

# Comparison of YOLO instance segmentation models in identifying forest species using leaf images

Isabela Yasmim Osowski\*, Eduardo Marcon Gonçalves Fadel\*, Roque Cielo Filho \*,  
Larissa De Bortolli Chiamolera Sabbi\*, Pedro Luiz de Paula Filho\*

\*UTFPR - Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil  
osowskiisabela@gmail.com, leduardofadel@gmail.com, cielofbr@gmail.com,  
larissasabbi@utfpr.edu.br, plpf2004@gmail.com

**Abstract**—The vast area of forests in Brazil, along with the biodiversity of the flora, poses a challenge to the identification of tree species. This process can be time-sensitive due to the nature of certain projects; therefore, the use of leaves for the identification is ideal in such cases. In order to assist this process, automated tools that use deep learning can be an aid in the conservation efforts. This study evaluates and compares two YOLO (You Only Look Once) models, the nano versions of YOLOv11 and YOLOv12, for the task of identifying species using their leaves. For this, a custom dataset was created with 26 species under varied conditions; the images also have a complex background to guarantee the model's efficiency in fieldwork. Overall, YOLOv11 achieved higher recall (56.746%) and mAP@50 (72.634%), YOLOv12 demonstrated superior precision (86.054%) and better per-class performance. The results show that both models' performance varies due to the dataset annotation and the leaf morphology, as both models struggled with species exhibiting compound leaves. This work contributes to the development of practical tools for biodiversity monitoring and supports future advancements in automated species identification, highlighting the potential of lightweight models for tree species identification.

**Keywords**—Instance segmentation; YOLO; species identification.

## I. INTRODUCTION

Brazil currently has the second largest forest area in the world, corresponding to 12% of the world's forest [1]. According to the latest country report on forest genetic resources, submitted by Brazil to the Food and Agriculture Organization of the United Nations (FAO), the nation faces a significant challenge in protecting its forest area while cataloging its biodiversity. This issue stems from the large diversity of forest species in the country, with many being unidentified despite research [2].

This happens because there are many dense forest areas that are difficult to access, coupled with the rigorous process of correctly identifying a species. Precision in this step is crucial for conservation, where the management of forest resources is essential [3]. The standard botanical identification process usually involves collecting samples, such as leaves, flowers,

and fruits. This is followed by the verification with herbarium specimens, which are dried and already identified. Lastly, if there is no certainty about the species, specialists in that group or area are called for consultation.

Given that Brazil encompasses a vast territory with six major biomes, each presents unique species compositions and conservation concerns. This biodiversity presents a significant challenge in species identification, particularly due to the scarcity of specialized personnel with both the taxonomic expertise and the availability to conduct extensive fieldwork [4]. This lack of human resources makes it difficult to monitor and catalog biodiversity at the scale needed, particularly in remote or understudied areas.

In this context, technological tools have emerged to assist in the effort to support species identification. Recent advances in instance segmentation enable not only the classification of plant species but also the precise delineation of key morphological features, such as leaves. They are known as vegetative characters and are available most of the time; therefore, using them is best for time sensitive projects. [5]. These tools can serve as a bridge between fieldwork and expert validation, aiding both specialists and non-specialists in accelerating initial identification and focusing expert attention where it is most needed.

However, most existing models are trained on datasets with controlled conditions, limiting their applicability in real-world scenarios with complex backgrounds and high interspecies variability. To address this gap, this work evaluates multiple YOLO-based instance segmentation models trained on a locally curated dataset containing native and foreign species under varied environmental conditions.

The study proposes a comparison of different YOLO (You Only Look Once) models trained for instance segmentation. These models can then be used for deployment in practical applications, such as a mobile identification tool, which could assist both researchers and conservationists in the field. Additionally, the dataset and model comparisons will serve as a

foundation for future work, facilitating further improvements in automated species recognition. As such, this research aims to support future efforts in biodiversity documentation and forest conservation.

## II. RELATED WORK

Various approaches have been proposed for instance segmentation applied to plant leaves. Yang *et al.* [6] introduced a method for segmentation and classification in images with complex backgrounds. Their approach locates and segments leaves in images. Their results demonstrate a 1.89% improvement in average precision compared to the original Mask R-CNN.

