International Research Journal of Multidisciplinary Scope (IRJMS), 2026;7(3):138-152 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i03.09519

Hybrid Wildlife Classification and Detection Using EfficientNet-B4 Custom CNN and YOLOv10

Sushant Suresh Kothari*, Aditi Chhabria, Gargi Phadke

Department of Computer Science and Engineering, Ramrao Adik Institute of Technology, D. Y. Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra, India. *Corresponding Author's Email: sk619kothari@gmail.com

Abstract

In constantly changing natural habitats, the conservation of wildlife and their monitoring pose serious issues that reflect over the ecological study as well as in preservation of biodiversity. For prompt interventions, well-informed choices of policy and also automated ecological surveys, accurate categorization of all species along with proper detection are very crucial. This research makes use of visual data in several different environmental settings that provides a viable deep learning system for animal species identification and also their categorization. To achieve accurate localization and species-level classification, two advanced deep learning models are combined in order to handle both spatial localization within the given image or frame and also for accurate classification: (i) a Custom EfficientNet-B4 Convolutional Neural Network (CNN) optimized for species-level classification and (ii) YOLOv10 for detection of the animal's position with accurate bounding box localization The models make use of an extensive collection of engineering characteristics that includes contextual signals from the surrounding environment, color and texture data along with multi-scale picture representations. For improving resilience to changes in illumination, occlusion as well as background clutter, techniques such as data augmentation and normalization are adapted and used. Varied types of applications including automated camera trap analysis, monitoring of ecology and also animal tracking is made possible by the proposed framework's that emphasises on modularity, interpretability as well as deployment readiness factors. This study provides an intelligent tool that is also scalable for helping conservation efforts and for the evaluation of biodiversity by making use of a combined feature-rich categorization along with precise detection system.

Keywords: Custom CNN, Deep Learning, EfficientNet-B4, Object Detection, Wildlife Classification, YOLOv10.

Introduction

Human encroachment, habitat fragmentation and climate change have created major challenges for wildlife monitoring and biodiversity protection. Accurate species identification is essential for understanding ecosystem health, population dynamics, habitat usage and conservation planning. However, traditional monitoring methods such as field surveys, camera-trap review and manual observation are labour-intensive, time-consuming and prone to human error, limiting their scalability in real-world environments.

Wildlife monitoring is further complicated by environmental variation. Illumination changes, vegetation density, occlusion, animal motion and cluttered backgrounds reduce the reliability of manual inspection and make rapid analysis difficult when large image collections must be processed efficiently. These challenges have encouraged the use of automated computer vision systems for ecological monitoring. Deep learning, particularly CNN-based approaches, can identify subtle visual patterns, while YOLO-based detectors can localize animals in real time within complex scenes. Combining detection and classification therefore provides a more complete wildlife-monitoring framework.

Recent studies have demonstrated strong progress in this field. Improved Top-1 and Top-5 species-recognition performance was reported using a three-branch CNN with VGG16 and VGG19 backbones trained on approximately 11k annotated wildlife images from Ergaki National Park (1). A two-stage CNN pipeline for animal-presence filtering and 48-species classification was trained on nearly 1.2 million volunteer-annotated images (2). Strong precision and recall were obtained in YOLOv8-based wounded-animal detection, although highly imbalanced labels remained challenging (3). EfficientNet-B0 demonstrated

(Received 04th December 2025; Accepted 02nd June 2026; Published 02nd July 2026)

improved computational efficiency compared with VGG16, DenseNet121 and InceptionV3 for Big-4 snake detection (4). Real-time wildlife–vehicle accident prevention was supported through YOLOv5 with MS-COCO transfer learning, where mAP@0.5 near 94% was reported (5). Detection, classification, IoT sensing and ultrasonic deterrence were integrated into a wildlife-protection pipeline with detection performance near 95% and species accuracy near 92% (6). Strong intrusion-detection capability was demonstrated with YOLOv9, where precision, recall and mAP values near 0.986, 0.964 and 0.981 were reported across 15 classes (7). On a large camera-trap dataset, animal-presence detection and multi-class identification were evaluated using Lite-AlexNet, VGG-16 and ResNet-50, where VGG-16 showed strong detection performance (8). A systematic review indicated that Top-1 accuracy can range from around 35% to 96.6%, while Top-5 accuracy can reach approximately 98.4% depending on architecture and deployment conditions (9). Integrated environmental response systems have also been introduced for wildlife relocation, fire detection, mitigation and alert generation using deep learning and IoT-based architectures (10).

Low-power and edge-deployable wildlife-monitoring systems have also gained attention. Long-range, low-latency animal monitoring with GPS tagging was enabled through an Efficient Det-based LoRa and Raspberry Pi system (11). Real-time low-power edge deployment was achieved by modifying YOLOv10 for NVIDIA Jetson Nano, where mAP@0.5 of 0.934 and accuracy of 0.928 were obtained (12). Differentiation among animal, domestic and human targets was enabled through a YOLOv5 pipeline with PIR and ultrasonic sensing while maintaining low-latency CPU inference (13). Spatial feature extraction and temporal modeling were combined in a CNN–BiLSTM–TCN architecture for smart-farming intrusion detection, where field accuracy near 98% was achieved (14). A lightweight depth-wise separable CNN produced classification accuracy near 99.6% for IWildCam images (15). Wildlife detection, tracking and behavioral analysis across six species were supported by YOLOv8m, DeepSORT and LSTM-based behavior modeling (16). Wildlife detection and recognition in digital images were benchmarked with YOLOv3 and YOLOv3-Tiny, where the larger model achieved higher mAP while the tiny variant provided faster inference (17). Conventional CNNs with standard augmentation were also shown to achieve high wildlife-classification performance (18).

Additional advances have been reported with two-stage and compressed architectures. Wildlife monitoring based on Faster R-CNN achieved detection accuracy near 92.78% using region proposal networks and multi-scale anchoring (19). Infrared wildlife recognition was evaluated across VGG, ResNet50, Xception, MobileNet and DenseNet, where Xception and ResNet50 provided strong test accuracy while MobileNet emphasized efficiency (20). Deer-crossing Road detection with roadside LiDAR achieved accuracy near 99.8% (21). Efficient wildlife detection was further improved through a lightweight Conv-Swin backbone, where Faster R-CNN inference speed increased by 17.5% (22). Real-time wildlife tracking and anomaly detection were also supported through YOLOv8-based systems deployed on low-cost devices (23). Wildlife re-identification was improved using a lightweight deformable-transformer framework with multi-image feature fusion (24). Wildlife containment and human-safety monitoring were enabled through a Flask-integrated YOLOv5 boundary-monitoring system (25). Sensitivity improvement in SDR telemetry was reported for small-wildlife tracking (26). Night-time wildlife detection was strengthened by CLAHE-Retinex-YOLOv5 preprocessing (27), while real-time alert generation through email and messaging was enabled in edge-based wildlife-monitoring systems (28). A low-power FPGA-based detection process was developed for wildlife surveillance in constrained environments (29). Fine-grained bird classification was achieved through a combination of YOLOv8x and EfficientNet-B7, where mAP near 0.92 and F1 near 0.87 were reported (30). High-efficiency progressive transmission and automatic recognition of wildlife monitoring images were supported through compressed Faster R-CNN processing in wireless image sensor networks (31). TinyML-based edge IoT frameworks were further shown to safeguard crops from wildlife threats with compact model footprints and high inference efficiency (32). Environmental awareness related to biodiversity protection was investigated through wetland conservation studies

(33), while bioacoustics approaches contributed to biodiversity monitoring and conservation (34). Real-time dynamic wildlife monitoring was further supported by YOLOv11n, where mAP near 97.4% and inference speed near 213 FPS were reported (35). For UAV-based infrared wildlife detection, CE-RetinaNet improved average precision by more than 11% (36).

