Instance Segmentation for Automated Pineapple Counting

*A Deep Learning Approach for Automated Fruit Enumeration, using MASK RCNN

Rodrigo Herera*
Project Manager
Cenfotec Research
University Cenfotec
Costa Rica
roherrera@ucenfotec.ac.cr

Andrés Calvo*
Cenfotec Research
University Cenfotec
Costa Rica
acalvov@ucenfotec.ac.cr

Daniel Guzman*
Cenfotec Research
University Cenfotec
Costa Rica
dguzmanh@ucenfotec.ac.cr

Isabell Galeno*
Cenfotec Research
University Cenfotec
Costa Rica
lgaleanoh@ucenfotec.ac.cr

Jeffry Valverde*
Cenfotec Research
University Cenfotec
Costa Rica
jvalverdef@ucenfotec.ac.cr

Joaquín Pappa*
Cenfotec Research
University Cenfotec
Costa Rica
jpappal@ucenfotec.ac.cr

Nayarit Lopez*
Cenfotec Research
University Cenfotec
Costa Rica
nlopezsa@ucenfotec.ac.cr

***Abstract*—This paper presents an innovative instance segmentation approach for automated pineapple flower counting using drone imagery. The proposed methodology leverages Mask R-CNN trained on a custom dataset captured in Costa Rican pineapple fields to enhance yield estimation and harvest logistics. Our innovation lies in three key contributions: (1) the creation of a proprietary dataset collected under real agricultural conditions, (2) the development of a Python conversion script to transform bounding boxes into polygonal masks for instance segmentation, and (3) the implementation of a robust data augmentation pipeline tailored to high-density crop imagery. Experimental results show that the best-performing model (V4) achieved 99.2% Precision, 82.3% Recall, 90.0% F1-Score, and 82.3% Accuracy, demonstrating robust performance even under variable lighting and occlusion.. These findings show the feasibility of integrating computer vision for practical agricultural applications, reducing manual labor and improving logistic forecasting. Future work includes exploring ensemble models (Mask R-CNN + YOLOv8), adapting Segment Anything (SAM) architectures, and expanding the dataset to include diverse climatic and growth conditions.**

***Index Terms*—Computer vision, precision agriculture, instance segmentation, Mask R-CNN, pineapple counting, data augmentation, agricultural innovation.**

I. Introduction

Pineapple production represents one of the key agricultural exports of Costa Rica, with strict logistical demands and high perishability. For producers and exporters, the lack of accurate visibility on the number of fruits in the field generates significant challenges for harvest planning, cold-chain management, and maritime transportation. Manual counting methods remain the norm in most farms, but these are time-consuming, costly, and error-prone, especially in large-scale plantations.

The ability to estimate yield in advance—specifically during the flowering stage—is crucial for optimizing harvest cycles and export logistics. This stage determines the future production volume months in advance, enabling better coordination of resources, container reservations, and shipment scheduling.

Recent advances in computer vision and deep learning have opened new possibilities for automating agricultural monitoring. Models such as Mask R-CNN, YOLO, and U-Net have demonstrated strong capabilities in object detection and segmentation, yet their application in tropical crops such as pineapple remains limited. In collaboration with AGTIVA, a Costa Rican agro-technology company, this project seeks to close that gap by developing an automated pineapple flower counting system based on instance segmentation using aerial imagery.

This work introduces three primary innovations:

- The creation of a custom, high-resolution pineapple dataset, collected by drones under real agricultural conditions.
- A conversion script that transforms bounding-box annotations into polygonal masks, enabling compatibility with instance segmentation models.
- A data augmentation pipeline specifically tailored to high-density, high-occlusion environments typical of pineapple fields.

The proposed system aims to support precision agriculture

University Cenfotec.

by improving yield estimation accuracy and providing actionable data for logistical decision-making. The outcomes contribute to strengthening the agricultural technology ecosystem in Costa Rica, demonstrating how artificial intelligence can enhance efficiency, sustainability, and competitiveness in tropical fruit production.

II. Related work

Deep learning has significantly advanced agricultural automation, enabling reliable detection, counting, and phenotyping of crops under real-world conditions. In the specific case of pineapple flowering, Hobbs et al. [1], who applied density-estimation techniques to count pineapple inflorescences using aerial imagery, presented one of the earliest large-scale deep learning approaches. Although effective at a global density level, their method struggled with occlusion and lacked instance-level separation.

A complementary effort by the same authors [2] demonstrated that the estimation of flowering density from drone images could be automated, highlighting the potential of computer vision for pineapple phenotyping. However, that work also relied on pixel-level estimation rather than explicit instance detection, which limited its ability to distinguish overlapping or partially visible flowers.

