Raunak Advani, Austin Barish, Matene Zamaninia
ANLY-677: Image Mining and Vision Analytics
Dr. Nakul R. Padalkar
9 August 2023

Group 3 Final Project

1. Introduction

This project aims to employ computer vision techniques to classify images that contain birds. To be more precise, we would like to count the number of birds present in various images taken in the Amazon Rainforest, specifically focusing on the challenging task of being able to identify birds in the rainforest environment where they are often camouflaged or hidden. By leveraging advanced image processing and machine learning techniques, we would like to enable automated bird identification, which can have various applications in ecological studies, conservation efforts, and wildlife monitoring. The majority of studies conducted in this area utilize Machine Learning and Deep Learning techniques. Convolutional Neural Networks (CNNs) are especially useful for finding patterns in images to recognize objects, classes, and categories.

2. Background/Literature Review

Past literature on the identification of living creatures in images has demonstrated the use of deep learning techniques for automatically identifying, counting, and describing wild animals in images. For example, although the focus is not solely on birds, the methodologies and algorithms developed in the work by Norouzzadeh et al. (2018) can provide valuable insights into handling complex environmental conditions, including occlusions and varying lighting, which are relevant to our project's challenges in identifying birds in the forest. In their study, Berg & Belhumeur (2013) address the challenge of fine-grained categorization, which is relevant to our project because of the difficult scenarios in which birds need to be identified. Given the success of this project, its methods can be expanded to other living organisms as well.

3. Analysis/Methods

Data Gathering and Preliminary Methods Review

Data Gathering
For this project, we obtained a comprehensive dataset from Brian Griffiths, a faculty member in the Environmental Science Department at Georgetown University. The dataset comprises approximately 40,000 images captured in 75 different sites within the Amazon rainforest. Each

image contains one or more bird species, including various types of pigeons and parrots. The images are in .jpg format and cover a wide range of lighting conditions and background complexities commonly encountered in forest environments.
The dataset's large size and diverse set of bird species provide a solid foundation for training and evaluating our computer vision algorithms. However, due to the challenging nature of bird identification in the forest, the data must be carefully preprocessed, and this will be essential to ensure accurate and reliable results.

Research Questions

- What are some of the conditions and scenarios that need to be considered when identifying the presence of birds in images?
- Are some conditions and scenarios more susceptible to inaccuracies in identifying than others?
- What is an ‘acceptable’ accuracy rate when identifying the presence of birds?
- Do different models provide different accuracy rates? What is the ‘best’ model to address this problem?

Preliminary Methods Review
Thus, this project aims to leverage computer vision techniques to address the challenging task of bird identification in forest environments. By utilizing the dataset provided by Brian Griffiths, we can develop and evaluate novel algorithms that accurately identify the presence of birds, despite their camouflage and hiding behavior. The project's outcomes hold the potential to contribute to ecological research, conservation efforts, and the development of automated tools for bird identification.

Empirical Design

Data Description

For our project, the dataset we will be analyzing was obtained from Brian Griffiths, a faculty member at Georgetown University, teaching in the Environmental Science Department. The dataset is the result of extensive research conducted in the Amazon Rainforest, specifically in 75 different sites within the region of South America. The dataset comprises over 14,000 images. The primary focus of our analysis is to develop a bird segmentation model to count the number of birds present in a given image. Therefore, the dataset contains a wide range of images showcasing the diverse bird species found in the Amazon Rainforest, namely parrots and pigeons which can be disguised by the parrots’ green color and the pigeons’ Earth toned gray. Within these, there are approximately twelve different species of parrots and pigeons featured in the images. However, we will focus on the number of total birds rather than the species' differentiation. We will access this dataset through a private gitlab folder.

Prior to the analysis, the images will be preprocessed to ensure consistency in size and format, allowing for accurate model training and evaluation.

Regarding the acquisition method, the dataset was collected through extensive fieldwork conducted by researchers in the Amazon Rainforest. The researchers employed cameras strategically placed in various locations to capture images of birds in their natural habitats. This approach ensured a comprehensive representation of bird species and their behaviors within the region. As for the training/test strategy, we will follow an 80:20 ratio for splitting the collected images into training and testing sets. Approximately 80% of the dataset will be used for training the model, allowing it to learn patterns and features for accurate bird segmentation and identification. The remaining 20% will be reserved for evaluating the model's performance and assessing its generalization capabilities. This train/test split ratio has been selected to strike a balance between model training and evaluation to achieve a high accuracy score.

Dependent Variable/Feature/Object

The dependent variable or feature/object of our analysis is the detection and segmentation of birds within the given images. The purpose of the analysis is to develop a model that can accurately identify the presence and counts of birds in images captured in the Amazon Rainforest.

The specific objective is to perform bird detection and segmentation, which involves two main tasks:

1) Bird Detection: The model should be able to identify whether a bird is present in an image or not. This involves analyzing the visual characteristics and patterns associated with birds, such as their shapes, colors, and textures. The model needs to distinguish birds from other objects or backgrounds in the images.

