ICPA
17th INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE
ConBAP 2026
CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO E DIGITAL
13-16 July 2026 | Porto Alegre, Brazil
ISPA AsBraAP

Detection of Banana Bunches and Peduncles in the Prata Catarina Cultivar Using Faster R-CNN With Transfer Learning

Yuri Costa G. da Silva[2], Emannuel Diego G. de Freitas[1,2], Paulo Ricardson S. Costa[2], Davi Vieira P. Beserra[2], Danielo G, Gomes[1]

[1]Department of Teleinformatics Engineering, Federal University of Ceará, Pici campus, Fortaleza-CE, Brazil. [2]Federal Institute of Education, Science and Technology of Ceará, Iguatu campus, Iguatu-CE, Brazil.

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract.

Banana is one of the most produced and consumed fruits worldwide, being strategic for precision agriculture, especially in applications aimed at intelligent management and automated harvesting. Its economic and social relevance in tropical countries reinforces the need for technological solutions that increase productive efficiency and reduce losses in the field. In this context, the automatic detection of bunches and stalks in a natural environment represents a relevant challenge due to occlusion, lighting variations, and the visual similarity between the stalk and the pseudostem. In addition, the complexity of the cultivation environment and the morphological variability of the plants make it even more challenging to obtain robust and generalizable models for field applications. Therefore, this work aimed to evaluate the performance of the Faster R-CNN model with different backbones in the detection of bunches and stalks of the Prata Catarina banana plant, in order to identify the architecture with the best balance between performance and operational feasibility. The results indicated that ResNet-50 presented the highest mAP@50 (47.4% ± 3.7) and mAP@50–95 (25.76% ± 1.84), but with indications of overfitting. DenseNet achieved an mAP@50 of 46.81% (± 1.65) and mAP@50–95 of 25.02% (± 0.74), with lower variability between folds, in addition to precision of 72.35% (± 1.94) and recall of 58.08% (± 1.8), demonstrating a better balance between the metrics. ConvNeXt presented similar performance (mAP@50 of 46.26% ± 5.03; mAP@50–95 of 24.09% ± 2.74), but with greater standard deviation. EfficientNet presented intermediate performance (mAP@50 of 38.18% ± 3.81; mAP@50–95 of 19.49% ± 1.49). MobileNetV3 achieved the highest precision (88.03% ± 2.38), but low recall (32.35% ± 1.55), resulting in an mAP@50 of 25.27% (± 1.71) and mAP@50–95 of 14.37% (± 0.3). Considering the balance between precision, recall, mAP, and statistical analysis, DenseNet was selected due to the lower variability between folds (± 0.74) and low tendency toward overfitting, indicating greater robustness and generalization capability for the proposed task. The findings of this study contribute to the advancement of computer vision–based solutions applied to fruit farming, providing support for the development of embedded systems and intelligent harvesting platforms that are more efficient and reliable.

Keywords

Banana detection; Convolutional Neural Networks, Automation in Agriculture.

Introduction

Produced in more than 120 countries, bananas stand out globally as one of the most consumed fresh fruits and the one with the highest production volume, occupying about 8.8% of the world's cultivated fruit-growing area (de Fátima Vidal, 2024). It is a crop of great nutritional and commercial relevance, whose growing demand for intelligent management, productivity estimation, and automated harvesting has strongly driven the advancement of precision agriculture (Fu et al., 2022; Guru & N, 2025).

Faced with the demands for automation and increased efficiency in the field, computer vision based on Artificial Intelligence is configured as a strategic approach. In this context, deep learning algorithms outperform traditional machine learning methods in the extraction of complex visual features (Fu et al., 2022). Thus, the accurate identification of bunches and stalks by these models constitutes a critical step for the implementation of robots and automated systems in fruit farming (Fu et al., 2022).

However, the rapid and real-time detection of specific plant structures directly in the natural environment remains a relevant technical challenge due to factors of occlusion, irregular lighting, and, critically, the strong visual and textural similarity between the stalk and the banana pseudostem (Fu et al., 2022).

From a practical point of view, the detection of these two structures serves distinct and complementary purposes in the context of banana agricultural production. While the identification of bunches acts as an essential indicator for monitoring development and estimating crop productivity, the precise detection of the stalk is fundamental for harvesting execution in the context of precision agriculture, as it accurately defines the point of intervention in the field (Fu et al., 2022).

