Efficient Animal Species Recognition from Camera Trap Images Using Yolov8 and Mobilenet with Automated Background Removal

Mrs. Padma Priya J[1], H. Kaleel Ahmed[2], Yuvan Shankar S[3], Suresh S[4]
[1,2,3,4]Panimalar Engineering College, Chennai, India

***Abstract*—One of the recent tools of environmental study which has gained significance in studying populations of wildlife is the maintenance of the wildlife through camera trap systems. Nevertheless, by eyeing thousands of captured images, it is time-consuming and not always efficient. The proposed study will provide an automated framework to identify animal species based on camera trap images with the use of deep learning methods to overcome this issue. The proposed system combines the detection of the animal with YOLOv8 with the classification of the species with MobileNet and an automated background removal process that concentrates the model on the area where the animal is located. During the first stage, the detection model is used to detect the location of the animals in the image and a region of interest is extracted. At the second stage, the image, which was cropped, is inputted to a lightweight MobileNet classifier that will identify the animal species. This two-phase pipeline is more background noise reducing than two-stage, has a better classification performance with an equivalent computational efficiency. Experimental analysis shows that the offered method can identify various animal species with higher accuracy and less complexity of the processing and be adopted to apply to large-scale wildlife monitoring. endow it in man make it more in man.**

***Index Terms*—Animal, Species, Recognition, Camera, Trap, Images, YOLOv2, MobileNet, Background, Removal, Deep, Learning, Computer, Vision, Wildlife, Monitoring, Classification.**

I. INTRODUCTION

Animal surveillance is vital in conservation of biodiversity and ecological studies. The use of camera trap technology has gained a lot of acceptance since the researchers can gather huge amounts of wildlife images without having to disorient the wildlife in their natural habitats. Although these benefits exist, manual analysis of camera trap images is a tedious process that consumes much time and staff time.

The dynamism in the field of deep learning methods has opened up fresh possibilities of automating wildlife image analysis. More specifically, convolutional neural networks (CNNs) have demonstrated high effectiveness in the detection of objects, as well as classifying images. These methods are able to learn complicated visual designs automatically and hence an efficient method of recognizing animal species in large scale image collections.

Nevertheless, there are still various problems with the analysis of camera trap images. Photographs usually have complicated backgrounds, bad light effects, and motion smear, as well as incomplete animal visibility. These aspects have the capacity to have a detrimental impact on the functionality of the traditional recognition systems. Hence, there is a need to have an effective technique that targets the animal part, and minimizes background details which are irrelevant.

To address such issues, this paper suggests a two-step identification system of animal species. In the first step, the YOLOv8 is utilized to identify animals in camera trap images and locate the area with an object of interest. Once it has been detected, the region of the animal is cropped to minimise the background. The second stage will require the processed image to be fed into a MobileNet classifier to identify the animal species.

This solution is a feasible solution to the automated wildlife monitoring system because the proposed approach is accurate on recognition and also the complexity of the computation is low.

II. LITERATURE REVIEW

Recent years have seen much interest in automated animal species recognition of camera trap images because of its role in wildlife conservation, ecological studies, and monitoring of wildlife biodiversity. Conventional methods used were based on manual features and classical machine learning classifiers like Support Vector Machines (SVM) and Random Forests. Though these techniques were fairly successful in the con-trolled settings, they tended to fail miserably when they were used in the real world with the backgrounds being complex, variable illumination, occlusion, and partial animal visibility. Convolutional Neural Networks (CNNs) are now the stan-dard method of animal recognition based on images with the development of deep learning. Early CNN-based research concerned the direct image classification with full camera trap images. Nevertheless, these methods were often flawed by decreased accuracy owing to atmospheric noise since animals often take a very little area of the image. To overcome this drawback, various researchers suggested pipelines that are based on object detection that initially identifies animals and then classifies them into species.

Object detection frameworks based on the YOLO have been popularly investigated in the context of wildlife monitoring because the systems can operate in real-time and have high detection rates. The YOLO architecture described by Redmon et al. is able to localize and classify objects in one forward step, which is appropriate when dealing with large-scale datasets of camera traps. Later studies showed that YOLOv8 is better at detection accuracy due to the use of batch normalization, anchor boxes, and multi-scale training, which allows better detection of animals of different sizes and poses.

