



ISSN:1984-2295

Revista Brasileira de Geografia Física

Homepage: <https://periodicos.ufpe.br/revistas/rbgf>



Identificação de Palmeiras (*Arecaceae*) Nativas em Áreas de floresta tropical baseado em Rede Neural Convolutacional com imagens de VANT

Airton Gaio Júnior¹, Rodrigo Pinheiro Ribas²

¹ PhD Candidate in Territorial Planning and Socioenvironmental Planning Graduate Program, Department of Geography, Santa Catarina State University, Av. Madre Benvenuta, 2007, Itacorubi, Florianópolis - SC, 88035-901, Brazil. (48) 3664-8548. Corresponding author, e-mail: airton.gao@edu.udesc.br. ² Professor in Territorial Planning and Socioenvironmental Planning Graduate Program, Department of Geography, Santa Catarina State University, Av. Madre Benvenuta, 2007, Itacorubi, Florianópolis - SC, 88035-901, Brazil. (48) 3664-8548. rodrigo.ribas@udesc.br.

Artigo recebido em 03/04/2023 e aceito em 06/05/2023

RESUMO

As palmeiras são componentes importantes para a manutenção da biodiversidade e ecossistemas em florestas tropicais. Além disso, elas são amplamente utilizadas por comunidades extrativistas para diversos fins, como alimentação, medicina e comércio. No entanto, o conhecimento sobre identificação e sua distribuição realizado por abordagens tradicionais apresentam baixa taxa de acurácia relatada e alto custo financeiro e operacional. Para solucionar esse problema, as redes neurais artificiais, especialmente as convolucionais, estão sendo utilizadas para reconhecimento de padrões em imagens, principalmente se coletadas por equipamentos remotos de baixo custo, como drones. Tais redes têm apresentado altas taxas de precisão na identificação de espécies florestais. Esse estudo propõe um método para classificar palmeiras nativas da família *Arecaceae* em áreas de floresta tropical, utilizando imagens adquiridas por um veículo aéreo não tripulado de baixo custo operacional e uma rede neural convolutacional. O método alcançou resultados mais precisos do que as abordagens convencionais, com uma acurácia de 95,86% e métricas de precisão de 99,57% e Recall de 95,95%. Desta forma, os mapas derivados desses sistemas de baixo custo podem ser úteis para apoiar projetos de manejo e monitoramento florestal comunitário na Amazônia.

Palavras-chave: segmentação, aprendizado profundo, CNN, palmeiras, sensoriamento remoto.

Identification of native palms (*Arecaceae*) in tropical forest areas based on Convolutional Neural Network with UAV Images

ABSTRACT

Palm trees are important components for maintaining biodiversity and ecosystems in tropical forests. Additionally, they are widely used by extractive communities for various purposes, such as food, medicine, and commerce. However, traditional approaches to identifying and mapping their distribution have reported low accuracy rates and high financial and operational costs. To address this problem, artificial neural networks, especially convolutional neural networks, are being used for pattern recognition in images, particularly those collected by low-cost remote equipment such as drones. Such networks have shown high accuracy rates in identifying forest species. This study proposes a method to classify native palm trees of the *Arecaceae* family in tropical forest areas using images acquired by a low-cost unmanned aerial vehicle and a convolutional neural network. The method achieved more accurate results than conventional approaches, with an accuracy of 95.86%, precision metrics of 99.57%, and recall and precision metrics of 95.95%. Thus, maps derived from these low-cost systems can be useful for supporting community forest management and monitoring projects in the Amazon.

Keywords: segmentation, deep learning, CNN, palm trees, remote sensing.

Introduction

The Amazon Basin holds the largest continuous area of tropical forest in the world, covering approximately 4.2 million km², of which 49.3% are located within Brazilian territory,

representing about 30% of all remaining tropical forests in the world. The Amazon biome is formed by a great diversity of environments, predominantly characterized by dense and open ombrophilous forests (SFB, 2010). This biome houses a large part

of the planet's biodiversity, with approximately 33,000 species of plants, of which at least 10,000 have potential in the fields of medicine, cosmetics, pest control, as well as a source of food for wildlife and for humans (Mittermeier et al., 2005). Among these species, palms comprise about 35 genera and more than 170 species (Alvez-Valles et al., 2018).

The inhabitants of the region explore various palm species as a source of economic order and consumption, being used for various purposes such as construction, roofing, furniture, wood, landscaping, handicrafts, oil production, food, beverages, among other uses (Lorenzi et al., 2010; Macia et al., 2011). It is observed, therefore, that palms contribute to the subsistence of traditional Amazonian populations (indigenous, extractivists and subsistence farmers who live near the forests) (Ferreira, 2006; Balslev et al., 2011; Macia et al., 2011; Gomes et al., 2016).

Despite their importance, information about the number of palms and their distribution in different scenarios is difficult to obtain, thus limited. A classical approach to solve this problem is to conduct field inventories, however, mapping palm species *in loco* in tropical forest areas is labor-intensive and involves planning, human resources, adequate infrastructure, and logistics since they are difficult to access areas with great diversity of flora, which hinders their identification and, consequently, generates a high cost.

The analysis of forest biodiversity requires accurate and reliable data, which can be obtained through various sensors, including satellite images, radar images, and, more recently, images captured by drones. The increasing use and popularization of drones have driven research in the identification of forest species, making this task more accessible and cost-effective. With the advantage of offering high spatial resolution images and low operational costs, these sensors allow for a more precise and efficient analysis of forest biodiversity, contributing to environmental conservation.

The increasing use of drones equipped with high spatial resolution sensors, capable of capturing images with details of less than 1 meter per pixel, has enabled more precise and detailed object mapping in the image. However, with the decrease in pixel size due to spatial resolution, a single object can occupy more than one pixel, making identification and classification more complex and challenging. Additionally, it is important to note that the information contained in a single pixel may be

insufficient to identify the object of interest in the image. Therefore, the development of classification algorithms with more sophisticated approaches has become essential to accurately identify and classify these elements, taking into account the complexity of the information generated by high spatial resolution sensors (Lang, 2008).

Thus, the application of new techniques and concepts from both the Geographic Information Science (GIScience) and Artificial Intelligence fields has led to the emergence of the emerging field of Object-Based Image Analysis (OBIA). The popularization of these new image classification methodologies has driven research in forest species identification, resulting in high accuracy rates and cost reduction.

As described by Peck et al., (2012); Otero et al., (2018), images captured by unmanned aerial vehicles (UAVs), although not free, have the potential to discriminate tree species in tropical environments. While these images are often captured with only three Red, Green, Blue (RGB) channels, providing limited spectral information compared to satellite images, drone images often have high resolutions (< 10 cm/pixel), allowing for clear visualization and extraction of structural characteristics (shape, size, and texture) of land objects, which can favor palm tree identification, as the crowns of palm trees have distinct morphological characteristics, according to (Ferreira et al., 2019).