Despite these advances, many existing systems remain either detection-centric or classification-centric and robust performance under cluttered, occluded and highly imbalanced field conditions is still not consistently achieved. Detection-oriented methods such as YOLOv3, YOLOv5, YOLOv8, YOLOv10 and Faster R-CNN offer strong localization but often struggle with fine-grained species recognition in complex ecological scenes. In contrast, CNN-based classifiers frequently rely on cropped or pre-segmented animal images, limiting their usefulness in unconstrained environments. To address these limitations, this study proposes a hybrid detect–crop–classify framework that combines YOLOv10n-based real-time localization with a custom EfficientNet-B4 CNN for refined species-level classification. The framework integrates transfer learning, multi-scale feature extraction, EMA stabilization and augmentation-based regularization to improve robustness, generalization and deployment readiness across 90 animal species while maintaining near real-time inference speed.

Methodology

This methodology describes a hybrid wildlife recognition framework that integrates fine-grained classification and real-time detection to address common field challenges such as occlusion, class imbalance and background clutter. The pipeline covers dataset preparation, preprocessing and augmentation, hybrid model architecture (classification + detection), training strategy, evaluation protocol, model technical specifications, fusion logic, deployment/export pathways and visualization (figures). The main focus is over stable training of deep-network, computational efficiency for edge deployment, transfer learning (ImageNet initialization), fusion of multi-scale feature along with a dual-stream inference-based approach which gives the labels of given species as well as boxed localization for each instance that is detected.

The dataset has about 5,400 verified images of 90 varied species. The images were checked for integrity as well as for any empty class folders before proceeding with the processing step; the count of all class distributions along with usual picture sizes were looked at in order to aid in sampling. Stratified sampling is used for splitting the dataset into a ratio of 80:20 train/validation split (4,320 train/1,080 val) that helps to preserve frequency per-class and also to make sure about the representation of rare species in both of the splits. A consistent input format is applied to each and every image.

By including species from a variety of biological groups, including terrestrial mammals, birds, reptiles, marine animals and insects, the dataset enhances species-level diversity during training. The framework can learn robust visual representations under realistic wildlife-monitoring scenarios thanks to the collection of images that depict a variety of environmental conditions, such as dense vegetation, open habitats, illumination variation, occlusion, pose diversity, scale variation and cluttered natural backgrounds. Before training and validation, every picture was manually arranged into species-specific labeled folders to ensure annotation uniformity and repeatability. While reducing annotation discrepancies throughout the dataset, the structured labeling procedure guaranteed dependable class-wise supervision during the detection and classification phases.

The pipeline is mirrored in preprocessing and augmentations: training transforms use RandomResizedCrop to 380x380, flips horizontally and vertically, rotations up to ±20°, affine transforms, ColorJitter/brightness/contrast /saturation/hue adjustments and normalization to ImageNet mean/std; validation uses resize to 380x380 plus normalization. Label smoothing (0.1) is applied to the classifier loss after MixUp is incorporated in-batch to create convex blends of pictures and labels (β-sampled).

Conv-block dropout (p=0.3), an extra dropout (p=0.5) prior to the final linear layer, batch normalization, optional random erasing to simulate occlusion and an Exponential Moving Average (EMA) of parameters (decay=0.999) for smoother test-time weights are examples of regularization and training stabilizers. The training script makes use of batch size=16,

num_epochs=35, CosineAnnealingLR scheduler (T_max matched to epoch cycle, e.g., 35), AdamW (lr=3e-4, weight_decay=1e-4) and more. In order to identify overfitting or underfitting, training/validation curves (loss/accuracy) are monitored and checkpointing is based on validation F1 improvement.

Since EfficientNet-based models have outperformed VGG16, DenseNet and InceptionV3 in relevant wildlife classification tests and have shown high fine-grained classification performance in animal-recognition tasks, EfficientNet-B4 was chosen as the classifier backbone. Timm's tf_efficientnet_b4_ns backbone (features only=True) with ImageNet-trained weights is the focal point of the classification stream. By generating multi-scale feature maps at various resolution levels, the backbone extracts complex hierarchical representations, allowing the model to concurrently capture global semantic signals, mid-level structures and fine-grained textures. This architecture helps to make sure that downstream layers receive a balanced blend of local as well as global features which are essential in order to provide robust decision making by doing an aggregate of only selected intermediate feature maps via a fusion of multi-scale mechanism that does integration of a broader contextual information and then also preserves spatial detail. In order to stabilize gradients and safeguard pretrained knowledge, the training pipeline employs a two-phase optimization strategy in which the backbone is first frozen for a number of epochs. This is followed by a gradual unfreezing for complete end-to-end fine-tuning, allowing the network to adjust its representations to dataset-specific features.

The classification model's system architecture shown in Figure 1, how 380x380 pre-processed images go into the EfficientNet-B4 backbone, whose multi-stage feature outputs feed a multi-scale fusion module before going into the custom CNN head, which is made up of a fully connected layer, adaptive pooling, a 512-channel reduction block, a 1024-channel convolution block and SoftMax-based prediction. It also shows the optional backbone-freezing step, the interface that exposes contextual aspects to the detection stream and the connection to EMA/test-time stabilized weights.

YOLO-family detectors have been widely used in wildlife surveillance because they provide real-time localization and deployment-friendly inference in monitoring, intrusion detection and edge-based conservation systems. For effective multi-instance bounding-box localization, the detection stream uses YOLOv10 (yolov10n). To enhance detection across varied scales and partial occlusion, YOLOv10 is trained using class-aware sampling, multi-scale jittering and mosaic-like augmentation. YOLO generates boxes, objectness scores and (class-agnostic) localization during inference; boxes that surpass a certain level of confidence are cropped and sent to the classifier. By separating localization from identification recognition, this two-stage detect-crop-classify method improves per-instance classification accuracy in crowded settings and permits reliable various animal handling within single frames.