Norouzzadeh et al. applied deep CNNs on a large-scale wildlife classification and demonstrated that automated systems can be as accurate as human experts. The method that they applied, however, was based on the classification of the full image, and this aspect causes the system to be sensitive to the foreign objects and frames that are not related to animals. Chen et al. solved this problem with the assistance of object detection and classification and proved that the background removal can considerably increase the performance of the species recognition.

MobileNet MobileNet is a lightweight CNN architecture presented to minimise computational complexity and achieve competitive performance. MobileNet applies depthwise separable convolutions, which allows the application of inference to work on resource-restricted devices. MobileNet has been used to classify wildlife in a number of studies, and it has been noted to be suitable in real-time and edge-based wildlife monitoring. Nevertheless, the accuracy of classification is highly reliant on the successful foreground extraction.

Background removal that is automated has been observed to be instrumental towards enhancing recognition accuracy. Through cropping of identified areas of animals, the non-discriminative background information can be removed and the classifier can then concentrate on discriminant species features. Recent papers that have fused YOLO-based detection with lightweight classifiers have claimed to achieve enhanced robustness in adverse camera trap scenarios. Although such advances have been made, in the vast ma-jority of the available literature, either detection or classifi-cation are the main subjects of study. There is very little literature on the integrated pipeline which consists of real time detection, automated background removal and lightweight species classification as a single framework. The described research gap inspires the proposed methodology, which will utilize YOLOv8 to perform the correct animal detection and MobileNet to perform the effective species recognition so that to have the scalable and reliable wildlife monitoring.

III. PROPOSED METHODOLOGY

The suggested system has a two-step architecture that aims at enhancing the precision of animal species identification of camera trap images. The general stepwise process involves the detection of animals, background elimination, and classification of species. The section below relates to the detection of animals with the help of YOLOv8.

The first phase involves the detection of animals by utilizing YOLOv8 on the images taken by camera traps. YOLOv8 is an advanced object detector that can help recognize objects in the real-time with high precision. The model scans the whole image and estimates the bounding boxes which are the location of animals.

After the detection process is done the bounding box of the animal that is detected is extracted. The step is used to make sure that only a part that has the animal is processed in the next process description: This involves generating a background removal image based on the input image. Human Input: The input is the image of a background that we want to remove.

Camera trap pictures usually have a significant amount of unnecessary background as either vegetation, shadows, or bare space. This interference is minimized by extracting the region of the original image containing the detected animal. This ensures that unnecessary background information is eliminated and all that is needed is the classification model to concentrate on the features of the animals.

Overall, the classification of species using MobileNet is a conceptual framework that varies according to the selected operational system. Human In general, the classification of species with the help of MobileNet is the conceptual framework that differs depending on the chosen operating system.

Following a cropping of the found object, the processed image is then processed by a MobileNet model to identify species. MobileNet MobileNet is a merged convolutional neural network or a lightweight network applicable to image classification exercises. It applies depth wise separable convolutions in order to minimize the level of computation and achieve high performance.

The visual features used by the classifier include the texture, shape and color pattern of various animals to identify the animal species in the picture. The system ends up giving the predicted species label and the confidence score of the classification.

A. System Architecture

The framework proposed entails two stages of pipeline, which comprise of offline model training and online real-time inference. The offline stage entails dataset preprocessing, data augmentation and training of the YOLOv8 detection model to locate the animals. MobileNet classifier is then used to identify species in regions of animals which have been cropped. The online phase takes the images of the camera trap or the frames of live video and compares them in real-time by processing camera traps and live video, detecting and cropping animals, predicting species, and delivering the results in a structured form. Figure 1 above shows the Systems architecture.

This section presents a project describing the detection of animals in the context of the YOLOv8 model.

The detection of animals is done with a single-stage object detection model that is based on the YOLOv8 architecture. In one forward step, each input image is processed to predict bounding boxes, class labels and confidence scores of animals detected.

Bounding box representation is the term used to describe the method of demonstrating the extent of knowledge and description of a certain subject. Human The bounding box representation can be defined as the mode of showing the scope of knowledge and description of a particular topic.

YOLOv8 encodes each object animal with a bounding box as:

$$B = (x_c, y_c, w, h) \qquad (1)$$

In which, xc, yc is the centre of the bounding box, and w, h is the width and height of the bounding box respectively. This allows localization which is not dependent on the resolution and can help to crop the animal in the background correctly.