It is observed that a considerable portion of research addressing this issue uses classical models in their solutions to identify and classify species. However, these models do not perform satisfactorily. In more recent research, artificial neural networks have been applied and have shown excellent pattern recognition rates from images collected by remote sensors, becoming an interesting alternative to solving these problems (Fernandes, 2013). They have evolved intensely in recent years as a result of their constant application in various research areas and activities. One advantage of using a neural network, in contrast to classical geoprocessing techniques, is that the neural network "learns" to automatically extract target characteristics such as predominant color, geometry, texture, etc., with the aim of identifying and classifying the target object (Weinstein et al., 2019).

Convolutional Neural Networks (CNNs), which have deep learning architectures, have proved to be the most suitable for tree detection in forest scenes (Weinstein et al., 2019), urban areas (Branson

et al., 2014), and agricultural lands (Saldana Ochoa e Guo, 2019). The reason is that CNNs achieve remarkable performance in object detection tasks, as they are complex enough to extract high-level intrinsic features to learn to spatially identify and label objects (Zhu et al., 2018). In computing, convolutional neural networks with deep learning is an emerging field that has many opportunities with good results for identifying physical characteristics of plants based on images (Sun et al., 2017).

The research by Culman et al., (2020) demonstrated success using convolutional neural networks with deep learning when applied to the challenging problem of locating and classifying individual Phoenix palm trees from high-resolution (RGB) aerial images, achieving accuracy results of 86%. It is worth noting, however, that in this cited research, the palms were from cultivation and not native to the forest, having a regular distribution, which in theory would facilitate the location of the species. In the work by (Li et al., 2019), two-stage convolutional neural networks were used to detect oil palm trees on a large scale based on high-resolution spatial images collected by the Quickbird satellite, resulting in a reported reciprocal average F1 score of 94.99%. However, the two-stage method adopted in this work, i.e., one for classifying ground cover and another for classifying the object, makes it complex for reproduction.

In the study by Gibril et al. (2021), an automatic approach was presented for large-scale mapping of date palm trees using very high-spatial-resolution unmanned aerial vehicle (UAV) data. This was based on a U-Shape convolutional neural network for semantic segmentation of date palm trees. The generalization evaluation of the proposed model on a comprehensive and complex test set exhibited high classification accuracy, demonstrating that date palm trees can be automatically mapped from UAV images with F-scores, mean intersection over union, precision, and recall of 91%, 85%, 0.91, and 0.92, respectively.

According to Khaing et al. (2021), the objective of their study was to classify and count different types of palm trees, including Toddy palm, Coconut palm, and Palm oil, using a remote sensing drone video and deep learning architecture known as mask R-CNN with retuning hyperparameter strategy. Aerial images were collected by drones in Upper and Delta Coastal areas of Myanmar, which were used to prepare a dataset of over 12,000 images. The performance of the system was examined using a

Bayesian optimization algorithm to retune the hyperparameters of the model. The study concluded that tuning the learning can improve the performance of classification for local palm tree segmentation task. The system accurately defined the Toddy palm with better accuracy and counting.

Wahed et al. (2022) aimed to develop a model for detecting the maturity of sago palm trees using drone images in their article. The methodology used was the combination of the architecture of three existing CNN models: AlexNet, Xception, and ResNet. The proposed model, called CraunNet, achieved 85.7% accuracy in 11 minutes of learning time and was two times faster than existing models. The study showed that CraunNet could be a more efficient and computationally cost-effective solution for detecting the maturity of sago palm trees.

In the research proposed by Mohamed et al. (2022), a deep learning-based instance segmentation framework was developed to detect and map individual date palm trees from UAV imagery. The framework involved converting image tiles and vector data into Common Objects in Context annotation format and evaluated various instance segmentation models using different network backbones. The study assessed the performance and generalizability of the evaluated models on testing datasets with varying spatial resolutions. Results showed that Mask R-CNN models with Swin Transformer backbones outperformed those with ResNets in detecting and segmenting date palm trees, achieving high mAP50 scores and F-measures. The proposed framework offers an efficient tool for mapping date palm trees from multi-scale UAV-based images, suitable for individual tree crown delineations and other earth-related applications.

The study by Letsoin et al. (2022) aimed to detect sago palms based on their physical morphology using unmanned aerial vehicle (UAV) RGB imagery. The authors used three pre-trained networks (SqueezeNet, AlexNet, and ResNet-50) and collected data from nine different groups of plants. The ResNet-50 model was found to be the preferred base model for sago palm classifiers, with a precision of 75%, 78%, and 83% for sago flowers, sago leaves, and sago trunk, respectively. However, the models tended to perform less effectively for sago palm and oil palm detection due to their physical similarity. The authors recommended improving the optimized parameters and providing more varied sago datasets for better detection and classification.

In the article by Marin et al. (2022), an integrated aerial system was proposed that uses UAV-captured images to identify the Amazonian Moriche palm in dense forests. The Mask R-CNN deep learning model was trained with 478 labeled palms using the transfer learning technique based on the well-known MS COCO framework. Comprehensive field experiments were conducted, resulting in a precision identification of 98%. The model is fully automatic and suitable for identifying and inventorying this species above 60 m under complex climate and soil conditions.

Considering the importance of palms and the challenges involved in identifying species, as highlighted above, this study proposes the development of a method for classifying native palm species *Arecaceae* in tropical forest areas, based on high spatial resolution images (4 cm/pixel) captured by an unmanned aerial vehicle (UAV), using a convolutional neural network - CNN, in an area belonging to the Santa Luzia Directed Settlement Project (PAD) in the municipality of Cruzeiro do Sul, Acre state.

Methods

Study area

The study area is a fraction of one of the various properties belonging to the Santa Luzia Directed Settlement Project (PAD), located in the western region of the state of Acre, with a central coordinate of 7°58'15"S, 72°25'34"W, in the municipality of Cruzeiro do Sul (Figures 1a, 1b, 1c, 1d).

The area is composed of diverse tropical forest, which presents a density and frequency of tree palm species. The mosaic of ortho-images in the form of a rectangle used in this investigation has an area of 2.72 hectares and measures 124 m by 220 m in size (Figure 3a).

According to data from (INMET, 2020), the region is inserted in the rainy equatorial subclimate,

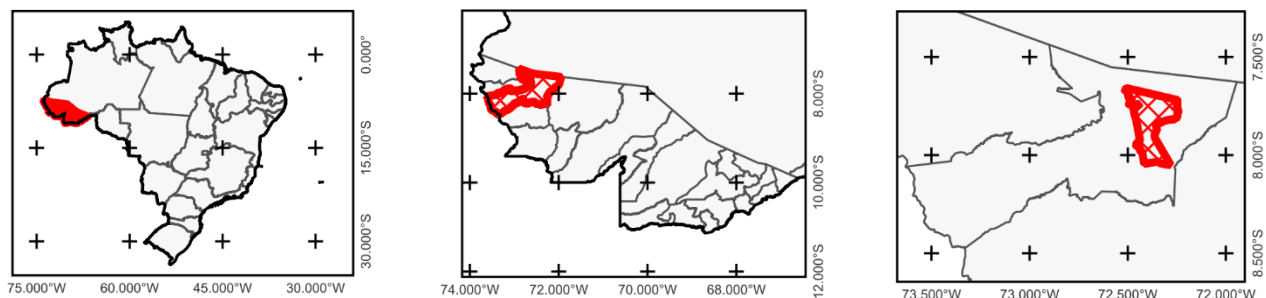
registering approximately 1,950 millimeters of rain annually, with an average annual temperature of 24.8°C. The driest month of the year is July, with a minimum climatological average of 60 mm, while March is the month with the highest rainfall, with a climatological average of 299 mm (INMET, 2020).

