



## Individual tree detection and species classification of Amazonian palms using UAV images and deep learning



Matheus Pinheiro Ferreira<sup>a,\*</sup>, Danilo Roberti Alves de Almeida<sup>b</sup>, Daniel de Almeida Papa<sup>c</sup>, Juliano Baldez Silva Minervino<sup>d</sup>, Hudson Franklin Pessoa Veras<sup>e</sup>, Arthur Formighieri<sup>d</sup>, Caio Alexandre Nascimento Santos<sup>c</sup>, Marcio Aurélio Dantas Ferreira<sup>f</sup>, Evandro Orfanó Figueiredo<sup>c</sup>, Evandro José Linhares Ferreira<sup>g</sup>

<sup>a</sup> Cartographic Engineering Section, Military Institute of Engineering (IME), Praça Gen. Tibúrcio 80, 22290-270 Rio de Janeiro, RJ, Brazil

<sup>b</sup> Forest Sciences Department, University of São Paulo (USP), Av. Pádua Dias 11, Piracicaba, SP, Brazil

<sup>c</sup> Embrapa Acre, Rodovia BR-364, km 14, 69900-056 Rio Branco, AC, Brazil

<sup>d</sup> Federal University of Acre, Rodovia BR 364, Km 04 - Distrito Industrial, 69920-900 Rio Branco, AC, Brazil

<sup>e</sup> Department of Forestry, Federal University of Paraná (UFPR), Pref. Lothário Meissner Avenue 900, 80210-170 Curitiba, PR, Brazil

<sup>f</sup> Technology Foundation of Acre State, Distrito Industrial, Rio Branco - State of Acre, 69920-202 Rio Branco, AC, Brazil

<sup>g</sup> National Institute of Amazonian Research (INPA), Estrada Dias Martins, 3868, Bairro Chácara Ipê, 69917-560 Rio Branco, AC, Brazil

### ARTICLE INFO

#### Keywords:

Drone  
DeepLabv3+  
Amazon forest  
Açaí  
*Euterpe precatoria*

### ABSTRACT

Information regarding the spatial distribution of palm trees in tropical forests is crucial for commercial exploitation and management. However, spatially continuous knowledge of palms occurrence is scarce and difficult to obtain with conventional approaches such as field inventories. Here, we developed a new method to map Amazonian palm species at the individual tree crown (ITC) level using RGB images acquired by a low-cost unmanned aerial vehicle (UAV). Our approach is based on morphological operations performed in the score maps of palm species derived from a fully convolutional neural network model. We first constructed a labeled dataset by dividing the study area (135 ha within an old-growth Amazon forest) into 28 plots of 250 m × 150 m. Then, we manually outlined all palm trees seen in RGB images with 4 cm pixels. We identified three palm species: *Attalea butyracea*, *Euterpe precatoria* and *Iriarteia deltoidea*. We randomly selected 22 plots (80%) for training and six plots (20%) for testing. We changed the plots for training and testing to evaluate the variability in the classification accuracy and assess model generalization. Our method outperformed the average producer's accuracy of conventional patch-wise semantic segmentation (CSS) in 4.7%. Moreover, our method correctly identified, on average, 34.7 percentage points more ITCs than CSS, which tended to merge trees that are close to each other. The producer's accuracy of *A. butyracea*, *E. precatoria* and *I. deltoidea* was  $78.6 \pm 5.5\%$ ,  $98.6 \pm 1.4\%$  and  $96.6 \pm 3.4\%$ , respectively. Fortunately, one of the most exploited and commercialized palm species in the Amazon (*E. precatoria*, a.k.a, Açaí) was mapped with the highest classification accuracy. Maps of *E. precatoria* derived from low-cost UAV systems can support management projects and community-based forest monitoring programs in the Amazon.

### 1. Introduction

Palms (Arecaceae) are among the most conspicuous and diverse groups of trees with 181 genera and ca. 2600 species (Baker and Dransfield, 2016). In the Amazon basin, the palm flora comprises about 35 genera made up of more than 170 species (Alvez-Valles et al., 2018), from which at least 96 species (Bernal et al., 2011) are managed to provide a variety of products such as fruits and palm hearts (*Euterpe*

spp., *Astrocaryum* spp.), fiber (*Iriarteia deltoidea*, *Attalea* spp.) and oil (*Attalea butyracea*). The majority of palm species are spread out over large tracts of forests, pastureland, and agricultural fields. A few exceptions are *Mauritia flexuosa* and *Oenocarpus bataua* that form oligarchic forests (Peters, 1992).

