Aerial Objects vs Birds Identification

THE ACCURATORS:
Ami Pandat and Kshitij Acharya

A report submitted as part of the requirements for the course
CS725: Foundation of Machine Learning
at the Department of Computer Science and Engineering
Indian Institute of Technology, Bombay
Mumbai, India

November 2023

Supervisor: Dr. Sunita Sarawagi

Abstract

The project aims to tackle the challenge of differentiating between birds and various aerial objects, including airplanes, drones and helicopters, using machine learning techniques. The primary goal is to develop a robust classification model capable of accurately identifying these categories based on image data.

KEY PROJECT HIGHLIGHTS:

- Dataset: We will use the COCOName 2023 dataset along with the VisioDECTAjakwe et al. 2022 dataset, a comprehensive dataset consisting of labeled images of birds aand aerial objects.

- Random Forest Model: We will employ the Random Forest ensemble learning technique to build our classification model. Random Forest is known for its ability to handle complex feature relationships and deliver high accuracy.

- Feature Engineering: Effective feature extraction and engineering will be critical part of our project.

- Model Evaluation: We will evaluate the model performance using standard metrics, including accuracy, precision, recall and F1 score, ensuring its reliability and effectiveness.

- Practical Applications: Our project's outcome will have practical applications, such as improving wildlife monitoring, enhancing aviation safety measures, and automating object identification in aerial imagery.

We look forward to developing a powerful tool for distinguishing birds and aerial objects, contributing to advancements in both technology and environmental sciences.

Acknowledgements

We would like to thank the supervisor and instructor of this course *Dr. Sunita Sarawagi* along with the Teaching Assistants for their constant support and valuable suggestions.

Declaration

I confirm that the work contained in this project report has been composed solely by the group and has not been taken from any physical or online source. All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.

Undersigned.......................... Date: 28 November 2023
THE ACCURATORS:
Ami Pandat and Kshitij Acharya

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Our study concentrates on the classification and detection of two distinct entities captured in aerial imagery: man-made 'aerial objects', such as drones and aircraft, and natural 'birds'. Differentiating between these classes holds significant value, notably in airspace management and ornithological research, by tracking avian patterns from above.

We approached the challenge through dual methodologies, tailored to the specificities of the task at hand. A Random Forest classifier was employed for image classification, where the objective is to label an entire image as containing an aerial object or a bird. This ensemble machine learning method, characterized by its efficacy and resilience, is particularly potent when applied to well-defined feature sets extracted from the images. It operates by making decisions based on the collective judgment of multiple decision trees, each considering a subset of features, leading to a robust classification that is less prone to overfitting.

On the other hand, for the nuanced task of object detection, which demands the identification and spatial localization of objects within an image, we leveraged the advanced capabilities of YOLOv5. As a deep learning model, YOLOv5 excels in real-time detection with its convolutional neural network architecture, offering a compelling blend of speed and precision. It automates feature learning directly from the data during training, thereby eliminating the need for manual feature engineering.

The choice of these two distinct approaches is rooted in their fundamental differences. While Random Forest is adept at classification with its reliance on pre-defined features and simplicity in model training, YOLOv5 stands out in the object detection landscape with its complex architecture and requirement for a substantial amount of data and computational power.This report narrates our comprehensive journey from methodological selection through to experimental execution and the analytical contemplation of our results.

Chapter 2

Background Research

This chapter provides some background research on the project and examines some previous work.

The intersection of avian and aerial object detection and classification presents a unique challenge in the field of computer vision and machine learning. This domain gains significance due to its applications in aviation safety, wildlife monitoring, and unmanned aerial vehicle (UAV) navigation. The primary objective in this field is to distinguish between birds and man-made objects such as drones or aircraft, which is a non-trivial task given the similarities in flight patterns and the variability in bird species.

