Detection of Cacao Trees in Orthomosaic Images Using YOLOv8 and SAHI

Thiago Mantovani Martins*, Matheus Arroyo de Lima*,
Henrique Montemor Junqueira†, Victor Hugo Campos Ferreira†,
David Barral Santos‡, Giovanni Bruno Papa‡
*Barão de Mauá University Center, Ribeirão Preto, Brazil
Emails: thiagomantovanimartins@outlook.com, matheusarroyodelima@gmail.com
†REGES University, Ribeirão Preto, Brazil
Emails: hjunqueira77@gmail.com, vhcf1405@gmail.com
‡Altamap, Ribeirão Preto, Brazil
Emails: dbarralgeo@gmail.com, giovanni.b.papa@gmail.com

***Abstract*—This study explores the application of drone-based orthomosaic imagery combined with the YOLOv8x segmentation model and the Slicing Aided Hyper Inference (SAHI) method for detecting cacao trees. High-resolution orthomosaic images were processed by slicing them into smaller tiles to facilitate object segmentation. Although the YOLOv8x model was trained for segmentation, the evaluation was conducted using bounding boxes for ease of comparison. The model achieved a precision of 93.49% and an accuracy of 81.10%, demonstrating its potential for monitoring large-scale plantations. However, the system is best suited for estimation purposes rather than precise, real-time decision-making.**

***Index Terms*—Cacao trees, orthomosaic images, drone survey, computer vision, segmentation, precision agriculture.**

I. Introduction

The cacao tree (*Theobroma cacao*) has been an integral part of human society for centuries, primarily used in human consumption, particularly in the production of chocolate. This perennial crop is native to Latin America and thrives in humid tropical climates. When properly managed, cacao trees can live for over 100 years [1]. Brazil ranks as the seventh-largest cacao producer globally, with the state of Bahia being the primary producing region in the Northeast, covering an area of approximately 403 hectares dedicated to cacao cultivation [2]–[4].

Cacao production in Brazil is predominantly concentrated in the states of Bahia and Pará, which together account for nearly 95% of the country's output. The cacao industry plays a significant role in the local economies of these states. In 2021, cacao production generated around BRL 2.2 billion in gross revenue, or approximately USD 400 million [5]. However, the sector has faced several challenges, particularly in Bahia, where cacao production has declined significantly due to pests, diseases, and unfavorable climatic conditions.

Since the 1990s, cacao production in Bahia has seen a sharp decrease, with a 62.1% drop in output and a 24.7% reduction in the cultivated area, according to data from the Brazilian Institute of Geography and Statistics (IBGE) [6]. Aguiar and Pires (2019) identified the primary factors behind this decline as the infestation of fungi, such as *Crinipellis perniciosa* (which causes witches' broom disease) and *Phytophthora palmivora* (responsible for brown rot), which have significantly impacted cacao plantations [7].

Despite these challenges, Pará has experienced growth in its cacao production, driven by investment in more resilient crop varieties and improved management practices. Additionally, in 2024, cacao was the agricultural commodity that appreciated the most, with prices increasing by 191% from $4.20 to 12.26$ per ton, benefiting Brazil's economy as African countries face ongoing challenges due to disease outbreaks [8].

The future of cacao cultivation is further threatened by climate change. Studies by Schroth et al. (2016) and Läderach et al. (2013) predict that by 2050, many areas in West Africa, a major cacao-producing region, will no longer be suitable for cultivation due to rising temperatures and changing precipitation patterns. Such environmental changes are expected to reduce the availability of land suitable for cacao cultivation, increasing the vulnerability of plantations worldwide [9], [10].

To mitigate these threats, modern technologies such as drones and computer vision systems offer innovative solutions for monitoring cacao plantations. High-resolution images captured by drones can be processed using specialized software for the early detection of diseases, assessment of plant health, and precision agriculture practices. This proactive approach enables rapid intervention, potentially reducing crop losses and improving plantation management efficiency [4].

In this context, the use of drone-based orthomosaic surveys provides a valuable tool for improving cacao plantation management. This study aimed to develop high-resolution orthomosaic maps of cacao fields by capturing detailed aerial imagery. The subsequent geospatial analysis of these maps helps to assess the spatial variability of plantations, identify areas requiring intervention, and support decision-making processes to enhance cacao production and ensure sustainable agricultural practices.

II. Related Works

Several studies have explored the use of deep learning models and image processing techniques in the cacao industry, particularly for fruit detection, classification, and yield estimation.

