Atta Leafcutter-Ant Image Classification with Machine Learning

Completed for the Certificate in Scientific Computation and Data Sciences
Spring 2024

Steven Doan
Bachelor of Science in Economics
Department of Economics
College of Liberal Arts

Fatma Tarlaci, Ph.D.
Adjunct Assistant Professor of Computer Science
Department of Computer Science

Introduction

Rapid ecological changes and pressing needs for ecological studies underscore the importance of advancing entomological research, specifically through the adoption of artificial intelligence (AI) and deep learning technologies. The identification of species through image recognition stands as a testimony to one of the most successful applications of biological image data, bridging the expertise of taxonomists and computer scientists. This synergistic collaboration has catalyzed the evolution of species identification and classification from traditional manual taxonomy to sophisticated use of convolutional neural networks (CNNs). Such technological advancements have not only simplified processing of extensive ecological datasets, but have also significantly boosted accuracy and efficiency—critical facets for effective species monitoring and ecological studies.

Despite advancements, challenges persist. The meticulous process of labeling extensive training datasets for machine learning purposes and the deployment of intricate models necessitate considerable computational resources. These challenges, particularly acute in the field of entomology, emphasize the importance of interdisciplinary collaboration between computer science and entomology to devise innovative data acquisition and labeling strategies. Moreover, the variability of environmental conditions under which these subjects exist demands careful consideration when applying these models in real-world scenarios.

This project leverages the capabilities of CNNs for detailed classification and detection of characteristics of *Atta* leafcutter ants, utilizing a comprehensive dataset from iNaturalist. This dataset, which includes meticulous observations of *Atta* reproductive ants or alates, offers a rich source of information to investigate mating-flight patterns and interactions with regional rainfall, shedding light on phenology and the environmental factors that influence lifecycle events.

Through this endeavor, it is expected to gain valuable insights into the ecological roles and behaviors of *Atta* leafcutter ants. Addressing specific challenges posed by *Atta* ant imagery—such as the presence of multiple individuals within a single frame or the variability in their physical states—the aim is to test and expand the current capabilities of machine learning techniques in entomological research. Consequently, this project is set to provide instrumental resources and discoveries to the scientific community.

Anticipating the outcomes, it is predicted that addressing the pivotal challenges of data annotation and model training will lead to significant advancements in the application of AI for biodiversity studies. Such progress promises to nurture the development of more sophisticated and integrative approaches to scientific inquiries. The ongoing exploration of the intersection between AI and biodiversity heralds a new era of projects poised to deepen the collective understanding and foster extensive studies of the natural world, with this research endeavor serving as a fundamental model for the application of computational tools in biological research and for navigating the complexities of processing large-scale ecological data.

Literature Review

The integration of artificial intelligence into entomological research signifies a substantial leap in scientific capabilities, as recent studies refine species identification and classification. The body of literature, particularly works by Xi et al. (2022), De Nart et al. (2022), Popkov et al. (2022), Rustia et al. (2020), and Omodior et al. (2021), demonstrates a trend towards the use of CNNs to enhance the accuracy and efficiency of classifying a broad spectrum of arthropod taxa.

Research by Xi et al. (2022) utilized the ResNet18 network to train a classifier that recognizes a wide array of butterfly species with a validation accuracy of 86%. This study not only foregrounds the value of image diversity, ranging from standard specimens to images captured in the wild, but also the nuanced approaches to dataset curation that significantly impact model performance. Such meticulous treatment of data sources and classifier training is crucial in entomology, where fine details of species morphology are paramount for accurate identification.

Concurrently, work by De Nart et al. (2022) focused on the application of CNNs to the classification of nocturnal Lepidoptera from automated light trap images, providing critical insights into automated species identification in field conditions. Research by Omodior et al. (2021) investigated the use of CNNs to discern disease vectors, particularly ticks, from ecological photographs, underlining the potential of deep learning in public health entomology by automating the detection of vectors in natural settings.

These examples underscore the diverse range of deep learning applications in entomology and the unique challenges addressed. Research by Popkov et al. (2022) showcased the capacity of machine learning to discriminate between highly similar species, while Rustia et al. (2020) introduced a cascaded model combining object detection with image classification for accurate pest identification.

The literature positions deep learning as a frontier technology in entomology, with the potential to revolutionize species monitoring and environmental management practices. As the fidelity of AI applications improves, the advancement of entomological research stands to benefit significantly.

Materials and Methods

In the realm of AI, two pivotal techniques—image recognition and object detection—play critical roles in interpreting and understanding visual data. Image recognition involves analyzing an entire image to identify and categorize objects, locations, actions, or other elements based on learned patterns. Object detection, a more complex subset of image recognition, not only classifies objects within an image, but also localizes them, providing bounding boxes that indicate the object's position and scale within the image. This allows for detailed and nuanced image analyses, which are fundamental in applications ranging from medical imaging to autonomous driving (Zhao et al., 2019; Zaidi et al., 2022).

Among the various models developed for object detection, YOLO (You Only Look Once) models stand out for their efficiency as single-stage detectors. Unlike traditional methods that involve separate stages for proposing candidate regions and classifying them, YOLO models integrate these steps, predicting bounding boxes and class probabilities simultaneously during a single pass through the neural network. This integration significantly enhances processing speed and accuracy, making YOLO models particularly suitable for scenarios where rapid and reliable object detection is crucial (Redmon et al., 2015; Srivastava et al., 2021).

For this project, the YOLOv8 model was selected due to its superior performance in handling diverse and challenging image contexts, such as those presented by *Atta* leafcutter ants. This choice was based on comparative studies like those by Srivastava et al. (2021), which found that YOLO models generally outperform other architectures in terms of speed and efficiency.