Regarding geomorphology, the study area is inserted in the geomorphology unit Depression of Juruá-Iaco. This unit presents a variable altitude between 150 and 440 m, its main forms of dissection are convex and sharp.

It is located in the Juruá/Liberdade River basin (Acre, 2006). Biotic environment studies produced by the Ecological-Economic Zoning of the State of Acre (ZEE) demonstrate that the research area is comprised of the following forest typology: Open Forest with Palms (FAP), which is generally found in areas near alluvial plains of rivers with high flow during the rainy season. This physiognomy is characterized by an open canopy forest with the presence of palms, and areas with lianas can also be found (Acre, 2006).

UAV images

The drone images were captured in December 2020 using a Phantom 4 Professional model, which has a 20-megapixel optical (RGB) camera with an automatic focus lens of 24 mm spatial resolution. To ensure a level view during image collection, the camera is attached to a three-axis electronic gimbal stabilization system. The UAV flew at 120 m above the forest canopy at cruising speed (13ms-1), and the flight was conducted between 09:00 and 12:00 local time (UTC -05:00). A total of 683 images were captured by the drone during its flight over the property, although, as mentioned earlier, the research uses only a fraction of the flight area (Figure 3a).



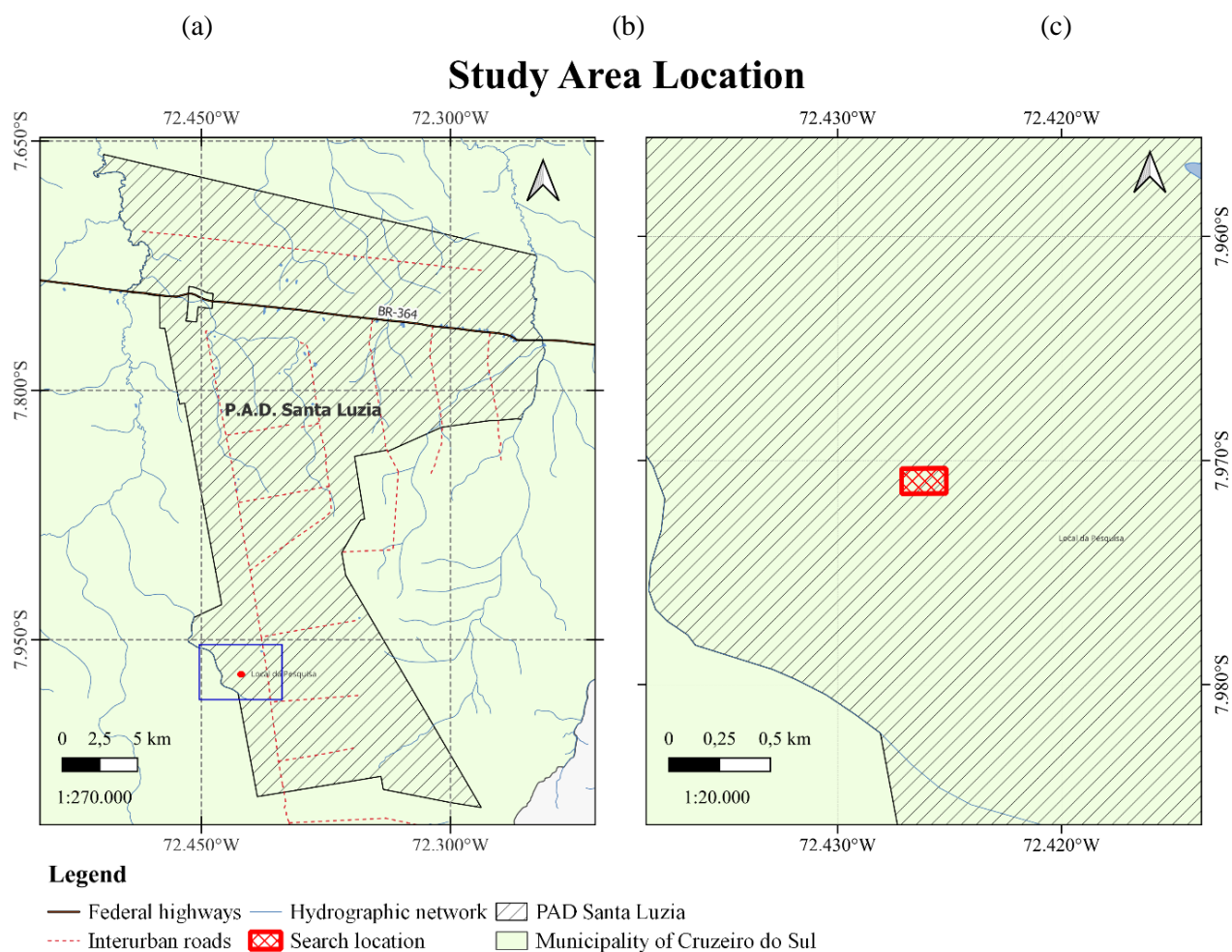


Figure 1 – (a) Location of the state of Acre in Brazil. (b) Location of the municipality of Cruzeiro do Sul in the state of Acre. (c) Location of the Santa Luzia Directed Settlement Project (PAD) in the municipality. (d) Location of the research area within the Santa Luzia Directed Settlement Project (PAD).

The images have a spatial resolution of 4.3 (cm/pixel), which were calibrated and generated the orthomosaic using Pix4Dmapper® software in the WGS 84/UTM zone 18S coordinate system.

Visual inspection of the study area revealed a significant number of palm trees (Figure 3a). According to the forest inventory of palm species conducted by the Acre Foundation of Technology - FUNTAC, the two most common species detected in the region are: - *Euterpe precatoria*; and - *Oenocarpus bataua*.

The forest inventory is a systematic and detailed process of collecting and analyzing data about a forest, including information on the composition, structure, and dynamics of the vegetation, as well as the physical environment and

socio-economic characteristics of the area in question (Husch, 2003).

The data from the forest inventory of palm species in the study area conducted by FUNTAC are presented in Figure 2. The Table 1 presents information about these species (Lorenzi et al., 2010).