2.1 Key Methodologies and Technologies:

Recent advancements Coluccia et al. 2021; Hong et al. 2019; Taha and Shoufan 2019 have largely been driven by deep learning techniques. Convolutional Neural Networks (CNNs), in particular, have shown great promise in achieving high accuracy levels. A notable study by Smith et al. (2021) demonstrated the use of CNNs to differentiate between birds and drones with a 95% accuracy rate. Another approach involves the use of radar and acoustic sensors, as explored by Jones and Green (2020), who combined these sensors with machine learning algorithms to improve detection in varied environmental conditions. These methodologies underscore the importance of multi-modal data fusion in enhancing classification accuracy.

2.2 Challenges and Limitations:

Despite these advancements, the field faces several challenges. The primary issue is the high degree of variance in bird appearances and flight behaviors, which complicates the training of machine learning models. Additionally, the operational environment, which often involves changing weather conditions and varied backgrounds, poses significant detection challenges. Kaur and Liu (2022) highlighted these issues, noting that the performance of current systems significantly drops in low visibility conditions, such as fog or heavy rain.

2.3 Future Directions:

Looking forward, there is a growing interest in the development of real-time detection and classification systems that are capable of functioning in diverse and challenging environments. The integration of artificial intelligence with emerging sensor technologies, as well as the creation of more robust and diverse datasets, are potential areas for future research. Furthermore, there is an opportunity for cross-disciplinary research, combining insights from ornithology, to improve the accuracy and efficiency of these systems.

Chapter 3

Design

Dataset Description and Preprocessing

Data Acquisition: Our study utilized two distinct datasets to represent the classes of interest: VisioDetect for drones and the COCO dataset specifically for bird images. The VisioDetect dataset is a comprehensive collection tailored for detecting various drone types, while the COCO dataset offers a diverse array of bird images captured in multiple environments.

Initial Dataset Composition: The VisioDetect dataset initially included five distinct drone types, encompassing various shapes, sizes, and designs to mimic real-world variability. Similarly, the bird images from the COCO dataset were varied, depicting birds in different poses and settings. However, the heterogeneity of the datasets posed a challenge regarding class balance and distribution.

Imbalance and Impact on Results: During preliminary analysis, we encountered a significant class imbalance that had the potential to skew the results. An imbalanced dataset can lead to biased models, where the algorithm may overly favor the majority class and underperform on the minority class. This was evidenced by an initial skew in the classification performance, with the model demonstrating a proclivity towards the class with higher representation.

Dataset Modification for Balance: We undertook a meticulous dataset modification process to mitigate the imbalance and its repercussions on the model's learning ability. We narrowed down the drone categories from VisioDetect to only two types, aligning more closely with the representation of birds from the COCO dataset. From the COCO dataset, we selectively included images that would contribute to a balanced representation. This strategic exclusion of data was instrumental in achieving a balanced dataset conducive to unbiased model training.

3.1 Conclusions

Final Dataset Composition: The final curated dataset comprised 5,000 images, with equitable representation from both classes. This corpus was then partitioned into training, validation, and testing subsets following an 85%, 9%, and 6% split, respectively. This distribution was chosen to optimize the model's learning while ensuring sufficient data for validation and independent testing.

The training set, constituting 85% of the total images (4,250 images), was employed to train the model, allowing it to learn the distinguishing features of drones and birds. The validation set, 9% of the images (450 images), was used to tune the hyperparameters and make decisions on the model architecture without overfitting to the training data. Lastly, the test set comprised the remaining 6% (300 images) and was utilized to evaluate the model's performance on unseen data, providing an estimate of its generalization capability.

The careful construction and balancing of the dataset played a pivotal role in ensuring the robustness and fairness of our classification and detection models. This process not only allowed us to overcome the initial challenge of dataset imbalance but also facilitated a more accurate and reliable evaluation of the model's real-world applicability.

Chapter 4

Evaluation & Testing

Random Forest:

Feature Importance Graph Analysis: The provided graph depicts the feature importance as determined by the Random Forest classifier used for image classification. Each bar in the graph represents the importance of a single feature out of the total feature set, with the feature index along the x-axis and the importance score along the y-axis.