The training strategy and hyperparameters are determined by the AdamW optimizer (lr = 3e-4, weight decay = 1e-4), CosineAnnealingLR scheduler with warm restarts (T_max tuned per run), batch size = 16, epochs = 35, EMA (decay = 0.999), CrossEntropyLoss with label smoothing = 0.1 and MixUp augmentation incorporated into the training loop. With validation findings of F1 = 0.9631, recall ≈ 0.9630, precision ≈ 0.9668 and final validation accuracy ≈ 0.9630, the runs continuously demonstrated steady convergence. Per-class metrics, confusion matrices and PR curves were assessed to detect inter-class confusion and direct targeted augmentation and data-collection enhancements. The hyperparameter settings were selected experimentally to achieve a balance between convergence stability, computational efficiency and generalization performance. A learning rate of 3e-4 with the AdamW optimizer provided stable gradient updates while avoiding unstable oscillations during optimization. Batch size 16 was selected to balance GPU memory utilization and gradient stability for high-resolution 380×380 wildlife images.

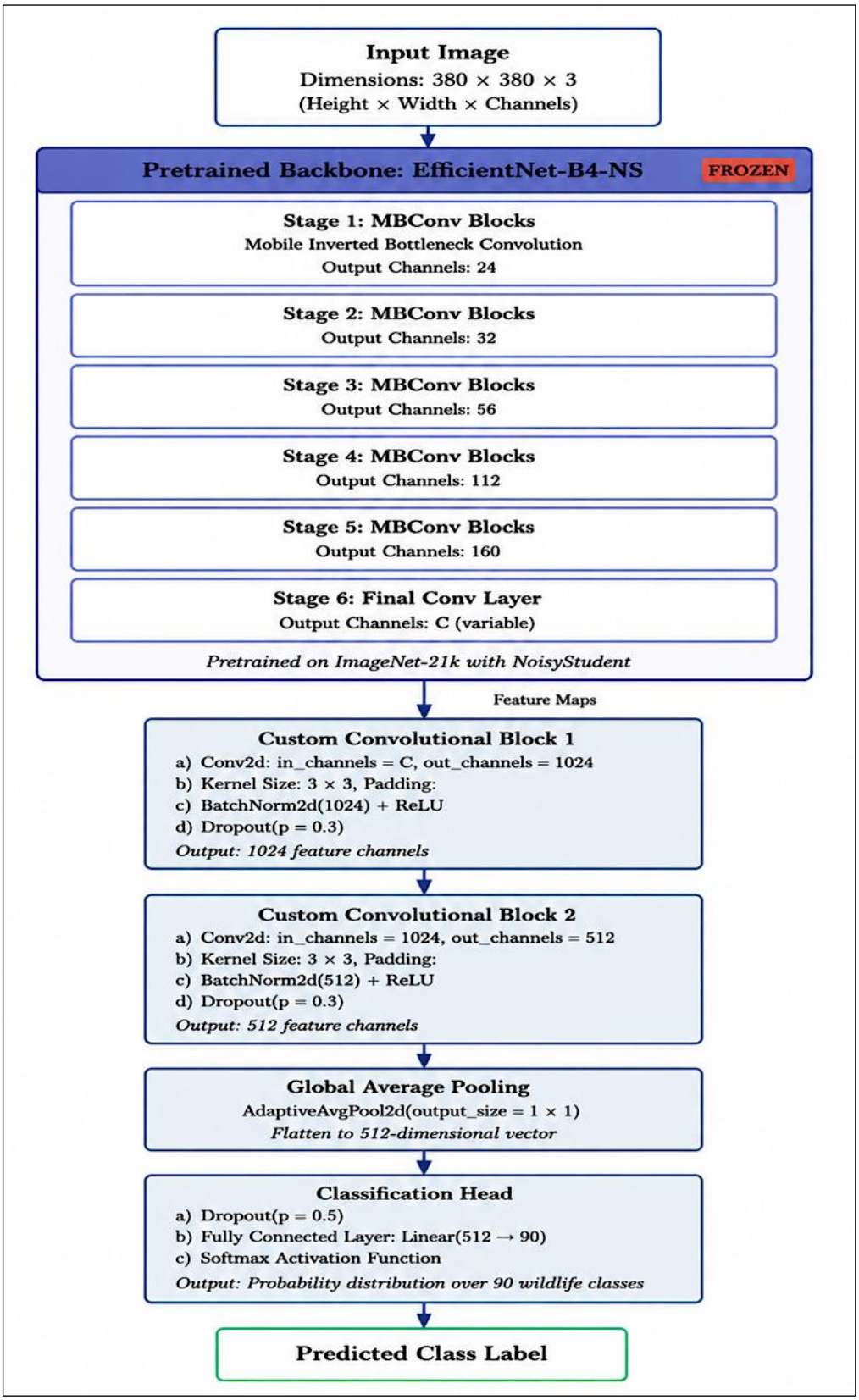

Figure 1: System Architecture of the Proposed Hybrid Wildlife Recognition Framework

The CosineAnnealingLR scheduler enabled gradual learning-rate reduction during later epochs, thereby improving convergence stability and reducing the risk of local minima trapping. EMA stabilization (decay = 0.999) was incorporated to smooth parameter updates and improve test-time

consistency, while label smoothing and MixUp augmentation were used to reduce overconfidence and improve robustness against class imbalance, occlusion and visually similar wildlife species.

The metric-based evaluation is explained in this section. Metrics such as accuracy, precision, recall and F-score. In this context, the terms true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are used.

Four level characteristics are as follows, based on the above four characteristics, accuracy, which measure the ratio of the total number of correct prediction Equation [1]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad [1]$$

Likewise, the precision is a measure of the number of correct classifications penalize by the number of incorrect classifications, as given by Equation [2]:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad [2]$$

Similarly, recall measures the number of correct classifications penalized by the number of recalls missed entries and is represented by the Equation [3]:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad [3]$$

F-score reflects the effect of accuracy and recall, there is no such solution that combines the effect of accuracy, accuracy and recalls in the same function as is given in the Equation [4]:

$$\text{F1-Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \qquad [4]$$

The hybrid wildlife recognition system's last decision layer is the fusion logic and post-processing step. A structured filtering pipeline is used to improve the findings once the detector and classifier have finished their own inference processes. While confidence thresholding eliminates predictions that fall below acceptable reliability margins, Non-Maximum Suppression (NMS) eliminates redundant or overlapping detection boxes by keeping just the highest-confidence areas. The classifier output is then coupled with each validated bounding box by matching its spatially cropped region to the class prediction with the highest degree of confidence. This helps to make sure that the detector produces a precise geographical localization and later the classifier's refined feature representations are then used to award the final identification of species. Important information including name of the identified species, bounding-box coordinates, confidence ratings, exact location metadata are then included within the given final combined output.

