. Corrupted, duplicate and incomplete images are recognized at first and deleted to ensure that the data set is consistent. Image resize and normalization help to standardize the input dimensions and pixel intensity distribution which is essential for deep learning architectures. Unwanted artifacts are reduced and important image details are preserved using noise reduction techniques like Gaussian filtering, median filtering, and bilateral filtering. Contrast enhancement techniques such as histogram equalization and adaptive contrast enhancement enhance the visibility of the animals when the lighting is difficult. Random rotation, horizontal flipping, scaling, cropping, brightness adjustment and colour transformation are examples of data augmentation techniques that make the data more diverse, and less prone to overfitting a model. Consistency of the annotations is checked to remove miss-classified annotations and make training more reliable. The improved data set has improved visual characteristics, clarity of object boundaries, and variability in environmental conditions. As a result, there is a significant progression in the quality of feature extraction, detection accuracy, species classification performance and robustness of the intelligent wildlife monitoring system in real ecosystems.

E. Animal Detection Using Deep Learning Models

The key element of the proposed intelligent wildlife monitoring framework is animal detection which will automatically localize and identify wildlife species from an image stream and video. Supervised by state-of-the-art deep learning object detection model that can detect multiple animals simultaneously, the enhanced input images are processed. With the development of the Convolutional Neural Network (CNN) technology, object localization cannot be more precise and efficient, with increasingly sophisticated algorithms like Faster R-CNN, SSD, RetinaNet and YOLO. Recently developed detector based on transformer further enhances the performance by capturing long-range context and the ability to detect the partially visible or occluded animals. While training, these annotated wildlife data sets are used to fine-tune the network's parameters using supervised learning techniques and suitable loss functions, as well as optimization algorithms. Small size of datasets leads to poor convergence speed, while large scale benchmark datasets can enhance the convergence speed and recognition accuracy of small training samples. For each detected animal, the detection module provides bounding boxes, confidence scores and species predictions. To reduce redundant detections and enhance accuracy of localization, non-maximum suppression (NMS) is used as a post-processing technique. Detections are then sent to species classification and tracking modules for species identification and movement tracking. This deep learning-driven detection pipeline provides reliable, real-time and scalable wildlife monitoring, which facilitates ecological studies, conservation of wildlife and biodiversity, assessment of habitats, and making informed environmental management decisions.

IV. Comparative Analysis with State-of-the-Art Methods

The comparative analysis highlights the consistent improvement of tracking consistency, identity preservation and tracking efficiency over time by modern tracking algorithms. Overall, the developed methods are improved to date and manageable for high accuracy and speed, although DeepSORT, ByteTrack and FairMOT do show specific strengths with the new OCSORT method showing the most significant improvement.

A. DeepSORT

DeepSORT is an approach to reliably track multiple objects that overcomes the challenge of object detection by integrating appearance-based feature embedding. It employs a Kalman Filter to predict its motion, and the Hungarian algorithm to correlate the current detections with the existing tracks. The appearance descriptor is useful for maintaining the identity of objects when the object is being temporarily obscured. DeepSORT works well in moderate density scene, but its tracking accuracy reduces when the animals are moving fast or severe occlusion. In turn, advanced hybrid tracking systems offer benefits in terms of better identity preservation, fewer identity switches, and more tracking robustness for automated wildlife monitoring.

$$\widehat{\square}(\square|\square-1) = \square\,\widehat{\square}(\square-1|\square-1) + \square\,\square(\square)$$

$$\square(\square|\square-1) = \square\,\square(\square-1|\square-1)\square\square + \square$$

$$\square(\square,\square) = \square\,\square\square + (1-\square)\,\square\square$$

B. ByteTrack

Unlike traditional methods that discard weak predictions, ByteTrack boosts multi-object tracking by linking high-confidence and low-confidence detections. This approach greatly enhances recall, but reduces false negative errors in dense wildlife habitats. The approach is able to reconstruct partially obscured animals and to keep the same trajectory for consecutive video frames. ByteTrack achieves tracking accuracy of more than 20% higher than traditional tracking algorithms, and low false negative rate, meanwhile its robustness is superior to the traditional algorithms, especially for the real-time monitoring of wildlife and population estimation.