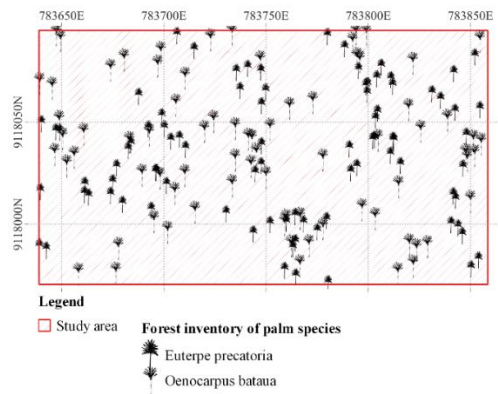


Figure 2 – Distribution of palm species in the study area conducted by FUNTAC.

Set of Images and Labels

Due to the limited size of the research area, a common technique in computer science was used to expand the image dataset. Thus, the original images were flipped horizontally and vertically (Flip and Flop) and combined (Flip/Flop), tripling the research area and significantly increasing the number of samples for building and testing the proposed model. This process was automatically performed using the ImageMagick® software for both images and labels, as described below.

In the first stage, automatic segmentation was performed to generate the necessary classes for supervised training of the model. For this task, the Mean Shift Clustering algorithm available in the Orfeo Toolbox® plugin of the QGIS® software was employed.

Mean-shift is a density-based clustering algorithm used to identify regions of high density of points in a feature space. It works by iteratively moving each point towards the mean of the points in its neighborhood, until the position of a point does not change significantly. Each region of high density is considered a cluster (Cheng, 1995; Comaniciu 2002; Meer, 2002).

This algorithm has the advantage of producing output in vector format, and dividing the processing into windows, defined in this process with a size of 1,024, which allows for segmentation of very large areas even with limited memory space. We set the minimum segment size to 5 pixels, while keeping the other parameters equal to the algorithm's default values. The segmentation process resulted in a total of 42,838 segments (Figure 3b), of which 4,146 ($\approx 1,900 \text{ m}^2$) belonged to the palm class, represented internally by the value 1, and 38,692 ($\approx 25,330 \text{ m}^2$) segments belonged to the forest class

without palm presence and/or background, represented by the value 9.

The assignment of classes to segments was performed by an analyst through visual inspection, using a natural color composition of the orthomosaic at a scale of 1:250. The assignment was carried out for all segments (Figure 7b), described in the previous section, that belonged to visible palm classes in the image, respecting the boundaries of the generated segments. The QGIS® software was used for this task.

The approach used in this study to securely assign segments corresponding to the palm class (value 1) by the analyst during visual inspection could be applied due to the high spatial resolution (4.3 cm/pixel) of the drone-captured images (Figure 4a, 4b), as well as having an inventory database of palms produced by FUNTAC as support. The study area was divided into 180 regular-sized 512×512-pixel parcels/files, with each parcel having an approximate size of 22 m². The cutting process was performed automatically using the ImageMagick® software.

CNN Architecture

The number of layers, i.e., the depth of the network, is directly related to the network's capacity to learn data features, which reduces classification errors, according to studies on deep networks (Shimodaira, 2000; Simonyan e Zisserman, 2014). This led researchers to increase the network depth by adding more layers in an attempt to improve results. However, some studies have shown that increasing network depth leads to an increase in training error (Simonyan e Zisserman, 2014).





To overcome the problem, He et al., (2016) proposed a deep residual learning structure called Residual Network (ResNet). A ResNet architecture is composed of several residual blocks that perform skip connections, meaning they forward the activations from a certain layer to a deeper layer. Common variations of ResNet include ResNet-18, ResNet-50, and ResNet-101, which differ in the number of residual layers. In this study, we used the ResNet-18 architecture 4, which provided a reasonable trade-off between processing time and accuracy in preliminary tests.

According to He et al., (2016), ResNet-18 is capable of achieving comparable or superior performance to other deep network architectures, even when trained with less data. This is due to the

use of the residual module, which allows the network to learn deeper and more robust representations of input data, and the use of normalization, which helps to avoid the problem of vanishing gradients and speeds up training (He et al., 2016). (Bengio; Boulanger-Lewandowski; Pascanu, 2012), on the other hand, describes that the vanishing gradient

problem can arise during the training of deep neural networks, where the gradient parameters (weights and biases) become too small, making it difficult for the network to learn.

Table 1 – Description of the native palm species present in the research region.

| Species | Description | UAV images | Ground View |
|--|--|---|---|
| <p><i>Euterpe precatoria</i></p> <p>Common Names: Forest Açai, Açai, Juçara.</p> | <p>Formed by a simple, erect stem, 3-20 m high and 4-23 cm in diameter, with a cone of visible roots at the base and a smooth palm heart at the top. Leaves pinnate, flat, in 10-20 contemporaneous, divergent, and occasionally pendulous; closed sheath 0.7-1.6m long, forming a tube with a green stem or sometimes green with vertical yellow stripes. Fruits globular, between 1.0 and 1.3 cm in diameter, purple-black in color. Its habitat occurs in the states of Acre, Amazonas, Pará, and Rondônia, in lowland humid tropical forests, usually along rivers in periodically flooded areas (Lorenzi et al., 2010).</p> |  |  |
| <p><i>Oenocarpus bataua</i></p> <p>Common name: Patawa.</p> | <p>It has a solitary stem with 5-25m height and 20-45 cm in diameter, with visible fasciculate roots at the base and without a smooth top. The leaves are in numbers of 10-20, erect and divergently arranged. It produces oblong fruits with 2.7-4.5 cm in length, with a dark purple color and ruminant endosperm. They are widely found in the Brazilian Amazon region and northern South America, in the humid forests of floodplains and gallery forests, both inundated and upland areas (Lorenzi et al., 2010).</p> |  |  |

As described by He et al., (2016), ResNet-18 consists of 18 layers divided into two types: *i*) residual layers and *ii*) transformation layers. The residual layers are composed of two convolutional

layers followed by a sum operation with the input, while the transformation layers are composed of a convolutional layer and a normalization operation.