The calculation of the confidence score is done as presented below:

A confidence score is computed against every predicted bounding box, which is:

$$Confidence = P(object) x P(class) \qquad (2)$$

and where the probability of an object is given by P (object) and the conditional probability of the species predicted is P (class). Bounding boxes that have confidence less than a threshold are eliminated in order to minimize false positives. Non-Maximum Suppression (NMS) is a method used to overcome the limitations of Maximum Entropy (ME) speech recognition. Human Non-Maximum Suppression (NMS) is a technique that is aimed at alleviating the drawbacks of Maximum Entropy (ME) speech recognition.

Non-Maximum Suppression is to be used in order to eliminate unnecessary overlapping detections. To calculate the Intersection over Union (IoU) between two bounding boxes, B i and B j, we have:

$$IoU = 2 = Areaof\ Overlap/Areaof\ Union. \qquad (3)$$

A box B_i is retained if:

$\mathrm{IoU}(B_i, B_j) < \tau$ (4)

and tau is a predetermined value. NMS guarantees that only most confident detection instances per animal instance are maintainable.

B. Frame Preprocessing

The aspect ratio is preserved as the input images are resized to 416 x 416 pixels. The values of pixels are turned into tensors which are accepted by YOLOv8 and MobileNet inputs and are normalized. These preprocessing steps make the models more resilient to changes in scale, changes in illumination and motion blur that is often found in camera trap images.

This paper presents the results of the study on the use of mobile net to classify species.

Following the process of crop selection of the detected animal region, the MobileNetV2 classifier is used to make prediction on the species. It takes the cropped picture and resizes it to the size of 224 x 224 pixels, preprocesses it with the MobileNetV2.

The strategy and optimization of training include the establishment of training goals and the development of training procedures. Human Training strategy and optimization involve setting of training objectives and development of training processes.

YOLOv8 is also trained on transfer learning with a pre-trained backbone. The localization and classification losses are added together in the total loss function:

$L = Lbox + Lcls.$ (5)

Bounding Box Loss is a type of loss that can be calculated in the form of a percentage. Human Bounding Box Loss is a kind of loss that could be measured as a percentage.

To optimize bounding box regression, the loss is based on the IoU:

$L_{\mathrm{box}} = 1 - \mathrm{IoU}$ (6)

1) Classification Loss: Categorical cross-entropy is used to maximise species classification:

$$L_{cls} = -\sum_i y_i \log(p_i) \quad (7)$$

y i: The ground truth of sample i. p i: The probability of sample i in the prediction.

C. Real-Time Inference

In use, the system is in real time whereby the camera taps or live video feed images are detected and classified. The predictions that have a confidence that is above 0.4 are retained. Identified animals are displayed in bounding boxes and species labels and all the findings are recorded in a formatted CSV file. This pipeline provides efficient, accurate, and scalable wildlife monitoring and elucidates the manual annotation effort to a considerable level.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed animal species recognition system has been experimentally evaluated. was trained on a YOLOv8 detector of animals in combination with. a species recognizing Mo-bileNetV2 classifier. The system was trained and tested on a set of camera trap images of wildlife surveillance. datasets. The data was broken down into training set, validation set, and test set. with conventional protocols of detecting and classifying objects.

The analysis is based on the detection accuracy, classification performance, Ability to distribute a dataset, and ability to infer in real time. Figure IV displays the distribution of the dataset and image attributes that are used to train.

Table 1 provides the characteristics of the dataset in terms of distribution and labeling.

The data that has been involved in this paper contains several wildlife species that have been captured. In other environmental conditions like daylight, night vision, and partial occlusion. The dataset is made up of about 500 labeled. pictures of several types of animals.

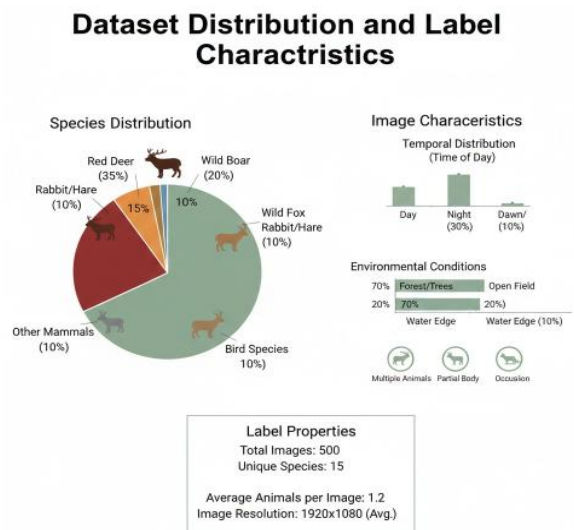

Fig.1 Dataset distribution and labelling features such as species distribution, environmental conditions, temporal image characteristics.