2) Bird Segmentation: Once a bird is detected in an image, the model should accurately segment and outline the boundaries of the bird. This segmentation task involves assigning a binary mask to each image, where the pixels belonging to the bird are marked as foreground and the rest as background. Precise bird segmentation is essential for subsequent analysis and species identification and will allow us to count the number of birds present in a given image.

Preprocessing Steps

All of the images feature a timestamp, location information, and logo on the bottom. This includes the weather, date, time, and the Bushnell logo. Bushnell is the equipment company used to take these images. Due to the artificial coloring of this part of the image, we will need to remove it in order to avoid it interfering with the actual photo data. Therefore, all of the photos will need to be cropped to exclude the Bushnell logo, the highest point on the timestamp. This will remove a small portion of the image itself, however, the crop is necessary to correctly process the images. Since the images are all the same proportions, this should not be too difficult. Furthermore, this will ensure that all of the images are of the same size and quality before they are entered into the model.

The dataset is also far larger than necessary in order to reach statistical significance within a training model. Given the excess of 40,000 images, we have more than enough data to test and train our data. The difference between a model with 40,000 images to train on as opposed to 1,000 images to train on is negligible; the thousands of additional data points will do little to improve the accuracy of the model due to the law of diminishing returns. Therefore, we will randomly sample the 40,000 images down to 6,000 images in order to save on computing time and space. In further analyses, given the time and processing power to operate on the entire dataset, we could seek to expand our analysis. For now, this sampling should not impact the accuracy of our models.

Training and Test Sets

The dataset will be split into a training set of 80% and a test and validation set of 20% respectively. The reason for creating a validation set is to prevent data leakage and to maintain a relatively substantial portion of the data for comparison on truly unseen data. Since our original dataset is not labeled, we will most likely need to use the pre-trained models on a secondary dataset for Bird Identification first; this is a form of unsupervised learning used to label our data. We will also be able to examine some of the images that the model has labeled in order to ensure its accuracy. The model will be modified until we are confident in the accuracy of its labeling methods. Our final labeled dataset has roughly 3000 labeled images. After the model has been trained, we will implement it on the obtained bird dataset in order to label the data. Alternatively, an annotation tool such as Makesense AI can be used for labeling the dataset. With the labeled data at hand, we could use basic Classification Evaluation methods such as Accuracy, Confusion Matrix as well as ROC and AUC to measure model performance (Sharma, Jain, & Mishra, 2018).

Deep Learning Methods

For the classification task a number of CNNs such as Resnet, ReLu, Inception-3, YOLOv5, and possibly the EfficientDet frameworks are considered (Mekhalfi et al., 2022). In further iterations, transfer learning methods will be implemented in order to improve model detection results.

Experiments

Table 1 Deep Learning Architectures

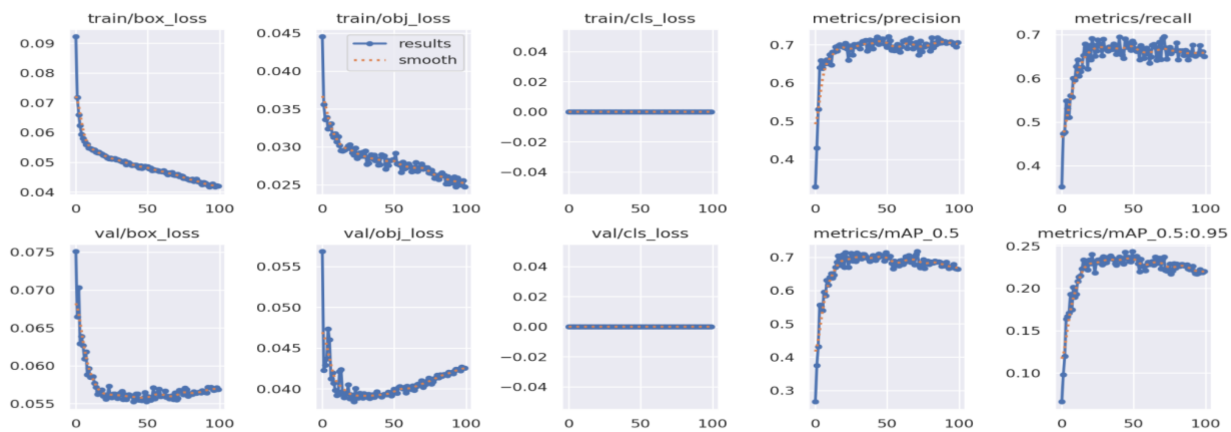
Architecture	**No. of layers**	**Short description**
YOLOv5	8	YOLOv5 is a cutting-edge, state-of-the-art (SOTA) model that builds upon the success of previous YOLO versions and introduces new features and improvements to further boost performance and flexibility. YOLOv5 is designed to be fast, accurate, and easy to use, making it an excellent choice for a wide range of object detection, instance segmentation and image classification task (Pytorch)