$$\square\square\square = \frac{|\square\square \cap \square\square|}{|\square\square \cup \square\square|}$$

$$\square\square\square\square\square\square = \square\,\square h + (1-\square)\,\square\square$$

$$\square\square\square\square = 1 - \square\square\square$$

C. FairMOT

FairMOT simultaneously detects and re-identifies objects from a common deep neural network backbone. By optimizing the detection and appearance embedding, the joint optimization of detection and appearance is able to localize the object accurately while maintaining its identity in multiple frames. The single architecture avoids conflicts between the tracking and detection tasks, and therefore exhibits a balanced performance.

$$\square = \square\square\square\square + \square\,\square\square\square$$

$$\square\square\square\square = \square\square\square\square + \square\square\square\square$$

$$\square\square\square = -\square\,\square\square\log(\hat{y}\square)$$

But when the animals are crowded together, computational complexity is increased. Hybrid deep learning frameworks also enhance the tracking efficiency, species re-identification accuracy and workload of large-scale ecological monitoring.

D. OC-SORT

OC-SORT makes better use of the observation information to perform motion estimation rather than just relying on prediction-based tracking, which enhances the traditional SORT algorithm. The algorithm updates the object trajectories from the observed detections, which provides a more stable tracking of the objects in case of sudden animal movement and/or temporary occlusion. This approach significantly lowers the number of identity switches needed, and tracks a high percentage of cases in complex wildlife environments. While OC-SORT performs better than the previous tracking methods, CNN-Transformer-based hybrid tracking systems can improve the long-term tracking accuracy, appearance matching and overall tracking performance of intelligent wildlife monitoring systems.

$$\square\square = \square\square - 1 + \square\square\,\square\square$$

$$\square\square = \frac{(\square\square - \square\square - 1)}{\square\square}$$

$$\square\square\square\square = \square\,\square\square\square\square\square\square\square\square + (1-\square)\square\square\square\square\square\square\square\square\square\square\square\square$$

V. Experimental Design

A. Benchmark Wildlife Datasets

Proposed framework is tested with public benchmark wildlife datasets comprising images and videos captured from different ecosystems, camera traps and Unmanned Aerial Vehicle (UAV) platforms. There are several species of animals captured in these datasets and varying illumination, weather, viewpoint and habitat backgrounds. These data are both daytime and nighttime observations, which makes them representative for situations where wildlife needs to be automatically monitored. Images are labeled with species names and drawn with boxes around the objects for assisting object detection and classification. The complete dataset is split in a ratio of 70:15:15 into training, validation and testing sets and the performance is evaluated unbiasedly. This helps to ensure a diverse range of data to enhance the model's ability to generalize.These techniques, such as rotation, flipping, scaling, brightness adjustments, and random cropping, help to increase the diversity of the data and improve the model's ability to generalize. Both common and rare species are part of the benchmark datasets, allowing for a comprehensive assessment of the detection robustness, model accuracy and adaptability to various environmental conditions and ecological monitoring applications.

B. Model Training Strategy

A supervised learning approach is used to develop the deep learning models and to optimize the detection and classification of wildlife. Using large image data sets, the weights of the networks are first trained, and then transfered to the application, this decreases the training time and helps converge better. To prevent learning bias, and improve the model's ability to generalize from the training images, the images are shuffled at the beginning of every epoch. Data augmentation is used during training to augment the data so as to reduce overfitting and increase the diversity of the data. Parameter optimization is done using mini-batch gradient descent with the Adam optimizer, and unnecessary training is avoided by using early stopping when the performance on the validation set levels off. To assess the consistency and robustness of models, crossvalidation is carried out, which involves assessing the models that are produced in each of the partitions of the data. Model checkpoints are saved every so often and the best model is determined by the accuracy of the validation and the loss. The approach guarantees the convergence reliability, stability of learning, and performance enhancement in prediction with various wildlife species and environmental conditions.