On the other hand, due to its reduced dimensions, its higher position on the plant, and the frequent occlusion by leaves or dead branches, the stalk becomes a highly difficult target for computer vision algorithms, presenting higher false negative rates than the bunch itself (Fu et al., 2022). Furthermore, lighting variations typical of the natural environment, such as direct sunlight, excessive brightness, or irregular shadows, and changes in viewing angles drastically affect classification accuracy, reinforcing the need to evaluate robust architectures capable of generalizing these adversities (Guru & N, 2025).

To address these challenges and automate monitoring, Convolutional Neural Networks (CNNs) have brought major advances to computer vision due to their high capacity to extract deep and abstract features, serving as the fundamental basis for object detection, which consists of the precise localization and classification of targets in an image (Shorten & Khoshgoftaar, 2019).

For CNNs to achieve high performance and avoid learning overfitting, they intrinsically depend on large volumes of training data with high-quality annotations. However, in the agricultural domain, the availability of well-structured images is a technical barrier; most public datasets on banana crops focus strictly on the individual fruit, neglecting the characteristics and complexity of the entire bunch or stalk in the real cultivation environment (Guru & N, 2025).

Given this limitation of samples, data augmentation approaches become indispensable tools for artificially expanding the visual diversity of training, helping the model to better generalize against strong variations in lighting, random angles, and occlusions typical of orchards (Guru & N, 2025; Mumuni & Mumuni, 2022).

Another important factor is that expanding data does not solve all challenges. The computational architecture must be robust enough to extract deep features and deal with targets of complex and irregular textures, such as banana plant structures (Fu et al., 2022; Neupane et al., 2019). For this application, CNNs are essential because they learn rich visual representations from topological information, resisting noise and transformations (Hakim et al., 2024). Moreover, in the detection of small and dense targets in natural environments, two-stage detectors present inherent advantages over single-stage detectors, ensuring greater accuracy in spatial localization (Li et al., 2025).

In this technological scenario, the Faster R-CNN algorithm stands out as a consolidated two-stage detection model that employs a Region Proposal Network (RPN) to generate bounding boxes and classify objects within the same pipeline, maximizing identification accuracy (Hakim et al., 2024; Li et al., 2025).

The effectiveness of this type of model, however, intrinsically depends on the backbone chosen for the image feature extraction stage, with several options such as ConvNext, DenseNet, EfficientNet, MobileNet, and ResNet-50, which offer different trade-offs between computational cost, the ability to generalize difficult spatial contexts, and the final model accuracy (Jeevan & Sethi, 2024; Hakim et al., 2024).

In addition to deep feature extraction, the implementation of these artificial intelligence systems directly in the field requires architectures that are efficient in terms of memory and processing (Fu et al., 2022). Evaluations have demonstrated that modern and lightweight convolutional architectures are often preferable to models requiring much greater computational resources, especially when access to a massive dataset is unavailable (Li et al., 2025).

Based on this context, selecting feature extraction bases (backbones) designed to operate with maximum efficiency is of great importance for achieving high real-time frame rates without sacrificing the sensitivity required to distinguish objects so well camouflaged within the foliage (Fu et al., 2022).

However, in the current literature, there is a lack of studies regarding the shortage of robust datasets at the whole-bunch level and, simultaneously, the lack of systematic analyses determining which lightweight backbones operate more efficiently during fine-tuning for such singular targets (Guru & N, 2025; Jeevan & Sethi, 2024).

Given this scenario, the objective of this work is to apply and evaluate the Faster R-CNN model using different backbones for the detection of bunches and stalks of the Prata Catarina banana plant, aiming to identify the architecture that provides the best performance and operational feasibility for the precision agriculture environment.

Materials and Methods

Input Dataset

For the development of this work, a dataset focused on the detection of bunches and stalks of Prata Catarina banana plants (Silva et al., 2025) was used, publicly available on the RoboFlow platform and accessible through the link: https://app.roboflow.com/detectionbunchstalk/detection_bunch_stalk-xd2v9/10. This dataset is composed of a total of 1,015 images in JPEG format encoded in the RGB color space, each with a resolution of 640 x 640 pixels, a value commonly adopted as a standard in the training of the chosen computer vision architecture. The objects present in the images of this dataset were classified into two classes: bunch and stalk (Figure 1). A total of 1,015 data samples.

Fig 1. Sample of images available in the dataset.

The fraction of images corresponding to the training set underwent a data augmentation stage, which resulted in increased variability of the image set and the duplication of the number of samples available for training. Data augmentation is established as an indirect regularization approach, widely adopted with the objective of mitigating overfitting and optimizing the performance of convolutional neural networks (Iqbal, 2025). The applied transformations included random scale resizing (multi-scale), random horizontal flipping, and slight photometric adjustments of brightness, contrast, and saturation.

