

Aerial Identification of Amazonian Palms in High-Density Forest Using Deep Learning

Willintong Marin , Ivan F. Mondragon  and Julian D. Colorado 

School of Engineering, Pontificia Universidad Javeriana Bogota, Cra. 7 No. 40-62, Bogota 110221, Colombia; imondragon@javeriana.edu.co (I.F.M.) coloradoj@javeriana.edu.co (J.D.C.)

* Correspondence: willintong.marinr@javeriana.edu.co

Abstract: This paper presents an integrated aerial system for the identification of Amazonian Moriche palm (*Mauritia flexuosa*) in dense forests, by analyzing the UAV-captured RGB imagery using a Mask R-CNN deep learning approach. The model was trained with 478 labeled palms, using the transfer learning technique based on the well-known MS COCO framework. Comprehensive in-field experiments were conducted in dense forests, yielding a precision identification of 98%. The proposed model is fully automatic and suitable for the identification and inventory of this species above 60 m, under complex climate and soil conditions.

Keywords: Amazon Palm Moriche (*Mauritia flexuosa*); dense forests; precision agriculture; Mask R-CNN model; deep learning; convolutional neural networks



Citation: Marin, W.; Mondragon, I.F.; Colorado, J.D. Aerial Identification of Amazonian Palms in High-Density Forest Using Deep Learning. *Forests* **2022**, *13*, 655. <https://doi.org/10.3390/f13050655>

Academic Editors: Yunchao Tang, Liang Gong, Lufeng Luo, Junfeng Gao, Ya Xiong, Chao Chen, Hao Gan and Huaibo Song

Received: 28 February 2022

Accepted: 15 April 2022

Published: 23 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mauritia flexuosa L.f. (Arecaceae) is a palm tree that occurs in wetlands called buritizais in Brazil, aguajales in Peru, morichales in Venezuela, and moritales in Ecuador [1,2], which are located in riverbanks or dense forests in Amazonia [3,4] and headwaters or margins of the headwaters of gallery forests in South America savannas [5,6]. It is an endemic and widely distributed palm of South America [2,7]. Figure 1 depicts the palm growing in dense forest areas. These palms can measure up to 40 m in height. They have a spherical crown and palmate leaves that are 2.5 m wide and up to 4.5 m in length [8].

For the economy of the Amazonian region, the palms represent a relevant source of income, allowing the indigenous communities to commercialize the Moriche fruits and several derivative products for the national and international industry, due to the highly nutritional bioactive compounds [8,9]. Additionally, the significant growth in the commercialization of palm-derived products is increasing the industrial valorization for this crops. Unsaturated fatty acids are extracted from the fruit and the shell, and bioproducts are considered to have great potential for the health industry. Furthermore, this variety is not only important for the human consumption, but also for other species, playing an important role for the conservation of several ecosystems in the neotropics [10].

Currently, the Amazonian Scientific Research Institute (SINCHI) inventories the aforementioned specie in a manual fashion [11], resulting in a labor-intensive and time-demanding task. In this regard, the use of computer vision methods enables more accurate and efficient approaches that facilitate the identification of the palms, by the means of unmanned aerial vehicles (UAVs) equipped with the necessary sensors.

Deep-learning approaches for the geo-location of objects using computer vision have gained significant traction for applications in crop remote sensing, plant phenotyping, and forest monitoring and conservation. An important body of work recently focused on novel methods based on R-CNN region proposals [12], Fast R-CNN [13], Faster R-CNN [14], and Mask R-CNN [15], which have rapidly evolved to become the state-of-the-art for crop feature detection and extraction in a wide range of agricultural scenarios. Additionally,

a high demand from the agriculture sector requires methods for the segmentation and identification of plant and leaf features to improve crop yield and management [16–19]. Convolutional neural networks (CNN) have become the fundamental basis for object detection algorithms in several fields of research [20]. In [21], convolutional neural networks (CNN) were used for the UAV-driven identification of conifer seedlings during the reforestation of areas deforested for mining. They reported a precision of 81% with Faster R-CNN, which was the best of the three DL models implemented. About 4000 tags in images were captured by the UAV; the flights were executed at 5 m height and used data augmentation.



Figure 1. The Amazonian Moriche palm (*Mauritia flexuosa*) specie growing in dense flooded forests. Its morphology and fruits are also shown.

Other works reported in [22,23] tackled the characterization of external perturbations for aerial crop imagery, such as changes in ambient illumination, soil noise, and image perspective corrections. In this regard, CNN models were integrated with an image pre-processing stage to counteract the external effects, which was helped by combining imagery showing different wavelengths (RGB and NIR) for extracting vegetation-index features that are not affected by the aforementioned external factors.