More recent advances include the lightweight model proposed by Yu et al. [3], which uses filter pruning to reduce computational cost while maintaining strong detection performance. Although efficient, the approach remains bounding-box–based, making it less suitable for environments where instance segmentation is required to handle overlapping crowns and dense flowering patterns.

Lin et al. [4], who conducted a systematic review highlighting the effectiveness of instance segmentation in dense-object environments, provides a broader perspective on deep learning for agricultural detection tasks. Their findings emphasize the importance of robust feature extraction and multiscale architectures—principles that align closely with the use of Mask R-CNN in this study.

Finally, generative modeling has emerged as a promising tool for addressing data set scarcity in agriculture. The work of Corrales-Garro et al. [5] demonstrates that synthetic data can help improve model generalization in scenarios where annotation is costly or limited, a challenge that is particularly relevant for pineapple flowering due to occlusion, lighting variability, and limited publicly available datasets.

Despite these advances, prior studies lack high-resolution instance-segmented datasets and specialized augmentation pipelines tailored to pineapple flowering. This research builds on those gaps by introducing a custom dataset, polygon-mask conversion script, and a segmentation model specifically designed for the visual complexity of pineapple fields.

III. Data

The dataset used in this research was developed specifically for the pineapple flowering stage, in collaboration with AGTIVA, a Costa Rican agricultural technology company. The imagery was collected in San Carlos, Alajuela, using a DJI FC6310R drone equipped with a high-resolution RGB sensor, flying at an altitude of 25 meters to capture consistent, large-scale views of pineapple fields.

A. Data Collection

A total of 288 aerial images were acquired under controlled flight paths, ensuring overlap and uniform coverage. Each image had a resolution of 5472×3648 pixels, allowing for precise object-level annotation of pineapple flowers. The data represent multiple days of observation during the flowering phase (approximately 61 days), capturing real environmental variations in lighting and plant growth.

B. Data Annotation and Conversion Script (Innovation 1)

Initial labeling was performed using bounding boxes via the Roboflow platform to test early detection models (YOLOv11, YOLOv12). However, for instance segmentation models such as Mask R-CNN, polygonal masks were required. To address this, we developed a Python conversion script that automatically transforms bounding-box annotations into polygon masks, adapting them for the COCO JSON format. This process significantly reduced manual annotation time and ensured compatibility with instance segmentation architectures. While the conversion is not perfect—particularly for overlapping regions—it provides a scalable annotation approach for high-density agricultural datasets.

C. Dataset Partitioning

The dataset was split following an 80–15–5 ratio for training, validation, and testing, respectively. Each image was subdivided into 4×4 tiles (1368×912 px) to improve detection performance on small and distant objects and to optimize GPU memory usage during training.

D. Data Augmentation Pipeline

Given the limited dataset size and high density of objects (averaging 70–80 pineapples per image), a comprehensive data augmentation pipeline was designed to improve generalization and simulate variable field conditions.

The augmentation strategy consisted of three levels:

Standard Augmentations:

- Horizontal/vertical flips ($p = 0.5$)
- Random rotations ±15°
- Brightness, contrast, and saturation adjustments within [0.8, 1.2]

Enhanced Augmentations:

- Increased flips (p = 0.6) and rotation ±20°
- Stronger lighting variation (scale = 0.15)
- Multiresolution scaling (shortest edge 1000–1400 px)
- Simulated occlusions (random crops 93–90%)

Specialized Configurations:

- For small objects, rotations limited to ±15°
- High resolution emphasized (1000–1300 px)
- Gentle color variation [0.8, 1.2]
- Gaussian blur (p = 0.3) in sparse scenarios

This pipeline allowed the model to learn from diverse spatial and environmental conditions, mimicking real-world lighting, camera tilt, and occlusion patterns typical of tropical crops.

IV. Methodology

The proposed methodology integrates a deep learning-based instance segmentation pipeline using Mask R-CNN to detect and count pineapple flowers in aerial images. Unlike prior works that relied solely on color thresholds or object detection, our approach performs pixel-level segmentation, enabling more accurate identification of overlapping and partially occluded fruits.

Fig. 1. Results of HSV and RGB transformation methods for pineapple flower detection. The examples illustrate how color-based transformations failed to generalize across different lighting conditions.

A. Model Architecture

Mask R-CNN extends the Faster R-CNN framework by adding a segmentation branch that predicts binary masks for each detected object. The architecture used in this work consists of:

- Backbone: ResNet-50 with Feature Pyramid Network (FPN) for multi-scale feature extraction.
- Region Proposal Network (RPN): Generates candidate regions of interest (ROIs) likely to contain flowers.
- ROI Align Layer: Performs bilinear interpolation to ensure precise pixel alignment.
- Segmentation Head: Produces a 28×28 pixel binary mask for each ROI, refined to match object contours.