The success of this model was heavily reliant on the preparation of a specialized dataset, a task undertaken by Dr. Ulrich Mueller, a professor from the Department of Integrative Biology at the University of Texas at Austin. Dr. Mueller gathered approximately 4700 images from iNaturalist, spanning the years 2012 to 2021, to create this unique dataset. Although the intention was to format and annotate all these images using Roboflow—a platform that supports the bounding box annotation critical for training YOLO models—time constraints led to only a subset being annotated. Specifically, about 1000 images were annotated for the first dataset and 2009 for the second. Each image was resized to 640 pixels and segmented into structured grids.

Within these grids, the model predicts multiple bounding boxes and assigns confidence scores, using non-maximum suppression (NMS) to enhance the accuracy of its predictions (Zou et al., 2023). NMS is a post-processing technique used to improve the accuracy of predictions by eliminating redundant bounding boxes. NMS scans through the boxes that predict the presence of an object. Each box is scored on the probability of object detection. The box with the highest score is selected, and all other boxes overlapping it beyond a certain threshold are discarded. This process is repeated for each object detected, effectively reducing the number of false positives and improving the overall precision of the model. This technique is particularly valuable in scenarios where objects are closely positioned or overlap significantly.

The training of YOLOv8 involved meticulous hyperparameter tuning, following a methodology akin to that documented by Ultralytics for YOLO models. This iterative process began with a set of baseline hyperparameters, which were adjusted based on ongoing training evaluations to optimize model accuracy and computational efficiency. Additionally, to enhance the model's ability to generalize across varied inputs, extensive preprocessing and augmentation techniques were applied to the dataset, including auto-orientation, resizing, and transformations like flips, crops, and rotations (Zou et al., 2023; Terven et al., 2023).

For the first dataset, each source image was augmented to produce seven variations, applying transformations with a 50% chance each for horizontal and vertical flips, random cropping up to 20% of the image, random rotations between -15 and +15 degrees, brightness adjustments of ±15%, a Gaussian blur up to 2.5 pixels, and a minimal application of salt and pepper noise to 0.1% of the pixels, resulting in a total of 5884 images. The second dataset underwent a similar augmentation process, which included additional transformations like horizontal and vertical shears between -10° and +10°, and exposure adjustments of ±10%, culminating in a total of 10445 images.

These augmentations were selected to bolster the model's robustness. Horizontal or vertical flips help the model adapt to different subject orientations, while cropping introduces variability in positioning and size, enhancing resilience to shifts in subject location and camera placement. Rotational adjustments mitigate the effects of camera roll. Shearing changes the perspective, aiding performance adjustments for different camera and subject angles. Adjustments in brightness and exposure prepare the model for various lighting conditions and camera settings. Incorporating Gaussian blur simulates out-of-focus scenarios, improving the model's ability to handle varying focus levels. Lastly, adding noise mimics real-world camera artifacts, enhancing the model's performance under typical operational imperfections.

The first dataset was utilized to configure a basic YOLOv8 model to establish a baseline performance. After seven iterations of hyperparameter tuning, the most optimal performance was recorded in the fifth iteration. This iteration achieved a best fitness score of 0.55894, with precision, recall, mAP50, and mAP50-95 recorded as 0.76529, 0.55259, 0.66344, and 0.54733, respectively. Additional metrics included box loss at 0.84083, class loss at 1.60947, and DFL loss at 1.5701. The optimal hyperparameters for this model were a learning rate of 0.00769, momentum of 0.98, and weight decay of 0.00049, among others.

Following this, the second dataset was used to refine the model through further tuning, extending over twenty-three iterations. The eighteenth iteration exhibited the highest fitness score of 0.61418, with significantly improved precision and recall rates of 0.64283 and 0.73693, and mAP50 and mAP50-95 values of 0.75427 and 0.59861. The hyperparameter setup for this iteration included a learning rate of 0.0107 and momentum of 0.9209, with slight adjustments in other parameters to optimize performance. This process highlighted the enhancements in model accuracy and efficiency through iterative tuning and adjustments.

The Gradio interface was deployed to provide a user-friendly web application for real-time object detection using the YOLOv8 model. This deployment enabled interactive image uploads where end-users could adjust detection settings such as the confidence threshold and Intersection of Union (IoU) threshold on-the-fly. This capability greatly facilitated demonstrations and experiments, allowing users to visually validate and interact with the model's outputs instantly. Similarly, the model is also available on Roboflow, which offers comparable features for adjusting detection parameters. Both platforms facilitate dynamic user engagement with real-time detection capabilities. The Roboflow deployment can be accessed here: https://universe.roboflow.com/sds-379r/atta-leafcutter-ants-object-detection.

Overall, this meticulous approach to selecting and optimizing the YOLOv8 model underscores the importance of carefully tailored dataset preparation and rigorous model tuning. Such detailed attention to methodology is essential for ensuring optimal performance in specialized application contexts, such as the analysis of biological specimens like *Atta* leafcutter ants

Results

In the course of developing a machine learning model to classify and detect characteristics of *Atta* leafcutter ants, two models were trained. The first model, which utilized less data, was outperformed by the second model. The performance of the second model improved significantly with the inclusion of more labeled image data. The second model's performance was rigorously evaluated using various data visualization techniques. These visualizations, central to understanding the model's strengths and weaknesses, offer a multifaceted view of the model's capabilities in classifying different states of the ants such as being alive or dead and by gender and class—alate or dealate.