Ayubi et al. [11] introduced a cacao ripeness detection and classification model based on an improved version of YOLOv5s integrated with the Convolutional Block Attention Module (CBAM). The study focused on enhancing the feature representation to improve detection accuracy while maintaining model compactness, which is crucial for real-world applications. Their approach showed promise in improving the productivity of cacao cultivation by accurately identifying ripe cacao fruits.

Bastidas-Alva et al. [12] proposed a system for recognizing and classifying the ripeness stages of Trinitario cacao fruits using the YOLOv5 algorithm. Their work employed the novel Mosaic-12 method for data augmentation, which resulted in a 60.2% accuracy, confirming the value of the improved dataset. The system is designed for real-time recognition and classification, showing potential for integration into automated harvesting systems.

Another relevant study by Justam et al. [13] used an Unmanned Aerial Vehicle (UAV) to detect and count the number of cocoa pods on trees. Their algorithm combined image segmentation, shape analysis, and overlap analysis to estimate the yield of stacked cocoa fruits. The system achieved a detection accuracy of 94.5%, demonstrating the effectiveness of UAV-based solutions for agricultural monitoring.

In a related domain, Alhichri et al. [16] proposed a deep-learning-based framework for the automated counting and geolocation of palm trees from aerial images. Their work utilized convolutional neural networks such as YOLOv4 and EfficientDet, achieving up to 99% mean average precision (mAP) and an average geolocation accuracy of 1.6 meters. This innovative approach combines object detection and geolocation, showing the potential for automating large-scale tree counting and geolocation in agriculture.

These studies demonstrate the growing interest in applying advanced object detection and segmentation models such as YOLO to the cacao industry, from fruit ripeness detection to yield estimation. The integration of additional techniques like CBAM and Mosaic-12 for data augmentation highlights the importance of model optimization for better accuracy and efficiency in real-world agricultural scenarios.

III. Methodology

A. *OrthoMosaic Images and Data Collection*

Orthomosaic images are geometrically corrected images that represent an accurate, map-like view of the area captured. These images are generated by stitching together multiple aerial images, ensuring that distortions caused by camera tilt and terrain variation are corrected. Orthomosaics are essential for precision agriculture as they provide high-resolution, georeferenced data that can be used to monitor plantation health and perform detailed analyses.

The data used in this study were collected by David Barral Santos (Altamap) using a DJI Phantom 4 Pro multirotor drone equipped with a high-resolution camera. The flights were conducted at an altitude of 90 meters, resulting in a Ground Sample Distance (GSD) of 3 cm, providing the necessary detail for small object detection such as cacao trees. After data acquisition, the aerial photogrammetry processing was performed using the Metashape software to generate high-resolution orthomosaics with georeferenced accuracy.

For this study, the orthomosaic images covered a cacao plantation area, with each image having a resolution of 26228x8772 pixels. The large size of these images posed significant computational challenges for small object detection. As a result, the images were preprocessed using the Slicing Aided Hyper Inference (SAHI) method, which divided them into smaller tiles of 640x640 pixels with a 20% overlap, enabling better detection of cacao trees.

In total, two orthomosaic images were used for training the segmentation model. These large images were divided into the following valid image counts with corresponding annotations in the datasets:

- **Validation dataset**: 35 images
- **Training dataset**: 126 images
- **Test dataset**: 20 images

It is important to note that several image slices were discarded during this process, as they did not represent areas containing cacao trees. These discarded images ensured that only relevant data was used for training and evaluating the model. Additionally, the dataset is not publicly available, but researchers interested in using it may contact Altamap to request access.

B. *Slicing Aided Hyper Inference (SAHI)*

To address the challenge of processing large-resolution orthomosaic images and detecting small objects such as cacao trees, we employed the Slicing Aided Hyper Inference (SAHI) method [14]. SAHI is an open-source framework designed to enhance small object detection by slicing large images into smaller sub-images and performing object detection or segmentation on each slice. This approach ensures that small objects, which might be missed in the full-resolution image, are detected by the model.

For this work, we applied SAHI by slicing each orthomosaic image into tiles of 640x640 pixels with a 20% overlap in both the x and y directions. This overlap is crucial to prevent missing detections at the borders of the slices. The framework was integrated with a YOLOv8x segmentation model, which had been pre-trained for cacao tree segmentation.

Figure 1 shows an example of an orthomosaic slice with the SAHI grid applied, highlighting how the image is divided for detection. Figure 2 illustrates the segmentation results from a single slice without the grid, showing the cacao trees identified by the model.