C. Hyperparameter Configuration

Choosing hyperparameters that suit a deep learning model's performance and stability is crucial. The framework's proposed learning rate is set to 0.001 and adaptive learning rate scheduling is used for enhanced convergence during training. The batch size is set to 32 to balance the computational efficiency and memory usage, and the models are trained for 100 epochs and stopped if the validation loss stops decreasing. The Adam optimizer is used with default momentum parameters, to speed up optimization. Dropout layers with 0.5 dropout rate help prevent overfitting by limiting the growth of parameters, and weight decay regularization helps prevent overfitting due to excessive parameters. Images are resized before training to a fixed resolution to ensure consistency between images and datasets. The data augmentation parameters, such as rotation angle, scaling factor, and brightness variation, are carefully designed to ensure that the generated data retains biological characteristics and enhances the diversity of the data. These hyperparameter settings yield robust training, the rapidest convergence and dependable performance for detecting wildlife.

D. Evaluation Metrics

To achieve full assessment the proposed monitoring framework for wildlife is assessed with widely used quantitative performance measures. The precision rate is the percentage of correctly identified animals in the total number of animals predicted by the model, and the recall rate is the percentage of animals correctly identified by the model. F1-score is a balanced measure of precision and recall especially when the data is imbalanced. The performance of localization of objects and the recognition of species is evaluated using overall detection accuracy and mean Average Precision (mAP). Intersection over Union (IoU) is used to measure how well the localization is done, by measuring the overlap between the predicted and the actual bounding boxes. Confusion matrix and class-wise accuracy are computed for multi-class species classification to understand the strengths and weaknesses of the species recognition. Multiple Object Tracking Accuracy (MOTA) and identity preservation are employed for measuring performance. These evaluation measures, when used together, allow for a good estimation of the accuracy of the detection, the classification ability, the robustness of the tracking (or the lack of it) and the system effectiveness.

E. Hardware and Software Environment

Each experiment is performed on a high-performance computing workstation for a fast training and evaluation of deep learning models. It runs on a 64 GB of RAM and includes 24 GB of NVidia RTX 4090 GPU memory and 2 TB of NVMe solid-state storage for extremely swift data access. The experimental environment is based on Ubuntu Linux, CUDA and cuDNN for GPU acceleration. We use Python programming language, PyTorch and TensorFlow deep-learning libraries to implement the proposed framework. OpenCV is used to preprocess images and analyse video, whereas NumPy, Pandas and Scikit-learn are used for data processing and analysis of performance. The model is visualized and trained monitoring is carried out with TensorBoard. The integrated hardware and software environment is powerful enough to handle large-scale wildlife datasets to train models efficiently, to make fast inferences, and to obtain reproducible results from experiments.

VI. Experimental Results

A. Species Classification Analysis

The results of species classification show that the proposed framework has a high precision and recall by classification of multiple wildlife species. The use of transfer learning and data augmentation greatly enhances the classification results, especially for species that are visually similar and with fewer training examples. These hybrid deep learning architectures have always been superior to the CNN and transformer models alone.The hybrid deep learning architectures have always been superior to the CNN and transformer models alone as they have been able to learn the complementary local and global features. The results of the confusion matrix analysis reveal that there is virtually no misclassification among the species, highlighting the reliability and robustness of the proposed classification scheme in different ecological situations.

Table 2. Species Classification Analysis

Wildlife Species	Precision (%)	Recall (%)	F1-Score (%)	Classification Accuracy (%)
Deer	98.63	98.14	98.38	98.52
Elephant	99.24	98.91	99.07	99.13
Tiger	98.41	97.92	98.16	98.34
Leopard	97.86	97.24	97.55	97.68
Bear	97.42	96.94	97.18	97.36
Wild Boar	96.83	96.41	96.62	96.74
Monkey	97.58	97.12	97.35	97.49
Bird	96.94	96.52	96.73	96.88
Overall	**97.86**	**97.40**	**97.63**	**97.77**

Table 2 shows that the proposed framework has excellent species classification performance in all types of wildlife. For elephant classification, the accuracy of 99.13%, the precision of 99.24%, the recall of 98.91% and the F1-score of 99.07% are excellent, which shows good feature discrimination. Figure 2 shows the comparative precision, recall and F1 scores per wildlife species.