The confusion matrix in Figure 1 provides a foundational understanding of the model's predictive accuracy by comparing true labels with predicted ones. The predominant values on the matrix's diagonal reflect correct classifications, notably for classes such as 'Alive-Female-Alate' and 'Alive-Male-Alate,' indicating a strong ability of the models to recognize these categories accurately. However, non-diagonal entries reveal misclassifications, such as 'Alive-Male-Alate' occasionally being misidentified as 'Alive-Female-Alate,' suggesting a struggle to distinguish between similar classes. This confusion may arise from overlapping features or insufficiently distinctive training data, as seen in the higher frequency of misclassifications for 'Dead-Male-Alate' and 'Dead-Female-Alate,' indicating difficulty in distinguishing finer morphological details that differentiate dead specimens.

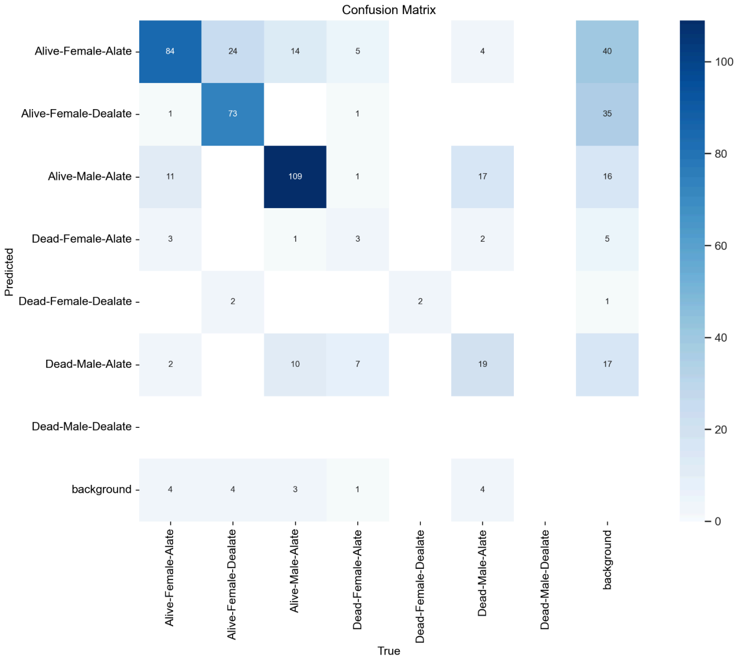

Figure 1: Confusion Matrix of the Second Model

The normalized confusion matrix in Figure 2 takes these results and scales them relative to the number of instances in each class, which allows for a fair comparison across classes that may not be equally represented in the dataset. The high relative values for 'Alive-Female-Dealate' suggest a robustness in the model's ability to identify this class. Notably, the lower relative values seen for 'Dead-Male-Alate' hint at the model's potential difficulty in differentiating features or an imbalance in the training data that skews towards other classes. This representation illuminated the precision with which 'Dead-Female-Dealate' was identified, achieving a perfect classification rate. However, this high performance may be slightly inflated due to lower class representation. The normalization process revealed that 'Alive-Female-Alate' had a relatively high misclassification rate with 'Alive-Female-Dealate,' signaling that even within a single sex and life state, the model faced challenges in differentiating between classes.

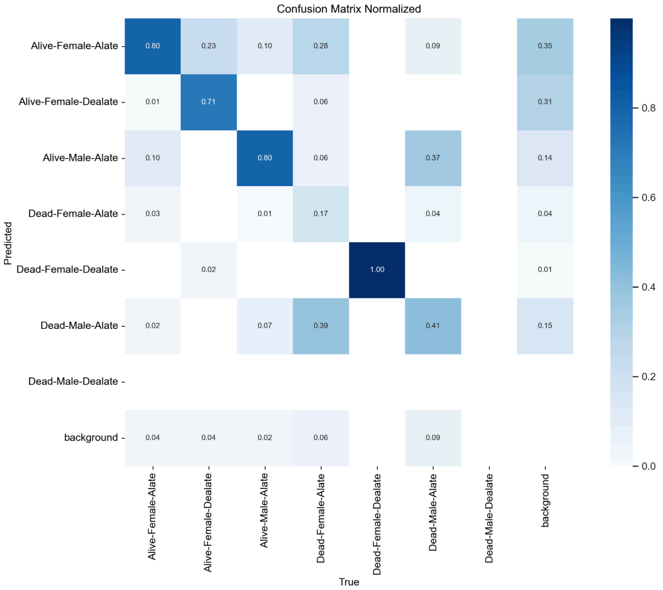

Figure 2: Normalized Confusion Matrix of the Second Model

The F1-Confidence curve in Figure 3 is essential for delineating the balance between precision and recall across various confidence thresholds, representing each class with a distinct curve that highlights the model's performance at optimal confidence levels. For instance, the 'Alive-Female-Alate' class curve plateaus, reflecting stable F1 scores over a consistent range of confidence thresholds, which underscores a reliable classification capability. In contrast, the curve for 'Dead-Female-Alate' exhibits a sharp peak at a particular confidence threshold, indicating that maximum precision and recall are achieved specifically at this point, beyond which the model's performance significantly declines. Collectively, these patterns in the F1-Confidence curve provide critical insights into the model's classification behavior, emphasizing the need for fine-tuning the confidence thresholds to enhance model reliability and effectiveness in operational settings.