Architecture of Faster R-CNN and the Evaluated Backbones

Convolutional Neural Networks, or CNNs (De Godoy et al., 2023), are deep artificial neural networks designed for processing grid-structured data, such as two-dimensional images, through convolutional operations. They are composed of multiple interconnected layers, constituting one of the most consolidated approaches in the field of Deep Learning.

In general terms, a CNN is composed of sets of convolutional, pooling, and fully connected layers, organized in successive layers so as to create a deep architecture capable of learning hierarchically structured descriptors of the patterns underlying the data (Nirthika et al., 2022). In this way, CNNs are able to automatically learn discriminative descriptors, thus mitigating the need for manual procedures for feature extraction, which makes them especially effective in computer vision tasks such as classification, segmentation, and object detection.

Faster R-CNN (Ren et al., 2015) (Figure 2) is a widely consolidated architecture in the object detection literature and was adopted in this study as the basis for comparing different feature extraction backbones. The following backbones were evaluated: DenseNet (Huang et al., 2016), EfficientNet (Tan & Le, 2019), MobileNetV3 (Howard et al., 2019), ResNet-50 (He et al., 2016), and ConvNext (Liu; Zhuang et al., 2022), selected for presenting different levels of depth and internal connection strategies. The analysis focused on verifying how these structural variations influence detector performance, considering the main evaluation metrics adopted in the literature.

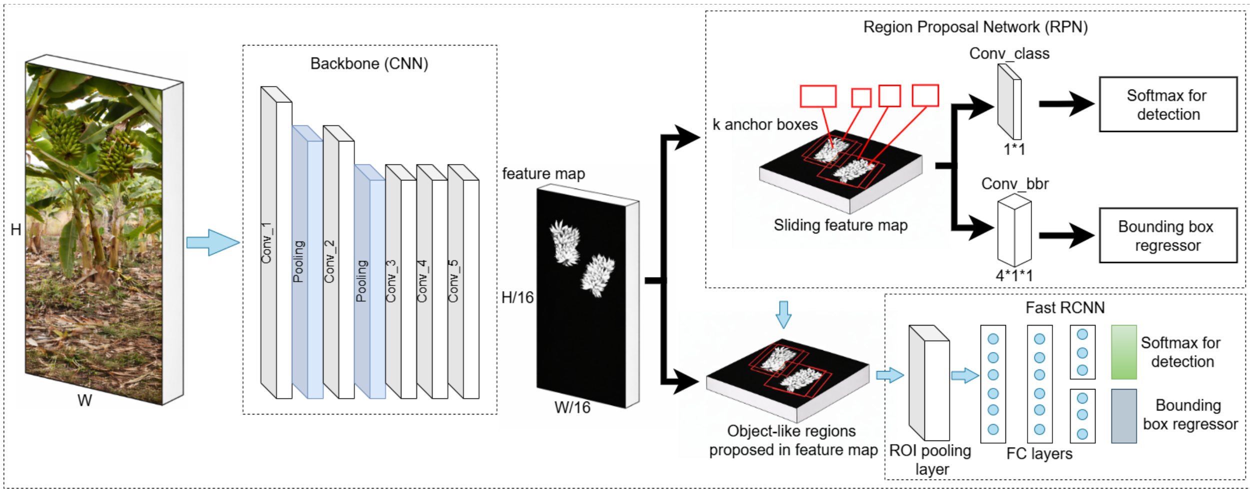


Fig 2. Faster R-CNN architecture (Deng et al., 2018).

Evaluation Metrics

Performance evaluation was carried out by comparing the bounding boxes predicted by the models with the reference annotations of the dataset. From this comparison, the following are obtained: true positives (TP), when the detection matches the annotation considering a minimum Intersection over Union (IoU) threshold; false positives (FP), when the detection is incorrect or poorly localized; and false negatives (FN), when annotated objects are not detected. From these measures, metrics such as Precision, Recall, mAP@0.5, and mAP@0.5–0.95 are derived.