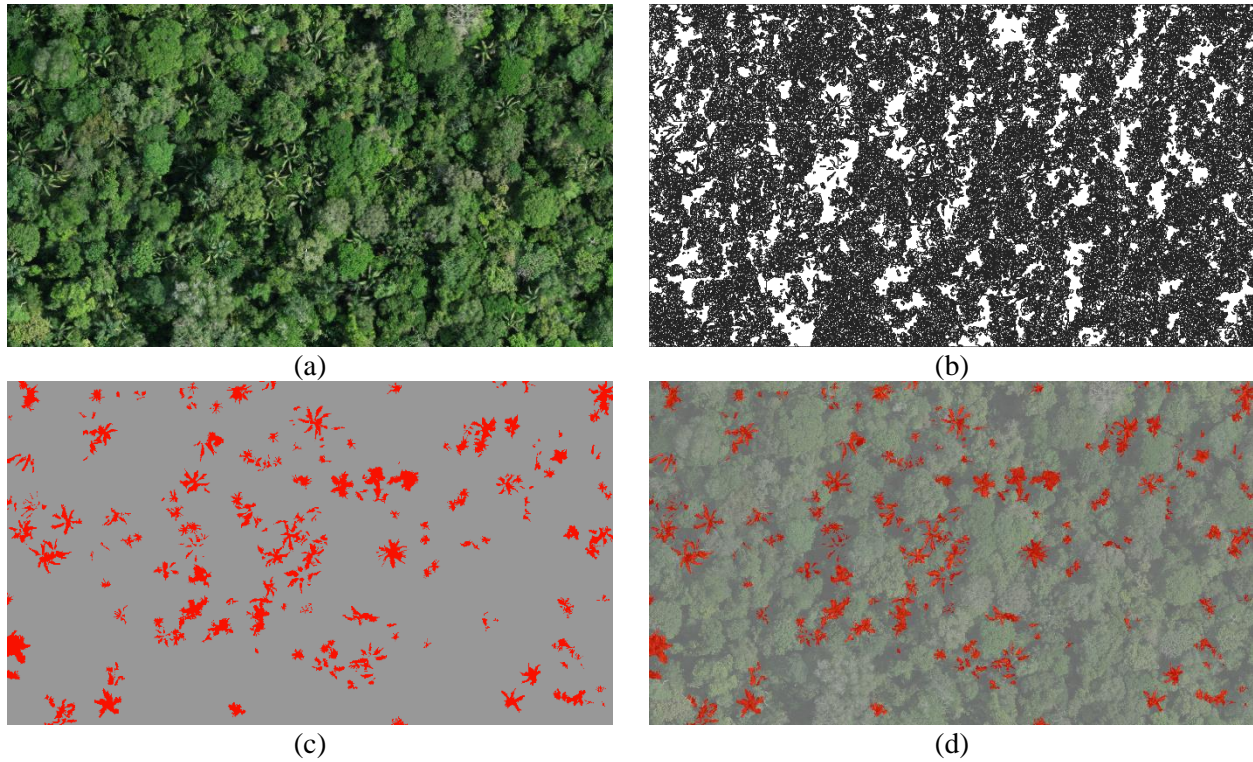


Figure 3 – (a) Study area: tropical forest with high incidence of native palms. (b) Vectorized automated segmentation of the study area using the Mean-Shift algorithm. Segments colored in red were assigned to the palm class [value 1], while gray segments were assigned to the background class [value 9]. (c) Pixel labels colored in red indicate the presence of palms, while the gray area indicates the forest without palm presence. (d) Overlay of data labels on the study area image.

The input layer of ResNet-18 is followed by a normalization layer and a convolutional layer, which serve as a feature extraction layer. The output of this layer is then passed through the residual and transformation layers, which allow the network to learn deeper and more robust representations of the input data. In summary, the network architecture presents different functions depending on the type of layer.

- Convolution: performs element-wise multiplication between input data and a set of learned filters, producing feature maps that capture specific characteristics of the input data;
- Batch Normalization: normalizes the activations of the previous layer by subtracting the mean and dividing by the standard deviation of activations in a batch of samples;
- Rectified Linear Unit - ReLU (activation): replaces all negative values in the input with zero, providing non-linearity to the network, thus helping

- to speed up the training process and increase the model accuracy;
- Addition: combines two or more inputs element-wise by adding corresponding elements from each input;
- Transpose Convolution: performs the reverse operation of a convolution, expanding the spatial resolution of the input;
- Cropping: crops a part of the input tensor to remove unnecessary information and keep only relevant information;
- Depth Concatenation: combines multiple inputs along the depth dimension by concatenating their feature maps to form a single output tensor; and
- Softmax (activation): converts input activations into a probability distribution, where the sum of the output values is equal to 1, it is commonly used in the final layer of a CNN for multi-class classification problems (He et al., 2016).



Figure 4 – (a) Details of a data input (512×512 pixels) of an image at a scale of 1:250 for class assignment to segments. (b) Segments with palm class [value 1] defined by the analyst in red color.

For semantic segmentation, the ResNet-18 was incorporated into the DeepLabv3+ architecture, which is considered a state-of-the-art deep learning model for semantic segmentation of images. DeepLabv3+ was proposed by (Chen et al., 2018) and consists of an encoder block and a decoder block, as shown in Figure 5.

The encoder module gradually reduces the spatial dimension of the input patch and captures high-level semantic information. The decoder module recovers the patch size by restoring spatial information to produce sharp segmentation results. DeepLabv3+ utilizes a powerful technique that allows for capturing multi-level features of the input image while controlling the resolution of the convolutional layers' output. Convolutions belonging to the block called Atrous Spatial Pyramid Pooling (ASPP) are applied in parallel with different rates and serve to capture multi-level characteristics of the input image.

The output generated by ResNet-18 with DeepLabv3+ are score maps for each class. The transposed convolutional layer performs up-sampling with five filters. Next, the softmax classifier is applied to produce pixel-wise maps where each pixel contains class association probabilities.

CNN Setup

Initially, the data was partitioned following an 80% ratio, which represents 144 parcels with a size of 512×512-pixels for training, and 20% (36 parcels) for testing randomly chosen to train and test the fully convolutional network model. This allowed for calculating the variability in classification accuracy, depending on the data used in the classification process. The image data does not contain edge pixels with a value of 0.

Thus, the model was trained with two classes, namely one class of palm trees, which comprises two species of palm trees: *Oenocarpus bataua* and *Euterpe precatoria*, in addition to the background class. The inclusion of the background class in the training process was necessary for the network to learn to differentiate palm trees from other types of trees.

The mini-batch size was 16, and the maximum number of epochs was 10. A mini-batch is a subset of the training data that is used by the Stochastic Gradient Descent with Momentum - SGDM algorithm, which aims to update the network's weight and bias parameters (Murphy, 2012). An epoch is a complete pass through the entire training set, consisting of 512 random patches. Each epoch, a different set of 512 patches was extracted from the images. During training, the input images are randomly rotated and transformed along their axes (X and Y).