For the evaluation process a wide range of metrics are taken into consideration of this hybrid model in order to guarantee performance that is reliable in both common as well as uncommon species. A stratified validation set is used for the computation of weighted accuracy, precision, recall and F1-score which directly helps to maintain class balance and also to avoid skewed assessment which is caused due to dominating species. Breakdowns of performance per-class along with confusion matrix assessments as well as with global metrics, which basically helps in class identification that might need additional augmentation or any modifications within the architecture. The model gives out many deployment routes that are optimized for various kinds of use cases, once the model has been trained and also verified. These include models that are fully precise especially for powerful cloud servers, lighter pruned as well as quantized models for low-power based edge devices such as traps or drones, ONNX models for wide interoperability and TensorRT versions for quick inference by using GPU. When combined, these export formats guarantee that the system may be effectively implemented in centralized, field and edge processing configurations.

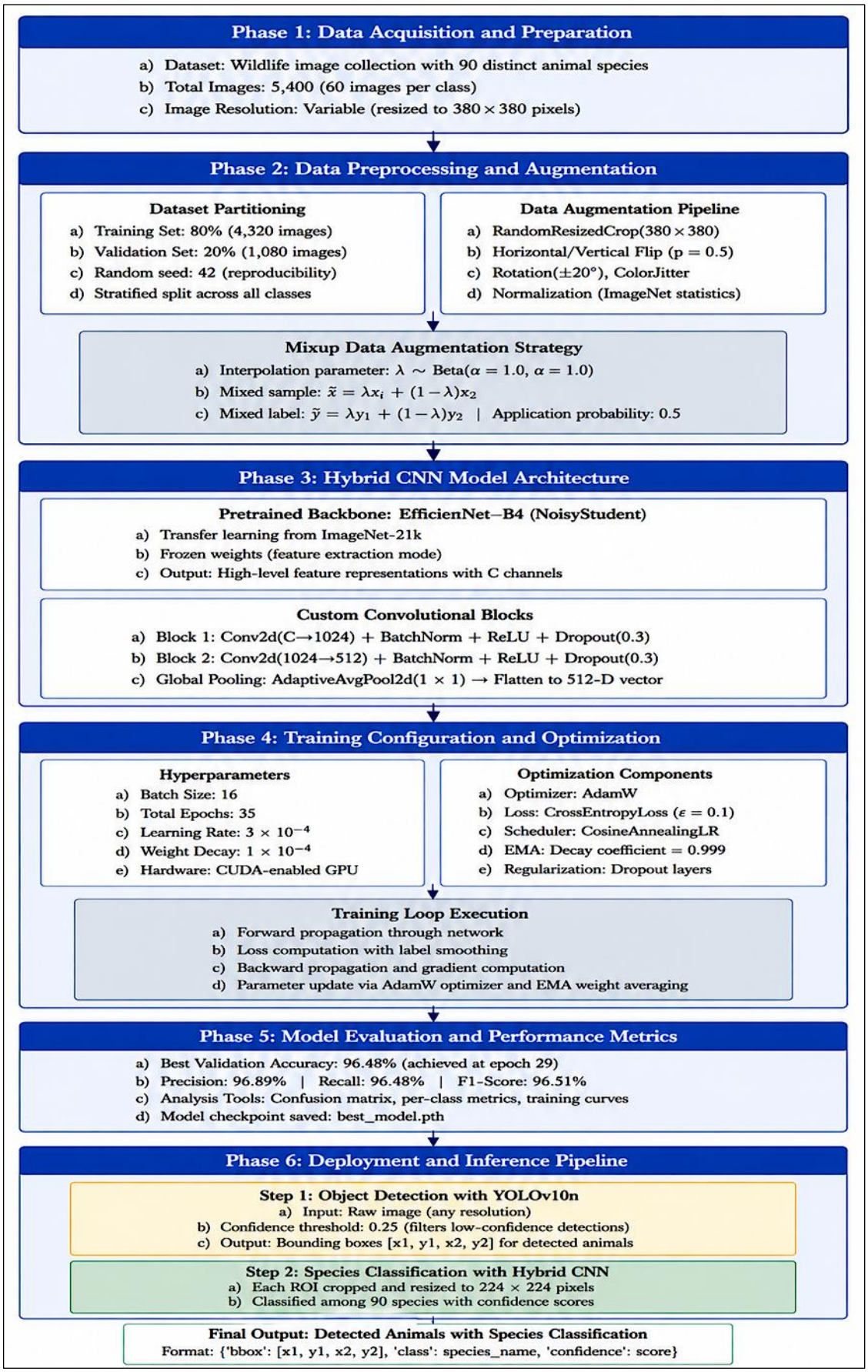

Figure 2: Overall Workflow of the Proposed Hybrid Wildlife Recognition and Detection Framework

Figure 2 presents a high-level end-to-end overview of the hybrid wildlife recognition pipeline, demonstrating the flow of raw multi-species photos through the system. In order to preserve balanced species representation, photos are first pre-processed using techniques such illumination normalization, resizing, denoising and augmentation. This is followed by stratified dataset partitioning. Following processing, the inputs go via two coordinated branches: an EfficientNet-B4-based classifier with bespoke CNN layers and a YOLOv10 detection module. Using multi-scale feature extraction and dropout-based regularization, YOLOv10 locates animals in real-time under a variety of sizes and environmental circumstances, generating bounding boxes that are fed into the classifier for species-level identification. AdamW optimization, cosine-annealing scheduling, EMA smoothing and other stabilizers that promote steady convergence are used in the training of both branches. Bounding boxes are aligned with anticipated labels, confidence scores are assigned and redundant detections are eliminated using NMS filtering in a unitary fusion module that combines their outputs. The overall workflow shows how detection, classification, fusion, evaluation and deployment are integrated into a single hybrid wildlife-recognition framework that can support real-world ecological monitoring. It includes logging for the purpose of the experiment tracking, monitoring tools for validation as well as automated update triggers for incorporating new field data.

Results

Using stratified sampling across all 90 species, a hold-out validation set of 1,080 entirely unseen images—representing 20% of the whole dataset—was used to thoroughly test the suggested hybrid framework. Throughout the training process, this validation set stayed completely apart from the 4,320 training photos, guaranteeing an objective evaluation of the model's performance on new data that it had never seen during optimization.

Throughout the optimization process, training dynamics were captured across 35 epochs and demonstrated distinct, progressive learning.

Validation accuracy increased quickly in the early stages of training, peaking at 86.48% after the first epoch. The accuracy then improved continuously and later peaked to 96.48% at epoch 29, apart from that at the very ending the accuracy was 96.30% by the final epoch. While the validation loss was maintained at 1.04, a steadily drop was seen in training loss that dropped from 3.21 to 1.82 which indicates a successful convergence without indications of overfitting. Thus, these patterns depict and helps us to understand that the model maintained dependable generalization, attained stable optimization as well as reacted well to the augmentation and regularization techniques while the training process was going on.