Meanwhile, Guo *et al.* [7] presented LeafMask, an anchor-free model specifically designed for delineating and counting plant leaves. Their work highlights the importance of precise leaf segmentation for plant phenotyping in agricultural applications, particularly for assessing leaf growth stages.

Focusing on forest species identification, studies often rely on features such as trunk, fruit, flower, and plant structure. The dataset introduced by Yang *et al.* [8] encompasses all these aspects, containing a total of 41,467 images of 50 tree species. Of these, only 9,763 are annotated for instance segmentation. The dataset includes images captured under varying lighting conditions and across different seasons and is available for use<sup>1</sup>.

Kumar *et al.* [9] proposed a hybrid approach combining Convolutional Neural Networks (CNNs) and edge-based segmentation to identify medicinal plants. Their model outperforms traditional methods, achieving 98% accuracy. The study uses the Flavia dataset<sup>2</sup>, which includes 33 species with a simple, white background. Their work explores the application of CNNs and segmentation techniques for species recognition, addressing challenges posed by morphological variations and reinforces the use of robust neural networks.

Wang *et al.* [10] modified the YOLOv5 neural network for forest species identification by incorporating a Squeeze-and-Excitation (SE) module. They used a custom dataset containing eight common Chinese species, featuring leaves in urban settings with seasonal, size, and lighting variations. The authors emphasize the importance of complex backgrounds for accurate feature learning and employ an object detection approach, leaving instance segmentation for future exploration.

These studies demonstrate the effectiveness of instance segmentation in both leaf localization and species identification, particularly in datasets featuring complex backgrounds and diverse environmental conditions. These works showcase the potential of deep learning-based segmentation for species recognition, particularly when handling morphological variations and challenging imaging conditions. The success of

earlier YOLO-based approaches also suggests opportunities for extending these frameworks to instance segmentation in future research.

## III. MATERIALS AND METHODS

With the goal of comparing two YOLO architectures, Figure 1 illustrates the methodological workflow adopted in this study, which consists of three main phases: dataset preparation, model training, and performance evaluation. The initial phase involves the creation and curation of a custom dataset containing annotated images of plant leaves. Subsequently, the two most recent YOLO architectures available at the time of experimentation were trained using this dataset. The final phase entails a comparative evaluation in which both models are tested on identical image sets to ensure fair benchmarking. Key performance metrics are collected and analyzed to assess the model's efficacy in instance segmentation tasks. This structured approach enables a systematic comparison of the model's performance within the study's specific context.

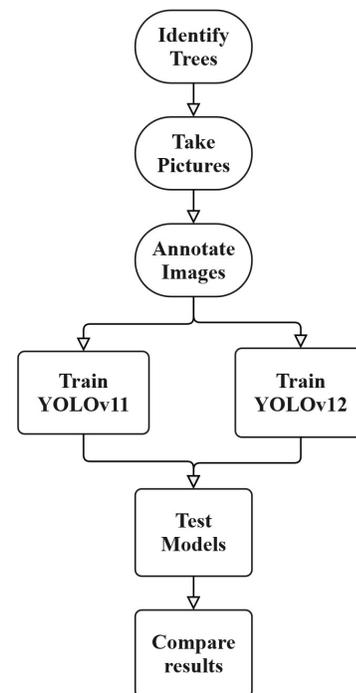


Fig. 1. Flowchart of the work's execution

### A. Dataset

The dataset creation process began with a survey of tree species located within the campus and urban forest area of UTFPR in Medianeira, Paraná, Brazil (25°17'43" S, 54°05'38"

<sup>1</sup>[https://ytt917251944.github.io/dataset/\\_jekyll/](https://ytt917251944.github.io/dataset/_jekyll/)

<sup>2</sup><https://flavia.sourceforge.net/>

W). This area contains an urban forest fragment of Seasonal Semideciduous Forest according to previous studies [11]. Due to constraints of time and accessibility, the current study focused on a subset of these species, utilizing existing taxonomic records where available and consulting botanical specialists for unidentified specimens to ensure accurate species identification.