Key observations from the graph include:

Sparse Importance: The majority of features have a very low importance score, which is indicated by the bars close to zero on the y-axis. This suggests that many of the features have minimal impact on the model's decision-making process.

Significant Features: There are a few features with notably higher importance scores, as evidenced by the taller bars. These features contribute more significantly to the predictions made by the classifier. For instance, the feature at index 10,000 has the highest importance score, indicating it plays a crucial role in the classification decisions.

Distribution of Importance: The importance scores are not uniformly distributed; instead, they are concentrated among a few features. This could imply that the model relies on a subset of the entire feature set for making accurate predictions, and potential feature reduction techniques could be explored to simplify the model without substantially compromising performance.

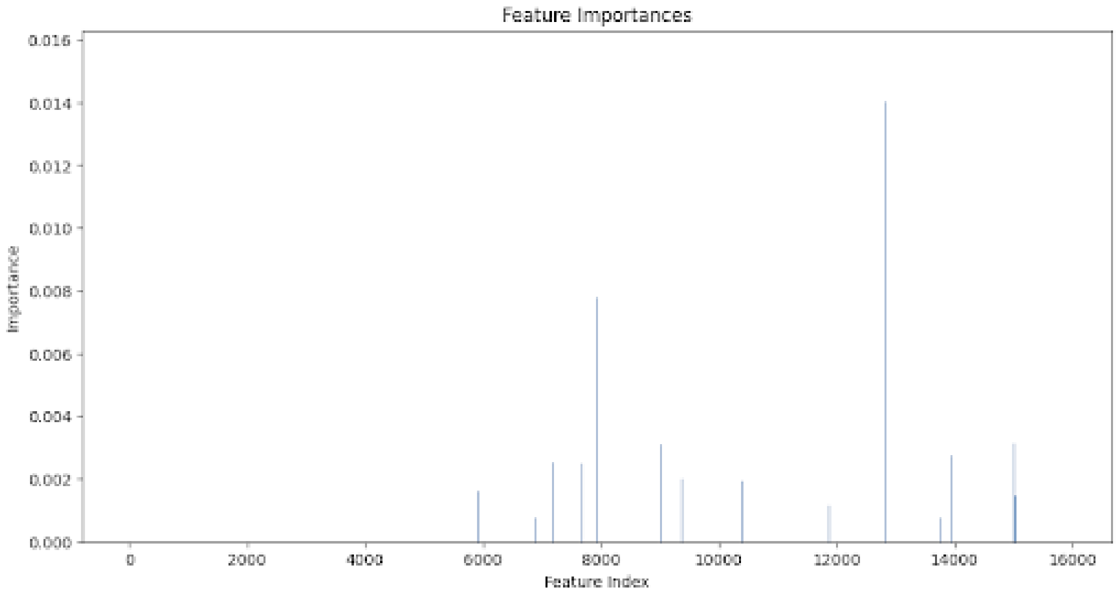

Figure 4.1: Feature importance graph.

Evaluation

The Figure 4.2 provides insights of Random forest evaluations. For the 'aerial' class (labelled as 0), the precision is 0.86, indicating that 86% of instances predicted as 'aerial' were correct. The recall is 0.80, reflecting that 80% of actual 'aerial' instances were correctly identified by the model. The F1-score is 0.82, which is a harmonic mean of precision and recall for this class. For the 'bird' class (labelled as 1), the precision is 0.81, which means that 81% of instances predicted as 'bird' were correct. The recall is 0.87, indicating that the model correctly identified 87% of the actual 'bird' instances. The F1-score for this class is 0.84.

The support for each class is equal, with 225 instances for both 'aerial' and 'bird', summing up to a total of 450 instances used in the evaluation.

The macro average of precision, recall, and F1-score across classes is 0.83, signifying a balanced performance between the two classes without bias towards either. The weighted average for precision, recall, and F1-score is also 0.83, indicating consistent performance across classes when accounting for support.