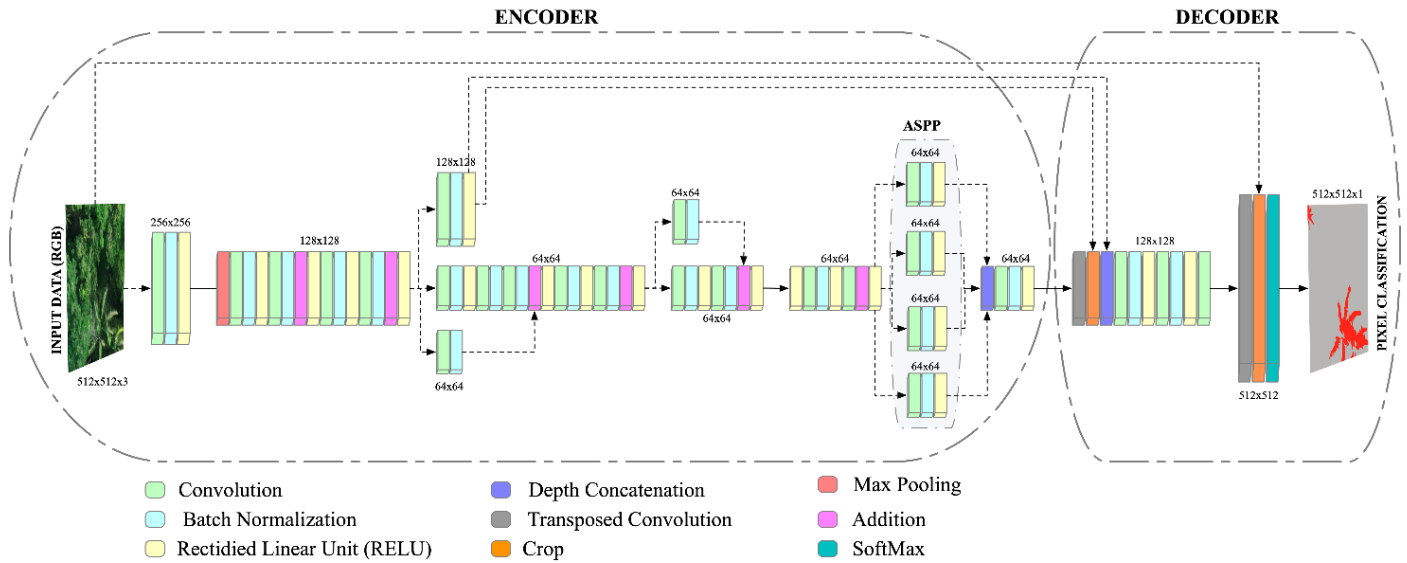


Figure 5 – Network architecture based on DeepLabv3+ CNN as described in (CHEN et al., 2018).

This procedure is another data augmentation technique, which is a common practice to avoid overfitting the network. The ResNet-18 model weights were initialized with pre-trained values from the ImageNet database (Deng et al., 2009), and the learning rate was 0.05. The class score maps in the testing phase were produced with a size of 512×512 pixels.

The training and inference were performed at the Geoprocessing Laboratory - GEOLAB, on a desktop workstation with Intel Xeon(R)® CPU ES-1650 v4 @ 3.6 GHz, 32 GB DDR4 main memory, and NVIDIA® Quadro K1200 GPU with 4 GB GDDR5 dedicated memory and 512 CUDA® cores. All image processing procedures were performed using the programming language MatLab R2022b®.

Results

The best performance evaluation of the proposed model achieved an average accuracy of 95.8%. During training, the validation frequency occurred 50 times, in which a portion of the training data is typically used exclusively for this purpose, allowing an evaluation of the network's ability to generalize its learning to new data.

The confusion matrix (Figure 6) shows that the CNN segmentation classified 976,442 pixels as palms, which represents 94.44% of the total palm pixels. For the background class, the segmentation correctly classified 13,181,512 pixels as background, representing 95.95% of this class. The result showed confusion in the segmentation that misclassified

556,880 background pixels as palms (3.8%) and 57,502 palm pixels were misclassified as background (0.40%).

Overall, the accuracy (Equation 1) achieved by the model was 95.84% of predictions were classified correctly, and 4.16% were wrong. However, the classes are not balanced, meaning that when the number of examples for each class is not approximately equal, recall and precision are more useful metrics to evaluate the model's performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Just like accuracy, the precision metrics (Equation 2) and recall (Equation 3) have widely publicized formulas. Where TP (True Positives) are the correct predictions of the positive class; FP (False Positives) are the incorrect predictions of the positive class. TN (True Negatives) are the correct predictions of the negative class, and FN (False Negatives) are the incorrect predictions of the negative class. The precision achieved by the model reached 99.57%, while the recall metric reached 95.95%.

The segmentation result with the classes assigned by the model was compared with the data

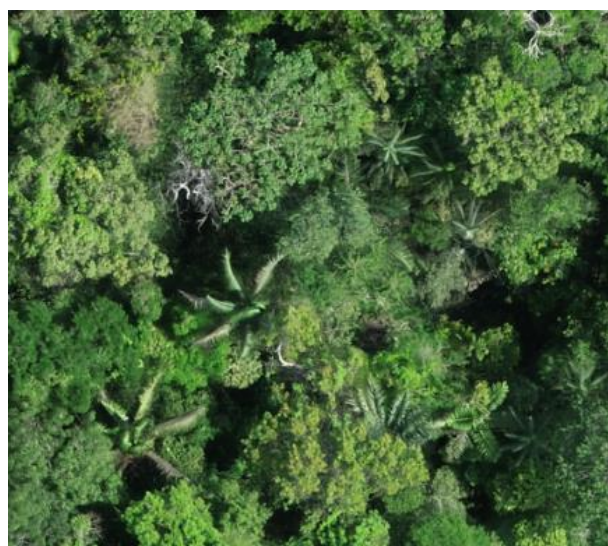
labeling performed by the analyst. The image difference operation is an image processing technique that allows the calculation of pixel-by-pixel difference between two images. This operation can be useful in various applications, such as detecting changes in images, identifying objects, etc (Gonzalez; e Woods, 2000).

The formula applied to obtain the difference image was: $D(x, y) = I1(x, y) - I2(x, y)$, where $D(x, y)$ is the pixel value in the resulting difference image at position (x, y) ; $I1(x, y)$ is the pixel value in the first image at position (x, y) ; and $I2(x, y)$ is the pixel value in the second image at position (x, y) .

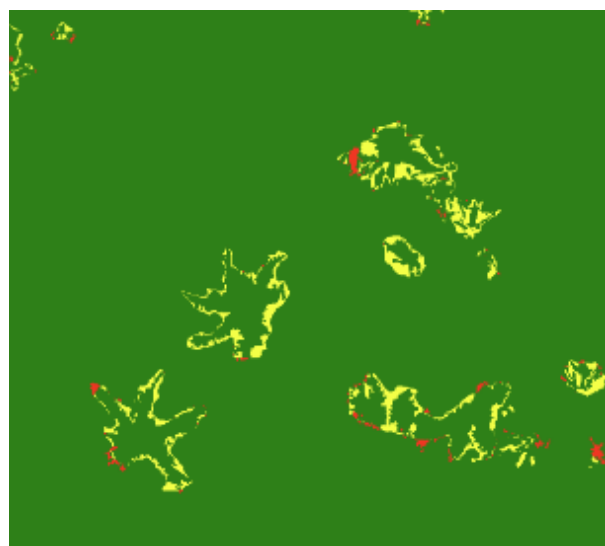
The resulting difference image was color-coded, with green representing correct classifications and yellow and red indicating classification errors.

| | Palm trees [1] | Background [9] | |
|----------------|-----------------------|--------------------------|-----------------|
| Palm trees [1] | 976442 6,6% | 57502 0,40% | 94,44% 5,56% |
| Background [9] | 556880 3,8% | 13181512 89,2% | 95,95% 4,05% |
| | 63,68% 36,32% | 99,57% 0,43% | 95,84% 4,16% |

Figure 6 – Confusion Matrix.