4. Discussion/Results

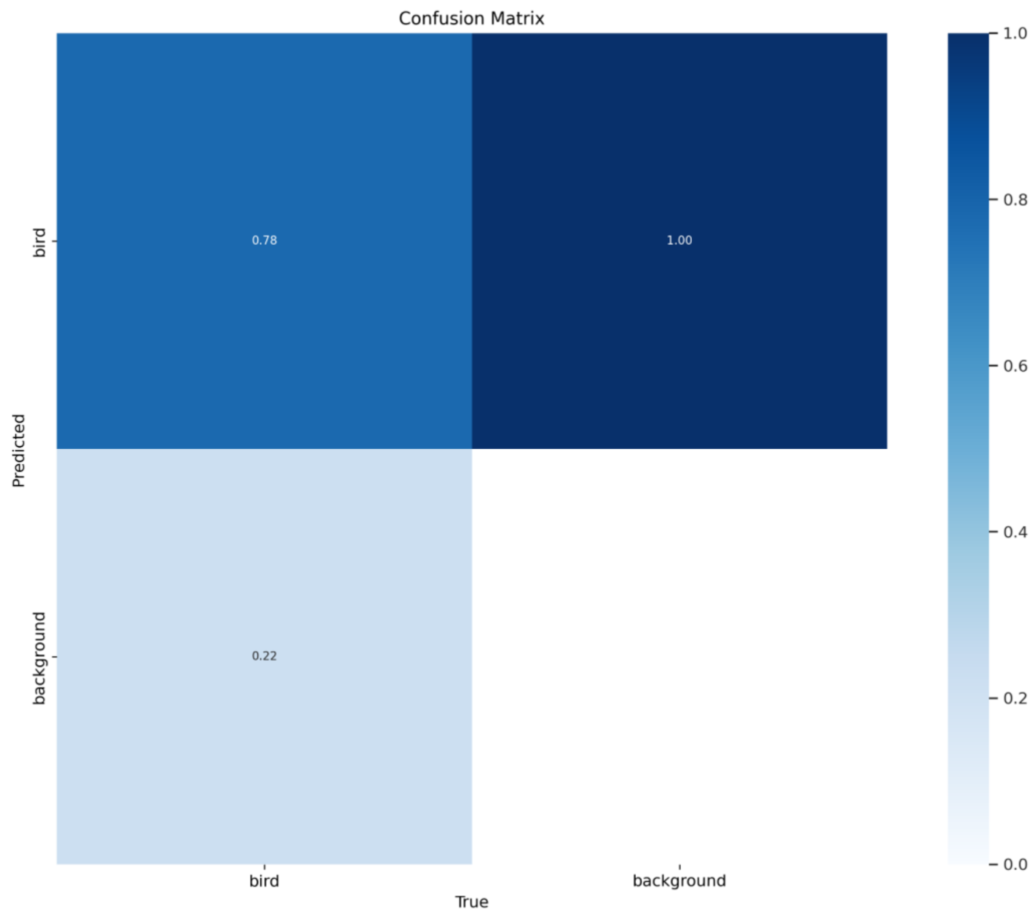
Since this problem is similar to a regression problem, instead of providing a simple accuracy measure, we will provide both accuracy as well as an RMSE value since it provides a clearer picture of the models overall performance The YOLOv5 model has an overall RMSE of 1.78. In this example and, for this experiment, this translates to an overall miscounting of 1.78 birds per image, which can be considered an acceptable number given that the project does not have security implications and does not require a non-fault model. In the next iterations, additional models will be implemented and compared to the YOLOv5 base model.

Table 2: Model Performance measures.

Architecture	RMSE	MAE	Accuracy
YOLOv5 M	1.78	1.22	33%



Graph 1: YOLOv5 Performance Graphs



Graph 2: YOLOv5 Confusion Matrix

The confusion matrix above is different from our normal confusion matrices. In the given task of counting the number of birds in an image, the confusion matrix is counting the number of times a given detected bounding box, in a single image, was in fact a bird, vs when it was not a bird. The overall model performance for all correctly identified numbers of birds in a given image still remains at 33%. One of our research questions sought to answer what an acceptable accuracy rate for counting the number of birds in an image is. Our results here indicate that it is likely far lower than the typical accuracy rates we look for in models (>90%) and likely something much lower as exact counts are extremely difficult to come by. In the future, we hope to incorporate more accuracy measures to determine the models ability to recognize more approximate numbers of birds so as to avoid dismissing it as ‘inaccurate’ when it miscounts the number of birds by one of 5, for example. Our current measure emphasizes the regressive nature of this task.

Citations

Berg, T., & Belhumeur, P. N. (2013). Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2013.128

Sharma, N., Jain, V., & Mishra, A. (2018). An analysis of convolutional neural networks for image classification. Procedia Computer Science, 132, 377-384. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2018.05.198

Raj, S., Garyali, S., Kumar, S., & Shidnal, S. (2020). Image-based Bird Species Identification using Convolutional Neural Network. International Journal of Engineering Research & Technology (IJERT), 9(6).

M. L. Mekhalfi, C. Nicolò, Y. Bazi, M. M. A. Rahhal, N. A. Alsharif and E. A. Maghayreh, "Contrasting YOLOv5, Transformer, and EfficientDet Detectors for Crop Circle Detection in Desert," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 3003205, doi: 10.1109/LGRS.2021.3085139.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716-E5725. doi: 10.1073/pnas.1719367115