C. *YOLOv8x for Object Segmentation*

In this work, we employed the YOLOv8x segmentation model for detecting and segmenting cacao trees in the orthomosaic images. YOLOv8x is one of the latest versions in the YOLO (You Only Look Once) family of real-time object detectors, renowned for its excellent trade-off between speed and accuracy. It has been widely adopted in various computer vision tasks, including object detection, instance

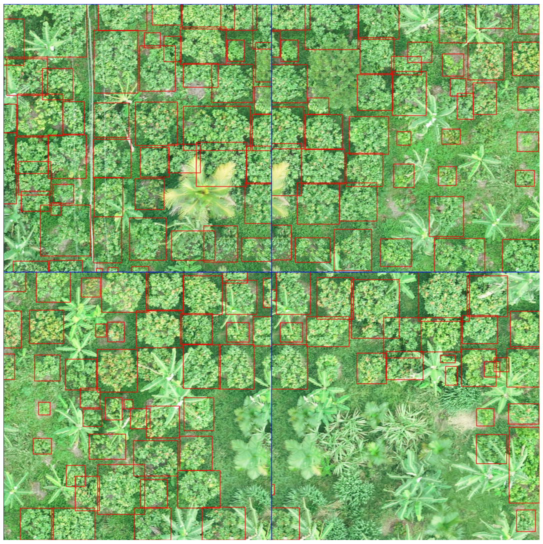

Fig. 1. Example of orthomosaic slice with SAHI grid applied for cacao tree segmentation.

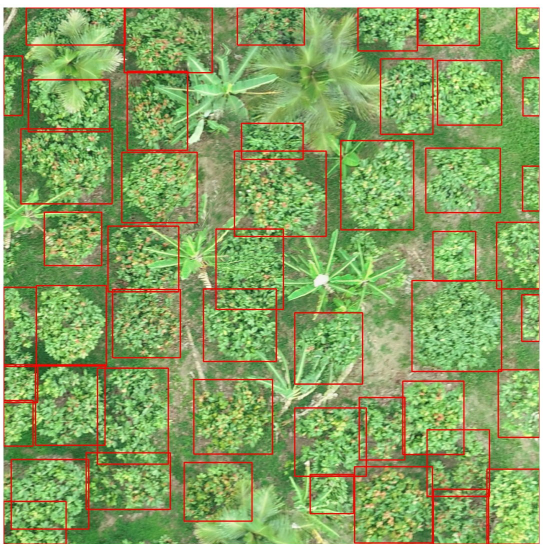

Fig. 2. Segmentation results from a 640x640 slice of the orthomosaic image after applying SAHI.

segmentation, and classification, due to its efficiency and cutting-edge performance [15].

Key advancements in YOLOv8x include:

- **Advanced Backbone and Neck Architectures:** YOLOv8x incorporates state-of-the-art backbone and neck architectures, allowing for improved feature extraction and enhancing overall segmentation performance [15].
- **Anchor-free Split Ultralytics Head:** The anchor-free design contributes to better accuracy, especially for small object segmentation, by avoiding the complexities associated with anchor-based approaches used in previous YOLO versions [15].
- **Optimized Accuracy-Speed Tradeoff:** With its design focus on balancing accuracy and speed, YOLOv8x is highly suitable for real-time applications, making it an ideal choice for agricultural monitoring using drone imagery [15].

The YOLOv8x model was selected due to its ease of integration with the Slicing Aided Hyper Inference (SAHI) method, which is crucial for handling large orthomosaic images and detecting small objects with high precision. Although the model was trained for segmentation, the evaluation of the results was conducted using bounding boxes for ease of comparison with object detection metrics.

D. Training Parameters

The YOLOv8x segmentation model was trained using the Adam optimizer, which was selected for its efficiency in handling sparse gradients and its adaptability in deep learning tasks. The learning rate was set to 10^{-3}, a value commonly recommended in the literature for object detection tasks, balancing fast convergence with stability.

A batch size of 8 was used, as it represented the optimal trade-off between computational efficiency and training stability for 640x640 pixel images. This batch size was the maximum supported by the GPU used in this experiment, a Tesla V100 with 16GB of VRAM, which allowed the model to be trained effectively without exceeding memory limitations.

The model was trained for 100 epochs to ensure convergence. However, the model selected for evaluation was based on the lowest training loss recorded across epochs, as this provided the best-performing version of the model. Notably, the training loss started to increase after epoch 90, indicating that further training beyond this point would likely result in overfitting rather than performance improvement. Therefore, the model saved at the epoch with the lowest loss was chosen for the final evaluation.