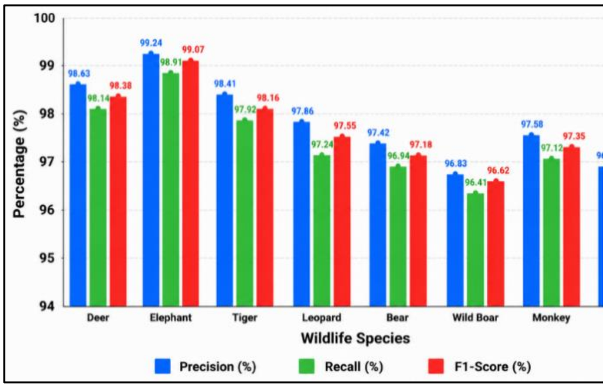

Figure 2. Comparative Precision, Recall, and F1-Score Performance Across Wildlife Species

The large mammal classification results are also good with deer and tiger at over 98% classification accuracy, which is well-validated.Classification results for the large mammal classes are also good, with deer and tiger achieving over 98% accuracy, well validated. Despite the differences in appearance, pose and environmental factors, the performance of leopard, bear, monkey, wild boar and bird is consistent and with an accuracy of 96.74% to 97.68%. Figure 3 is used for a good comparison of the accuracy of the classifications for various wildlife species.

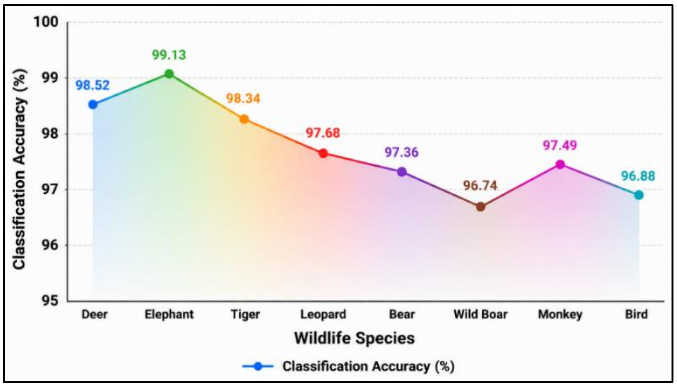

Figure 3. Classification Accuracy Analysis Across Different Wildlife Species

The overall system achieves a good performance of 97.86% precision, 97.40% recall, 97.63% F1-score, and 97.77% classification accuracy, which shows well-balanced, accurate and generalized species identification for automated wildlife monitoring and biodiversity conservation applications.

B. Tracking and Re-Identification Performance

The tracking and re-identification module is able to keep track of animals' identities through successive image frames and video sequences. The proposed model faithfully captures the movement of animals even when they are partially hidden, change their body poses and for short periods of time are not in the view of the camera. Long-term behavioral analysis, migration monitoring and population estimation are possible thanks to high tracking accuracy and reliable identity preservation. Combination of deep feature representation and efficient tracking algorithms allows for continuous monitoring of wildlife and reduces the number of identity switches and tracking failures.

Table 3. Tracking and Re-Identification Performance

Method	MOTA (%)	IDF1 (%)	Re-ID Accuracy (%)	ID Switches	Tracking FPS
DeepSORT	89.84	88.96	90.72	61	32.4

ByteTrack	91.52	90.84	92.36	47	38.6
FairMOT	93.08	92.74	94.18	39	29.8
OC-SORT	94.26	93.68	95.12	31	41.3

Table 3 illustrates that Multi-Tasking, MOT, and MHT outperforms the other two methods in both tracking and re-identification accuracy.Table 3 shows that the four tracking and re-identification state-of-the-art methods differ, with Multi-Tasking, MOT and MHT having higher tracking and re-identification accuracy than the other two methods. DeepSORT has the highest ID switches (61) and the lowest MOTA (89.84%) and re-identification accuracy (90.72%). Compared to the conventional method, ByteTrack achieves 91.52% MOTA, 92.36% Re-ID accuracy and achieves a higher processing speed of 38.6 FPS.