Alternatives besides CNN models, such as transfer learning, have been proposed by [24], enabling an identification precision of 77.30% in vineyard crops. In [25,26], the authors demonstrated that combining transfer learning methods with CNN models prior to the training stage (AlexNet and VGG-16) increased the precision by up to 90% in terms of fruit maturity detection in vineyards. Similarly, the work in [27] presented a method using Mask R-CNN to identify apples in real-time, obtaining a precision of 97.3% and a recall of 95.7%.

In [28], a novel CNN model was applied for citrus tree identification in dense orchards. The proposed method outperforms Faster R-CNN and RetinaNet in terms of precision, by training with multispectral imagery captured with a UAV. Additionally, in [29] Mask R-CNN was used for the identification and segmentation of oranges, using both RGB and the HSV color space combined. It achieved precision of 89% to 97%. Following a similar approach, in [30], multimodal RGB and NIR images were used for training a Faster R-CNN detector via transfer learning, achieving 83% precision for bell pepper detection. In [31] Faster R-CNN was used to identify passionfruit by applying the spacial pyramid matching (SPM) method as a feature extractor, in which a SVM classifier handled the maturity state identification. Its precision was 92.7%.

Other works reported in [32,33] also used Mask R-CNN to detect panicles in rice crops, by processing UAV-captured imagery at the canopy level. Several images at different UAV flight altitudes were combined in order to improve on the detection precision. In this regard, feature-extraction methods are key to defining proper input training data for machine learning models. In [34], the well-known YOLO algorithm was compared against the Mask R-CNN detector, by outperforming the former during grapefruit identification

tasks. Furthermore, several works [35–37] also demonstrated that fusing both RGB and NIR data improves the precision in fruit identification labors, even when the fruits overlap with each other. In this regard, the Mask R-CNN method was used in combination with the DBSCAN spatial clustering algorithm.

To the best of the authors' knowledge, two works were found in the specialized literature concerning the study of the Moriche fruit in the Amazon region. In [38,39], deep-learning models based on CNN were used in order to quantify and predict how the Moriche palms grow in the Peruvian Amazon. To that end, training data were collected during four consecutive years. The random forest algorithm was used for the proper labeling within orthomosaics models.

For the application at hand, the proposed system architecture is presented in Figure 2. Here, we propose the integration of Mask R-CNN models for detecting Moriche palms in high-dense forests located at the Colombian Amazon region. We used the Phantom 4 Multispectral UAV (P4M) manufactured by DJI to capture the images. Our AI-driven UAV solution allows for the aerial identification and counting of the palms, being a fundamental input for the inventory research conducted in the tropical forest zone of the Amazon.

Considering the development of different CNN-based algorithms, in this paper we focus on the region-based models, such as Faster R-CNN, which has a region proposal network (RPN) for segmentation. Additionally, Mask R-CNN specializes in the segmentation of instances, based on Faster R-CNN, by using the input image as a feature map to apply the region proposal network (RPN), thereby performing a process of alignment of features of the regions of interest (RoI) [15].

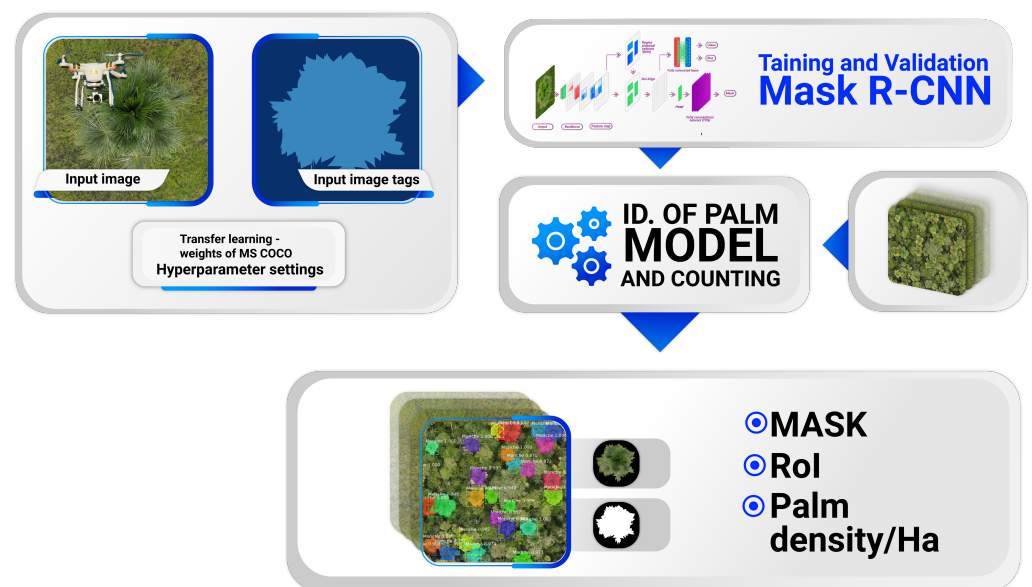


Figure 2. System architecture. Image capture protocol, data and label preparation, hyperparameter configuration, model training, validation, and testing.