(a)



(b)

Figure 7 – (a) Image detail at the 1:250 scale. (b) Result of the difference between the segmentation produced by the neural network and the data labels. Pixels in green represent correct classification, while pixels in yellow and red represent misclassification.

Figures (7a and 7b) presents a small sample of the difference image between these classifications. The green pixels were classified correctly, while the yellow pixels represent an error, where palm tree pixels (class 1) were classified as background (class 9), and the red pixels represent errors where the model classified background (class 9) as palm trees (class 1).

The Table 2 presents the results of the difference image $D(x, y) = I1(x, y) - I2(x, y)$, where $I1$ is the data labels of the study area produced by the analyst through visual inspection, and $I2$ is the segmentation and identification of the study area performed by the proposed convolutional neural network model.

Table 2 - Results of the difference image process.

| Category | # Pixel | Square meters | % |
|-----------------|-----------------|-------------------|---------------|
| True Positives | 14289414 | 26,384.267 | 96.67 |
| False Positives | 477401 | 881.483 | 3.23 |
| False Negatives | 13522 | 24.967 | 0.09 |
| Total | 14780337 | 27,290.717 | 100.00 |

The results of the difference image demonstrate that the highest percentage (96.67%) was correctly classified as True Positives. However, there was confusion of 3.23% of the palm tree class that was wrongly classified as background (False Positives). This percentage covers a total area of 881,483 square meters. Additionally, about 24,967 square meters (0.09%) that should have been classified as background were instead classified as palm trees (False Negatives).

Discussion and conclusion

This study was based on low-cost drone images acquired with only three channels (RGB), and it demonstrated the potential for the use of deep learning to map the spatial distribution of palm species in Amazonian forests.

The study did not differentiate between the two palm species (*Euterpe precatoria* and *Oenocarpus bataua*) present in the study area, but the network model used achieved excellent generalization by efficiently representing both species and separating them from the rest of the background. This was possible due to the network's ability to learn and generalize from the training data. Additionally, during the training process, the validation occurred 50 times, where a portion of the training data was exclusively used for this purpose. This allowed for an assessment of the network's performance and its ability to generalize its learning to new data, which is a critical aspect of developing robust and accurate models.

The proposed CNN segmentation model achieved an average accuracy of 95.8%, demonstrating its ability to effectively classify palm and background pixels in the image data. The confusion matrix analysis revealed that the model correctly classified 94.44% of palm pixels and 95.95% of background pixels, while misclassifying

only a small percentage of pixels (3.8% of background and 0.40% of palm pixels).

It is important to highlight that the classes in the dataset are not balanced, which means that the number of examples for each class is not equal. Therefore, it is essential to evaluate the performance of the model using metrics such as recall and precision, which take into account the class imbalance. These metrics are particularly important in situations where imbalanced classes can lead to biased or inaccurate results. The precision achieved by the model was 99.57%, and the recall metric reached 95.95%.

The high classification accuracy rate achieved for palms suggests that there is no need for spectral information in the classification process, which contradicts the results of previous studies that show that classification success usually depends on high spectral resolutions (Fassnacht et al., 2016).

The results demonstrate the potential of CNN segmentation for accurately identifying palm pixels in images and have practical implications for projects related to palm mapping and monitoring. Maps of palm spatial distribution can significantly assist management projects in the Amazon, providing a valuable tool to support decision-making and community forest monitoring programs. The approach developed in this study can be applied to other tropical forest areas where high spatial resolution drone images are available, with a resolution of at least 4 cm per pixel.

Based on the results, we can conclude that the use of convolutional neural networks is a viable and efficient alternative for the identification and classification of palm species in the Amazon Rainforest. This technique showed high accuracy in species identification, surpassing traditional methods used until now, as well as presenting a lower cost compared to these methods. Therefore, the use of object image analysis techniques such as convolutional neural networks may be a promising solution for mapping and studying the biodiversity of the Amazon Rainforest and other areas with high species diversity.

Further research is needed to evaluate the potential of such sensors for palm species mapping in tropical environments. One of our future investigations is to retrain the model by differentiating palm species and adding the 3D data (point clouds) obtained from the drone to the classification process, hypothesizing that this data could help identify a unique canopy architecture.

Acknowledgments

This research received funding from the Public Call Fapesc 48/2022 - Support for Infrastructure for Research Groups at UDESC. The authors would like to thank the State University of Santa Catarina - UDESC, through the Graduate Program in Territorial Planning and Socio-Environmental Development - PPGPLAN, which provides experienced and high-quality faculty, as well as all the physical and technological structure of the Geoprocessing Laboratory - GEOLAB, and the support of the Federal Institute of Acre - IFAC.

Conflicts of Interest

The authors state that they do not have any known conflicting financial interests or personal ties that may seem to have influenced the work disclosed in this study.

References

- Acre, Governo do Estado. (2006). Programa estadual de Zoneamento Ecológico-Econômico do estado do Acre. Zoneamento ecológico-econômico do Acre Fase II: Documento síntese – Escala, v. 1, n. 250.000.
- Alvez-Valles, C. M., Balslev, H., Garcia-Villacorta, R., Carvalho, F. A., & Menini Neto, L. (2018). Palm species richness, latitudinal gradients, sampling effort, and deforestation in the Amazon region. *Acta Botanica Brasilica*, 32(4), 527–539. <https://doi.org/10.1590/0102-33062018abb0091>.
- Balslev, H., Kahn, F., Millán, B., Svenning, J.-C., Kristiansen, T., Borchsenius, F., Pedersen, D., & Eiserhardt, W. L. (2011). Species diversity and growth forms in tropical American palm communities. *The Botanical Review*, 77, 381–425.
- Bengio, Y., Boulanger-Lewandowski, N., & Pascanu, R. (2012). Advances in optimizing recurrent networks. arXiv. <https://arxiv.org/abs/1212.0901>.
- Branson, S., Wegner, J. D., Hall, D., Lang, N., & Schindler, K. (2014). Beyond counting: New perspectives on optical remote sensing for conservation. *Remote Sensing*, 6(11), 11089–11102. <https://doi.org/10.3390/rs61111089>.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv. <https://arxiv.org/abs/1802.02611>.
- Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), 790–799. <https://doi.org/10.1109/34.400568>.
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619. <https://doi.org/10.1109/34.1000236>.
- Culman, M., Delalieux, S., & Van Tricht, K. (2020). Individual palm tree detection using deep learning on RGB imagery to support tree inventory. *Remote Sensing*, 12(21), 3476. <https://doi.org/10.3390/rs12213476>.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248–255). <https://doi.org/10.1109/CVPR.2009.5206848>.
- Fassnacht, F. E., Latifi, H., Sterenczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>.
- Fernandes, B. J. T. (2013). Redes neurais com extração implícita de características para reconhecimento de padrões visuais [Neural networks with implicit feature extraction for visual pattern recognition]. Universidade Federal de Pernambuco. <https://repositorio.ufpe.br/handle/123456789/12245>.
- Ferreira, E. J. L. (2006). Manual das palmeiras do Acre, Brasil [Manual of the palms of Acre, Brazil]. Instituto Nacional de Pesquisas da Amazônia. <https://www.nybg.org/bsci/acre/www1/manual%5Ctextunderscore%7B%7Dpalmeiras.html>.
- Ferreira, M. P., Wagner, F. H., Aragão, L. E. O. C., Shimabukuro, Y. E., & De Souza Filho, C. R. (2019). Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 119–131. <https://doi.org/10.1016/j.isprsjprs.2019.01.019>.
- Gibril, M. B. A., Shafri, H. Z. M., Shanableh, A., Al-Ruzouq, R., Wayayok, A., & Hashim, S. J. (2021). Deep convolutional neural network for