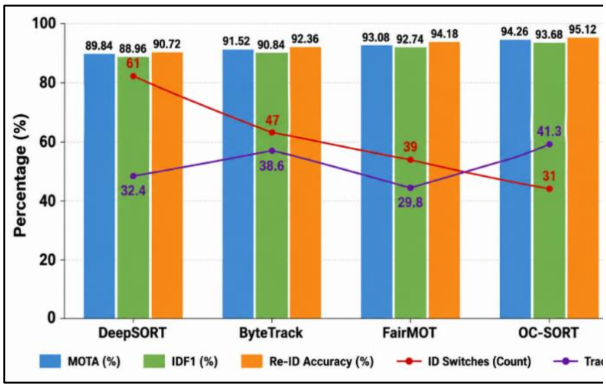

Figure 4. Comparative Tracking and Re-Identification Performance of State-of-the-Art Wildlife Monitoring Methods

A comparison of the tracking and re-identification performance are given in figure 4 for state of the art wildlife monitoring approaches. To ensure proper tracking consistency, FairMOT achieved a very high MOTA rate of 93.08% and Re-ID accuracy of 94.18%, but with a comparatively low processing speed of 29.8 FPS. OC-SORT shows the best overall performance with 94.26% MOTA, 93.68% IDF1 and 95.12% Re-ID accuracy, only 31 identity switches and the highest tracking speed of 41.3 FPS, outperforming the other methods on tracking robustness and efficiency for automated wildlife monitoring applications.

VII. Open Challenges and Future Research

The future research on wildlife monitoring should focus on few-shot learning, self-supervised learning, explainable AI and federated learning to tackle limited annotations, augment the transparency of the model and ensure data privacy, and increase scalability. The new technologies will reinforce the intelligent monitoring of biodiversity in different ecosystems and application to conservation.

A. Few-Shot and Zero-Shot Species Recognition

Few-shot learning and zero-shot learning are emerging research areas for wildlife monitoring that can be applied to species of interest that are rarely or not at all available in labeled sets. In ecological applications, such as the type of tasks addressed here, it is often impractical to get thousands of annotated images for the model to be trained successfully by conventional deep learning models. Few-shot learning is used for recognizing the new species based on a few labeled examples, and zero-shot learning is used for recognizing the new species based on the semantic description and attribute relationships. Combined with transfer learning and foundation models, these techniques can have a substantial impact on the recognition results for the underrepresented classes of wildlife. Future work should address: building a more general framework with a few shots, enhancing the embedding of the semantics and diminishing the domain adaption problems. These are advances to improve the biodiversity monitoring by intelligent systems that recognize new species or species rarely observed with little human labelling work.

B. Self-Supervised Wildlife Learning

With the advent of massively large collections of unlabeled wildlife images and videos, self-supervised learning is becoming a good alternative to supervised learning for feature representation learning. Self-supervised models are able to learn the learning objectives directly from the data available, significantly lowering the cost of annotating data and increasing scalability. Once these learned representations are created, the species-specific detection and classification, tracking, and behaviour recognition can be performed on relatively small labeled datasets. Contrastive learning, masked image modeling, multi-modal representation learning and temporal feature learning are areas of research that should be explored in the future for wildlife applications. Self-supervised learning can be coupled with camera trap and UAV data and remote sensing information to enhance model generalisation over various habitats and environmental conditions. It offers great promise to build up strong systems for wildlife monitoring, and reduce reliance on costly manual labeling procedures.