In this paper, our contributions are twofold: (i) we propose an integrated AI-driven method to identify the Moriche palm within dense forests of the Amazon, by assembling a comprehensive dataset with UAV-captured imagery that has been properly labeled, and (ii) the aforementioned dataset was validated in the field, by training a Mask R-CNN deep-learning method to identify Moriche palms with a precision of 98%.

2. Materials and Methods

A CNN is mainly composed of two parts: one stage that can be used as a feature extractor that generates a feature map based on the input data and a second stage that can be used as a layer classifier. The resulting feature map is lighter and easier to process. Figure 3 shows the general CNN architecture. Equation (1) describes the convolution

filters, Equation (2) the ReLU (Rectified Linear Unit) nonlinearity activation function, and Equation (3) the Max pooling or grouping of maximum pixels for each defined window, which reduces the feature map by applying the corresponding kernel or convolution filter.

$$S_{i,j}(I \times k)_{i,j} = \sum_m \sum_n I_{i,j} \times K_{i-m,j-n} \quad (1)$$

$$ReLU(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

$$Max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

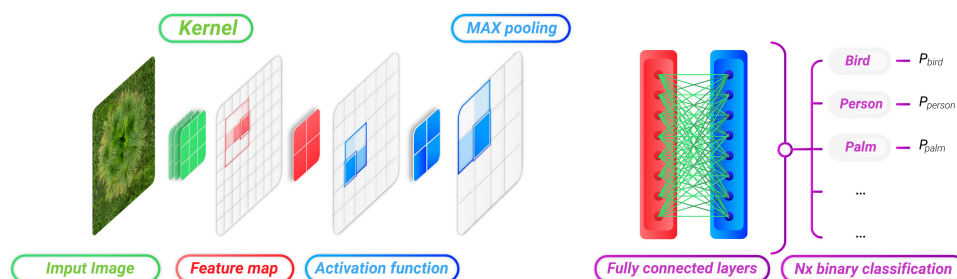


Figure 3. Convolutional neural network.

Our application requires the automatic identification of Moriche palms. The Mask R-CNN algorithm was implemented, as previously described by Figure 3. Additionally, the proposed system has been validated and tested by conducting in-field experiments at San Jose del Guaviare, located in the Amazon region of Colombia, as shown in Figure 4. At this location, the Moriche palms grow in flooded plots within dense forests with limited human access.

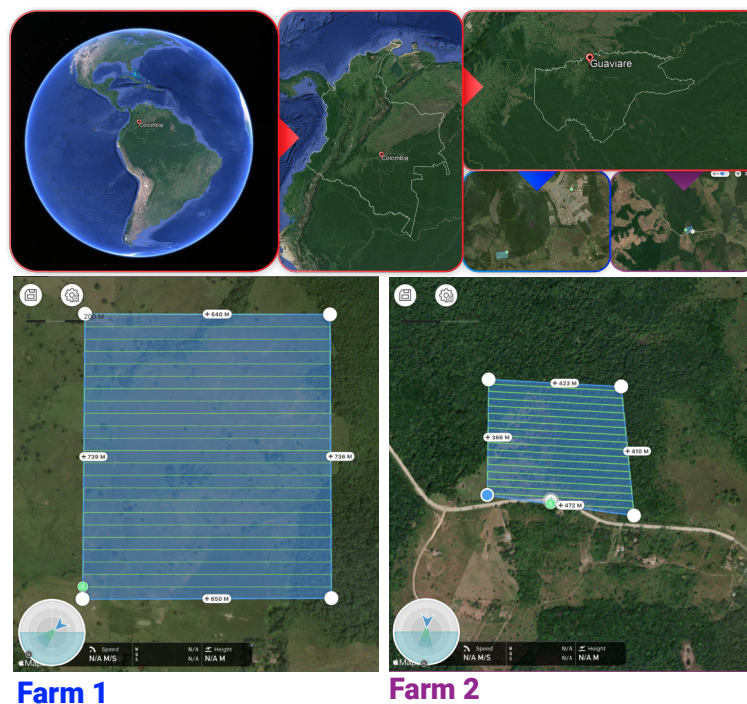


Figure 4. Geo-location of the Moriche palm plots. South America, Colombia, Guaviare Department, San José del Guaviare Municipality.