- large-scale date palm tree mapping from UAV-based images. *Remote Sensing*, 13(14), 2787. <https://doi.org/10.3390/rs13142787>.
- Gomes, J. P., Condé, T. M., Santos, R. L., Dionisio, L. F. S., Duarte, O. R., Miranda, D. L. C. de, & Silva, F. D. (2016). Efeitos de gradientes ambientais na fitossociologia de assembleias de palmeiras no Sudeste de Roraima, Brasil [Effects of environmental gradients on the phytosociology of palm assemblages in southeastern Roraima, Brazil]. *Nativa*, 4(5), 317–327. <https://doi.org/10.31413/nativa.v4i5.3581>.
- Gonzalez, R. C., & Woods, R. E. (2000). *Processamento de imagens digitais [Digital image processing]*. Editora Blucher.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE*.
- Husch, B., Beers, T. W., & Kershaw JR, J. A. (2003). *Forest mensuration*. John Wiley & Sons.
- INMET, Instituto Nacional de Meteorologia. (2020)
- Khaing, S. H., & Sein, M. M. (2021). Toddy Palm Trees Classification and Counting Using Drone Video: Retuning Hyperparameter Mask-RCNN. In 2021 7th International Conference on Control, Automation and Robotics (ICCAR) (pp. 196-200). *IEEE*. <https://doi.org/10.1109/ICCAR52225.2021.9463466>.
- Lang, S. (2008). Object-based image analysis for remote sensing applications: modeling reality – dealing with complexity. In T. Blaschke, S. Lang, & G. J. Hay (Eds.), *Object-Based Image Analysis* (pp. 3-22). Springer. https://doi.org/10.1007/978-3-540-77058-9_1.
- Letsoin, S. M. A., Purwestri, R. C., Rahmawan, F., & Herak, D. (2022). Recognition of Sago Palm Trees Based on Transfer Learning. *Remote Sensing*, 14(19), 4932. <https://doi.org/10.3390/rs14194932>.
- Li, W., Dong, R., Fu, H., & Yu, L. (2019). Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks. *Remote Sensing*, 11(1), 11. <https://doi.org/10.3390/rs11010011>.
- Lorenzi, H., Noblick, L. R., Kahn, F., & Ferreira, E. (2010). *Flora Brasileira Lorenzi: Arecaceae (Palmeiras)*. Nova Odessa.
- Macía, M. J., Armesilla, P. J., Cámara-Leret, R., Paniagua-Zambrana, N., Villalba, S., Balslev, H., & Pardo-de-Santayana, M. (2011). Palm uses in northwestern South America: a quantitative review. *The Botanical Review*, 77(4), 462–570. <https://doi.org/10.1007/s12229-011-9086-8>.
- Marin, W., Mondragon, I. F., & Colorado, J. D. (2022). Aerial Identification of Amazonian Palms in High-Density Forest Using Deep Learning. *Forests*, 13(5), 655. <https://doi.org/10.3390/f13050655>.
- Mittermeier, R. A., Fonseca, G., Rylands, A., & Brandon, K. (2005). Uma breve história da conservação da biodiversidade no Brasil. *Megadiversidade*, 1(1), 14-21.
- Mohamed Barakat A. Gibril, Helmi Zulhaidi Mohd Shafri, Rami Al-Ruzouq, Abdallah Shanableh, Faten Nahas, & Saeed Al Mansoori. (2023). Large-Scale Date Palm Tree Segmentation from Multiscale UAV-Based and Aerial Images Using Deep Vision Transformers. *Drones*, 7(2), 93. <https://doi.org/10.1080/10106049.2022.2142966>.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Otero, V., Van De Kerchove, R., Satyanarayana, B., Martínez-Espinosa, C., Fisol, M. A. B., Ibrahim, M. R. B., Sulong, I., Mohd-Lokman, H., Lucas, R., & Dahdouh-Guebas, F. (2018). Managing mangrove forests from the sky: Forest inventory using field data and Unmanned Aerial Vehicle (UAV) imagery in the Matang Mangrove Forest Reserve, Peninsular Malaysia. *Forest Ecology and Management*, 411, 35–45. <https://doi.org/10.1016/j.foreco.2017.12.049>.
- Peck, M., Mariscal, A., Padbury, M., Cane, T., Kniveton, D., & Chinchero, M. A. (2012). Identifying tropical Ecuadorian Andean trees from inter-crown pixel distributions in hyperspatial aerial imagery. *Applied Vegetation Science*, 15(4), 548–559. <https://doi.org/10.1111/j.1654-109X.2012.01196.x>.
- Saldana Ochoa, K., & Guo, Z. (2019). A framework for the management of agricultural resources with automated aerial imagery detection. *Computers and Electronics in Agriculture*, 162, 53–69. <https://doi.org/10.1016/j.compag.2019.03.028>.
- SFB, Serviço Florestal Brasileiro. (2010). *Florestas do Brasil em resumo - 2010: Dados de 2005-2010*. Ministério do Meio Ambiente.
- Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2), 227–244. [https://doi.org/10.1016/S0378-3758\(00\)00115-4](https://doi.org/10.1016/S0378-3758(00)00115-4).

- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. <https://arxiv.org/abs/1409.1556>.
- Sun, Y., Liu, Y., Wang, G., & Zhang, H. (2017). Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, 2017, 7361042. <https://doi.org/10.1155/2017/7361042>.
- Wahed, Z., Joseph, A., Zen, H., & Kipli, K. (2022). Sago Palm Detection and its Maturity Identification Based on Improved Convolution Neural Network. *Pertanika Journal of Science & Technology*, 30(2). <https://doi.org/10.47836/pjst.30.2.20>.
- Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sensing*, 11(11). <https://doi.org/10.3390/rs11111309>.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2018). Deep learning in remote sensing: a review. arXiv. <https://doi.org/10.48550/arXiv.1710.03959>.