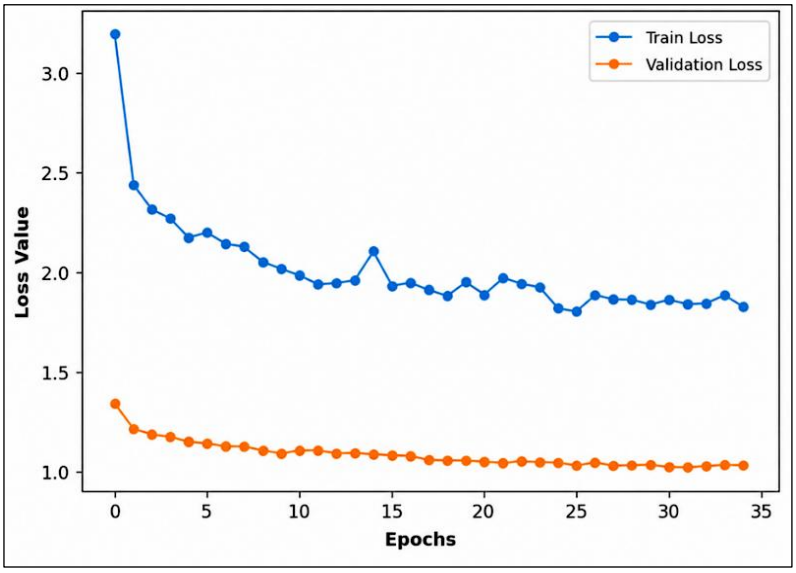

Figure 3: Training And Validation Loss

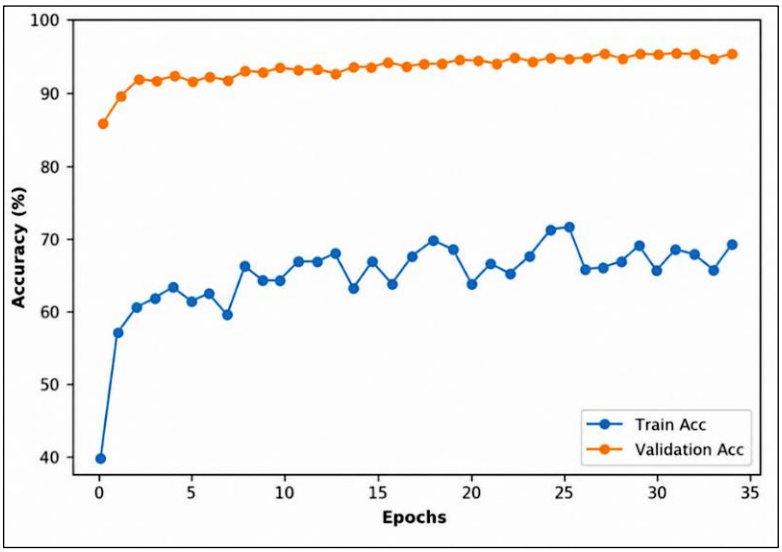

Figure 4: Training And Validation Accuracy

Figures 3 and 4 work together to show the training dynamics in a complimentary way. Training loss is still declining in Figure 3, but validation loss stabilizes at 1.04, this showcases a consistent convergence without overfitting even when heavy augmentation is used. Figure 4 illustrates effective optimization as well as the effects of regularization along with augmentation. In the early epochs, the accuracy rises rapidly before tapering toward a ~96.3% plateau.

The deliberate training difficulty, which results in a lower training accuracy (69.03%) compared to a significantly higher validation accuracy (96.30%), is a noteworthy feature of the learning regime. This is caused by aggressive augmentations like mixup, random crops, rotations and color jittering, which compel the model to learn robust, generalizable features rather than memorize data. Strong transferability to pristine, unseen pictures is indicated by this advantageous gap. Label smoothing (0.1) and EMA (decay 0.999) are two more methods that further stabilized optimization and enhanced model calibration.

Table 1 summarizes the main performance measures (such as accuracy, precision, recall, F1-score, validation picture count and classification error rate) for the hold-out validation set.

Table 1: Performance Evaluation Metrics of the Proposed Hybrid Wildlife Recognition Framework

Performance Metric	Score
Accuracy	0.9630
Precision (weighted)	0.9668
Recall (weighted)	0.9630
F1-Score (weighted)	0.9631

The consistently high weighted accuracy, recall and F1-score values show that both dominating and comparatively underrepresented animal species exhibit balanced categorization behavior. Stable generalization capabilities and decreased prediction bias across the stratified validation dataset are further demonstrated by the tight agreement between these assessment criteria. Weighted assessment measures were examined across the stratified validation dataset that included all 90 animal species in order to further assess statistical reliability. With weighted accuracy, recall and F1-score values of 0.9668, 0.9630 and 0.9631, respectively, the suggested framework demonstrated extremely consistent prediction performance across several species categories. These measures exhibit balanced categorization behavior with little class-specific bias and consistent generalization capabilities, as seen by their modest variance. Additionally, steady optimization and less overfitting throughout the training phase are shown by the training and validation curves' smooth convergence tendencies and the lack of notable variations in validation performance across subsequent epochs.

While the sparse off-diagonal entries in Figure 5 indicate that sporadic errors mostly occur between taxonomically related or visually similar species with similar fur patterns, shapes, or habitat cues, Figure 5 displays a strong diagonal pattern across all 90 species, confirming consistently high per-class accuracy. This kind of localized uncertainty is frequent in fine-grained wildlife recognition and shows that no species undergoes systematic misclassification; rather, ambiguities are caused by natural elements like as illumination, position variation, occlusion and partial animal visibility that are typical in camera-trap situations.