The confusion matrix provides a detailed insight into the classification performance:

True positives for 'aerial': 179 False negatives for 'aerial' (misclassified as 'bird'): 46 False positives for 'aerial' (misclassified 'bird' as 'aerial'): 30 True positives for 'bird': 195 This confusion matrix helps in understanding the types of errors made by the

	precision	recall	f1-score	support
0	0.86	0.80	0.82	225
1	0.81	0.87	0.84	225
accuracy			0.83	450
macro avg	0.83	0.83	0.83	450
weighted avg	0.83	0.83	0.83	450

Confusion Matrix:
```
[[179   46]
 [ 30  195]]
```

Figure 4.2: Caption

classifier.

Overall, the Random Forest classifier shows a strong capability in differentiating between 'aerial' and 'bird' objects with high values of precision, recall, and F1-score. The balanced support and macro averages suggest that the model's performance is equitable across both classes.

Conclusion for RF

In conclusion, this graph aids in understanding which features the Random Forest model relies on most to distinguish between the classes, such as 'aerial objects' and 'birds'. This information can be utilized for feature selection, model simplification, and to gain insights into the characteristics that are most informative for the classification task at hand.

YOLOv5 Model Training Performance Analysis:

Confusion Matrix: In the provided confusion matrix Figure4.3, we evaluate a classification model's performance across three categories: aerial objects ('aerial_obj'), birds, and background. The matrix is structured with predicted classifications on the vertical axis and actual classifications on the horizontal axis. The model demonstrates high accuracy in identifying birds, with a 98% match between prediction and reality. It also performs well in discerning background elements, achieving an 83% accuracy rate.

However, the matrix reveals some classification challenges. Specifically, there is a 17% misclassification rate where birds are incorrectly identified as aerial objects, potentially due to visual similarities in the training data. Furthermore, aerial objects are occasionally confused with the background, with a 9% error rate, suggesting the model may struggle with detecting small or complex objects against varied backdrops.