2.1. Experimental Protocol

Two Moriche farms were characterized in this study. The former was enclosed in a polygon of 47 ha and the latter in a polygon of 17 ha. Overall, a training and validation dataset of 478 palms with 132 labeled imagery was assembled. UAV-based imagery was captured at an altitude of 60 m above the ground. The assembled dataset for training and validation is the result of several trial flights over the two polygons, in order to consider different weather conditions, during the months of October and November 2020. The UAV (P4 DJI Phantom) comes with an integrated multispectral camera of 2 megapixels for each band. Based on the flying altitude, the UAV captures the corresponding images, avoiding overlapping. The data resolution corresponds to 3.2 cm/pixel. Only the RGB bands were used for this research. The other images were kept for other purposes. The images were not preprocessed, nor were orthomosaics constructed.

On the other hand, two sets of images with complex backgrounds were prepared to be used for model testing. Set 1 corresponds to a flooded zone which casts multiple shadows, and set 2 corresponds to a dense forest, in which the palms are overlapped or occluded with each other, and accompanied by other types of palm and tree species. Additionally, a testing dataset was assembled by capturing imagery within a range of UAV altitudes (40–150 m above the ground), with the aim of comparing extracted features at different field of view scales. All images (around 500) for testing were captured on different flights, at different times of day and in different weather conditions.

The annotation process was performed by using a VGG Image Annotator approach, which is a semi-automatic and interactive tool for labeling images, using the polygon shape for that purpose [40]. Overall, we obtained 478 labels for both farms, a sufficient number to train our model, according to the experimental results reported in the forthcoming Section 3. It is important to highlight that the Mask R-CNN model was trained using the Keras 2.3.1 API for the Tensorflow framework, running on a 10th generation Intel core i7 processor with 16 GB of RAM and 6 GB NVidia Geforce RTX 2060 CPU.

2.2. Hyperparameters and Input Data

The ResNet 50 architecture was used as a feature extractor [41]. Image dimensions were set to 512×512 , having a learning rate of 0.001 with 500 steps per epoch. Overall, 20 epochs were required, and minimum confidence thresholds were 20, 60 and 85. Transfer learning was used to train the CNN model with few labeled images, resulting in less data collection time and computational costs associated with the model training. Additionally, pre-trained weights from the MS COCO dataset were used for this task [42]. The set of images for training and validation was split into 80% for training (354 labels in 106 images) and 20% for validation (124 labels in 26 images).

2.3. Palm Segmentation and Identification

The architecture presented in Figure 3 enables the detection, localization, and classification of Moriche palms, by applying a semantic segmentation of instances that generate a mask with the region of interest (RoI). Only two clusters or classes were needed: the palm and the background. A comprehensive ground truth was assembled by combining imagery from dispersed palms plots that can be identified with ease and precision, and also with imagery from palms located at dense forests. The model is able to help in the inventory of the species.

2.4. Performance Metrics

The model training was evaluated based on precision, recall, *F1*-score and the intersection over Union (*IoU*). The precision (*P*) is defined in Equation (4) as an indicator of correct predictions according to the true positives (T_p) among all identified instances, denoting quality. Recall (*R*) in Equation (5) is an indicator of the correct predictions among all the elements contained in the ground truth (T_p plus (F_n) false negatives). The *F1*-score

in Equation (6) is the harmonic mean between precision and recall, and the *IoU* metric in Equation (7) is the sum of expected pixels over the sum of pixels predicted by the model.

$$Precision = \frac{T_p}{T_p + F_p} \quad (4)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (5)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (6)$$

$$IoU = \frac{Areaofoverlap}{AreaofUnion} = \frac{T_p}{T_p + F_p + F_n} \quad (7)$$

3. Results

To determine the corresponding density of palms per hectare, the proposed model must identify, classify, and count the detected Moriche palms. To this purpose, several Mask R-CNN hyperparameters were tuned during training. In this regard, Table 1 and Figure 5 show the error curves resulting from training the algorithm for 10 to 100 epochs. The learning rate was set to 0.001, since higher rates generated deficient learning, whereas lower rates resulted in over-fitting.

Table 1. Training and validation loss for Mask R-CNN.

Epoch	Loss Training	Loss Val
10	21.75	48.69
20	14.11	10.5
40	10.65	8.13
60	9.09	14.54
100	7.64	9.69