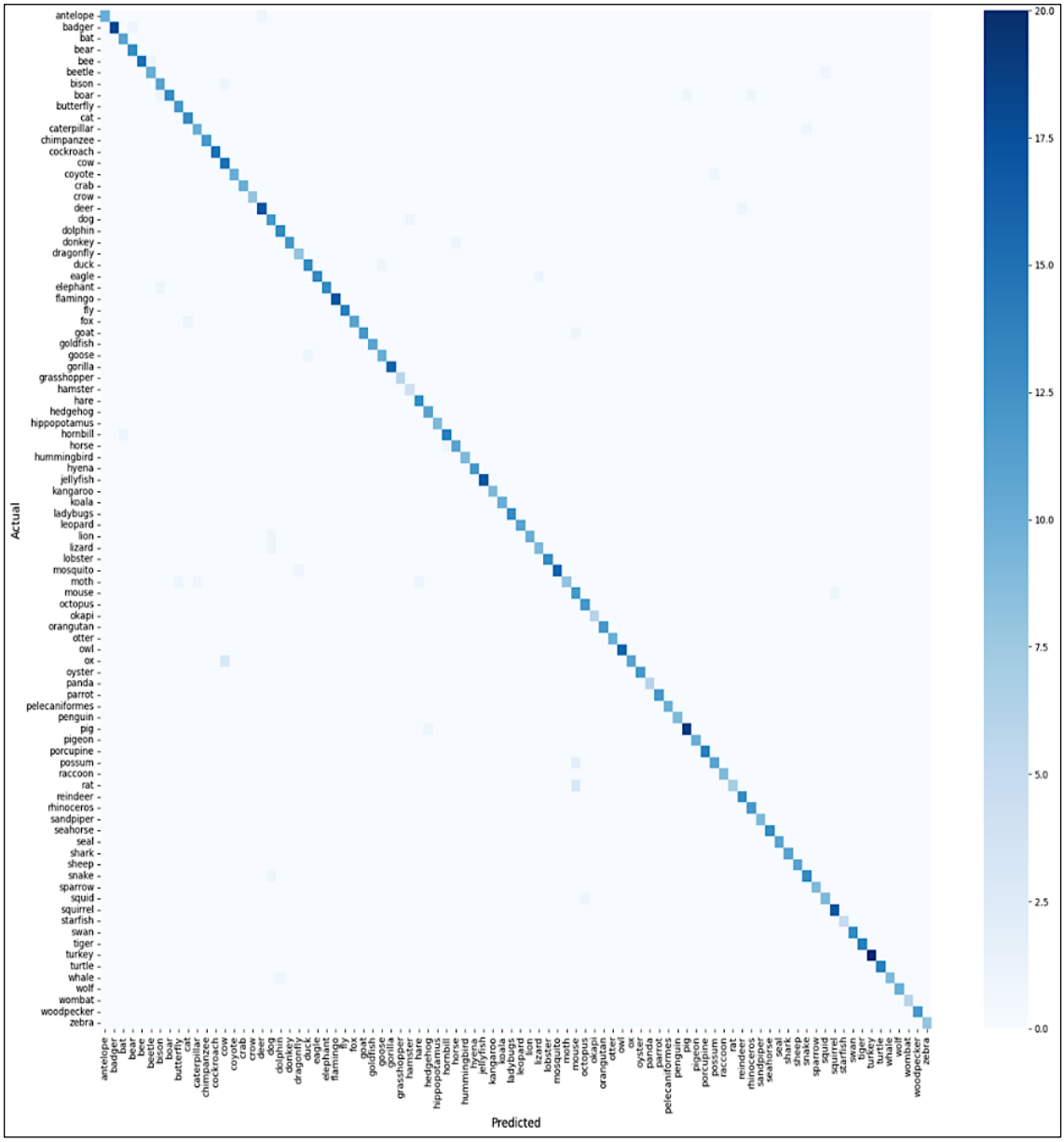

Figure 5: Confusion Matrix Showing Per-class Classification Performance Across 90 Wildlife Species

To offer complete animal identification from uncropped field photos, the hybrid detection + classification pipeline combines the trained EfficientNet-B4 classifier with YOLOv10n for object location. With inference speeds of about 42.7 milliseconds per image (≈23 frames per second), YOLOv10n showed strong detection capability, allowing for near real-time localization; the custom CNN cropped and classified detected animal regions, usually yielding classification confidence scores above 95%. The proposed pipeline is very much capable of handling a wide variety of environmental conditions which are dynamic that includes different lighting conditions, partial occlusion, complex backgrounds and multiple poses which is demonstrated by a

qualitative analysis of random samples on the field. This makes sure about the practical robustness which is beyond conditions that are present within controlled dataset and also supports the applicability in a given range of ecological monitoring scenarios.

Discussion

The results clearly depict excellent accuracy with 96.30% on 90 species which suggests that hybrid EfficientNet-B4 Custom CNN and YOLOv10 framework achieves remarkable performance for large-scale wildlife classification and detection. The results are a significant improvement over previously used methodologies which handled same kind of complexity. Better generalization over the data that is entirely is made possible by employing the combination of sophisticated augmentation techniques, transfer learning as well as hybrid architecture design, all of which helps to preserve real-time inference capacity that is basically appropriate for implementation within real-world scenarios used in automated wildlife monitoring systems. These findings clearly demonstrate the framework's intelligence along with its scalability for applications regarding ecological research that requires accurate identification of all species along with their localization in real environment, as well as conservation initiatives and evaluation of biodiversity. The obtained weighted F1-score and validation accuracy are comparable to and in some cases better than, previously published wildlife-monitoring and species-recognition frameworks covered in the literature.

Compared with previously reported wildlife-monitoring frameworks, the proposed hybrid architecture demonstrates improved capability for simultaneous localization and fine-grained species recognition in complex ecological environments. Detection-oriented systems such as YOLOv5, YOLOv3 and YOLOv10 primarily emphasized real-time localization and surveillance performance. Approximately 94% mAP was previously achieved using YOLOv5 for wildlife–vehicle monitoring, while mAP@0.5 of 0.934 was reported for edge-based wildlife detection using YOLOv10 (5, 12). Effective wildlife localization capability was also demonstrated using YOLOv3-based frameworks, although the primary emphasis remained on detection performance (17). In contrast, strong species-recognition accuracy was achieved in CNN-based wildlife-classification systems, although these approaches generally relied on manually cropped or controlled animal inputs (8, 15). By integrating YOLOv10n-based localization with a customized EfficientNet-B4 classification stream, the proposed framework combines accurate detection with refined species-level categorization within a unified detect–crop–classify pipeline, thereby improving robustness, scalability and deployment suitability for practical wildlife-monitoring and biodiversity-assessment applications. In comparison to several previously reported wildlife-monitoring approaches discussed in the literature, the proposed framework achieved competitive or superior performance with a weighted F1-score of 96.31% and validation accuracy of 96.30%.

Apart from the strong empirical results, certain limits are mentioned along with the possibilities for improvement of the model which are also highlighted. Claims regarding out-of-distribution performance are limited by the evaluation's use of a hold-out validation set without a separate test set; to more thoroughly confirm generalization, future research should employ cross-dataset evaluation or three-way train/validation/test splits. Although balanced, the dataset's 60 photos per species are still a small sample size for each class and could not adequately represent intra-class variety across age, sex, or seasonal morphologies. The observed misclassifications usually tend to cluster among species that are visually very similar. This indicates the possibility of improving discrimination as well as reducing confusions within the classes that are closely related, this is done by utilizing temporal consistency across various frames or by adding the additional modalities (thermal imaging, audio signatures, or behavioural patterns). Furthermore, model generalization across geographically varied biological regions may be impacted by domain-shift effects introduced by differences in habitat features such as plant density, background composition, lighting conditions, seasonal fluctuations and environmental context.

Future research can further improve the proposed framework through the integration of multimodal ecological data such as thermal imagery, infrared sensing, motion patterns and bioacoustic signals to enhance recognition under challenging

environmental conditions. Cross-dataset validation and larger-scale wildlife datasets may further improve domain generalization and robustness across geographically diverse habitats. In addition, transformer-based architectures, continual learning strategies, lightweight model compression and edge-oriented optimization techniques can be explored to improve scalability, inference efficiency and long-term deployment capability for real-time biodiversity monitoring and automated ecological surveillance systems.