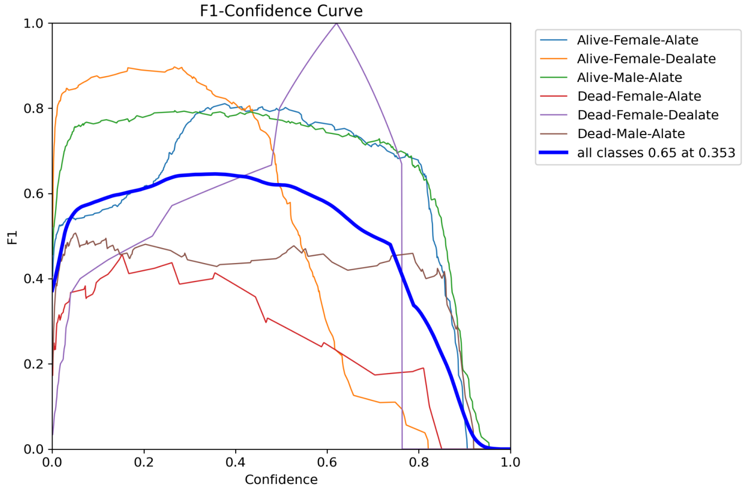

Figure 3: F1-Confidence Curve of the Second Model

The class distribution bar chart in the top left of Figure 4 illustrates the frequency of instances for each class within the dataset, highlighting a pronounced imbalance where classes such as 'Alive-Female-Alate' and 'Alive-Male-Alate' significantly outnumber others, like 'Dead-Male-Dealate.' This disparity suggests that these dominant classes may benefit from more robust feature representation, enabling the model to learn from them more effectively, but potentially at the expense of less represented classes. Such an imbalance poses a risk of overfitting, where the model's predictive power may be unduly influenced by the more common classes, thus affecting its ability to generalize to new, unseen data. To address this, balancing the dataset is critical. Techniques like resampling or employing weighted loss functions could be instrumental in promoting equitable learning and improving overall predictive performance across all classes.

In the top right of Figure 4, the correlogram visualizes the correlation between bounding box coordinates across the dataset. The concentric square patterns suggest a consistent placement of objects within the images, indicating a lack of diversity in the dataset regarding the spatial distribution of the subjects. This lack of variability could affect the model’s ability to generalize to new data where subjects may appear in different positions.

The scatter plot in the bottom left of Figure 4 presents the typical positions (x, y coordinates) of the bounding boxes in the images. The dense clustering in the center suggests that most ants are centrally located within the images. This central bias in object placement may limit the model's ability to detect ants situated towards the edges of the image frame.

The heatmap in the bottom right of Figure 4 displays the distribution of bounding box aspect ratios, with 'width' on the x-axis and 'height' on the y-axis. The elongated cluster along the x-axis implies that many of the bounding boxes are wider than they are tall, which might be characteristic of the way ants are typically imaged or an attribute of their natural shape.

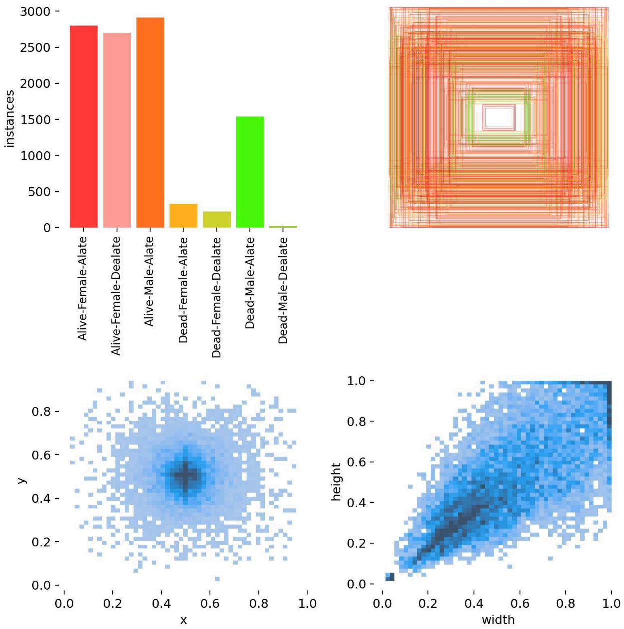

Figure 4: Label Distribution of the Second Model

The correlogram and bounding box distribution plots in Figure 5 collectively visualize the correlation and distribution of bounding box dimensions, revealing a notable concentration of bounding boxes in specific areas of the image space. This suggests that ants are frequently imaged in particular orientations or sections of the frame. Such a cluster indicates a homogeneity in spatial distribution, which provides an opportunity to refine data collection protocols. By ensuring a diverse range of ant placements, it is possible to enrich the dataset and inform targeted data augmentation strategies. These improvements aim to bolster the model's spatial invariance and enhance its detection capabilities across varying scenarios.