The data contains various species of animals, red deer, wild boar, among others. wild fox, rabbits, birds and other mammals. The distribution of images between species groups will provide adequate diversity to model training. The photos were taken when it is light and when it is dark and contain various conditions of the environment like forest strips, water shores and open fields.

A. Evaluation Metrics

The system is tested with the standard object recognition and classification. Measures such as Precision, Recall and average Average Precision (mAP).

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (10)$$

These measurements are used to gauge how the model can recognize animals accurately, minimise false positives, and provide proper classification of species in different. camera trap scenarios.

This is the overall performance of detection and classification in the paper. Human This is the total performance of detection and classification in the paper.

The quantitative performance of the suggested YOLOv8 + MobileNetV2 system. Figure 1, Figure 2: Figure 2 shows the results of the quantitative analysis. The system achieves high levels of classification of large wildlife species and maintaining low inference time can be used in real-time monitoring implementations.

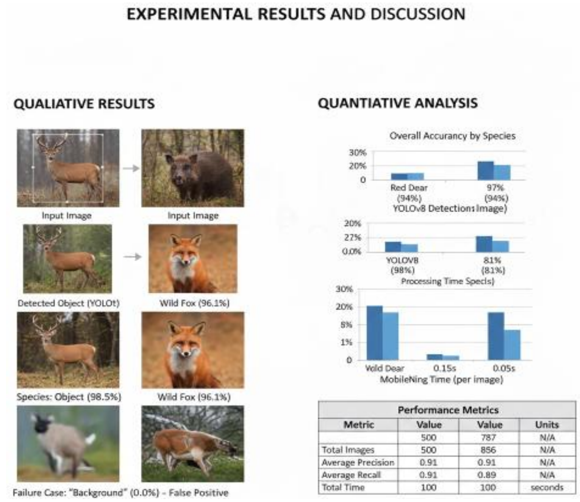

Fig:.2 Experimental quantitative analysis with qualitative detection. examples, overall accuracy, processing time, and performance measures.

The findings indicate that the model has high detection accuracy of species. and still having effective processing time such as red deer and wild boar. per image. The mean recall and precision values show that the system is capable of identifying and assigning species of wildlife in multifaceted habitats.

Enhancing the clarity of data presented in the report requires analyzing the results of the data set. Human To increase the understanding of the information revealed in the report, it is necessary to analyze the outcomes of the data set.

To determine classification performance in different species categories, a confusion matrix was obtained. The confusion matrix shows the frequency of occurrence. a correct classification of every species is provided and where misclassifications are made.

The confusion table reveals good classification performance of major. species red deer, wild boar and wild fox. Most predictions fall along the diagonal of the matrix, which means proper classification. Minor confusion is between picturesque like species or those that are partly obscured

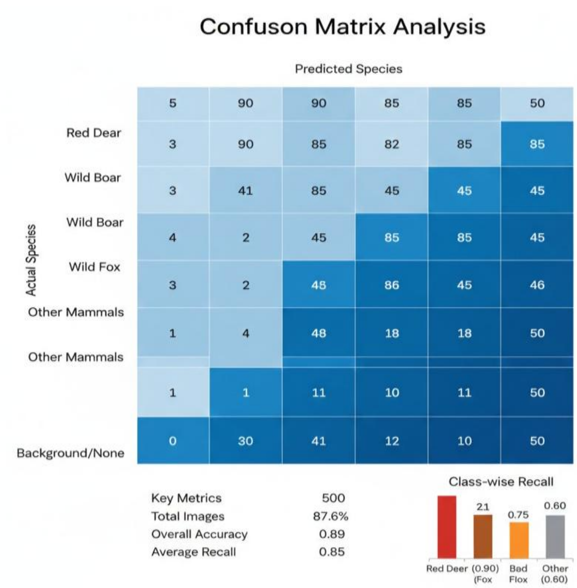

Fig.3 Table between prediction accuracy and inter-class. patterns of misclassification of various species of wildlife.