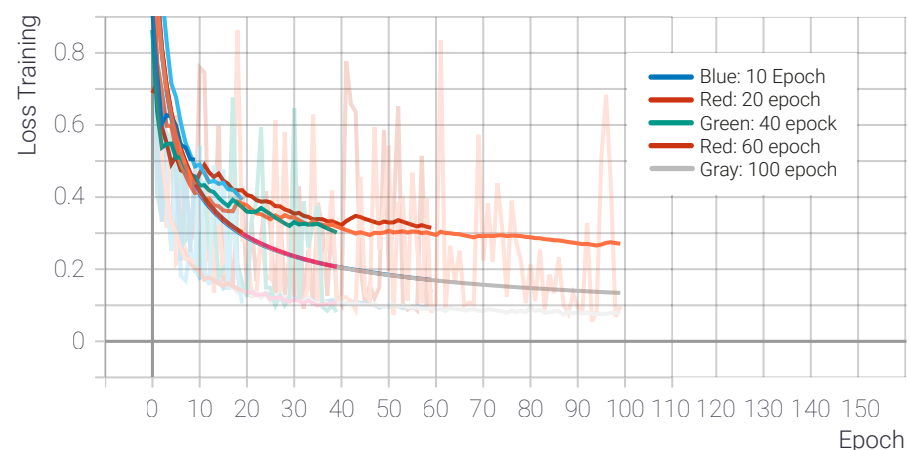


Figure 5. Training and validation loss curves for Mask R-CNN.

According to Table 1, the lowest error for training was obtained at 100 epochs. Nonetheless, the lowest error for validation was obtained at only 40 epochs of iterations. Table 2 shows the numerical results for testing each model by varying the confidence threshold among 20, 60, and 85. As mentioned, the validation set contained 124 palms labeled in 26 images.

Table 2. Validation of models with varying confidence thresholds. Quantitative results for palm identification and counting at different epoch iterations.

Epoch	Confidence 20		Confidence 60		Confidence 85	
	Counting	IoU	Counting	IoU	Counting	IoU
10	125	81.76	124	81.91	118	83.56
20	126	84.33	125	84.94	121	85.60
40	126	85.81	124	85.63	120	85.17
60	119	87.23	118	87.38	118	87.38
100	119	88.55	119	88.55	119	88.55

Table 2 shows quantitative results for palm identification and counting at different epoch iterations. Accurate results were obtained from those models trained with 10, 20, or 40 epochs, and confidence thresholds of 20 and 60. Additionally, the IoU tended to improve by increasing the training epochs; however, the improvement was not significant, being only about 5%.

Table 3 shows the results of using the trained model with 20 epochs on the validation set. With this setup, we expect to obtain an intermediate IoU in order to generalize the implementation of the model to other types of input data, since upcoming work will be oriented towards fruit maturity prediction. As observed from the reported data, all tests resulted in outstanding metrics, which validates the proposed model.

Table 3. Metrics on validation data.

Confidence	Precision	Recall	F1 Score	Count Model
C-20	0.9683	1	0.9839	126
C-60	0.9680	1	0.9837	125
C-85	0.9917	1	0.9958	121

The two sets of images available for testing were implemented. Tables 4 and 5 show the precision, recall, and F1-score obtained. According to the results, there is no parity between precision and recall; however, the best performance can be identified by the F1-score of 91.52%, a recall of 99.58%, and a precision of 84.67%, which were achieved with a threshold of 20. The best performance on set 2 was given by a confidence threshold of 60, with an F1-score of 91.19%, a recall of 88.22%, and precision of 94.37%. With set 2, the recall was quite low compared to the set 1 results and those for the validation set. It can also be interpreted that the confidence threshold had a significant impact depending on the background of the images, yielding an average F1-score of 90.

Table 4. Metrics on data from set 1, flooded bottom zone.

Set 1	Precision	Recall	F1 Score
C-20	0.8467	0.9958	0.9152
C-60	0.8462	0.9665	0.9023
C-85	0.8750	0.9363	0.9046

Table 5. Metrics on data from set 2, dense forest zone.

Set 1	Precision	Recall	F1 Score
C-20	0.9437	0.8662	0.9033
C-60	0.9437	0.8822	0.9119
C-85	0.9712	0.8172	0.8876

Finally, in Figure 6 we characterize how the Moriche palm identification process performs depending on the UAV altitude. The numerical data report optimal classifications

in a range between 40 and 90 m, maintaining proper precision and recall metrics, and Figure 7 shows a sample of the model classification results in different contexts, dense forest, scattered palms, flooded ground, and dense forest without the presence of the palm under study.

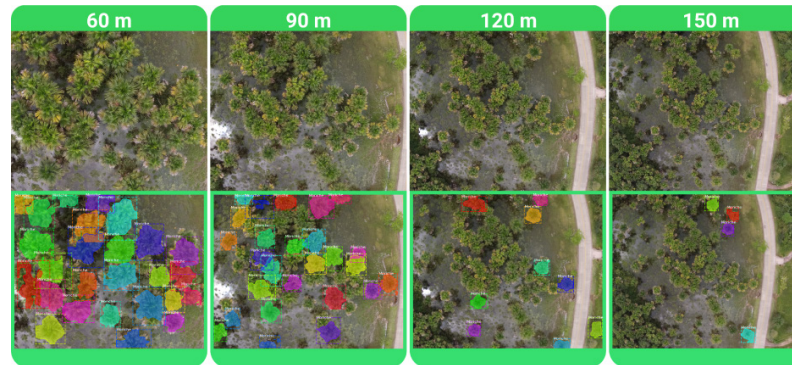


Figure 6. The Moriche palm identification process at different UAV altitudes: 60, 90, 120, and 150 m.