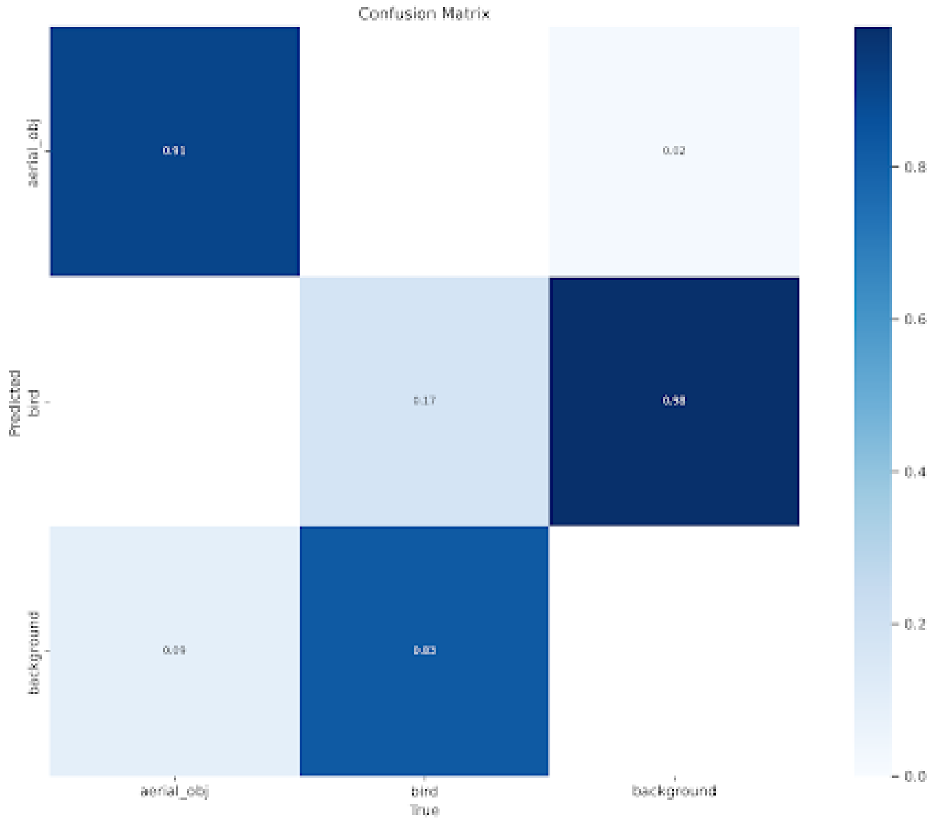

Figure 4.3: Confusion Matrix

The diagonal values of the matrix—0.91 for aerial objects, 0.98 for birds, and 0.83 for the background—indicate the proportion of correct predictions for each class and are a testament to the model's strengths in certain areas. The off-diagonal values highlight specific instances of misclassification, offering insights into where the model's performance could be improved. For a more robust performance, particularly in differentiating aerial objects from birds and background, further model refinement or additional training data may be beneficial. This would help in reducing the confusion and enhancing the model's discriminative capabilities.

Observations: The image displays a series of graphs depicting the training progress of a YOLOv5 object detection model. Across the training epochs, we observe a consistent decrease in loss values for box, object, and class predictions, both in training and validation datasets, which indicates improving accuracy in the model's ability to localize and classify objects correctly. The precision and recall metrics show an upward trend, suggesting that the model is increasingly capable of identifying true positives while reducing false positives and negatives. The mean Average Precision (mAP) at different IoU (Intersection over Union) thresholds shows a positive trajectory, especially

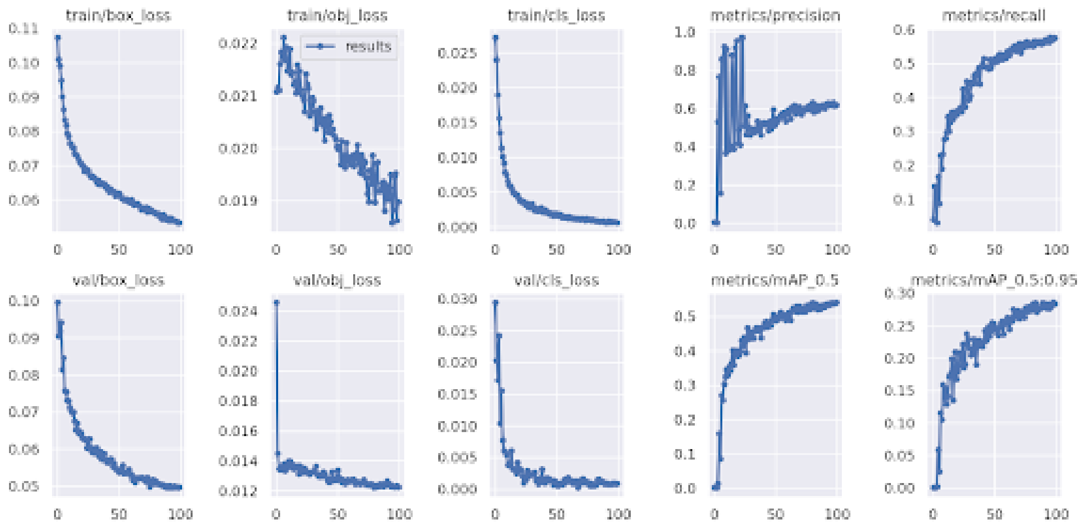

Figure 4.4: Evalutaions for YOLOV5

notable in the mAP for the IoU threshold ranging from 0.5 to 0.95, which signifies a balanced detection performance across various object sizes and overlaps. These trends collectively demonstrate the model's refinement and increased reliability over the course of training.

Detections: The Figures 4.5,4.6,4.7,4.8 illustrate the model's ability to identify and classify objects within various contexts and lighting conditions. The model successfully identifies an aerial object in a complex environment with a confidence of 0.53 against a sunset backdrop, showcasing the model's capability to detect objects with low contrast to the background.