The image acquisition process followed a systematic protocol designed to capture natural variability in both subject and environmental conditions. To account for environmental variability, the images composing the dataset were captured across different days with varying natural lighting conditions and weather. The process stretched for one year to account for the variation in appearance across seasons. For species with multiple individual specimens, photographs were taken from different trees to capture phenotypic variations caused by the environment, such as the soil composition where the tree was planted.

The photographs were taken using smartphone cameras, with particular attention to maintaining multiple leaves within each frame to ensure instance-rich images. The collection strategy intentionally incorporated three distinct image types: leaves on a branch, handheld leaves, leaves on grass, and high-contrast backgrounds, like concrete, to enhance model robustness. These different image types are showcased in Figure 2.

Image annotation was performed using the Computer Vision and Annotation (CVAT) tool, where each leaf instance was precisely delineated with polygonal masks. In the case of compound leaves, the smallest morphological unit (leaflet) was annotated. With the exception of *Parapiptadenia rigida* and *Cojoba arborea*, whose leaflets were too small to annotate individually. Each mask was assigned to its corresponding species class, ensuring accurate taxonomic representation throughout the dataset. Note that some images don't have all the leaves annotated, due to the extensive manual work required for the task; the annotation prioritized more prominent leaves.

The final dataset was partitioned into training (70%), validation (20%), and test (10%) subsets using a stratified approach that maintains proportional representation of all classes across all partitions, preventing any class imbalance that could affect model performance evaluation. In total, there are 1,415 images distributed among 26 species. Table I shows the distribution of species and their corresponding image count.

### B. YOLO

Ultralytics released You Only Look Once (YOLO), an architecture for solving object detection problems in real-time execution. The original YOLO follows a straightforward pipeline: input images are resized to 448×448 pixels, processed by a CNN for feature extraction, and refined using Non-Maximum Suppression (NMS) to filter detections. This design achieves

TABLE I  
SPECIES AND IMAGE RELATION

| Scientific name                  | Image amount |
|----------------------------------|--------------|
| <i>Anacardium occidentale</i>    | 72           |
| <i>Annona squamosa</i> sp.       | 64           |
| <i>Aspidosperma polyneuron</i>   | 80           |
| <i>Brunfelsia uniflora</i>       | 135          |
| <i>Campomanesia xanthocarpa</i>  | 76           |
| <i>Casearia decandra</i>         | 101          |
| <i>Cecropia</i> sp.              | 103          |
| <i>Citrus</i> sp.                | 51           |
| <i>Coffea arabica</i> sp.        | 102          |
| <i>Cojoba arborea</i>            | 89           |
| <i>Cordia trichotoma</i>         | 63           |
| <i>Crescentia cujete</i>         | 73           |
| <i>Grevillea robusta</i>         | 113          |
| <i>Guarea kunthiana</i>          | 78           |
| <i>Inga</i> sp.                  | 71           |
| <i>Machaerium paraguayense</i>   | 64           |
| <i>Nectandra lanceolata</i>      | 67           |
| <i>Nectandra megapotamica</i>    | 89           |
| <i>Parapiptadenia rigida</i> sp. | 60           |
| <i>Paubrasilia echinata</i>      | 136          |
| <i>Persea americana</i>          | 76           |
| <i>Plinia cauliflora</i>         | 79           |
| <i>Prunus serrulata</i>          | 67           |
| <i>Psidium cattleianum</i>       | 82           |
| <i>Quercus robur</i>             | 84           |
| <i>Syzygium cumini</i>           | 123          |

faster inference speeds while maintaining higher accuracy than other real-time detection systems [12].

Since its release, the YOLO architecture has undergone significant evolution, with twelve major versions released at the time of this study. A key advancement was introduced with YOLOv5, which offered a scalable framework with multiple model sizes (e.g., nano, small, medium, large, and extra-large) to balance speed and accuracy. Smaller variants, such as the nano model, prioritize computational efficiency with faster inference times but reduced precision, while larger models (e.g., extra-large) maximize detection accuracy at the cost of increased computational resources. Subsequent versions have maintained this scalable design philosophy, further refining the trade-offs between performance and efficiency across different application requirements.