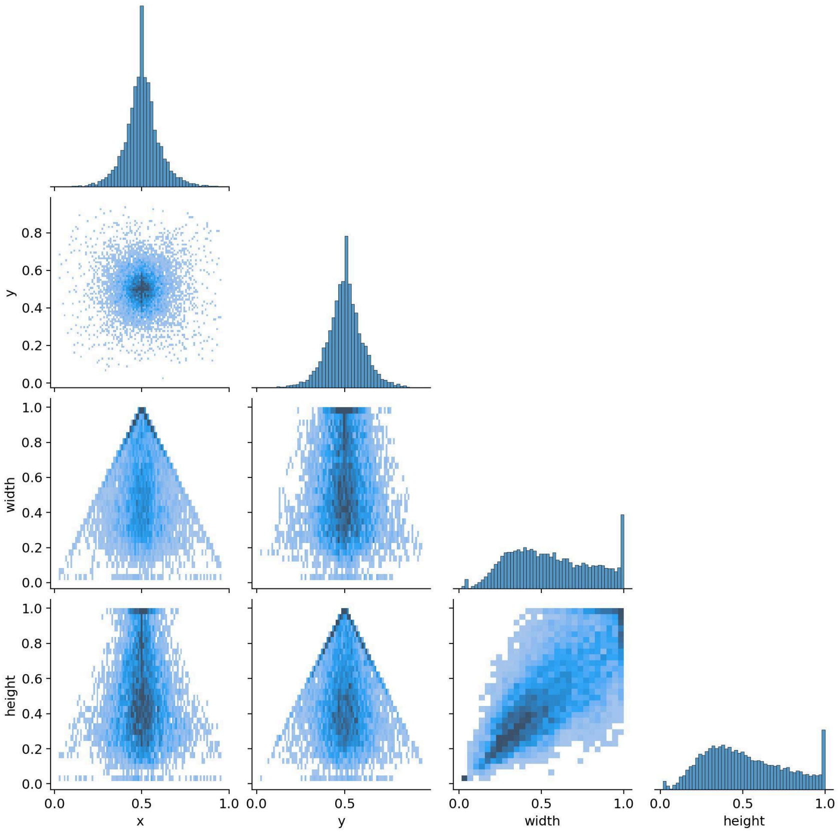

Figure 5: Label Correlogram of the Second Model

The Precision-Confidence and Recall-Confidence curves in Figures 6 and 7, respectively, elucidate the nuanced relationship between precision and recall at varying confidence thresholds for the model. The precision curve indicates that the model maintains a high degree of reliability for certain classes over a wide span of confidence thresholds, showing confidence in its predictions. Conversely, the recall curve portrays the model's ability to identify all relevant instances, with the caveat that an increase in confidence threshold often results in reduced recall, signifying a conservative stance at higher levels of confidence. This interplay suggests a need to balance the precision and sensitivity of the model, especially as the recall for classes like 'Dead-Male-Alate' tends to drop significantly at elevated thresholds, highlighting the model's challenge in capturing all true positives without compromising the confidence in its predictions.

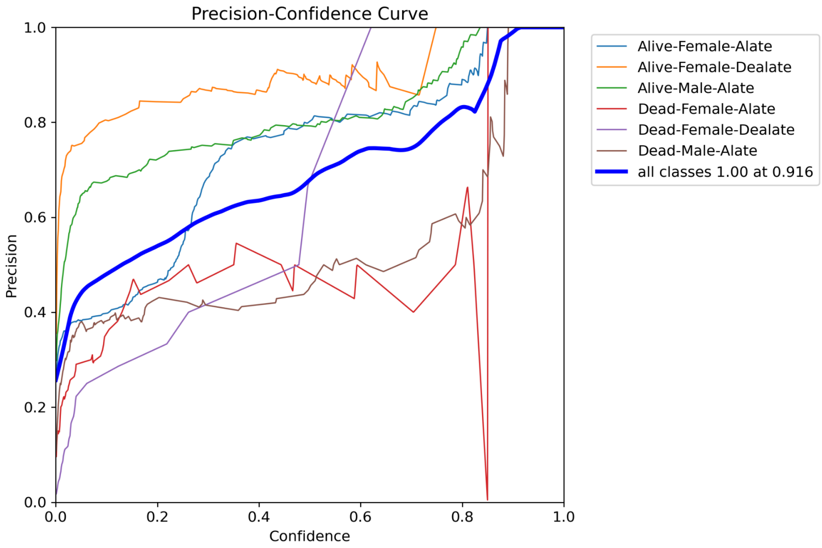

Figure 6: Precision-Confidence Curve of the Second Model

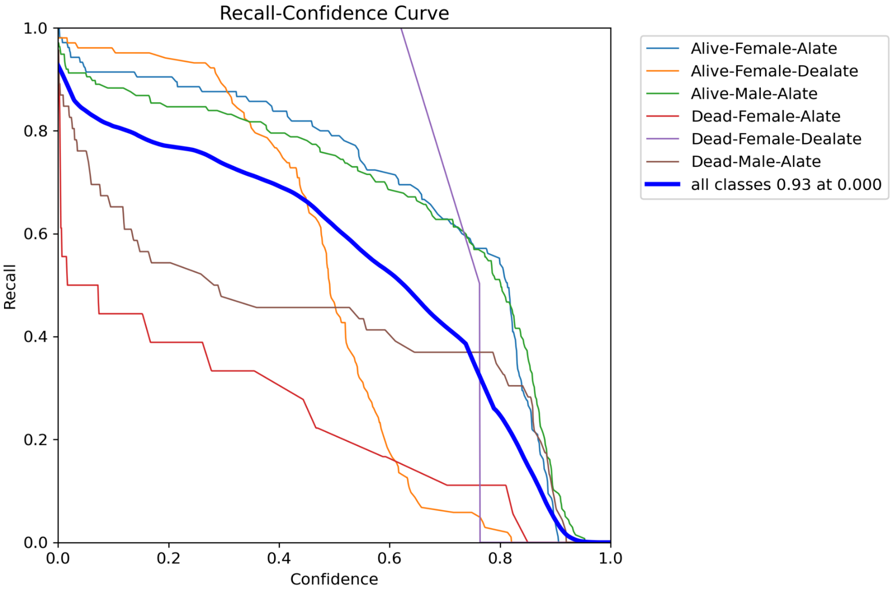

Figure 7: Recall-Confidence Curve of the Second Model

The Precision-Recall curve in Figure 8 provides an aggregate measure of the model's performance, offering insight into the balance between precision and recall across various classes within an imbalanced dataset. For classes like 'Alive-Female-Dealate,' the model demonstrates a strong ability to reliably identify and classify instances with high precision and recall, a characteristic evident from its position closer to the top right corner of the graph. The area under the curve serves as a single-value metric that can be used to compare the performance of different classes. For example, while 'Dead-Female-Dealate' also achieves an exceptional balance between precision and recall, the 'Dead-Female-Alate' class's position near the bottom left suggests that the model requires improvements to enhance both precision and recall for this class.