The combination of detection and segmentation enables the model to differentiate between individual pineapple flowers, even under dense foliage or overlapping structures—conditions where object detectors typically fail.

B. Training Configuration

Training was conducted on a workstation equipped with an NVIDIA RTX 3070 GPU (8 GB), using automatic mixed precision (AMP) for optimized performance.

Parameter	***Configuration***
Learning Rate	0.0002
Batch Size	2
Warmup Iterations	3,000
Total Iterations	60,000
Input Resolution	1200–1400 px
Anchors	[8, 9, 11, 12, 14]
Architecture	ResNet-50 + FPN
Framework	Detectron2 (PyTorch)

This configuration was determined experimentally, balancing computational efficiency with detection accuracy. Lower learning rates (0.0002) were found to yield more stable convergence and prevent overfitting on the limited dataset.

C. Model Versions and Iterative Refinement

Five model variants (V1–V5) were trained using different hyperparameter combinations, learning rates, and anchor scales. Performance was evaluated using Precision, Recall, F1-score, and Accuracy.

The V4 model obtained 99.2% Precision, 82.3% Recall, 90.0% F1-Score, and 82.3% Accuracy, representing the best balance of detection quality and generalization. This improvement can be attributed to:

- Fine-tuned anchor sizes optimized for pineapple flower dimensions.
- Multi-scale image training (1200–1400 px).
- Stable learning rate and extended warmup for smoother optimization.

D. Inference and Counting Logic

During inference, Mask R-CNN outputs segmented instances of pineapple flowers per image tile. Post-processing steps include:

1) Merging detections from subdivided 4×4 tiles into the original image.
2) Filtering overlapping masks using an Intersection-over-Union (IoU) threshold of 0.5.
3) Instance counting, aggregating valid detections per image.

The final output provides both the visual segmentation maps and the quantitative count, supporting practical applications such as yield estimation and harvest planning.

V. Results and Visualization

The Mask R-CNN model demonstrated consistent learning behavior in training iterations, showing stable convergence and clear improvement in both precision and loss metrics.

A. Quantitative Evaluation

The best-performing configuration (Model V4) achieved the following performance:

Metric	Value (%)
Precision	99.2
Recall	82.3
F1-Score	90.0
Accuracy	82.3

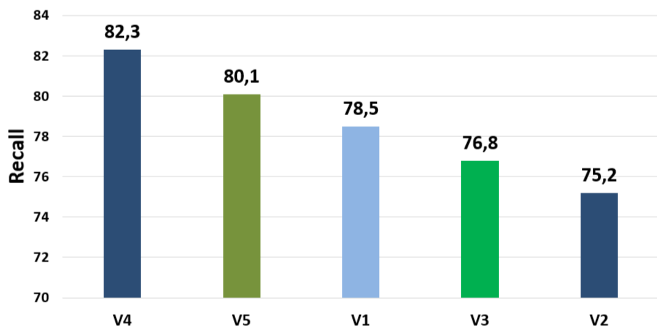

Fig. 2. Metrics Comparison Chart.

Compared to earlier models (V1–V3), V4 improved precision by more than 6 % and reduced overfitting through an extended warm-up and optimized anchors. Although recall remains moderate, this reflects the inherent complexity of detecting small, partially occluded flowers in dense agricultural scenes.

B. Precision–Recall and Loss Curves

Figure 1 illustrates the Precision–Recall curve, showing stable precision up to a recall threshold of 0.2, after which performance decreases due to occlusion and illumination noise. Figure 2 displays the training and validation loss curves, where the model converges after approximately 40 000 iterations, maintaining a consistent gap between both losses, confirming proper generalization without divergence.

C. Visual Analysis of Detections

To further validate performance, visual inspections were carried out on successful and failed detections (Figure 3).

- Successful detections occur under uniform lighting and limited overlapping, producing well-defined polygon masks.
- Failure cases are mainly associated with shadows, sun glare, or overlapping crowns that reduce contrast.
- A heatmap visualization (Figure 4) indicates that the highest detection activations concentrate near the central regions of the tiles, while border areas exhibit decreased confidence scores.

D. Practical Interpretation

The strong performance of the model—characterized by 99.2% Precision, 82.3% Recall, and an F1-Score of 90.0%—provides actionable and reliable detection capability for operational use. This enables producers to estimate flowering density with an approximate error margin of ±10%, sufficient for planning container loads and harvest schedules. While not yet suitable for fully autonomous inventory estimation, the system shows significant potential to replace manual sampling and accelerate decision-making during the export season.