Precision (Eq. 1) corresponds to the ratio between TP and the total number of detections performed. Recall (Eq. 2) is the proportion of TP relative to the total number of real objects. mAP@0.5 (Eq. 3) represents the mean Average Precision across all classes with IoU fixed at 0.5. Meanwhile, mAP@0.5–0.95 (Eq. 4) is a more rigorous metric, calculated as the mean Average Precision over multiple IoU thresholds, ranging from 0.5 to 0.95, with increments of 0.05.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{VP}{VP + FN} \quad (2)$$

$$\mathrm{mAP}_{50} = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (3)$$

$$\mathrm{mAP}_{50\_95} = \frac{1}{10}\sum_{t=0.5}^{0.95} mAP_t \quad (4)$$

In addition, model training was evaluated through the loss function (Eq. 5), which quantifies the prediction error on the training data. Faster R-CNN uses a multi-task loss function, combining classification loss (log loss) and bounding box regression loss based on Smooth L1, enabling the joint training of the Region Proposal Network and the final detector.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{Ncls}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{Nreg}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (5)$$

Experiment Design

All models were trained on the Google Colab platform, running the Ubuntu operating system under an x86_64 architecture and KVM virtualization, in a Python 3.12.12 environment. The available hardware consisted of two logical CPU cores (Intel Xeon 2.20 GHz), 12 GB of RAM, approximately 113 GB of temporary disk storage, and an NVIDIA Tesla T4 GPU with 15 GB of VRAM, using CUDA 12.0 for computational acceleration.

During training, the image dataset was partitioned into five subsets of equal size through the cross-validation procedure known as k-fold, with the objective of evaluating model performance more robustly. The process was repeated five times, and in each fold, one part (1/5) was assigned to validation, while the remaining parts were used for training, as illustrated in Figure 3. In each fold, training was conducted for 20 epochs, with a batch size of 4 images.

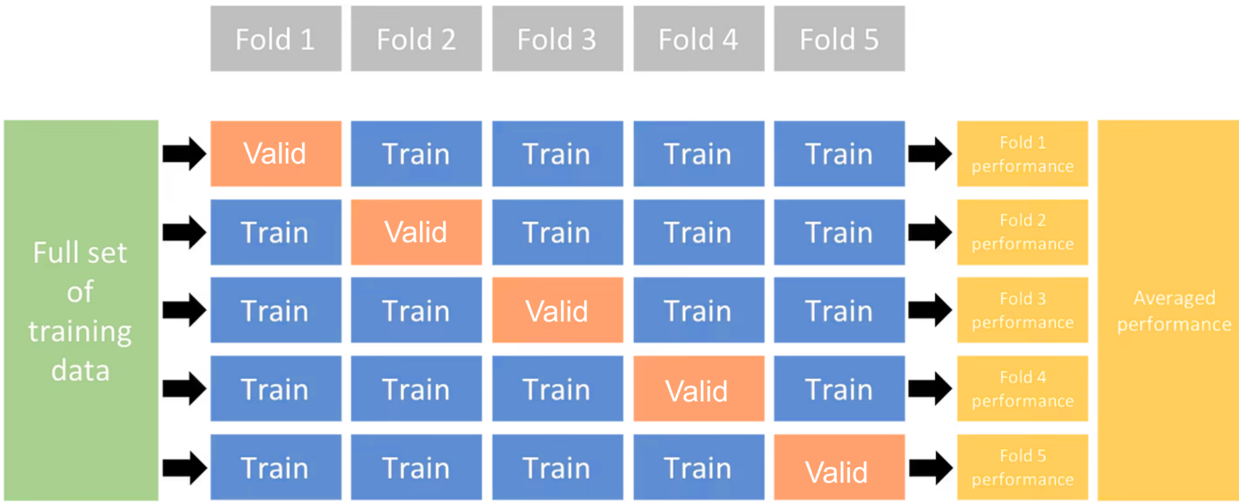


Fig 3. Scheme of the cross-validation method used (Ultralytics, 2023).

To optimize training time, pre-trained backbone weights were used, previously trained on the Common Objects in Context (COCO) dataset. In addition, an independent test subset was reserved and used exclusively for the final evaluation of the models, without participation in the hyperparameter tuning process. In this study, the following hyperparameters were adopted: a learning rate of 1×10^{-5} and a score threshold of 0.7, kept constant and applied uniformly across all evaluated models in order to ensure experimental comparability. The score threshold was defined empirically with the objective of reducing the occurrence of false positives in the detections.

Results and Discussion

This section presents and discusses the results obtained from the comparison between different backbones used in the Faster R-CNN architecture, namely: DenseNet, EfficientNet, MobileNetV3, ResNet-50, and ConvNeXt. In order to perform a fair and controlled comparison among the different backbones, the experimental methodology was structured to keep constant the Faster R-CNN detector base architecture, the initial conditions, and the main training parameters, varying only the backbone responsible for feature extraction. Thus, it becomes possible to analyze in isolation the impact of each backbone on the performance of the object detection model.