E. Model Validation and Metrics

To evaluate the performance of the segmentation model, we used several metrics designed to assess the accuracy and reliability of the object detection results (evaluated with bounding boxes). These metrics included:

- **True Positive (TP)**: A detection is considered a true positive if it correctly identifies a cacao tree, with the Intersection over Union (IoU) between the detection and ground truth annotation exceeding a threshold of 0.5.
- **False Positive (FP)**: A detection is considered a false positive if it does not overlap significantly with any ground truth annotation, i.e., the IoU is below the threshold.
- **False Negative (FN)**: A false negative occurs when the model fails to detect a cacao tree that is present in the ground truth annotations.

We also calculated the following performance metrics based on the above definitions:

Accuracy measures the proportion of correct detections (true positives and true negatives) out of all detections:

$$\text{Accuracy} = \frac{TP}{TP + FP + FN}$$

Precision measures the proportion of true positive detections out of all positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall measures the proportion of actual positive cases (cacao trees) that were correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score is the harmonic mean of precision and recall, providing a single metric that balances both:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics provided a comprehensive assessment of the model's performance in detecting cacao trees in the orthomosaic images, with the F1-Score offering a balanced view between precision and recall.

IV. Results and Discussion

The performance of the cacao tree detection model was evaluated using a dedicated test dataset. The following metrics were calculated to assess the effectiveness and reliability of the model: Precision, Accuracy, Recall, F1-Score, True Positive (TP), False Positive (FP), and False Negative (FN). Table I summarizes these metrics.

TABLE I
Performance Metrics on Test Dataset

Metric	**Value**
Precision	93.49%
Recall	85.94%
F1-Score	89.28%
Accuracy	81.10%
True Positive (TP)	618
False Positive (FP)	43
False Negative (FN)	101

A. Analysis of Detection Metrics

The model achieved a Precision of 93.49%, indicating that a high proportion of the detected cacao trees were correctly identified. Additionally, the Recall of 85.94% demonstrates that the model successfully identified a significant portion of the trees present in the dataset. The balance between these two metrics is reflected in the F1-Score of 89.28%, showing that the model maintains a good trade-off between precision and recall.

However, when analyzing the Accuracy the model's performance is slightly more moderate, with an accuracy of 81%. This means that while the model can reliably detect cacao trees when it does, there are still some errors present in the detection process.

The True Positive count of 618 highlights the model's ability to detect a substantial number of cacao trees. However, the presence of 43 False Positives (FP) indicates that there were a number of incorrect detections, where objects were identified as cacao trees but were not. Additionally, 101 False Negatives (FN) show that some cacao trees in the test dataset were not detected by the model. These false positives and negatives suggest that further refinement of the detection process is needed to minimize these errors and improve overall performance.

B. Implications of the Results

While the model demonstrates strong performance in terms of precision (93%) and F1-Score (89.28%), the recall value of 85.94% suggests there is room for improvement in identifying all trees present. Additionally, the accuracy of 81% indicates that some errors, particularly false positives and false negatives, still occur.

Given these results, the model is promising for estimating cacao tree populations and supporting large-scale monitoring. However, the moderate recall and accuracy values highlight the need for improvements, particularly in reducing the number of false negatives, which can result in trees being overlooked. The inclusion of Recall and F1-Score metrics allows for a more comprehensive assessment of the model's performance, balancing precision with its ability to detect all relevant objects.

Moreover, the presence of false positives indicates the need for further refinement of the bounding box predictions or adjustments to the post-processing steps to filter out incorrect detections. The false negatives, on the other hand, suggest that future improvements could involve enhancing the model's ability to detect trees in challenging conditions, such as dense vegetation or shaded areas, which could be achieved through more diverse training data or advanced augmentation techniques.

In summary, while the current performance is not optimal for precise plantation management, the integration of the YOLOv8x segmentation model with SAHI shows promise as a scalable solution for cacao tree detection. However, its application should be understood within the context of its limitations, and further developments would be necessary to enhance its reliability for more critical agricultural tasks.

C. Future Work

One limitation of this study is the lack of cross-validation, which would provide a more robust estimate of model performance across different data splits. The small dataset size and computational constraints limited the ability to implement this approach. In future studies, cross-validation and statistical tests will be conducted to better assess model reliability and generalizability.

To further enhance the detection performance, future research could explore the following avenues:

- Model Optimization: Fine-tuning the YOLOv8x model with a larger and more diverse dataset to reduce false negatives and improve overall detection accuracy.
- Advanced Post-processing: Implementing more sophisticated post-processing techniques to minimize FP instances and refine bounding box predictions.
- Multispectral Imaging: Incorporating multispectral or hyperspectral imagery to provide additional features for

the model, potentially improving detection under varying environmental conditions.

