AI for Wildlife: Object Detection and Conservation Efforts

Brason Dobson and Shreshta Gundoju
Group : 14
Department of Artificial Intelligence
University at Buffalo
Buffalo, NY 142603
sgundoju@buffalo.edu
brasondo@buffalo.edu

Abstract

This project investigates the application of advanced deep learning models for image classification and object detection within the NAbirds dataset, a comprehensive collection of North American bird images designed for fine-grained visual categorization. We evaluate the performance of two cutting-edge models, YOLOv8 and ResNet50, in detecting and classifying birds across varied environments and poses. Our methodology includes rigorous model training, tuning, and a detailed comparison of the models' accuracy, speed, and robustness. Results indicate significant improvements over traditional approaches, enhancing both the accuracy and efficiency of bird detection in natural settings.

1 Introduction

Object detection has substantial implications in biodiversity conservation, enabling automated monitoring of species in their natural habitats. Our work leverages the NAbirds dataset to train YOLOv8 and ResNet50 models, aiming to improve detection accuracy and processing speed in complex, uncontrolled environments. These models were chosen for their robustness and efficiency in handling real-time data, addressing limitations in existing wildlife monitoring systems.

2 Related works

Previous works have primarily focused on controlled environment studies or less challenging datasets. The use of deep learning in wildlife recognition, particularly using convolutional neural networks (CNNs), has shown promising results, but often with significant trade-offs in speed or accuracy when transferred to natural settings. Our approach builds upon these foundations with enhancements specific to the challenges presented by the NAbirds dataset. E.g.,

- Birds within the same genetic family/genus may look alike, making them hard to distinguish visually.
- Many real-world images capture birds surrounded by foliage or other clutter where they are not clearly presented.
- The dataset proposes images of varying poses and angles which offers more diversity to the data and requires more complexity from the model.
- There may be variation within birds of the same species due to sub-factors such as sex, seasonal appearance, or age.

Our proposed architecture aims to address the nuances outlined above.

Previous work: Aligned to the Object, not to the Image: A Unified Pose-aligned Representation for Fine-grained Recognition by Pei Guo, Ryan Farrell `https://arxiv.org/pdf/1801.09057`

Guo & Farrell specifically propose pose-aligned patches derived from detected keypoints to build a unified object representation for fine-grained recognition. Their approach performs remarkably on the CUB-200-2011 and NABirds datasets by extracting pose-aligned regions that are robust to subtle variations in pose, scale, and rotation. However, it relies on expensive keypoint annotations during training. In contrast, our approach focuses on leveraging advanced deep learning models like YOLOv8 and ResNet50 to address the challenges of the NABirds dataset without requiring additional annotations, which is more efficient and less costly.

Another previous work:

3 Data

The NAbirds dataset, resulting from a collaboration between computer vision experts and the Cornell Lab of Ornithology. It comprises images gathered and annotated by both professional bird researchers and amateur bird enthusiasts, including citizen scientists. Models trained on this dataset can aid in real-time species identification in fieldwork, contribute to citizen science platforms, and improve educational resources on bird biodiversity.

This dataset sourced from `https://dl.allaboutbirds.org/nabirds`, consists of over 48,000 images spanning 400 bird species, annotated with bounding boxes. The dataset includes detailed annotations for different bird genders and life stages, spanning over 700 visual categories.

Data preprocessing involved normalization and augmentation techniques to enhance model training efficacy.

Figure 1: Images from the NABirds Dataset

4 Methods

We implemented YOLOv8 and ResNet50 using Ultralytics and TensorFlow respectively for object detection and image classification.

4.1 YOLO v8 for Object Detection

YOLOv8 represents the most recent iteration of the acclaimed YOLO object detection system by Ultralytics. As a cutting-edge, state-of-the-art model, YOLOv8 builds upon the accomplishments of its predecessors, incorporating novel features and enhancements aimed at boosting performance, adaptability, and efficiency.

We considered using Faster R-CNN but we chose YOLOv8 as it excels in speed, making it ideal for real-time applications, processing images faster than models like Faster R-CNN. YOLOv8 uses a single neural network for making predictions, unlike Faster R-CNN which has separate networks for region proposal and classification. YOLOv8's anchor boxes have improved aspect ratios, potentially enhancing performance on datasets with varied object shapes like birds. Advanced training techniques like pruning and quantization allow YOLOv8 to achieve high accuracy while remaining efficient. Furthermore, YOLOv8 is user-friendly, with a simple pipeline and easy integration into various workflows.