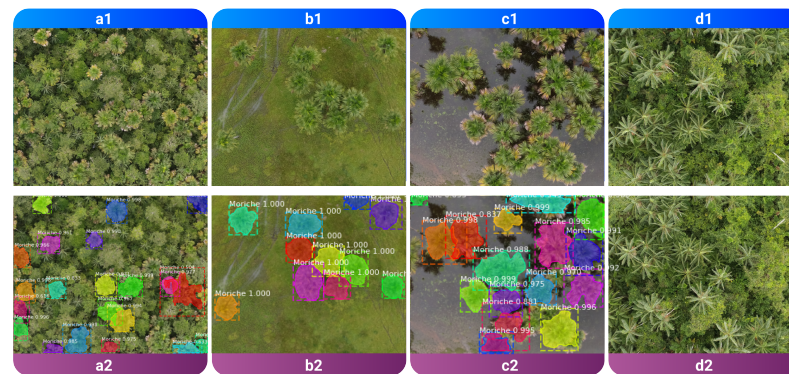


Figure 7. Moriche palm identification. Line one (a1–d1) shows the different contexts, dense forest, scattered palms, flooded ground, and dense forest without the presence of the Moriche palm. Line two (a2–d2) shows the model's results.

4. Conclusions and Discussion

In this work, we presented a deep-learning approach based on the Mask R-CNN mathematical model aimed at the identification of Amazonian Moriche palms (*Mauritia flexuosa*) in dense forests, by processing UAV-captured RGB imagery that enabled the training of the model. A structured dataset was assembled with 132 images containing 478 labeled palms. It is important to highlight that due to the use of the transfer learning technique, the model achieved an overall precision $>96\%$ with a reduced input dataset, i.e., without the need for adding data augmentation algorithms into the model.

By conducting several in-field experiments in dense forests located at the Colombian Amazon region, the proposed method was tested and validated, allowing us to identify and count the number of Moriche palms per image, as shown in Figure 6. In this regard, the proposed model is fully automatic and suitable for the identification and inventory of this specie under complex climate and soil conditions, such as the dense forest of the Amazon, since the model was trained to consider different weather conditions, and image segmentation ensured adequate extraction of the RoI. As reported in Tables 4 and 5, two scenarios were considered: scattered palms in flooded terrains and palms in highly-dense vegetation. The former imposed several challenges to the proposed model, since water reflections tend to add noise and uncertainty during training. Nonetheless, the model was able to reduce external noise, mainly caused by shadow reflections.

As future work, it would be interesting to explore graph-based networks in combination with the Mask R-CNN deep-learning approach proposed herein, since graph

techniques allow for the extraction of rich feature information that can be associated with the phenological stage of the Moriche, enabling an accurate estimation of the fruit maturity and other health-related information of the palms.