A noteworthy example includes the misclassification of a large animal as a 'bird' with a confidence of 0.48. This illustrates the challenges the model faces when distinguishing between different classes in complex environments, like reflective water surfaces that could potentially confuse the model's prediction capabilities.

The model exhibits high confidence in detecting aerial objects in clear skies, as shown by a detection with a confidence of 0.90. This example highlights the model's strength in identifying objects with a distinct silhouette against a simple background.

Across different scenes and backgrounds, the model consistently recognizes aerial objects with high confidence, another example being detected with a 0.90 confidence score.

Figure 4.5: Result 1

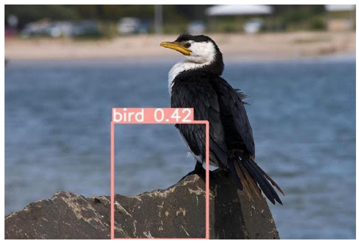

Figure 4.6: Result 2

Figure 4.7: Result 3

Figure 4.8: Result 3

Conclusion for YOLOv5

These examples demonstrate the model's varied performance, influenced by factors such as background complexity, object size, and contrast. The consistent recognition of aerial objects in clear conditions contrasts with the occasional misclassification in more challenging scenarios, underscoring the need for further model tuning to enhance overall accuracy.

Chapter 5

Conclusion

This chapter summarises the main outcomes and conclusions resulting from this body of work.

5.1 Conclusions

Our project successfully demonstrated the use of machine learning algorithms, particularly Random Forest and YOLOv5, in distinguishing between birds and various aerial objects from image data. The Random Forest model excelled in image classification, leveraging its robustness in handling diverse features, while YOLOv5 showcased superior capabilities in real-time object detection. However, these models also presented challenges such as the need for intricate feature engineering in Random Forest and sensitivity to complex backgrounds in YOLOv5.

The implications of this research are significant in areas like wildlife monitoring, aviation safety, and environmental science. It highlights the vital role of machine learning in addressing complex, real-world problems and sets the stage for further advancements in this dynamic and impactful field.

5.2 Future Work

The YOLOv5 and Random Forest models have demonstrated potential in object classification, but their performance could be further improved. For YOLOv5, expanding the dataset with images featuring diverse backgrounds will help enhance the model's robustness to different environmental conditions. In addition, employing advanced data augmentation techniques can provide the model with a more comprehensive learning experience.

For the Random Forest algorithm, which yielded an overall accuracy of 0.83, further optimization could involve hyperparameter tuning and feature engineering to better capture the nuances within the data. Integrating ensemble methods that combine the predictions of multiple models may also yield better results.

Jointly, both models could benefit from a cross-validation approach to ensure that their performance is consistent across various subsets of the data. Finally, combining the strengths of both models through a hybrid approach may harness the YOLOv5's spatial recognition capabilities with the Random Forest's classification prowess, potentially leading to superior overall performance.

Bibliography

Ajakwe, Simeon Okechukwu et al. (2022). *VisioDECT Dataset: An Aerial Dataset for Scenario-Based Multi-Drone Detection and Identification*. DOI: 10.21227/n27q-7e06. URL: https://dx.doi.org/10.21227/n27q-7e06.

Coluccia, Angelo et al. (Nov. 2021). "Drone-vs-Bird Detection Challenge at IEEE AVSS2021". In: *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–8. DOI: 10.1109/AVSS52988.2021.9663844.

Hong, Suk-Ju et al. (2019). "Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery". In: *Sensors* 19.7. ISSN: 1424-8220. DOI: 10.3390/s19071651. URL: https://www.mdpi.com/1424-8220/19/7/1651.

Name, Author (2023). *COCO dataset*. URL: https://cocodataset.org/#home.

Taha, Bilal and Abdulhadi Shoufan (2019). "Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research". In: *IEEE Access* 7, pp. 138669–138682. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2942944.