- Integration with Other Data Sources: Combining orthomosaic imagery with other geospatial data, such as soil quality and climate information, to create a more comprehensive agricultural management system.

These improvements could lead to even higher precision and accuracy, making the system more robust and versatile for various agricultural applications.

V. Conclusion

This study explored the use of drone-based orthomosaic imagery combined with the YOLOv8x segmentation model and the SAHI method for detecting cacao trees. The model achieved good precision (93%), but its accuracy of 81% indicates the presence of errors, particularly in false positives and false negatives. As a result, the system is best suited for providing population estimates and identifying areas for further investigation rather than for precise, real-time decision-making.

While promising for large-scale monitoring, improvements are needed to enhance the model's reliability, including reducing detection errors and incorporating additional data sources. Future work should focus on optimizing the model for more accurate and comprehensive plantation management, making it a more robust tool in precision agriculture.

References

[1] KOBLITZ, M. G. B. *Café, cacau e chá*. In: KOBLITZ, M. G. B. (Ed.). Matérias-primas alimentícias: composição e controle de qualidade. Rio de Janeiro: Guanabara Koogan, 2011. p. 122-146.

[2] SEQUEIRA, Andreia Filipa Caeiro. *CACAU: DO FRUTO AO CHOCOLATE*. 2016. 27 f. Tese (Doutorado) - Curso de Biologia, Escola Secundária Rainha Santa Isabel, Estremoz, 2016.

[3] SENAR, Serviço Nacional de Aprendizagem Rural. *Cacau: produção, manejo e colheita*, 2018.

[4] BRAINER, M. S. C. P. *Produção de Cacau*, Caderno Setorial ETENE. Banco do Nordeste do Brasil, ano 6, nº 149, 2021.

[5] SANTOS FILHO, L. P. et al. *Cacauicultura baiana: considerações econômicas e tecnológicas, anos agrícolas de 1975/76 a 2011/12*. Ilhéus, BA: CEPLAC/CEPEC, 2021. Boletim Técnico, nº 220, 23p.

[6] IBGE - Instituto Brasileiro de Geografia e Estatística. *Produção Agrícola Municipal - PAM*. Disponível em: https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9107-producao-agricola-municipal.html. Acesso em: 2023.

[7] BAHIA DE AGUIAR, P. C.; PIRES, M. M. *A região cacaueira do sul do estado da Bahia (Brasil): crise e transformação*. Cuadernos de Geografía: Revista Colombiana de Geografía, vol. 28, núm. 1:192, 2019. Disponível em: http://dx.doi.org/10.15446/rcdg.v28n1.67437. *Relatório Anual de Produção de Cacau*, 2021.

[8] International Cocoa Organization. *Cocoa Market Report*. April 2024.

[9] SCHROTH, G., et al. Cocoa in a Changing Climate: Vulnerability and Adaptation. *Climatic Change*, 2016.

[10] LÄDERACH, P., et al. Predicting the Impact of Climate Change on Cocoa Production. *Agricultural Systems*, 2013.

[11] Ayubi, A., Faiz, M., Situmorang, G. B., Ramadhani, K. N., Utama, N. P., A Cocoa Ripeness Detection and Classification Model Based on Improved YOLOv5s, *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2023, pp. 1-6, doi: 10.1109/ICAICTA59291.2023.10390119.

[12] Bastidas-Alva, R. A., Paitan Cardenas, J. A., Bazan Espinoza, K. S., Povez Nuñez, V. K., Quincho Rivera, M. E., Huaytalla, J., Recognition and classification system for trinitario cocoa fruits according to their ripening stage based on the Yolo v5 algorithm, *2022 Asia Conference on Advanced Robotics, Automation, and Control Engineering (ARACE)*, 2022, pp. 138-142, doi: 10.1109/ARACE56528.2022.00032.

[13] Justam, Indrabayu, Zainuddin, Z., Basri, Detection and Counting of the Number of Cocoa Fruits on Trees Using UAV, *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2023, pp. 257-263, doi: 10.1109/IAICT59002.2023.10205752.

[14] Akyon, F. C., Altinuc, S. O., Temizel, A., Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection, *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 966-970, doi: 10.1109/ICIP46576.2022.9897990.

[15] G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics YOLO, version 8.0.0, Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics. [Accessed: Sep. 08, 2024].

[16] ALHICHRI, H.; AL-KHUSHI, W.; BOUKERCH, I.; et al. *Deep-Learning-Based Automated Palm Tree Counting and Geolocation in Large Farms from Aerial Geotagged Images*. In: IEEE Access, 2021, p. 1-16.