When deploying automated wildlife monitoring systems in the real world, potential ethical issues should also be taken into account. If monitoring data are poorly accessed or maintained, ongoing monitoring in ecologically sensitive settings may inadvertently disrupt animal behavior or provide location-specific information about endangered species. Furthermore, observational bias may be introduced by habitat-specific sampling and dataset imbalance, which might affect monitoring reliability in a variety of habitats and ecological interpretation. Therefore, in order to guarantee sustainable and ethically responsible environmental monitoring practices, AI-based wildlife-monitoring frameworks should be implemented responsibly with suitable data-security measures, ecological sensitivity and cooperation with conservation agencies.

Conclusion

This paper introduces a hybrid deep learning system that makes use of a combination by using YOLOv10 for detection in real-time along with EfficientNet-B4 and a bespoke CNN head for refined categorization of species. Across 90 unknown animal species, the system was able to achieve 96.30% as validation accuracy, 96.68% as weighted precision, 96.30% as weighted recall and a 96.31% F1-score, indicating models' capability of better as well as higher generalization and also for validating the viability of scaled, automated wildlife monitoring.

The performance of the framework is very well because of its trainable convolutional layers that basically allows domain-specific refinement while om the other hand frozen EfficientNet-B4 backbone provides a strong and high-level feature extraction. Strong augmentations that addressed occlusion, illumination variance and background clutter, such as mixup, randomly scaled cropping, geometric perturbations and photometric modifications, strengthened robustness. The AdamW is used as an optimizer, cosine-annealing restarts, exponential moving average and label smoothing were used in order to further improve the stability during training.

At around 42.7 ms per picture (≈23 FPS), YOLOv10n provided quick and accurate localization, while the bespoke CNN classifier generated species predictions with high confidence. Because of these features, the pipeline is ideal for contexts with limited resources, such drones, embedded devices and remote camera traps, where computing efficiency is crucial.

Compared to single-model systems, the integrated detection-classification pipeline has definite advantages: detection by itself cannot give species-level identification and classification by itself cannot pinpoint animals. The framework achieves strong precision–recall performance, low false positives and real-time throughput appropriate for automated wildlife monitoring and biodiversity assessment by combining YOLOv10 with an EfficientNet-B4 classifier to provide precise bounding-box detection along with fine-grained recognition in a single workflow.

The suggested hybrid architecture still has certain drawbacks despite its excellent performance. A thorough examination of generalization across geographically varied habitats and environmental circumstances may be limited by the current evaluation's hold-out validation technique, which lacks comprehensive cross-dataset testing. Furthermore, the amount of the dataset per species is still rather small, making it difficult to capture more extensive intra-class variations including seasonal changes in appearance, variations in light, occlusion and intricate behavioral patterns. To further enhance scalability, robustness and practical applicability for automated wildlife monitoring systems, future research can concentrate on larger multi-environment wildlife datasets, multimodal ecological sensing using thermal or bioacoustics information, transformer-based architectures, continuous learning strategies and lightweight edge-optimized deployment techniques.

All things considered, the suggested hybrid EfficientNet-B4 and YOLOv10 architecture shows great promise for next-generation intelligent wildlife-monitoring systems by fusing effective

real-time detection with fine-grained species recognition inside a single pipeline. Deployment-oriented optimization, transfer learning, robust augmentation techniques and deep hierarchical feature extraction are all integrated to provide dependable performance in challenging ecological settings. For biodiversity evaluation, automated ecological surveillance, conservation planning and real-world wildlife monitoring applications, the framework therefore offers a scalable and useful solution.

Abbreviations

CNN: Convolutional Neural Network, EMA: Exponential Moving Average, FN: False Negative, FP: False Positive, FPS: Frames Per Second, IoU: Intersection over Union, mAP: Mean Average Precision, TN: True Negative, TP: True Positive, YOLO: You Only Look Once.

Acknowledgements

We thank all our colleagues with full heartfelt support for their support in the completion of this research work**.**

Author Contributions

All authors contributed equally to this work.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the content of this article.

Data Availability

A publicly available dataset was used in this study and can be accessed upon request or through its official source.

Declaration of Artificial Intelligence (AI) Assistance Process

Generative AI tools were used only for language refinement; all scientific content and conclusions are the authors' own. The authors take full responsibility for the content's originality, interpretation and accuracy

Ethics Approval

Not Applicable.

Funding

None.

References

1. Favorskaya M, Pakhirka A. Animal species recognition in the wildlife based on muzzle and shape features using joint CNN. Procedia Comput Sci. 2019; 159:933-42. doi:10.1016/j.procs.2019.09.260
2. Chandrakar R, Raja R, Miri R, Tandan SR, Ramya Laxmi K. Detection and identification of animals in wildlife sanctuaries using convolutional neural network. Int J Recent Technol Eng. 2020;8(5):181-185. doi:10.35940/ijrte.E4579.018520
3. Sami M, Devi SVS. A deep learning approach to detect wounded animals. In: Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2025). IEEE; 2025. doi:10.1109/ICICT64420.2025.11005180
4. Ahuja NJ, Pasi N, Naz H. A deep learning-based framework for the detection of Big-4 snakes. In: Proceedings of the 2024 International Conference on Computing, Sciences and Communications (ICCSC). IEEE. 2024:1-6. doi:10.1109/ICCSC62048.2024.10830347
5. Teng CY, Connie T, Choo KY, Goh MKO. A visual approach towards wildlife surveillance in Malaysia. In: 2022 10th International Conference on Information and Communication Technology (ICoICT). IEEE. 2022:374-79. doi:10.1109/ICOICT55009.2022.9914861
6. Preethi K, Vinitha A, Vinothiga V, Mahalakshmi V, Kumararaja V. AI-driven wildlife detection and management system for agricultural protection. In: 2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA). IEEE. 2025:1-4. doi:10.1109/ICAECA63854.2025.11012427
7. Sree MR, Tejaswi K, Shalini S, MS A. Animal intrusion detection using YOLOv9. In: 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). IEEE. 2025:62-67. doi:10.1109/ICSSAS66150.2025.11080962
8. Nguyen H, Maclagan SJ, Nguyen TD, Nguyen T, Flemons P andrews K, Ritchie EG, Phung DQ. Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In: 2017 International Conference on Data Science and Advanced Analytics (DSAA). IEEE. 2017:40-49. doi:10.1109/DSAA.2017.31
9. Palanisamy V, Ratnarajah N. Detection of wildlife animals using deep learning approaches: A systematic review. In: 2021 21st International Conference on Advances in ICT for Emerging Regions (ICTer). IEEE. 2021:153-158. doi:10.1109/ICter53630.2021.9774826
10. TG KK, Ajay M, Bhavana N, Chaithra BC, Ram CB. Eco-Harbor: An integrated approach for forest fire detection, mitigation, wildlife relocation and tribal alert using deep learning and IoT. In: 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N). IEEE. 2024:1688-93. doi:10.1109/ICAC2N63387.2024.10895796