(a) Leaves on a branch

(b) Handheld

(c) Leaves on grass surface

(d) High-contrast background

Fig. 2. Example of image types on *Campomanesia xanthocarpa*

The YOLO architecture has undergone significant improvements in its recent iterations. YOLOv11 introduces an enhanced backbone network for more efficient feature extraction, achieving superior accuracy across multiple tasks while utilizing fewer parameters compared to previous versions [13]. In turn, YOLOv12 incorporates an innovative area attention mechanism that partitions feature maps into uniform regions, thereby reducing computational complexity while maintaining detection performance. This version further optimizes computational efficiency through modifications to the attention mechanism and the integration of Residual Efficient Layer Aggregation Networks (R-ELAN) [14].

### C. Experiments

The models were trained using Python as the programming language, with all computational experiments conducted on the Google Colaboratory<sup>3</sup>. The dataset was strategically partitioned to optimize model development and evaluation, with 70% of the original samples allocated for training, 20% for validation during the model optimization phase, and the remaining 10% reserved for final testing. This distribution ensured sufficient data for robust model training while maintaining adequate samples for hyperparameter tuning and unbiased performance assessment.

The cloud-based Colab environment provided access to GPU acceleration, significantly reducing training times for the computationally intensive deep learning models. As such, there was liberty to choose from different-sized models of each version. For both versions, the nano model was selected to explore the trade-offs between model size, accuracy, and training cost.

<sup>3</sup><https://colab.research.google.com/>

Focusing on a lightweight approach, more viable for application in devices with hardware limitations.

Notably, YOLOv11 provides access to pretrained instance segmentation weights derived from the COCO dataset, which facilitated faster convergence and improved generalization during training. This pretrained foundation significantly reduced the number of epochs required to achieve stable performance. On the other hand, the YOLOv12 model, at the time of experimentation, had no available pretrained instance segmentation weights.

Consequently, the model had to be trained from scratch, which demanded more training epochs and fine-tuning to achieve performance levels comparable to its pretrained counterpart, which was facilitated by the lower computational cost of the nano version. Since this version is presented as a lightweight architecture, and therefore is inherently optimized for speed and low-resource inference, it's ideal for deployment in field applications where computational resources are limited.

## IV. RESULTS AND DISCUSSION

Both models were evaluated on the same test dataset using a confidence threshold of 70%, an IoU threshold of 60%, and a non-maximum suppression (NMS) algorithm to handle overlapping predictions. Table II summarizes the overall metrics for YOLOv11 and YOLOv12, while Table III provides precision scores for each of the 26 target species.

Although YOLOv11 achieved higher values in recall (56.746%) and mAP@50 (72.634%), YOLOv12 slightly outperformed in overall precision (86.05% vs. 84.21%). This difference suggests that YOLOv11 with pretrained COCO weights was better at identifying a broader range of leaf instances

(as reflected in recall), while YOLOv12, despite being a nano model trained from scratch, made more confident and correct predictions on average.

TABLE II  
DIRECT MODEL COMPARISON

| Metric    | YOLOv11 | YOLOv12 |
|-----------|---------|---------|
| Precision | 85.549% | 86.054% |
| Recall    | 56.746% | 40.333% |
| mAP50     | 72.634% | 64.093% |

The recall metric measures the model's ability to identify true positives, so it checks how many of the classes' annotated instances were classified correctly in the given image. While this metric is largely used in predictive models, its importance comes from applications where false negatives are more critical, such as disease detection. Recall calculates the true positives divided by the total number of true positives plus the false negatives. The false negatives correspond to the number of annotated instances that the model failed to recognize, whether from picking the wrong class or not identifying it at all.

The main impact observed in both yolo models comes from the nature of the images; the multiple leaves or leaflets present a challenge for the image annotation, as it raises the complexity of detecting the individual instances. Figure 3a and Figure 4a depict an example of two different species with compound leaves.

This presents a problem while utilizing the recall as an evaluation metric, given that the extra leaves detected in Figure 3b and Figure 3c are classified as false positives. Combined with the non-identification of one or more annotated leaves, false negatives decrease both the precision and recall of the model. This problem is largely due to the way the dataset was annotated. However, the standard YOLO evaluation metrics are not able to accurately assess the ability of the model to solve the current problem of identifying species.