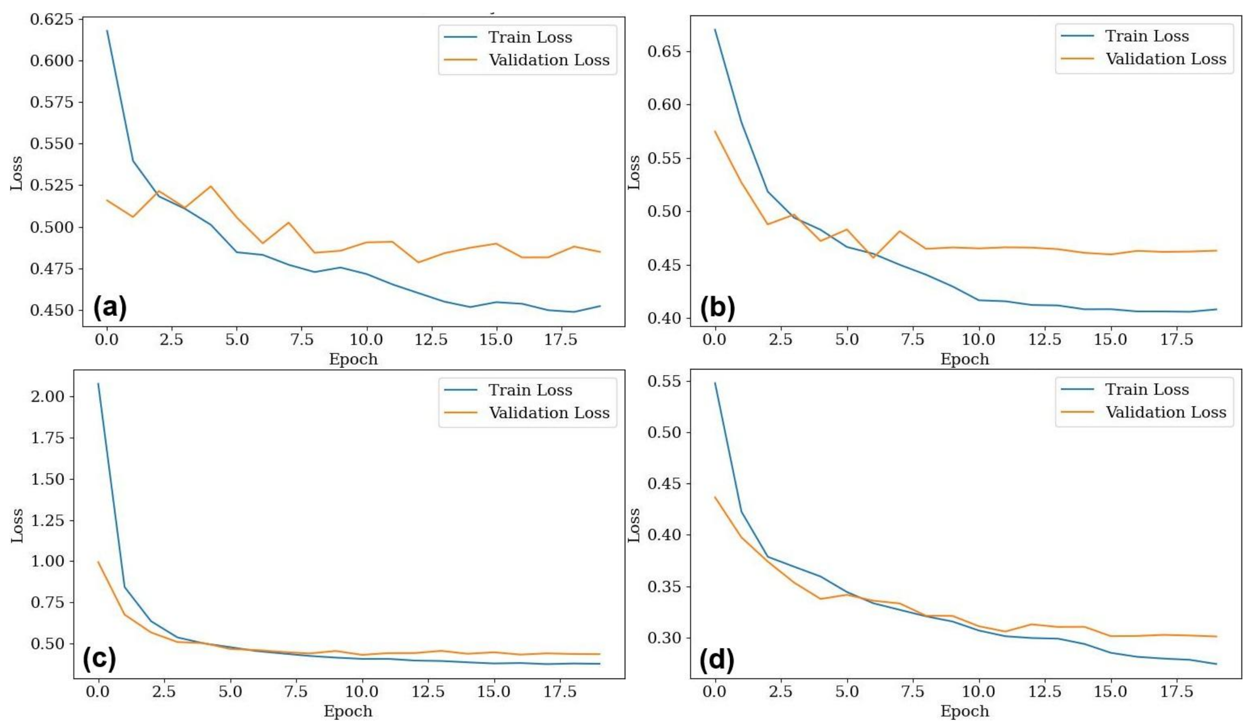
Information regarding the training of the Faster R-CNN model using the ConvNeXt backbone can be observed in Figure 4(a). In this configuration, the training loss curve showed a continuous reduction throughout the epochs, indicating that the model was able to progressively learn the characteristics of the training set. On the other hand, the validation loss curve showed early stabilization, with small oscillations and no significant reduction after the first epochs. This behavior suggests a possible onset of overfitting.

The model with the DenseNet backbone generated the loss graph presented in Figure 4(b). In it, it is observed that the loss curve related to the training phase shows a continuous reduction throughout the epochs, indicating that the model was able to learn the characteristics of the training set. On the other hand, the loss curve related to validation shows an initial reduction, followed by early stabilization and slight oscillations, remaining at levels higher than the training curve. This behavior indicates low generalization, suggesting that the model has limitations when dealing with unseen data.

Figure 4(c) shows the training losses of the model with the EfficientNet backbone, where a rapid convergence is observed in the first epochs, with pronounced reductions in both training and validation losses. Both curves exhibit almost parallel declines, stabilizing at similar levels until the end of training. This behavior indicates a progressive and stable learning process, suggesting a possible generalization capability of the model.

Similarly to the previous model, the model with the MobileNetV3 backbone shown in Figure 4(d) presented a consistent reduction in training and validation losses, with both curves stabilizing at very close values. The small difference between the curves indicates a balanced training process, with no significant evidence of overfitting, demonstrating good generalization capability.