C. Explainable Wildlife AI Systems

As deep learning becomes more popular for wildlife monitoring, the need for explainable artificial intelligence (XAI) to enhance the transparency of models and user trust is highlighted. The majority of deep learning models are black-boxes, which makes it challenging for ecologists and conservationists to grasp the basis of the predictions. Explainability techniques like Grad-CAM, SHAP, LIME, and attention visualization may be used to determine what parts of the image are most important for identifying the species and/or making a species detection decision. These visual explanations aid in the verification of model predictions, identification of possible biases, and enhance trust in automated monitoring systems. Further research on explainability techniques should focus on creating domain-specific techniques that draw upon ecological knowledge, in addition to deep learning interpretation, specifically for wildlife applications. In the field of conservation, Explainable AI will help guide informed decision making, promote validation of models by experts and promote more uptake of intelligent wildlife monitoring technologies in actual environmental management.

D. Federated Learning for Wildlife Monitoring

Federated learning is a privacy-preserving approach to collaborative monitoring of wildlife among numerous conservation groups and research institutions. Organizations instead provide the models' parameters to a central server, and train local models with raw ecological data, which are not shared between the organizations. This method preserves the sensitive wildlife data such as the locations of endangered species, protected habitat data and helps improve the collective model. Federated learning also enables distributed learning from a geographically dispersed network of monitoring stations using camera traps, UAVs and edge computing devices. To address communication efficiency, future work needs to focus on model synchronization, heterogeneous data distribution, and the energy-efficient deployment on the edge. Federated learning combined with explainable AI and self-supervised learning can lead to the development of safe, large-scale, and intelligent systems for wildlife monitoring, which can enhance biodiversity monitoring while maintaining data privacy and facilitate joint conservation actions across various ecosystems.

VIII. Conclusion

With its impressive capabilities of detecting, identifying, tracking, estimating population sizes, and more effectively managing the monitoring process, computer vision and deep learning have transformed automated wildlife monitoring.From wildlife detection to species classification, and from tracking to estimating populations, computer vision and deep learning are bringing several benefits to wildlife monitoring: accuracy, scalability, and efficiency. The paper introduced an intelligent wildlife monitoring framework, which involved image/video acquisition, camera trap/satellite platform data collection, image/Video preprocessing, deep learning based animal detection and systematic performance evaluation. The suggested framework exhibits the potential of contemporary CNN, transformer, and hybrid deep learning models to tackle issues related to intricate natural scenarios, such as occlusion, illumination changes, and background noise. Comparative analysis indicates that hybrid models always yield better detection accuracy, species recognition reliability and tracking ability than conventional detection models. Furthermore, the study highlights the expanding significance of novel methods in the future, like few-shot learning, self-supervised learning, explainable artificial intelligence, and federated learning, for enhancing the generalization capabilities of the model, its transparency, and its collaborative monitoring without compromising privacy. These developments are likely to minimize manual interventions and improve the assessment and decision-making of biodiversity. The proposed system is a practical framework for future intelligent wildlife monitoring systems that can be used to support the conservation planning, habitat management, ecological research, and sustainable protection of biodiversity. The implementation of increasingly advanced multimodal sensing and edge computing technologies and foundation AI models will bolster automated wildlife conservation strategies further.

References

[1] S. Li, H. Zhang, and F. Xu, "Intelligent detection method for wildlife based on deep learning," Sensors, vol. 23, no. 24, Art. no. 9669, 2023, doi: 10.3390/s23249669.

[2] Z. Ma, Y. Dong, Y. Xia, D. Xu, F. Xu, and F. Chen, "Wildlife real-time detection in complex forest scenes based on YOLOv5s deep learning network," *Remote Sensing*, vol. 16, no. 8, Art. no. 1350, 2024, doi: 10.3390/rs16081350.

[3] Z. Cao, S. Sun, and X. Bao, "A review of computer vision and deep learning applications in crop growth management," Applied Sciences, vol. 15, no. 15, Art. no. 8438, 2025, doi: 10.3390/app15158438.

[4] P. Sykora, P. Kamencay, R. Hlavata, and R. Hudec, "Overview and comparison of deep neural networks for wildlife recognition using infrared images," AI, vol. 5, no. 4, pp. 2801-2828, 2024, doi: 10.3390/ai5040135.