11. Aditya LM, Kumar DK, Sivakumar KR, Ramji T. EfficientDet-based IoT system for wildlife monitoring using LoRa and Raspberry Pi. In: 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI). IEEE. 2025;3:1-6. doi:10.1109/IATMSI64286.2025.10985501
12. Nishanth I, Rakeshkumar R, Thenmozhi V. Real-time wild animal detection using YOLOv10 algorithm. In: 2025 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE. 2025:1-6. doi:10.1109/WiSPNET64060.2025.11005222
13. Pujitha K, Revathi T, Darla DB, Siddham A. Smart wildlife surveillance leveraging deep learning and sensor data analysis. In: 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET). IEEE. 2024. doi:10.1109/ICICET59348.2024.10616360
14. Renuka N, Kannan N, Dhanushram S, Abinesh S, Sowmiya T. Temporal convolutional network-based animal intrusion detection model for smart farming. In: 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN). IEEE. 2024:63-6. doi:10.1109/ICPCSN62568.2024.00018.
15. Sreedevi KL, Edison A. Wild animal detection using deep learning. In: 2022 IEEE 19th India Council International Conference (INDICON). IEEE. 2022:1-5. doi:10.1109/INDICON56171.2022.10039799
16. Cybi M, Joy E, Evangeline S, Paul J, Sanjai S. Wildlife detection and behavioral analysis in farmland using YOLOv8m and LSTM. In: 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). IEEE. 2025:246-51. doi:10.1109/ICSSAS66150.2025.11080886
17. Gabriel M, Cha S, Al-Nakash NYB, Yun D. Wildlife detection and recognition in digital images using YOLOv3. In: 2020 IEEE Cloud Summit. IEEE. 2020: 170-1. doi:10.1109/IEEECloudSummit48914.2020.00033
18. Mane V, Nikude P, Patil T, Tambe P. Wildlife classification using convolutional neural networks. In: 2024 7th International Conference on Inventive Computation Technologies (ICICT). IEEE. 2024: 1046-53. doi:10.1109/ICICT60155.2024.10544702
19. Deng C, Zhou G, Cai Y. Wildlife monitoring and identification based on Faster R-CNN. In: 2023 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA). IEEE.2023:638-42. doi:10.1109/AEECA59734.2023.00119
20. Sykora P, Kamencay P, Hlavata R, Hudec R. Overview and comparison of deep neural networks for wildlife recognition using infrared images. AI. 2024;5(4): 2801-28. doi:10.3390/ai5040135
21. Chen J, Xu H, Wu J, Yue R, Yuan C, Wang L. Deer crossing road detection with roadside LiDAR sensor. IEEE Access. 2019;7:65944-54. doi:10.1109/ACCESS.2019.2916718
22. Yang G, Pan Y, Sui C, Zang A, Jiang F, Hu J. Lightweight Conv-Swin Transformer for wildlife detection. In: 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE). 2022:1-5. doi:10.1109/ICARCE55724.2022.10046623
23. Parkavi K, Ganguly A, Kejriwal K, Sharma S, Banerjee A. Enhancing road safety: Detection of animals on highways during night. IEEE Access. 2025;13:36877-96. doi:10.1109/ACCESS.2025.3545490
24. Govardhan M, Lavanya MN, Chandu Reddy KG, Kishore Kumar A, Koushik Reddy KK. Real-time wildlife tracking and anomaly detection using YOLOv8. In: 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). 2024:1-7. doi:10.1109/ICSES63760.2024.10910740.
25. Li Z, Yan Z, Tian W, Zeng D, Liu Y, Li W. ReDeformTR: Wildlife re-identification based on light-weight deformable transformer with multi-image feature fusion. IEEE Access. 2024;12:106321-32. doi:10.1109/ACCESS.2024.3436813
26. Richard BY, Kumar SS, Sindhu V, Thiyagaraj K, Thangarajan CV. Smart boundary monitoring system for wildlife containment and human safety. In: 2025 International Conference on Data Science and Business Systems (ICDSBS). 2025:1-5. doi:10.1109/ICDSBS63635.2025.11031612
27. Shafer MW, Flikkema PG. Tracking small wildlife with minimal-complexity radio frequency transmitters: Near-optimal detection. IEEE Access. 2023;11:40029-37. doi:10.1109/ACCESS.2023.3268631
28. Vidhya K, Nagarajan B, Thiyagu TM, Archpaul J, Shirley CP. Enhancing wildlife monitoring: real-time alerts through email and messaging. In: 2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). 2025:1779-85. doi:10.1109/ICICV64824.2025.11085661
29. Abhishek KV, Megalingam RK. FPGA-based wild animal detection system using Verilog: Advancing surveillance and mitigating human–wildlife conflicts. In: 2025 8th International Conference on Circuit, Power & Computing Technologies (ICCPCT). 2025:335-40. doi:10.1109/ICCPCT65132.2025.11176500
30. Teigen H, Ahmed A, Imran AS, Ullah M, Azad RMA, Soylu A. From blurs to birds: Localization and classification of hard-to-see bird species in Norwegian wilderness camera trap images. IEEE Access. 2025;13:169534-53. doi:10.1109/ACCESS.2025.3613068
31. Feng W, Ju W, Li A, Bao W, Zhang J. High-efficiency progressive transmission and automatic recognition of wildlife monitoring images with WISNs. IEEE Access. 2019;7:161412-23. doi:10.1109/ACCESS.2019.2951596
32. Reddy KV, Reddy BK, Goutham V, Mahesh M, Nisha JS, Palanisamy G, Golla M, Purushothaman S, Reddy KR, Ramkumar V. Edge AI in sustainable farming: Deep learning-driven IoT framework to safeguard crops from wildlife threats. IEEE Access. 2024; 12:77707-23. doi:10.1109/ACCESS.2024.3406585
33. Wang J, Zhu M, He Y, Xie M. Assessment of environmental awareness to biodiversity protection of wetlands in Northwest Yunnan. In: 2010 2nd

International Conference on Education Technology and Computer (ICETC). IEEE. 2010;3:175–179. doi:10.1109/ICETC.2010.5529570
34. Maina CW. Bioacoustic approaches to biodiversity monitoring and conservation in Kenya. In: IST-Africa 2015 Conference Proceedings. IEEE. 2015:1–8. doi:10.1109/ISTAFRICA.2015.7190558
35. Mughal AB, Khan RU, ur Rehman A, Bermak A. Deep learning for dynamic wildlife monitoring: a real-time approach. IEEE Access. 2025;13:147422-48. doi:10.1109/ACCESS.2025.3600625
36. Zhang Y, Cai Z. CE-RetinaNet: A channel enhancement method for infrared wildlife detection in UAV images. IEEE Transactions on Geoscience and Remote Sensing. 2023;61:4104012. doi:10.1109/TGRS.2023.3299651

How to Cite: Kothari SS, Chhabria A, Phadke G. Hybrid Wildlife Classification and Detection Using EfficientNet-B4 Custom CNN and YOLOv10. Int Res J Multidiscip Scope. 2026;7(3):138-152.
DOI: 10.47857/irjms.2026.v07i03.09519