VI. Discussion

The results obtained demonstrate that instance segmentation using Mask R-CNN provides a reliable foundation for automating pineapple flowering detection in real agricultural settings. However, as with most computer vision systems applied to natural environments, several factors influence the overall performance and applicability.

A. Technical Interpretation

The high precision of the model (99.2%) indicates very few false positives, while the Recall (82.3%) reflects the strong ability to detect flowers even under occlusion. The resulting F1-Score of 90.0% confirms a balanced performance suitable for real agricultural environments. This trade-off suggests that the model is conservative—prioritizing accuracy over completeness—which can be acceptable in practical scenarios where false positives are more costly than false negatives (e.g., overestimating yield or container load).

From a computational perspective, the model achieved stable convergence after 60,000 iterations with an average training loss below 0.25, confirming the effectiveness of the chosen learning rate (0.0002) and anchor configuration. These parameters improved generalization compared to baseline versions (V1–V3), demonstrating that hyperparameter tuning is critical for small agricultural datasets.

B. Practical Implications

In operational terms, this performance level allows for an approximate 10% error margin in estimating flowering density—sufficient for planning harvest logistics, optimizing container usage, and improving export scheduling. When integrated with geographic information systems (GIS) or drone automation, this approach can support near–real-time monitoring and reduce manual labor costs associated with traditional field counting.

For producers and exporters, the system offers three key advantages:

1) Faster decision-making: Automated counts are generated within minutes after drone flights.
2) Better logistic forecasting: Early visibility of flower density supports container allocation and shipping route optimization.
3) Sustainability: Reduces waste by aligning harvest volume with logistical capacity, preventing fruit spoilage.

C. Limitations

The primary limitations of this study include:

- Dataset size and variability: Only 288 images were used, all from similar altitudes and environmental conditions.
- Low recall: Many undetected instances in areas of high occlusion or shadow suggest the need for additional training data and more robust architectures.
- Computation cost: High-resolution segmentation requires substantial GPU resources, limiting scalability in low-resource settings.

Nevertheless, these challenges can be mitigated through data diversification, ensemble modeling, and semi-supervised learning, which could expand detection coverage without compromising precision.

VII. Model Mask RCNN

Mask R-CNN (He et al., 2017) extends Faster R-CNN by integrating a pixel-level segmentation branch capable of producing precise object masks. Its multi-task design—combining classification, localization, and segmentation—makes it particularly effective for dense agricultural environments, where occlusion and irregular lighting complicate detection.

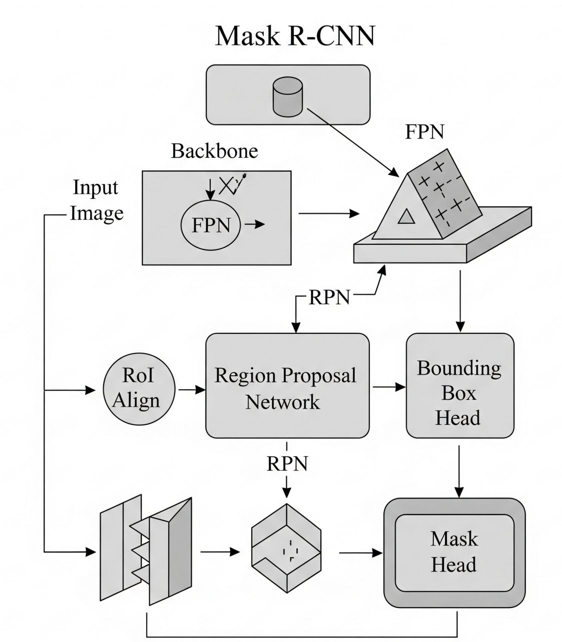

Fig. 3. Overall architecture of Mask R-CNN, including feature extraction backbone, region proposal network (RPN), ROI Align, and mask prediction branch.

Compared to U-Net, which performs semantic segmentation without instance separation, and YOLOv8, which focuses on bounding-box detection, Mask R-CNN achieves superior spatial precision. This capability enables the identification of individual pineapple flowers even under overlapping crowns.

The model's architecture, optimized with a ResNet-50 backbone and Feature Pyramid Network, provides multi-scale feature extraction while maintaining reasonable inference time. Its segmentation masks facilitate accurate counting and can be post-processed to estimate flowering density and spatial distribution across the field.

VIII. Future Work

While the results of this study demonstrate the feasibility of using Mask R-CNN for pineapple flowering detection, several research directions can further enhance accuracy, scalability, and real-world applicability.