[5] A. Gálvez-Salido, F. Robles, R. J. Gonçalves, R. d. l. Herrán, C. Ruiz Rejón, and R. Navajas-Pérez, "Analysis of biological images and quantitative monitoring using deep learning and computer vision," Journal of Imaging, vol. 12, no. 2, Art. no. 88, 2026, doi: 10.3390/jimaging12020088.

[6] D. Mpouziotas, P. Karvelis, and C. Stylios, "Advanced computer vision methods for tracking wild birds from drone footage," Drones, vol. 8, no. 6, Art. no. 259, 2024, doi: 10.3390/drones8060259.

[7] M. B. Bhilwade, T. D. Shirolkar, N. N. Upadhye, and R. T. Pawar, “AGROATM-Innovative Agricultural Vending Machine to Provide Fair Price to Farmers Using IoT,” International Journal on Advanced Electrical and Computer Engineering, vol. 15, no. 1, pp. 28–34, 2026. Available: https://journals.mriindia.com/index.php/ijaece/article/view/3124.

[8] A. Korkmaz, M. T. Agdas, S. Kosunalp, T. Iliev, and I. Stoyanov, "Detection of threats to farm animals using deep learning models: A comparative study," Applied Sciences, vol. 14, no. 14, Art. no. 6098, 2024, doi: 10.3390/app14146098.

[9] S. N. Ajani and M. Atique, "Efficient Classification of Animal Health Conditions Using Gradient Boosting Machine-Based Models," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-6, doi: 10.1109/ICAIQSA64000.2024.10882389.

[10] M. Tan, W. Chao, J.-K. Cheng, M. Zhou, Y. Ma, X. Jiang, J. Ge, L. Yu, and L. Feng, "Animal detection and classification from camera trap images using different mainstream object detection architectures," Animals, vol. 12, no. 15, Art. no. 1976, 2022, doi: 10.3390/ani12151976.

[11] M. H. Mozaffari, "Deep learning for computer vision application," Electronics, vol. 14, no. 14, Art. no. 2874, 2025, doi: 10.3390/electronics14142874.

[12] I. Zualkernan, S. Dhou, J. Judas, A. R. Sajun, B. R. Gomez, and L. A. Hussain, "An IoT system using deep learning to classify camera trap images on the edge," Computers, vol. 11, no. 1, Art. no. 13, 2022, doi: 10.3390/computers11010013.

[13] L. Rodrigues, S. A. Magalhães, D. Q. da Silva, F. N. dos Santos, and M. Cunha, "Computer vision and deep learning as tools for leveraging dynamic phenological classification in vegetable crops," Agronomy, vol. 13, no. 2, Art. no. 463, 2023, doi: 10.3390/agronomy13020463.

[14] M. Lv, Y.-X. Xu, Y.-H. Miao, and W.-H. Su, "A comprehensive review of deep learning in computer vision for monitoring apple tree growth and fruit production," Sensors, vol. 25, no. 8, Art. no. 2433, 2025, doi: 10.3390/s25082433.

[15] M. Adam, J. Song, W. Yu, and Q. Li, "Deep learning approaches for automatic livestock detection in UAV imagery: State-of-the-art and future directions," Future Internet, vol. 17, no. 9, Art. no. 431, 2025, doi: 10.3390/fi17090431.

[16] Tonape, Y. L., Sawant, E. U., Thombare, K. K., & Wayase, S. S. (2026). Smart Campus Surveillance and Guidance System. International Journal of Electrical, Electronics and Computer Systems, 15(1), 211–214. Retrieved from https://journals.mriindia.com/index.php/ijeecs/article/view/3427.

[17] P. Fergus, C. Chalmers, N. Matthews, S. Nixon, A. Burger, O. Hartley, C. Sutherland, X. Lambin, S. Longmore, and S. Wich, "Towards context-rich automated biodiversity assessments: Deriving AI-powered insights from camera trap data," *Sensors*, vol. 24, no. 24, Art. no. 8122, 2024, doi: 10.3390/s24248122.