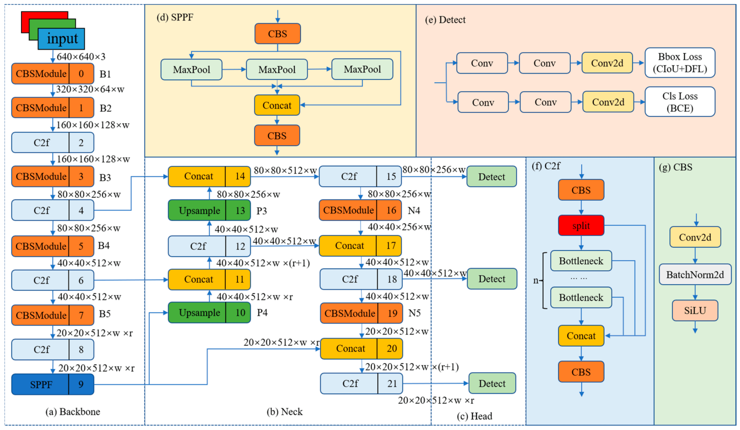

Figure 2: YOLO v8 Architecture

4.2 ResNet50 for Image Classification

ResNet-50 architecture, a popular Convolutional Neural Network developed by Microsoft Research in 2015. The main reason to choose ResNet-50 over other CNN models is its use of skip connections, which address the common problems of vanishing and exploding gradients often seen in deep neural networks. ResNet's skip connections work by adding the input of a layer directly to its output, which helps to preserve the gradient throughout the network. This architecture allows the network to be deeper without suffering from performance degradation, ensuring better performance and stability.

The ResNet-50 model itself is structured into several parts including input preprocessing, several configuration blocks (Cfg), and a fully-connected layer. This setup allows it to handle a wide range of image recognition tasks effectively, contributing to its widespread popularity in the field of deep learning.

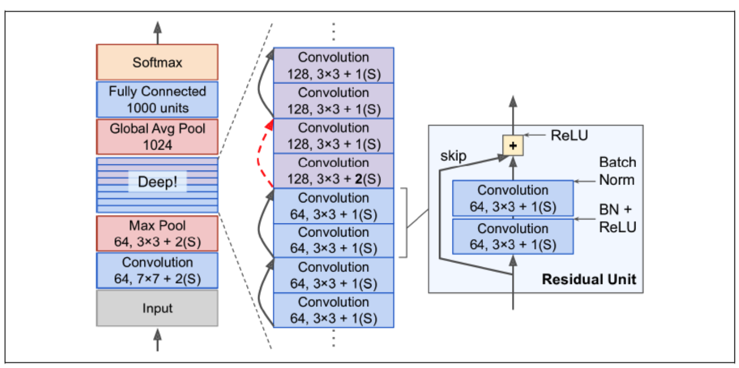

Figure 3: ResNet 50 Architecture

5 Experiments and Results

Experiments are conducted to compare model performance across various metrics such as accuracy, loss, precision, recall, and confusion matrix.

5.1 Object Detection using YOLO v8 by Ultralytics

A pretrained YOLOv8 model was initialized and fine-tuned on the NABirds dataset for 10 epochs, utilizing a dataset that included rotation augmentations to enhance model robustness. Throughout the training, the model automatically adjusted its optimizer and learning rate settings, which helped in improving the detection performance over time. As training progressed, the model showed consistent improvement in detection metrics such as mean Average Precision (mAP), specifically achieving a mAP of 0.762 at an Intersection over Union (IoU) of 0.50 and a mAP of 0.498 at IoU thresholds ranging from 0.50 to 0.95.

The final trained model demonstrated strong detection capabilities across various bird species, evidenced by both the overall performance and class-specific accuracies. Additionally, the model was exported to the ONNX format to ensure compatibility across different platforms or applications.