A. Ensemble Models

Future experiments will explore ensemble architectures, combining the strengths of Mask R-CNN and YOLOv8 to balance detection speed and segmentation precision. Ensembling models with complementary architectures could mitigate low recall while preserving high precision, particularly in complex visual conditions such as overlapping crowns and heterogeneous lighting.

B. Integration of New Architectures

Recent models such as Segment Anything Model (SAM) and Vision Transformers (ViTs) offer promising capabilities for generalizable segmentation without extensive retraining. Adapting SAM to agricultural imagery could significantly reduce annotation requirements and improve cross-domain performance, while transformer-based models could better capture long-range spatial relationships between plants.

C. Dataset Expansion and Environmental Diversity

The current dataset—although robust—remains limited to images captured under uniform weather and lighting conditions. Expanding it to include different climatic and phenological stages (e.g., morning vs. midday illumination, rainy vs. dry season) would improve the model's generalization and enable domain adaptation across farms and regions.

D. Semi-Supervised and Synthetic Data Generation

Incorporating generative models (e.g., GANs or diffusion-based approaches) to produce synthetic training images could address data scarcity. These techniques can simulate occlusion, illumination variation, and scale diversity, enriching the dataset without extensive manual labeling.

E. Operational Integration

Future deployments may include real-time inference on embedded hardware (e.g., NVIDIA Jetson) for in-field analysis, combined with georeferenced mapping of flowering density to support large-scale decision-making.

IX. Conclusions

This research presented an innovative computer vision pipeline for automated pineapple flower counting using Mask

R-CNN and aerial imagery. Through the development of a custom dataset, a Python-based annotation conversion script, and a data augmentation pipeline tailored to agricultural environments, the proposed approach demonstrated that deep learning can provide reliable yield estimation support for tropical crops.

The best-performing configuration (Model V4) achieved 99.2% Precision, 82.3% Recall, 90.0% F1-Score, and 82.3% Accuracy, demonstrating its effectiveness in real agricultural environments characterized by occlusion and lighting variability. Although recall remains a limiting factor, the model provides actionable data to estimate flowering density with an approximate error margin of ±10%, sufficient for operational planning in harvest and logistics.

The integration of Mask R-CNN with field-collected drone imagery demonstrates the viability of combining artificial intelligence and precision agriculture for real-world applications in Costa Rica. This work establishes a baseline for automated fruit phenotyping in the pineapple industry and provides a foundation for future research on model ensembles, adaptive segmentation architectures, and real-time field deployment.

Ultimately, the study reinforces that data-driven approaches can significantly enhance agricultural decision-making, reduce manual labor, and contribute to more sustainable and efficient production systems.

A. Authors and Affiliations

Rodrigo Herrera[1], **Andrés Calvo**[1], **Daniel Guzman**[1], **Jeffry Valverde**[1], **Joaquin Pappa**[1], **Isabell Galeno**[1], **Nayarit Lopez**[1]
[1]University Cenfotec, Costa Rica

Acknowledgment

We are grateful to the University Cenfotec for its support with fellowships and funding, and also the support provided by agtivita company throughout this project.

References

[1] J. Hobbs, P. Prakash, R. Paull, H. Hovhannisyan, B. Markowicz, and G. Rose, "Large-Scale Counting and Localization of Pineapple Inflorescence Through Deep Density-Estimation," *Front. Plant Sci.*, vol. 11, Art. no. 599705, 2021, doi: 10.3389/fpls.2020.599705.

[2] J. Hobbs and R. Paull and B. Markowicz and G. Rose, "Flowering density estimation from aerial imagery for automated pineapple flower counting," *AI for Social Good Workshop, in Proc. of the Int. Joint Conf. on Artificial Intelligence.*, Yokohama, Japan, 2020.

[3] G. Yu, R. Cai, Y. Luo, M. Hou, and R. Deng, "A-pruning: A lightweight pineapple flower counting network based on filter pruning," *Complex Intell. Syst.*, vol. 10, pp. 2047–2066, 2024, doi: 10.1007/s40747-023-01261-7.

[4] Z. Lin, R. Zhang, M. Wang, and H. Liu,"A Systematic Review of Deep Learning-Based Object Detection in Agriculture," Precision Agriculture Journal, *v*ol. 26, no. 2, pp. 315–343, 2025.

[5] F. Corrales-Garro, L. Zhang, and Y. Li,"Computer Vision Meets Generative Models in Agriculture: Technological Advances, Challenges and Opportunities," Artificial Intelligence in Agriculture, *v*ol. 18, pp. 101–122, 2025.