The evaluation demonstrated that the models' performance metrics vary with the confidence threshold. When the threshold is raised, the model only makes predictions when it is highly certain. This more rigorous approach means it's more confident in identifying the annotated leaves, with high recall, but generates fewer predictions overall. Consequently, a single false positive among these few predictions reduces precision as there are fewer instances per image. Furthermore, an excessively high threshold can cause the model to miss detections entirely in some images.

As observed in the *Paubrasilia echinata* examples, YOLOv11 in Figure 3b and YOLOv12 in Figure 3c have different capacities when looking for precise leaves. Looking

at the results in a test image, YOLOv12 has a larger ability to recognize multiple leaves. This shows that the model adapted better when learning the actual features of the species, and could be a result of training the model from scratch as opposed to using pre-trained weights like in YOLOv11.

Beyond that, a species-level analysis reveals several key insights. First, species such as *Cecropia sp.* and *Paubrasilia echinata*, which possess compound leaves and were only partially annotated in the dataset, exhibited relatively lower precision across both models. This is likely due to the models identifying additional unannotated leaflets as distinct instances, a consequence of incomplete annotations rather than model failure.

TABLE III  
PRECISION BY CLASS

| Species name                     | YOLOv11        | YOLOv12        |
|----------------------------------|----------------|----------------|
| <i>Anacardium occidentale</i>    | 91.429%        | 87.097%        |
| <i>Annona squamosa</i> sp.       | 85.714%        | 86.667%        |
| <i>Aspidosperma polyneuron</i>   | 82.500%        | 89.474%        |
| <i>Brunfelsia uniflora</i>       | 77.778%        | 85.366%        |
| <i>Campomanesia xanthocarpa</i>  | 82.500%        | <u>67.568%</u> |
| <i>Casearia decandra</i>         | 87.097%        | 90.476%        |
| <i>Cecropia sp.</i>              | 70.000%        | 73.611%        |
| <i>Citrus sp.</i>                | 90.476%        | 92.308%        |
| <i>Coffea arabica</i> sp.        | 87.500%        | 94.444%        |
| <i>Cojoba arborea</i>            | <b>100.00%</b> | 88.889%        |
| <i>Cordia trichotoma</i>         | 96.154%        | <b>100.00%</b> |
| <i>Crescentia cujete</i>         | 80.000%        | 84.848%        |
| <i>Grevillea robusta</i>         | 86.957%        | 80.769%        |
| <i>Guarea kunthiana</i>          | 88.000%        | 92.857%        |
| <i>Inga sp.</i>                  | 80.000%        | 92.593%        |
| <i>Machaerium paraguariense</i>  | 78.788%        | 76.000%        |
| <i>Nectandra lanceolata</i>      | 84.000%        | 90.000%        |
| <i>Nectandra megapotamica</i>    | 80.000%        | 88.235%        |
| <i>Parapiptadenia rigida</i> sp. | 77.778%        | 83.333%        |
| <i>Paubrasilia echinata</i>      | <u>67.045%</u> | 73.723%        |
| <i>Persea americana</i>          | 92.857%        | 92.857%        |
| <i>Plinia cauliflora</i>         | 91.304%        | 79.167%        |
| <i>Prunus serrulata</i>          | 90.909%        | 87.500%        |
| <i>Psidium cattleianum</i>       | 89.655%        | 82.353%        |
| <i>Quercus robur</i>             | 89.655%        | 86.957%        |
| <i>Syzygium cumini</i>           | 82.692%        | 90.323%        |

In contrast, *Parapiptadenia rigida* and *Cojoba arborea*, which also have compound leaves but were fully annotated, achieved better precision scores. However, a common issue observed in these cases was the prediction of multiple bounding masks over the same leaf area. In these cases, there are multiple class predictions referred to by their popular name. The