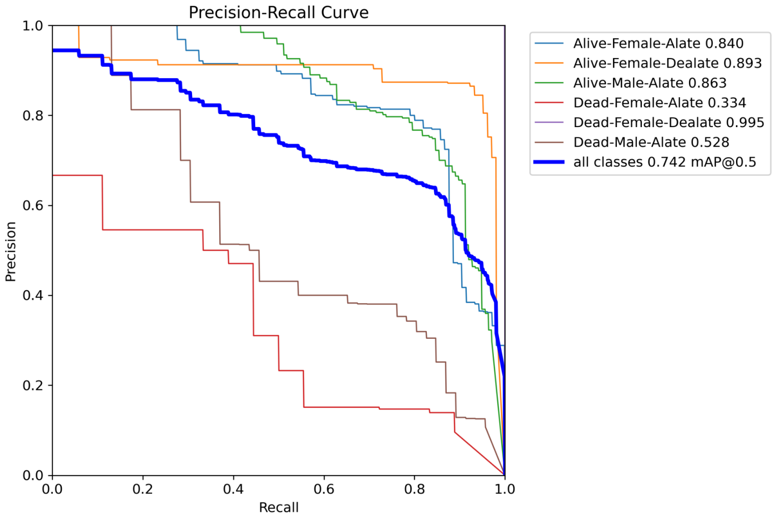

Figure 8: Precision-Recall Curve of the Second Model

In the comprehensive analysis of the model's learning trajectory, the plots of Figure 9 capture the evolution of predictive accuracy and classification capabilities over time. The box loss metrics, labeled as ‘train/box_loss’ and ‘val/box_loss’, quantify the model's precision in predicting the bounding boxes for object detection. A decreasing trajectory in these graphs is indicative of the model's improving accuracy in aligning predicted bounding boxes with the actual, ground truth coordinates across both training and validation datasets.

The class loss metrics, denoted as ‘train/cls_loss’ and ‘val/cls_loss’, measure the error in object classification. The descending trend across these plots signifies that the model is increasingly accurate in determining the correct class for each detected object. This improvement is consistent during both the training and validation phases, suggesting a solid learning transfer from the training data to unseen data in validation.

Additionally, the Distribution Focal Loss (DFL), captured in the ‘train/df1_loss’ and ‘val/df1_loss’ metrics, is a specialized loss function that addresses challenges in fine-grained localization within object detection models. The downward trend in DFL loss points to the model's enhanced capacity to focus on nuanced aspects of object localization, which is crucial for detecting objects accurately.

Precision, a metric shown in ‘metrics/precision(B)’, is the proportion of true positive predictions out of all positive predictions made by the model. An improvement in precision indicates that the model is becoming more adept at correctly identifying items as belonging to a particular class over the course of training. However, the observed fluctuations suggest variability that may need to be addressed to ensure the model's consistent performance.

Recall, measured by ‘metrics/recall(B)’, is the ratio of true positive detections to the total number of actual positive instances. An upward trend in recall demonstrates that the model is becoming more effective at identifying all relevant cases within the dataset, which is essential for comprehensive object detection. However, like precision, the observed fluctuations in recall suggests that there is room for the model's performance to be more consistent.

The mean Average Precision (mAP) metrics, encompassing ‘metrics/mAP50(B)’ and ‘metrics/mAP50-95(B)’, provide a cumulative evaluation of the model's precision across various thresholds of IoU. mAP50 refers to the average precision at an IoU threshold of 50%, which is a standard for moderate overlap in object detection tasks. The more rigorous mAP50-95 calculates the mAP across IoU thresholds from 50% to 95%, offering a nuanced view of the model's precision at different levels of strictness regarding bounding box alignment. Peaks in these metrics indicate moments where the model achieves a balance between precision and recall, yet the variability signals the need for further refinement.

The plots in Figure 9 demonstrate that the model is steadily improving in its ability to detect and classify objects, as evidenced by the decreasing loss metrics—box, class, and DFL—which indicate a progression toward greater accuracy in predicting the location and classifying the objects within its scope. Despite these positive learning signals, the variability observed in performance metrics such as precision, recall, and mean Average Precision (mAP) suggests potential issues such as overfitting, which highlights the need for improved regularization or adjustments to address challenges inherent to the data itself. This inconsistency could potentially undermine the model's reliability in practical applications. To enhance the model's performance stability and reliability, it may be beneficial to experiment with different regularization techniques, revise the learning rate schedule to better match the stages of learning, or enrich the dataset to provide a more robust foundation for training. These methodical enhancements aim to refine the model’s training regimen and stabilize its performance metrics, thereby maximizing its operational effectiveness in real-world scenarios.