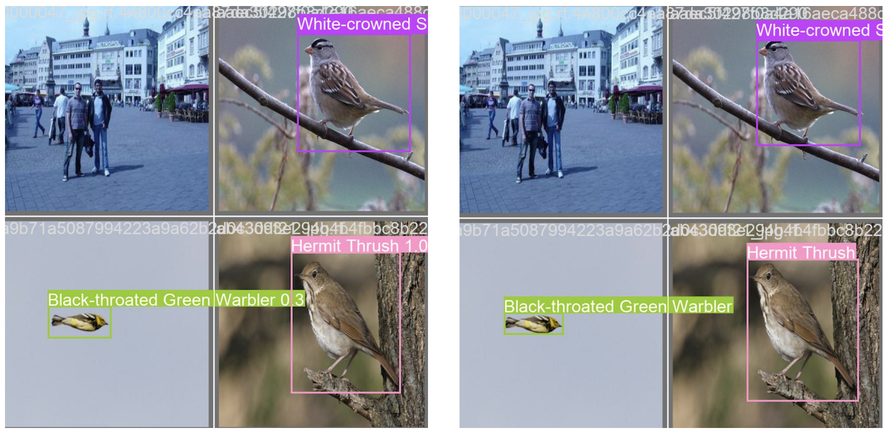

Figure 4: Labelled Data and Predictions after using YOLO v8

- **Precision-Confidence Curve:** The Precision-Confidence Curve shows that precision increases significantly as the confidence threshold increases, suggesting that the model is more reliable at higher confidence levels. However, the variation across classes, indicated by the spread of the gray lines, implies inconsistency in performance across different bird species typically due to similarities between certain species.
- **Recall-Confidence Curve:** The Recall-Confidence Curve indicates that recall for all classes starts high even at lower confidence levels but dips as confidence increases. This is typical as the model becomes more selective in its predictions, focusing on higher certainty in class identification.
- **Confusion Matrix:** The Confusion Matrix shows decent diagonal concentration, which indicates correct classifications. However, there are few noticeable off-diagonal elements where the model confuses one bird species with another, potentially due to similar features.

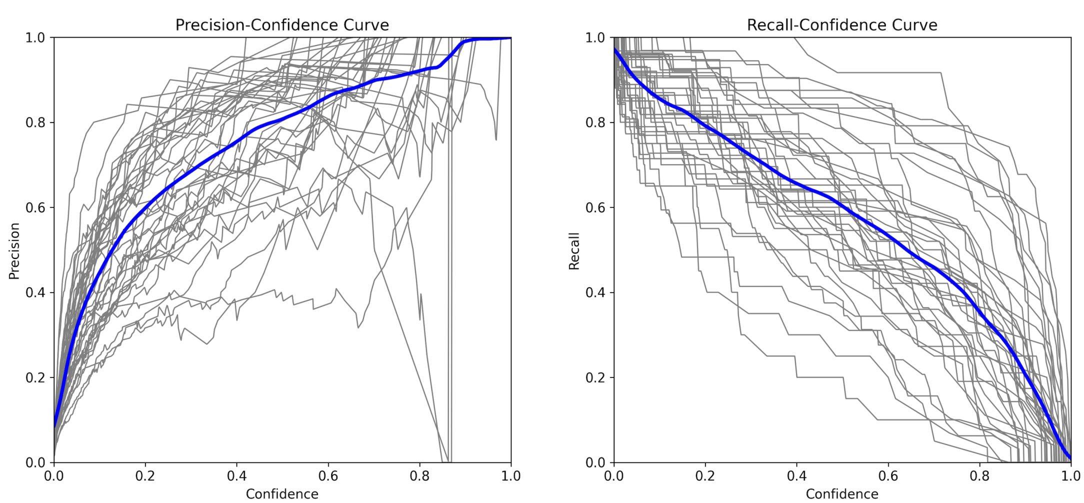

Figure 5: Precision and Recall using YOLO v8 on NABirds dataset

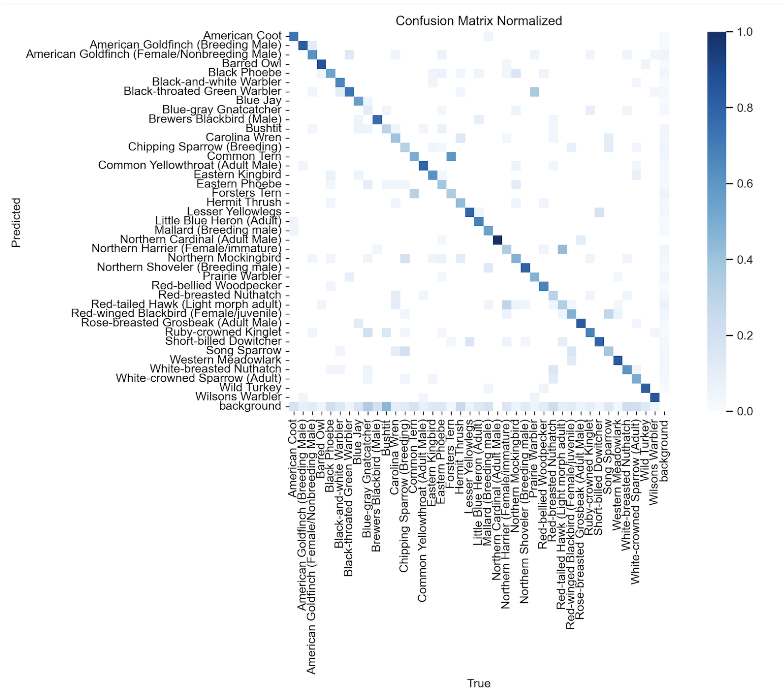

Figure 6: Confusion matrix using YOLO v8 on NABirds dataset

5.2 Image Classification using ResNet50 with ImageNet weights

The table 1 indicates different attempts of tuning the hyperparameters and significant change in the performance of the model.