Author Contributions: Conceptualization, W.M., J.D.C. and I.F.M.; methodology, W.M., J.D.C. and I.F.M.; software, W.M.; validation, W.M.; formal analysis and investigation, W.M., J.D.C. and I.F.M.; data curation, W.M., J.D.C. and I.F.M.; writing—original draft preparation, W.M.; writing—review and editing, J.D.C. and I.F.M.; supervision, J.D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the Department of Guaviare and the Ministry of Science, Technology and Innovation (MinCiencias) of Colombia. Call number 752 of 2016. High-level human capital training for the Department of Guaviare, 2016—II cohort, to finance doctoral studies. Additionally, it was partly funded by the OMICAS program: "Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (Infraestructura y validación en Arroz y Caña de Azúcar)," anchored at the Pontificia Universidad Javeriana in Cali and funded within the Colombian Scientific Ecosystem by The World Bank, the Colombian Ministry of Science, Technology and Innovation, the Colombian Ministry of Education, the Colombian Ministry of Industry and Tourism, and ICETEX, under grant ID: FP44842-217-2018 and OMICAS Award ID: 792-61187.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the SINCHI Amazon Institute for its support of the research. Likewise, the company Smart Life Technology SAS is thanked for the UAV equipment for capturing the images.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kahn, F. Palms as key swamp forest resources in Amazonia. *For. Ecol. Manag.* **1991**, *38*, 133–142. [\[CrossRef\]](#)
2. Navarro-Cruz, A.R.; Lazcano-Hernández, M.; Vera-López, O.; Kammar-García, A.; Segura-Badilla, O.; Aguilar-Alonso, P.; Pérez-Fernández, M.S. *Mauritia flexuosa* L. f. In *Fruits of the Brazilian Cerrado*; Springer: Cham, Switzerland, 2021; pp. 79–98. [\[CrossRef\]](#)
3. Galeano, A.; Urrego, L.E.; Sánchez, M.; Peñuela, M.C. Environmental drivers for regeneration of *Mauritia flexuosa* L.f. in Colombian Amazonian swamp forest. *Aquat. Bot.* **2015**, *123*, 47–53. [\[CrossRef\]](#)
4. Mendes, F.N.; de Melo Valente, R.; Rêgo, M.M.C.; Esposito, M.C. The floral biology and reproductive system of *Mauritia flexuosa* (Arecaceae) in a restinga environment in northeastern Brazil. *Brittonia* **2017**, *69*, 11–25. [\[CrossRef\]](#)
5. Furley, P.A. Tropical Forests of the Lowlands. In *The Physical Geography of South America*; Oxford University Press: Oxford, UK, 2007. [\[CrossRef\]](#)
6. Moreira, S.N.; Eisenlohr, P.V.; Pott, A.; Pott, V.J.; Oliveira-Filho, A.T. Similar vegetation structure in protected and non-protected wetlands in Central Brazil: Conservation significance. *Environ. Conserv.* **2014**, *42*, 356–362. [\[CrossRef\]](#)
7. Maciel, E.A.; Martins, F.R. Rarity patterns and the conservation status of tree species in South American savannas. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2021**, *285*, 151942. [\[CrossRef\]](#)
8. Hernández, M.S.; Castro, S.Y.; Giraldo, B.; Barrera, J. *Seje, Moriche, Asaí: Palmas Amazónicas con Potencial*, 1st ed.; Instituto Amazónico de Investigaciones Científicas SINCHI-Equilátero Diseño Impreso: Bogotá, Colombia, 2018; p. 123.
9. Quintero-Angel, M.; Martínez-Girón, J.; Orjuela-Salazar, S. Agroindustrial valorization of the pulp and peel, seed, flour, and oil of moriche (*Mauritia flexuosa*) from the Bitá River, Colombia: A potential source of essential fatty acids. *Biomass Convers. Biorefin.* **2022**, *1*, 1–9. [\[CrossRef\]](#)
10. van der Hoek, Y.; Solas, S.Á.; Peñuela, M.C. The palm *Mauritia flexuosa*, a keystone plant resource on multiple fronts. *Biodivers. Conserv.* **2019**, *28*, 539–551. [\[CrossRef\]](#)
11. Cárdenas López, D.; Arias G., J.C. *Manual de Identificación, Selección y Evaluación de Oferta de Productos Forestales no Maderables*; Instituto Amazónico de Investigaciones Científicas "SINCHI": Bogotá, Colombia, 2007.
12. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [\[CrossRef\]](#)
13. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 1440–1448. [\[CrossRef\]](#)
14. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)

15. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [\[CrossRef\]](#)
16. Coelho Eugenio, F.; Badin, T.L.; Fernandes, P.; Mallmann, C.L.; Schons, C.; Schuh, M.S.; Soares Pereira, R.; Fantinel, R.A.; Pereira da Silva, S.D. Remotely Piloted Aircraft Systems (RPAS) and machine learning: A review in the context of forest science. *Int. J. Remote Sens.* **2021**, *42*, 8207–8235. [\[CrossRef\]](#)
17. Orozco, Ó.A.; Llano Ramírez, G. Sistemas de Información enfocados en tecnologías de agricultura de precisión y aplicables a la caña de azúcar, una revisión. *Rev. Ing. Univ. Medellín* **2016**, *15*, 103–124. [\[CrossRef\]](#)
18. Ponce-Corona, E.; Guadalupe Sánchez, M.; Fajardo-Delgado, D.; Acevedo-Juárez, B.; De-La-Torre, M.; Avila-George, H.; Castro, W. A systematic review of the literature focused on the use of unmanned aerial vehicles during the vegetation detection process. *RISTI Rev. Iber. Sist. Tecnol. Inf.* **2020**, *2020*, 82–101. [\[CrossRef\]](#)
19. Urbahs, A.; Jonaite, I. Features of the use of unmanned aerial vehicles for agriculture applications. *Aviation* **2013**, *17*, 170–175. [\[CrossRef\]](#)
20. Naranjo-Torres, J.; Mora, M.; Hernández-García, R.; Barrientos, R.J.; Fredes, C.; Valenzuela, A. A review of convolutional neural network applied to fruit image processing. *Appl. Sci.* **2020**, *10*, 3443. [\[CrossRef\]](#)
21. Fromm, M.; Schubert, M.; Castilla, G.; Linke, J.; McDermid, G. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Remote Sens.* **2019**, *11*, 2585. [\[CrossRef\]](#)
22. Rzanny, M.; Seeland, M.; Wäldchen, J.; Mäder, P. Acquiring and preprocessing leaf images for automated plant identification: Understanding the tradeoff between effort and information gain. *Plant Methods* **2017**, *13*, 97. [\[CrossRef\]](#)
23. Kerkech, M.; Hafiane, A.; Canals, R. Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images. *Comput. Electron. Agric.* **2018**, *155*, 237–243. [\[CrossRef\]](#)
24. Pereira, C.S.; Morais, R.; Reis, M.J. Deep learning techniques for grape plant species identification in natural images. *Sensors* **2019**, *19*, 4850. [\[CrossRef\]](#)
25. Barré, P.; Stöver, B.C.; Müller, K.F.; Steinhage, V. LeafNet: A computer vision system for automatic plant species identification. *Ecol. Inform.* **2017**, *40*, 50–56. [\[CrossRef\]](#)
26. Altaheri, H.; Alsulaiman, M.; Muhammad, G.; Amin, S.U.; Bencherif, M.; Mekhtiche, M. Date fruit dataset for intelligent harvesting. *Data Brief* **2019**, *26*, 104514. [\[CrossRef\]](#)
27. Jia, W.; Tian, Y.; Luo, R.; Zhang, Z.; Lian, J.; Zheng, Y. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Comput. Electron. Agric.* **2020**, *172*, 105380. [\[CrossRef\]](#)
28. Osco, L.P.; de Arruda, M.d.S.; Marcato Junior, J.; da Silva, N.B.; Ramos, A.P.M.; Moryia, É.A.S.; Imai, N.N.; Pereira, D.R.; Creste, J.E.; Matsubara, E.T.; et al. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 97–106. [\[CrossRef\]](#)
29. Ganesh, P.; Volle, K.; Burks, T.F.; Mehta, S.S. Deep Orange: Mask R-CNN based Orange Detection and Segmentation. *IFAC-PapersOnLine* **2019**, *52*, 70–75. [\[CrossRef\]](#)
30. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors* **2016**, *16*, 1222. [\[CrossRef\]](#)
31. Tu, S.; Xue, Y.; Zheng, C.; Qi, Y.; Wan, H.; Mao, L. Detection of passion fruits and maturity classification using Red-Green-Blue Depth images. *Biosyst. Eng.* **2018**, *175*, 156–167. [\[CrossRef\]](#)
32. Lyu, S.; Noguchi, N.; Ospina, R.; Kishima, Y. Development of phenotyping system using low altitude UAV imagery and deep learning. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 207–215. [\[CrossRef\]](#)
33. Neupane, B.; Horanont, T.; Hung, N.D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLoS ONE* **2019**, *14*, e0223906. [\[CrossRef\]](#)
34. Santos, T.T.; de Souza, L.L.; dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron. Agric.* **2020**, *170*, 105247. [\[CrossRef\]](#)
35. Liu, Z.; Wu, J.; Fu, L.; Majeed, Y.; Feng, Y.; Li, R.; Cui, Y. Improved Kiwifruit Detection Using Pre-Trained VGG16 with RGB and NIR Information Fusion. *IEEE Access* **2020**, *8*, 2327–2336. [\[CrossRef\]](#)
36. Ge, Y.; Xiong, Y.; From, P.J. Instance Segmentation and Localization of Strawberries in Farm Conditions for Automatic Fruit Harvesting. *IFAC-PapersOnLine* **2019**, *52*, 294–299. [\[CrossRef\]](#)
37. Liu, X.; Hu, C.; Li, P. Automatic segmentation of overlapped poplar seedling leaves combining Mask R-CNN and DBSCAN. *Comput. Electron. Agric.* **2020**, *178*, 105753. [\[CrossRef\]](#)
38. Morales, G.; Kemper, G.; Sevillano, G.; Arteaga, D.; Ortega, I.; Telles, J. Automatic Segmentation of *Mauritia flexuosa* in Unmanned Aerial Vehicle (UAV) Imagery Using Deep Learning. *Forests* **2018**, *9*, 736. [\[CrossRef\]](#)
39. Casapia, X.T.; Falen, L.; Bartholomeus, H.; Cárdenas, R.; Flores, G.; Herold, M.; Coronado, E.N.H.; Baker, T.R. Identifying and Quantifying the Abundance of Economically Important Palms in Tropical Moist Forest Using UAV Imagery. *Remote Sens.* **2019**, *12*, 9. [\[CrossRef\]](#)
40. Dutta, A.; Zisserman, A. The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19), Nice, France, 21–25 October 2019; ACM: New York, NY, USA, 2019; 4p. [\[CrossRef\]](#)

41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
42. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2014; Volume 8693, pp. 740–755. [[CrossRef](#)]