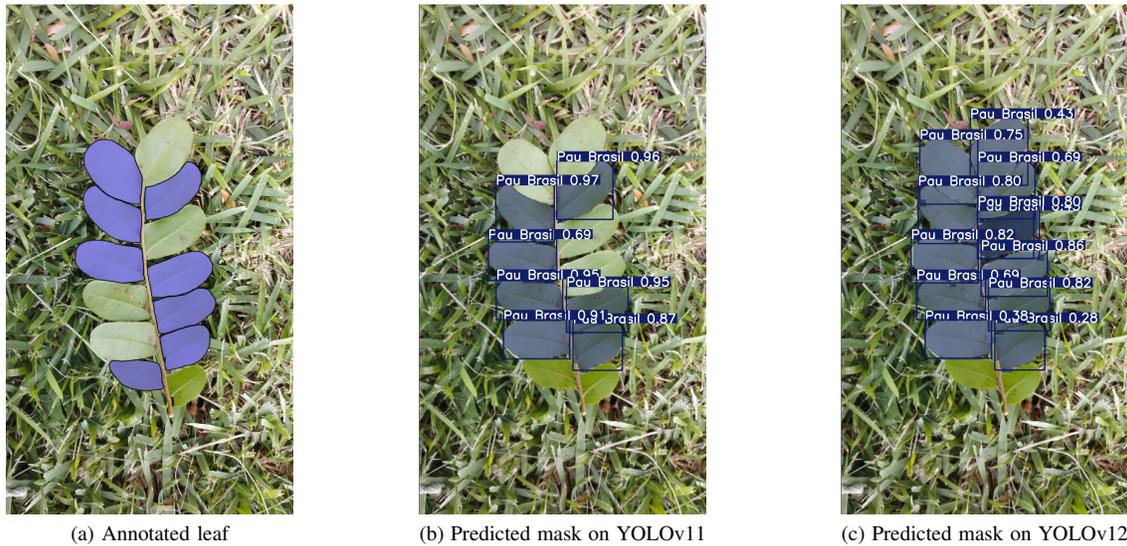


Fig. 3. Difference between *Paubrasilia echinata* annotation and prediction

species *Parapiptadenia rigida*, with the popular name 'Angico Vermelho', is often identified more times than necessary. An example is in a test image, in Figure 4, there is a side-by-side comparison of the annotated picture and the predicted mask by the YOLOv11 model; even with an NMS technique, it identifies two leaves on a similar mask. This over-detection behaviour was minimized using an NMS algorithm to reduce redundant instance predictions, particularly in compound leaved species.

Interestingly, YOLOv12 outperformed YOLOv11 in precision for 16 of the 26 species, despite lacking pretrained weights. This highlights the nano architecture's efficiency and its capability to generalize well when trained thoroughly on a specialized dataset. Notable examples include *Coffea arabica*, *Cordia trichotoma*, and *Persea americana*. These results suggest that YOLOv12's compact architecture, combined with targeted training, can yield competitive or superior results for specific classes in resource-constrained scenarios.

Nevertheless, YOLOv11 maintained superior performance in mAP and recall, indicating that its larger architecture and access to pretrained segmentation features facilitated a better overall understanding of the object distribution in complex forest imagery. For example, it demonstrated high precision for species like *Prunus serrulata* and *Grevillea robusta*.

Overall, while both models have merits, YOLOv12 nano showed remarkable effectiveness given its size and training from scratch. These findings suggest that, for applications requiring lower computational overhead or on-device inference, the YOLOv12 nano architecture can be a viable and accurate solution. Meanwhile, YOLOv11 nano, particularly when pre-

trained, remains a strong choice for achieving higher recall and broader detection coverage in richly annotated datasets.

## V. CONCLUSION

This study compared two YOLO-based instance segmentation models, YOLOv11 medium and YOLOv12 nano, for leaf segmentation and species classification in a dataset of forest species for the west region of Paraná in Brazil. The results demonstrate that while YOLOv11, benefiting from pretrained weights, achieved higher recall and mAP, YOLOv12 nano, despite its smaller size and training from scratch, exhibited superior precision in most species. This highlights the potential of lightweight models for deployment in field applications where computational resources are limited.

Key challenges emerged in segmenting compound leaves, where incomplete annotations and overlapping predictions affected the performance metrics. These findings underscore the importance of high-quality datasets and refined non-maximum suppression techniques for improving model accuracy. Future work should focus on expanding the dataset to include more species and refining annotation strategies, particularly for complex leaf structures. Additionally, integrating these models into mobile applications could enhance accessibility for researchers and field workers, facilitating real-time species identification in biodiversity monitoring efforts.