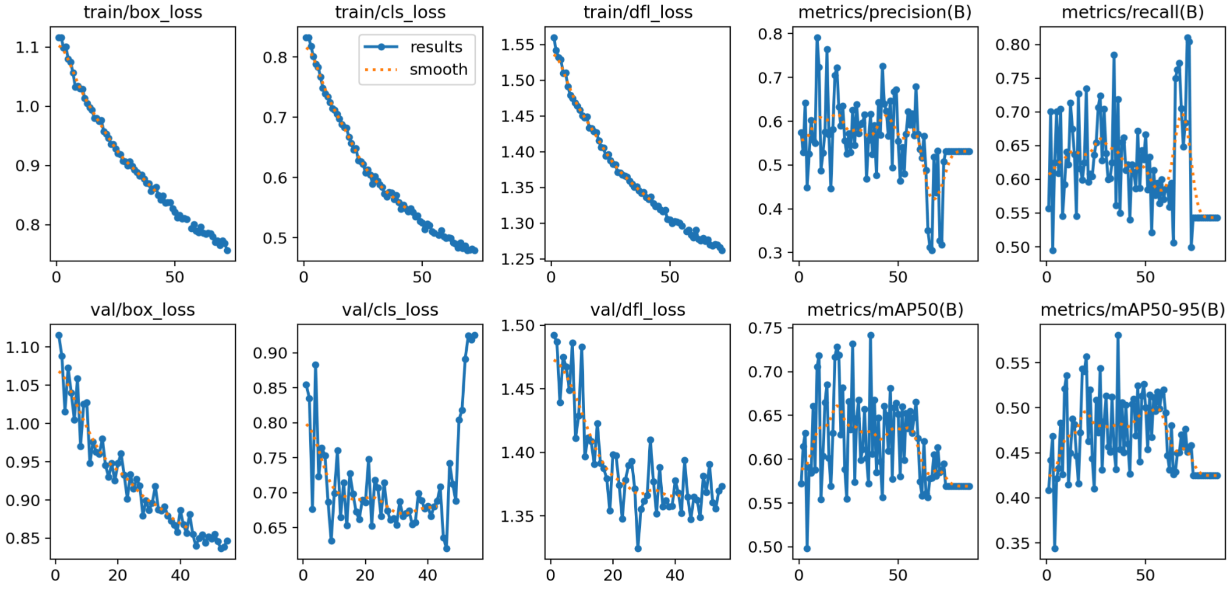

Figure 9: Training Metrics of the Second Model

Altogether, these visualizations compose a comprehensive narrative of the model's current capabilities and limitations. The high performance in certain classes is promising, but the variability across classes highlights the need for further refinement, possibly through more balanced training data, improved feature extraction methods, and more nuanced annotation strategies to better delineate closely related classes.

Discussion

The interpretation of the YOLOv8 model's results in classifying *Atta* leafcutter ants has exhibited notable precision and recall, although it has faced challenges in differentiating closely resembling classes such as 'Alive-Male-Alate' and 'Alive-Female-Alate'. This performance is a testament to the meticulous data augmentation and hyperparameter tuning strategies implemented, emphasizing the significance of intensive model training and validation in machine learning.

Contrasting with parallel studies, such as Xi et al. (2022) which achieved an 86% validation accuracy using ResNet18 for butterfly species recognition, this project has similarly integrated a variety of image datasets to improve model generalizability. The use of YOLOv8 for detecting and classifying ants in a single pass introduces distinct challenges and prospects. This diverges from works by Rustia et al. (2020) and Popkov et al. (2022) that utilized cascaded models for pest detection and discrimination between highly similar species. The focus here has been prioritizing efficiency, which could lead to the practical deployment of real-time detection in entomological research, contributing significantly to entomology.

The insights gained bear implications for further research and practical applications. The successful application of YOLOv8 for ant classification has the potential to be replicated and adapted for other entomological studies, enhancing species monitoring and ecological studies. The hurdles in class differentiation underscore the necessity for refined data annotation strategies to improve the accuracy of similar models. These advancements support the possibility that deep learning can substantially boost the efficiency and precision of biological monitoring, pivotal for entomological research, environmental management, and ecological studies.

The project's utilization of Ultralytics' YOLOv8 model has been fundamental in translating theoretical frameworks into practical, efficient, and fast object detection applications. This included both training and detection functionalities which were critical for the classification and detection of *Atta* leafcutter ants. The 'detect' task, for instance, can be triggered by using a command like ‘yolo detect data=coco.yaml weights=yolov8.pt source=images/’, where it processes input images or videos to output them with annotated bounding boxes. Similarly, the training function can be initiated with ‘yolo train data=dataset.yaml model=yolov8.yaml imgsz=640 batch=16 epochs=300’, where settings such as batch size and epochs are meticulously calibrated.

This sophisticated computational process applies CNNs that perform forward propagation to extract features, apply bounding box regression, and object classification, followed by backpropagation to update weights based on loss calculations. NMS is then used to refine predictions, ensuring high precision in annotated outputs or detection coordinates.

A robust data preprocessing pipeline begins with the collection of images using `iNAT_scraper.py` from the iNaturalist repository, ensuring high-quality inputs suitable for YOLOv8’s requirements. Images undergo resizing, normalization, and extensive augmentations—like rotations, flips, and color space adjustments—to enhance dataset variability while maintaining ecological authenticity. These preparations are critical in simulating the natural variations encountered in the ants' outdoor environments.

Hyperparameter optimization is pivotal, as detailed in `yolo_hyper_optimization.py`, where iterative training sessions refine learning rate, weight decay, and other settings to optimize model performance. Augmentation settings adjusted include hsv_h, hsv_s, hsv_v, degrees, translate, scale, shear, and perspective, as defined in project configurations to maximize detection precision.

The project's computational infrastructure is designed to leverage the available hardware efficiently while adhering to computational limits. This includes the creation of a reproducible software environment as defined in the `environment.yml` file, which outlines all necessary library versions. The combination of `gpu_roboflow_setup.py` and Roboflow's tools for image annotation further enhances the robustness of data preparation.

However, a significant drawback of the Roboflow platform is its rigidity in exporting the dataset strictly as a train-valid-test split. Although users have the flexibility to manually adjust the proportions of the split and select specific annotated images for each set, the inescapable fact is that the dataset will be partitioned into these three subsets upon download. This inflexibility hinders critical preprocessing steps in machine learning pipelines. For example, Sklearn's GridSearchCV—an exhaustive search over specified parameter values for an estimator—relies on the ability to perform cross-validation to evaluate model performance comprehensively. Cross-validation requires splitting the dataset into different folds, which GridSearchCV systematically works through, training and validating the model on these different segments to optimize hyperparameters effectively. By pre-splitting the dataset, Roboflow inadvertently precludes the use of GridSearchCV's cross-validation, thus limiting the capacity to rigorously tune models and potentially leading to suboptimal generalization on unseen data.