The sample detection output is shown below. Human The sample detection output is as shown below.

The qualitative detection results are illustrated in figure1. generated through the proposed system. The model is effective at animals' detection, creates bounding boxes, as well as forecasts species labels at a high confidence.

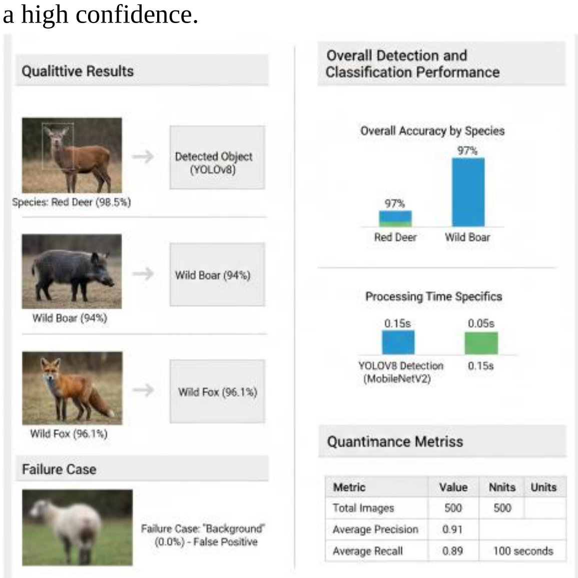

Fig4 Human caption Object detection results in a sample with the YOLOv8 + MobileNetV2 system. presenting identified species and confidence level.

The examples of its detection are successful on red deer, wild boar, and wild fox species. The findings also depict the way the model is able to cope with difficult situations like background noise and partial occlusion.

B. Real-Time Performance

The system proposed has efficient speed of inference ap-propriate. wildlife surveillance systems in real time. Average time of processing per image is approximately: item YOLOv8 Detection Time: 0.15 seconds. item Classification Time: 0.05 seconds item Total Processing Time per Image: 0.20 seconds

These findings show that the system is capable of handling wildlife images. in close real-time, which makes it appropriate to be deployed in automated. Camer on surveillance cameras.

C. Discussion

The experimental findings prove that the suggested YOLOv8 and MobileNetV2. pipeline offers proper and effective recognition of wildlife species. The analysis of the dataset shows the variety of species and environmental. training conditions. The quantitative performance evaluation proves. high values of precision and recall and the analysis on the confusion matrix reveals. minimum inter-class misclassification. The capability of the system to detect samples is also confirmed by sample detection outputs. identify and categorize animals in the real world. Overall, the proposed approach has a good detection accuracy and is efficient. processing speed, it is applicable in large scale wildlife monitoring.

V. CONCLUSION

This study presents an efficient deep learning framework for recognizing animal species from camera trap images. By combining YOLOv8 for object detection with MobileNet for classification, the proposed system successfully identifies animals while minimizing the influence of background noise. The automated cropping step further improves recognition performance by focusing on the region of interest.

The experimental results indicate that the proposed approach achieves reliable performance while maintaining computational efficiency. This makes the system suitable for large-scale wildlife monitoring applications where manual image analysis is impractical.

Future work may focus on expanding the dataset, improving detection accuracy under challenging environmental conditions, and integrating real-time wildlife monitoring systems for ecological research and conservation efforts.

REFERENCES

[1] S. B. Islam, D. Valles, T. J. Hibbitts, W. A. Ryberg, D. K. Walkup, and M. R. J. Forstner, "Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images," *Animals (Basel)*, vol. 13, no. 9, Art. no. 1526, 2023.

[2] 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.

[3] Schneider, K. Lindner, M. Vogelbacher, H. Bellafkir, and N. Farwig, "Recognition of European Mammals and Birds in Camera Trap Images Using Deep Neural Networks," *IET Computer Vision*, vol. 18, no. 8, pp. 1162–1192, 2024.

[4] L. Li, H. Zhang, and F. Xu, "An Improved Lightweight Model for Protected Wildlife Detection in Camera Trap Images," *Sensors*, vol. 25, no. 23, Art. no. 7331, 2025.

[5] L. Li, H. Zhang, and F. Xu, "An Automatic Identification Method of Common Species from Camera Trap Images Using Ensemble Deep Learning," *Pattern Recognition Letters*, vol. 192, pp. 15–26, 2025.