Ultimately, this research contributes to the growing field of automated plant recognition, offering insights into model selection and optimization for ecological studies. Such tools can bridge the gap between deep learning and botanical expertise,



(a) Annotated leaf



(b) Predicted mask on YOLOv11

Fig. 4. Issue on mask prediction on *Parapiptadenia rigida* compound leaves

offering knowledge to both specialists and other users, all the while being more accessible. In this context, these tools can play a crucial role in conservation, supporting the sustainable management of Brazil's rich forest resources.

#### ACKNOWLEDGMENTS

The authors would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq) for the Research Productivity Grant in Technical Development and Innovative Extension (DT-II – Process: 303279/2023-0) and the research scholarship provided via the Institutional Program of Technological Initiation Scholarships (PIBIT).

#### REFERENCES

- [1] Food and Agriculture Organization of the United Nations, "Global forest resources assessment 2020," Rome, 2020. [Online]. Available: <https://doi.org/10.4060/ca9825en>
- [2] —, "The second report on the state of the world's forest genetic resources," Rome, 2025. [Online]. Available: <https://doi.org/10.4060/cd4838en>
- [3] R. C. V. M. d. Silva, M. J. G. Hopkins, and I. S. Thompson, "Identificação botânica na amazônia: Situação atual e perspectivas," p. 81, 2003. [Online]. Available: <https://repositorio.inpa.gov.br/handle/1/36149>
- [4] N. Haider, N. Haider (2011). *Identification of Plant Species using Traditional and Molecular-based Methods*, pp. 1-62. In: *Wild Plants: Identification, Uses and Conservation* (Ed. RE Davis). Nova Science Publishers, Inc. Nova Science Publishers, 01 2011.
- [5] C. Urbanetz, J. Y. Tamashiro, and L. S. Kinoshita, "Chave de identificação de espécies lenhosas de um trecho de floresta ombrófila densa atlântica, no sudeste do Brasil, baseada em caracteres vegetativos," *Biota Neotropica*, vol. 10, no. 2, p. 349–398, 06 2010. [Online]. Available: <http://dx.doi.org/10.1590/s1676-06032010000200036>
- [6] X. Yang, A. Chen, G. Zhou, J. Wang, W. Chen, Y. Gao, and R. Jiang, "Instance segmentation and classification method for plant leaf images based on isc-mrcnn and aps-dcnn," *IEEE Access*, vol. 8, pp. 151 555–151 573, 2020. [Online]. Available: <https://doi.org/10.1109/ACCESS.2020.3017560>
- [7] R. Guo, L. Qu, D. Niu, Z. Li, and J. Yue, "Leafmask: Towards greater accuracy on leaf segmentation," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1249–1258. [Online]. Available: <https://doi.org/10.1109/ICCVW54120.2021.00145>
- [8] T. Yang, S. Zhou, Z. Huang, A. Xu, J. Ye, and J. Yin, "Urban street tree dataset for image classification and instance segmentation," *Computers and Electronics in Agriculture*, vol. 209, p. 107852, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169923002405>
- [9] P. Kumar, S. V. Kumar, and S. S. Pandi, "Cnn and edge-based segmentation for the identification of medicinal plants," in *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2024, pp. 89–94. [Online]. Available: <https://doi.org/10.1109/ICICV62344.2024.00021>
- [10] Z. Wang, X. Su, and S. Mao, "Single-species leaf detection against complex backgrounds with yolov5s," *Forests*, vol. 15, 6 2024. [Online]. Available: <https://www.mdpi.com/1999-4907/15/6/894>
- [11] L. D. B. C. Sabbi, C. D. Câmara, R. Cielo-Filho, A. L. d. Silva, and M. D. Thrun, "Áreas verdes urbanas: uma proposta para conservação da biodiversidade e educação ambiental, estudo de caso do bosque da utfpr, câmpus medianeira," *Tempo da Ciência*, vol. 31, no. 61, p. 16, nov. 2024. [Online]. Available: <https://e-revista.unioeste.br/index.php/tempodaciencia/article/view/34365>
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. [Online]. Available: <https://doi.org/10.1109/CVPR.2016.91>
- [13] G. Jocher and J. Qiu, "Ultralytics yolov11," 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [14] Y. Tian, Q. Ye, and D. Doermann, "Yolov12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, 2025.