Meanwhile, the training of the model with the ResNet-50 backbone presented the greatest discrepancy between the loss curves, as observed in Figure 5(e). The training loss progressively decreased, reaching approximately 0.34 at the end of training, while the validation loss remained higher, around 0.48. This behavior characterizes a typical case of overfitting, in which the model adjusts excessively to the training data, compromising its ability to generalize to unseen data.



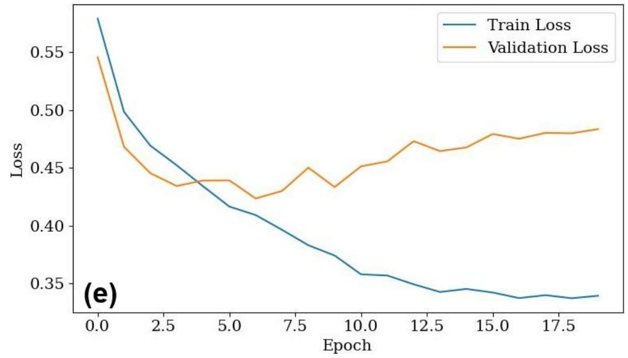


Fig 4. (a) Training and validation loss curves for ConvNeXt; (b) Training and validation loss curves for DenseNet; (c) Training and validation loss curves for EfficientNet; (d) Training and validation loss curves for MobileNetV3; (e) Training and validation loss curves for ResNet-50.

During validation, the comparative analysis of the backbones in Faster R-CNN (Figure 5) demonstrated relevant differences both in the behavior of the loss curves and in the final performance metrics.

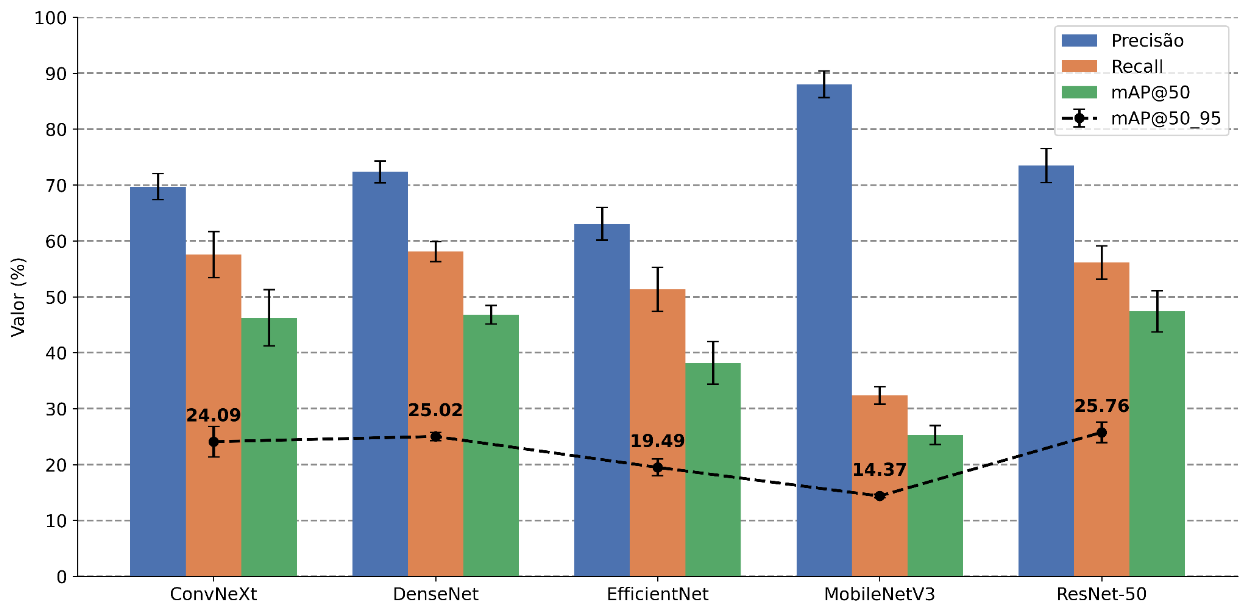


Fig 5. Precision, recall, mAP@50, and mAP@50–95 values of the evaluated backbones.

Overall, ResNet-50, DenseNet, and ConvNeXt presented the best results in terms of mAP@50 and mAP@50-95, with ResNet-50 standing out by achieving the highest average performance (mAP@50 of 47.4% ± 3.7 and mAP@50-95 of 25.76% ± 1.84), despite showing signs of overfitting. DenseNet demonstrated competitive performance and greater stability across folds (mAP@50 of 46.81% ± 1.65 and mAP@50-95 of 25.02% ± 0.74), while ConvNeXt presented results close to the best-performing models (mAP@50 of 46.26% ± 5.03 and mAP@50-95 of 24.09% ± 2.74), although with greater variability.