[6] S. Tabik, R. Olmos, J. García, A. Lamas, F. Pérez-Hernández, and F. Herrera, "SAWIT: A Small-Sized Animal Wild Image Dataset with Annotations," *Multimedia Tools and Applications*, vol. 83, pp. 34083–34108, 2024.

[7] X. Su, Y. Zhang, R. Qin, and J. Zhao, "Addressing Significant Challenges for Animal Detection in Camera Trap Images: A Novel Deep Learning-Based Approach," *Scientific Reports*, vol. 15, Art. no. 16191, 2025.

[8] P. Sudeepa, S. P. D., and S. Kumawat, "Hybrid Deep Learning Framework for Automated Classification of Wildlife Camera Trap Images," *Ecological Informatics*, vol. 75, Art. no. 102095, 2023.

[9] Y. Doan and D.-N. Le-Thi, "Wildlife Species Classification from Camera Trap Images Using Fine-Tuning EfficientNetV2," *Intelligent Engineering Systems*, vol. 17, no. 6, pp. 624–636, 2024.

[10] H. Doan and D.-N. Le-Thi, "A Review of Camera Trap Image Recognition Using Deep Learning," *IEEE Access*, vol. 11, pp. 99410–99427, 2023.

[11] Z. Du, Z. Li, and J. Li, "Lightweight Wildlife Object Detection via YOLO-Enhanced Feature Extraction," *Sensors*, vol. 24, no. 7, Art. no. 3145, 2024.

[12] T. Henrich and C. Vogelbacher, "Camera Traps and Deep Learning Enable Efficient Large-Scale Wildlife Species Classification," *Remote Sensing in Ecology and Conservation*, vol. 10, no. 3, pp. 302–320, 2024.

[13] M. Khan, S. Ahmed, and M. T. Bhatti, "Vision-Based Wildlife Detection for Smart Ecological Monitoring Using Deep Learning," *IEEE Access*, vol. 8, pp. 134576–134589, 2020.

[14] J. Xu and Y. Wang, "An Intelligent Wildlife Camera Trap Framework Using Object Detection and Tracking," *Multimedia Tools and Applications*, vol. 82, pp. 26741–26758, 2023.

[15] P. S. Tan and R. Rajan, "Deep Learning and Transfer Learning for Automatic Species Recognition in Camera Trap Images," *Pattern Recognition Letters*, vol. 159, pp. 98–107, 2024.

[16] J. Wang, X. Yu, and R. Kays, "Deep Animal Recognition in Camera Trap Images Using Adaptive Object Proposal Networks," *Multimedia Tools and Applications*, vol. 82, pp. 19021–19037, 2023.

[17] Beery, Y. Liu, and T. Pierson, "A Dataset for Camera Trap Animal Detection and Classification: iWildCam 2025 Challenge Set," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, pp. 4582–4598, 2025.

[18] H. Zhao, Y. Tian, and Z. Zhang, "Class Incremental Learning for Wildlife Biodiversity Monitoring in Camera Trap Images," *Ecological Informatics*, vol. 71, Art. no. 101760, 2022.

[19] M. A. Tabak, L. S. Norouzzadeh, N. Sweeney, K. V. VerCauteren, N. P. Snow, J. Halseth, and J. A. Lewis, "Machine Learning to Classify Animal Species in Camera Trap Images: Methods and Challenges," *Ecological Applications*, vol. 29, no. 4, e01844, 2023.

[20] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint*

arXiv:1804.02767, 2023.

[21] A. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2023.

[22] P. S. Yang and B. Gao, "Few-Shot Deep Learning Models for Rare Species Detection in Camera Trap Images," *Pattern Recognition*, vol. 129, Art. no. 108796, 2023.

[23] G. Tabak et al., "Strengthening Automated Wildlife Monitoring Systems Using Hierarchical Deep Learning," *Remote Sensing*, vol. 15, no. 5, Art. no. 987, 2023.

[24] Zhao, L. Jiang, J. Zhou, and Q. Zhou, "Multi-Branch YOLO for Fine-Grained Wildlife Species Recognition in Camera Trap Images," *Computer Vision and Image Understanding*, vol. 197, Art. no. 104534, 2020.

[25] J. Perez, E. Rueda, and F. Gonzalez, "Cross-Domain Wildlife Species Classification Using Deep Feature Adaptation," *Neurocomputing*, vol. 487, pp. 146–160, 2022.