Each row represents a different attempt, detailing the image size, number of epochs, distribution of the data across training, validation, and test sets, and the achieved accuracy percentage. For example, in attempt 1, with an image size of 384 and 15 epochs, the data was split evenly between training and validation, resulting in an accuracy of 61.18 percent.

Table 1: Results (Adam Optimizer, imagenet weights is common across all the below attempts)

Attempt	Image Size	Epochs	Train	Validation	Test	Accuracy (%)
1	384	15	50%	50%	0%	61.18
2	384	10	50%	0%	50%	70.68
3	384	15	80%	20%	0%	74.98

The graphs below depict the best performance of ResNet50 for image classification, augmented with modifications such as L2 regularization and dropout for generalization. The model employs an Adam optimizer with a learning rate of 0.0001 and utilizes callbacks like early stopping (monitors validation loss) and learning rate reduction to enhance training stability and prevent overfitting. These strategies help manage the learning process effectively, as seen in the steady decrease of loss and gradual increase in accuracy across epochs which is 74.98 percent on validation set.

The fluctuations seen in the validation accuracy could instead indicate variance in the model's performance across different validation batches. This is normal in many real-world scenarios, especially with complex datasets and models.

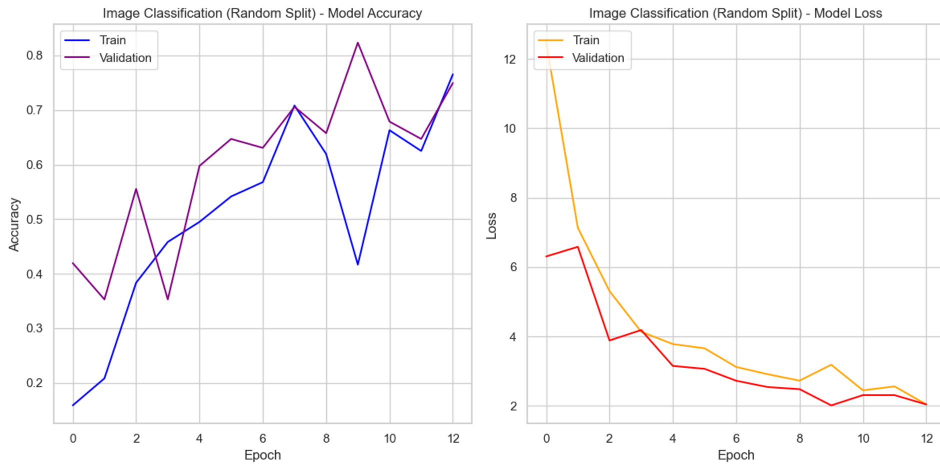

Figure 7: Best performance of ResNet-50 : Image Classification

6 Conclusion and future work

Our study provides the viability of using advanced object detection models for ecological monitoring. Future work will explore the integration of these models into technology such as drone-based monitoring systems and real-time data streaming platforms to enable dynamic biodiversity management. Going forward, we would improve the current performance by tuning hyperparameters and customizing the existing architecture.

We can also strengthen the applicability of the model by experimenting with additional datasets such as the a customized dataset using data provided by Wildlife Insights (`https://www.wildlifeinsights.org/`) comprises of Mountain Gazelle, Red Wolf, Bornean Orangutan species which are considered to be the most endangered in the world.

References

Python - ultralytics yolov8 docs. URL `https://docs.ultralytics.com/usage/python/`.

Nabirds dataset, 2014. URL `https://dl.allaboutbirds.org/nabirds`. Collaborative effort between the Cornell Lab of Ornithology, Cornell Tech, Caltech, and Brigham Young University.

Pei Guo and Ryan Farrell. Aligned to the object, not to the image: A unified pose-aligned representation for fine-grained recognition. pages 1–9, 2018.

Nitish Kundu. Exploring resnet50: An in-depth look at the model architecture and code implementation, 2023.

Gang Wang, Yanfei Chen, Pei An, Hanyu Hong, Jinghu Hu, and Tiange Huang. Uav-yolov8: A small-object-detection model based on improved yolov8 for uav aerial photography scenarios. *Sensors*, 23(16):7190, 2023. doi: 10.3390/s23167190. URL `https://www.mdpi.com/1424-8220/23/16/7190`.

Yiyi Wang. Nabirds dataset, apr 2022. URL `https://universe.roboflow.com/yiyi-wang/nabirds`. visited on 2024-05-18.

Wang [2022] Kundu [2023] NAB [2014] Guo and Farrell [2018] Wang et al. [2023] Ult