EfficientNet achieved intermediate performance, even with stable training. MobileNetV3, on the other hand, reached the highest precision (88.03% ± 2.38), but with considerably reduced recall (32.35% ± 1.55), resulting in the lowest mAP values, indicating conservative behavior and limitations in detecting all objects.

Thus, considering the balance between precision, recall, and mAP, ResNet-50 and DenseNet proved to be the most suitable options for the task, while MobileNetV3 may be more appropriate in scenarios with computational constraints and when Type II error is not a critical concern.

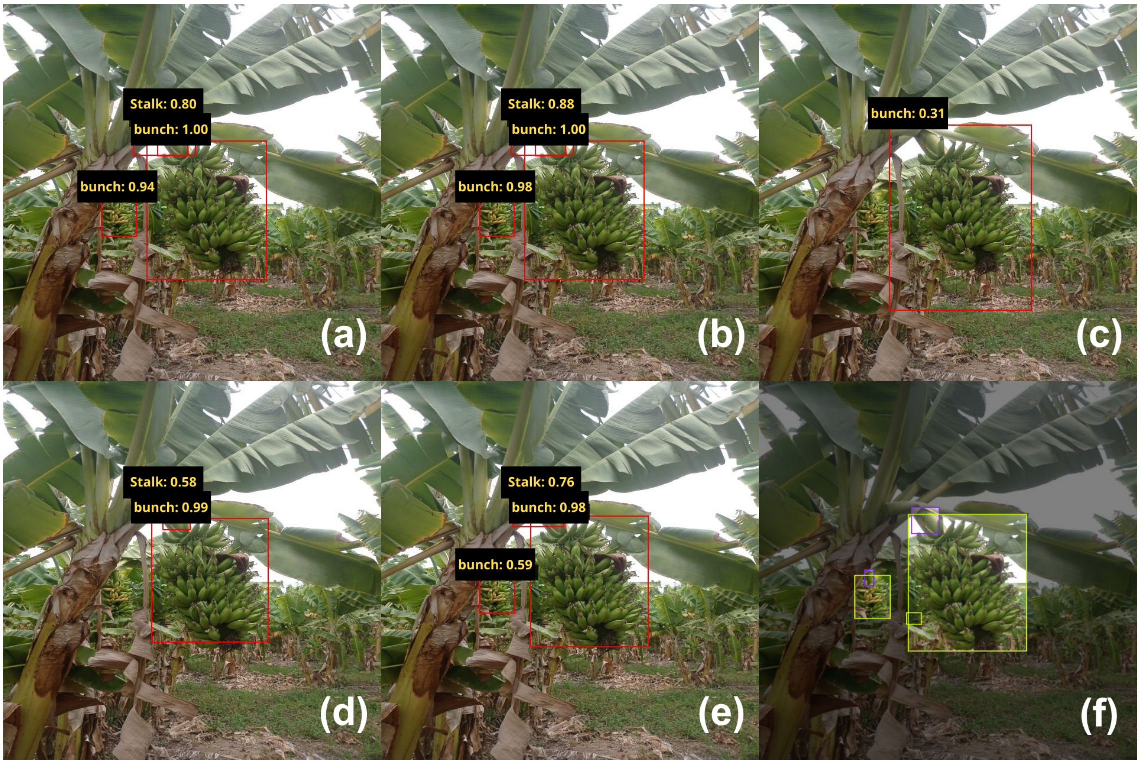


Fig 4. (a) ConvNeXt inference image; (b) DenseNet inference image; (c) EfficientNet inference image; (d) MobileNetV3 inference image; (e) ResNet-50 inference image; (f) reference annotation.

Conclusion

This work investigated the performance of the Faster R-CNN model with different backbones in the detection of bunches and stalks of the Prata Catarina banana plant, aiming to identify the most suitable architecture for precision agriculture applications. Based on a controlled methodology, keeping the main hyperparameters constant and varying exclusively the backbone, it was possible to isolate and analyze the impact of feature extraction on the final performance of the detector.

The results showed that the choice of backbone directly influences both training behavior and detection metrics. Among the evaluated architectures, DenseNet presented a balance between precision, recall, and mAP, standing out as the most suitable alternative for the proposed task because it did not demonstrate a strong tendency toward overfitting, unlike ResNet-50 (although the latter achieved the highest average mAP).