The models were trained using an NVIDIA GeForce RTX 3050 Ti Laptop GPU, which provided significant processing power and enabled the execution of complex neural network operations required for object detection. The choice of GPU played a pivotal role in facilitating rapid model iteration and experimentation, which is critical for the swift development and refinement of detection algorithms.

However, for a project with such intensive computational needs, the Texas Advanced Computing Center (TACC) at UT Austin would have been an invaluable resource. TACC specializes in high-performance computing resources tailored for intensive data analysis and large-scale computational tasks, features that are particularly advantageous for training complex models like YOLOv8. The access to more powerful computing nodes and extensive data storage solutions at TACC could potentially reduce model training times and increase the efficiency of hyperparameter optimization processes. Furthermore, TACC's advanced systems are engineered to manage substantial workloads, which would allow for the extension of training datasets and iterations, thereby potentially enhancing the model's predictive accuracy and robustness. Leveraging TACC's facilities could have provided a more optimal computational environment for the project, aligning with its requirements for high-performance computing and intricate data processing.

All in all, the integration of detailed configuration settings, strategic hyperparameter tuning, and advanced data filtering has culminated in a system that not only supports the detailed study of biodiversity, but also sets benchmarks for future computational entomology. These methodical adjustments ensure the model's adaptability to varied data scenarios, significantly enhancing predictive performance on test datasets and providing valuable insights for ecological studies. This comprehensive approach demonstrates a meticulous balance between theoretical innovation and practical execution, crucial for advancing machine learning applications in real-world ecological monitoring.

Acknowledgements

I would like to thank Dr. Fatma Tarlaci for her supervision and expertise in machine learning. I would also like to thank Dr. Ulrich Mueller for initiating this collaborative project and providing the dataset.

Bibliography

De Nart, D., Costa, C., Di Prisco, G., & Carpana, E. (2022). Image recognition using convolutional neural networks for classification of honey bee subspecies. *Apidologie*, *53*(5). https://doi.org/10.1007/s13592-022-00918-5

Omodior, O., Saeedpour-Parizi, M. R., Rahman, M. K., Azad, A., & Clay, K. (2021). Using convolutional neural networks for tick image recognition – a preliminary exploration. *Experimental and Applied Acarology*, *84*(3), 607–622. https://doi.org/10.1007/s10493-021-00639-x

Popkov, A., Konstantinov, F., Neimorovets, V., Solodovnikov, A. (2022). Machine learning for expert-level image-based identification of very similar species in the hyperdiverse plant bug family Miridae (Hemiptera: Heteroptera). *Systematic Entomology*, *47*(3), 487-503. https://doi.org/10.1111/syen.12543

Redmon, J., Divvala, S. K., Girshick, R. B., & Farhadi, A. (2015). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. https://doi.org/10.1109/CVPR.2016.91

Rustia, D. J. A., Chao, J., Chiu, L.-Y., Wu, Y.-F., Chung, J.-Y., Hsu, J.-C., & Lin, T.-T. (2021). Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method. *Journal of Applied Entomology*, *145*(3), 206-222. https://doi.org/10.1111/jen.12834

Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, *8*(66). https://doi.org/10.1186/s40537-021-00434-w

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and

YOLO-NAS. *Machine Learning and Knowledge Extraction*, *5*(4), 1680–1716. https://doi.org/10.3390/make5040083

Xi, T., Wang, J., Han, Y., Lin, C., & Ji, L. (2022). Multiple butterfly recognition based on deep residual learning and image analysis. *Entomological Research*, *52*(1), 44-53. https://doi.org/10.1111/1748-5967.12564

Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, *126*. https://doi.org/10.1016/j.dsp.2022.103514

Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, *111*(3), 257-276. https://doi.org/10.1109/JPROC.2023.3238524

Appendix: Code Repository on GitHub

The code files referenced in this report are hosted on a GitHub repository and can be accessed at the following URL: https://github.com/WarpedReflections/SDS-379R---Undergraduate-Research

This repository contains all the necessary code files used for the analysis and results presented in this document, including environment setup, data scraping, model training, and hyperparameter optimization scripts.

Reflection

Throughout the course of this project, I have acquired a thorough understanding of the complexities involved in applying machine learning to biological data. The task of classifying and detecting the intricate characteristics of *Atta* leafcutter ants demanded not only a deep dive into the subtleties of CNNs, but also an appreciation for the nuance of ecological datasets.

My role in this endeavor was multifaceted, encompassing the initial gathering and preparation of the dataset, iterative model training, and the meticulous tuning of hyperparameters. Each image was rigorously annotated, augmented, and processed to create a robust training set that would challenge and refine the model's capabilities. Not only did this task hone my technical skills, but it also imbued me with a methodical approach to problem-solving and the application of theoretical knowledge to real-world scenarios.

The collaborative nature of the project was critical to its success. Working closely with Dr. Fatma Tarlaci and Dr. Ulrich Mueller, our joint effort focused on the intersection of computer science with biology. Their guidance provided a strong foundation for my work, while regular bi-weekly discussions ensured that the project stayed on course towards its objectives. This synergy of skills and knowledge was not only instrumental in achieving the project's goals, but also served as a vital learning experience in interdisciplinary collaboration.

Overall, this project was a significant learning journey. It presented challenges that pushed the boundaries of my understanding and skills, and through this, I gained insights into the potential and real-world application of machine learning in ecological studies. The project contributed to my personal development as I navigated the demands of this rigorous interdisciplinary project. My experiences underscored the importance of adaptability, continuous learning, and the ability to synthesize information across fields to advance our understanding of the natural world.