In general, the findings confirm that deeper architectures with greater representational capacity tend to offer better performance in the detection of visually similar and partially occluded structures, such as the stalk and pseudostem. On the other hand, lighter models may be considered in scenarios with computational constraints, provided that a possible reduction in system sensitivity is acceptable.

As a contribution, this study provides a systematic comparative analysis of backbones applied to whole-bunch-level detection, a topic that remains underexplored in the literature. For future work, it is recommended to expand the dataset, investigate additional regularization strategies and hyperparameter fine-tuning, as well as evaluate models under real field conditions and on embedded systems, aiming at their integration into robotic platforms for automated harvesting and monitoring.

Acknowledgments

This study was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil – Finance Code 001. Danielo G. Gomes acknowledges the support of the National Council for Scientific and Technological Development (CNPq) under grant numbers 311845/2022-3 and 403996/2025-2.

References

de Fátima Vidal, M. (2024). FRUTICULTURA (BANANA): v. 9 n. 362, novembro, 2024. *Caderno Setorial ETENE*, 9. https://www.bnb.gov.br/revista/cse/article/view/3104

de Godoy, M. F., Chatkin, J. M., Rodrigues, R. S., Forte, G. C., Marchiori, E., Gavenski, N., Barros, R. C., & Hochhegger, B. (2023). Artificial intelligence to predict the need for mechanical ventilation in cases of severe COVID-19. *Radiologia Brasileira*, *56*(2), 81–85. https://doi.org/10.1590/0100-3984.2022.0049

Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., & Zou, H. (2018). Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, *145*, 3–22. https://doi.org/10.1016/j.isprsjprs.2018.04.003

Fu, L., Yang, Z., Wu, F., Zou, X., Lin, J., Cao, Y., & Duan, J. (2022). YOLO-banana: A lightweight neural network for rapid detection of banana bunches and stalks in the natural environment. *Agronomy (Basel, Switzerland)*, *12*(2), 391. https://doi.org/10.3390/agronomy12020391

Guru, D. S., & N, S. (2025). Banana bunch image and video dataset for variety classification and grading. *Data in Brief*, *60*(111478), 111478. https://doi.org/10.1016/j.dib.2025.111478

Hakim, L., Hendrawan, A., & Khoiriyah, R. (2024). Traffic vehicle detection using Faster R-CNN: A comparative analysis of backbone architectures. *International Journal of Artificial Intelligence and Science*, *1*(1), 50–62. https://doi.org/10.63158/ijais.v1.i1.5

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. Em *arXiv [cs.CV]*. https://doi.org/10.48550/ARXIV.1512.03385

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. Em *arXiv [cs.CV]*. https://doi.org/10.48550/ARXIV.1905.02244

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2016). Densely connected convolutional networks. Em *arXiv* [cs.CV]. https://doi.org/10.48550/ARXIV.1608.06993

Iqbal, A. (2025). Comprehensive AI framework for automated classification, detection, segmentation, and severity estimation of date palm diseases using vision-language models and generative AI. *Frontiers in Plant Science*, *16*(1710188), 1710188. https://doi.org/10.3389/fpls.2025.1710188

Jeevan, P., & Sethi, A. (2024). Which backbone to use: A resource-efficient domain specific comparison for computer vision. Em *arXiv [cs.CV]*. https://doi.org/10.48550/ARXIV.2406.05612

Li, M., Hu, Y., Chen, H., & Li, X. (2025). Feature-enhanced faster R-CNN for small object detection in complex scenes. *IEEE Access: Practical Innovations, Open Solutions*, *13*, 174297–174315. https://doi.org/10.1109/access.2025.3618019

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array (New York, N.Y.)*, *16*(100258), 100258. https://doi.org/10.1016/j.array.2022.100258

Neupane, B., Horanont, T., & Hung, N. D. (2019). Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PloS One*, *14*(10), e0223906. https://doi.org/10.1371/journal.pone.0223906

Nirthika, R., Manivannan, S., Ramanan, A., & Wang, R. (2022). Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing & Applications*, *34*(7), 5321–5347. https://doi.org/10.1007/s00521-022-06953-8

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Em *arXiv [cs.CV]*. https://doi.org/10.48550/ARXIV.1506.01497

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Silva, Y. C., Da, Freitas, E., Diego, G., De, Costa, P. R. S., Davi, P., & Danielo, G. (2025). *Detecção de Cachos de Banana em Campo Usando YOLOv8 e um Novo Dataset da Variedade Prata Catarina*.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional Neural Networks. Em *arXiv [cs.LG]*. https://doi.org/10.48550/ARXIV.1905.11946