Information regarding the spatial distribution of palms is crucial for commercial exploitation and management (Muñiz-Miret et al., 1996). This information is usually obtained with field inventories that include,

\* Corresponding author.

E-mail address: [matheus@ime.eb.br](mailto:matheus@ime.eb.br) (M.P. Ferreira).

<https://doi.org/10.1016/j.foreco.2020.118397>

Received 14 April 2020; Received in revised form 1 July 2020; Accepted 2 July 2020

0378-1127/ © 2020 Elsevier B.V. All rights reserved.

for example, *in situ* counting of suitable individuals (Rocha, 2004). Due to the high costs and effort involved, field inventories cover small areas (< 1 ha) and are not adequate to obtain spatially continuous knowledge on palms occurrence over large spatial extents. For this purpose, a combination of remote sensing and field data proved to be effective. Particularly the use of very-high-resolution (VHR) images (pixel size  $\leq 1$  m) has been hailed as a processing way to identify tree species, including palms, at the individual tree crown (ITC) level.

In tropical forests, ITC mapping has been performed with hyperspectral (Clark et al., 2005; Feret and Asner, 2013; Ferreira et al., 2016) and multispectral (Cho et al., 2015; Omer et al., 2015; Ferreira et al., 2019) images acquired by airborne or spaceborne sensors. While conventional aero-photogrammetric surveys in tropical regions are costly and often operationally prohibitive, images collected by satellite-based systems are more affordable, but still, show some limitations. For example, satellite images of VHR sensors (WorldView, SkySat, and Ikonos) are not free of charge and usually show clouds that limit their use.

Aerial images captured by unmanned aerial vehicles (UAVs), although not free of charge, have the potential to discriminate canopy tree species in tropical environments (Peck et al., 2012; Otero et al., 2018) and support community-based forest monitoring programs (Paneque-Gálvez et al., 2014). Unlike satellite images, UAV images are typically acquired with only three (RGB) channels, thus providing limited spectral information. However, they usually feature hyperspatial resolutions (pixel size <10 cm), which allows a clear visualization and extraction of structural characteristics (shape, size, and texture) of ground objects. Individual tree crowns of palms have distinctive morphological traits. These include pinnate leaves concentrated at the top of the stem, which may produce peculiar texture patterns or fine-scale variations in gray image levels. When spectral information is poorly available, in the case of RGB images, the quantification of texture may improve the classification accuracy of tree species (Ferreira et al., 2019).

Convolutional Neural Networks (CNNs), a type of deep learning method, automatically extract texture patterns of images and has been producing outstanding results for remote sensing image classification (Zhang et al., 2016). CNNs constitute a class of deep Artificial Neural Network (ANN) that rely on convolutions (local linear operations) followed by non-linear transformations, creating different input data representations. The convolutional layers act as feature extractors of the input images. CNNs were initially developed to recognize the presence of objects in images and produce outputs without any spatial dimension. A comprehensive overview of CNNs and deep learning can be found in Ponti et al. (2017).

In many forestry applications, it is essential to retrieve the location of ITCs in an image. Object detection methods based on deep learning proved useful for this task. Li et al. (2017) developed a framework for oil palm tree detection and counting with VHR images (pixel = 0.6-m) from the QuickBird satellite, achieving 96% of accuracy. Weinstein et al. (2019) proposed a semi-supervised deep learning pipeline for detecting tree crowns based on RGB and Light Detection and Ranging (LiDAR) data and obtained 69% and 60% of recall (producer's accuracy) and precision (user's accuracy), respectively. The methods employed in these studies produced a bounding box that enclosed the ITCs, thus neither delineating its shape nor identifying the species.

Of great interest in the context of tree species mapping is to retrieve the location of pixels encompassing ITCs. Semantic segmentation is the procedure of assigning a label to each pixel of an image. Fully convolutional networks are one of the first approaches that successfully employed deep networks for semantic segmentation. It is based on the transformation of fully connected layers into upsampling or transposed convolutional layers that perform dense pixel predictions. The pioneering work of Long et al. (2015) adapted well-known CNNs models such as AlexNet (Krizhevsky et al., 2012) and VGG (Simonyan and Zisserman, 2014) for semantic segmentation tasks. More recently, Chen

et al. (2018) proposed a model called DeepLabv3+ that uses encoder-decoder (Badrinarayanan et al., 2017) and spatial pyramid pooling (Lazebnik et al., 2006) structures to capture contextual information at multiple scales while preserving object boundaries.

The potential of fully convolutional networks to map tree species in tropical environments with remote sensing images is poorly known. A few pioneering studies successfully employed the U-net architecture (Ronneberger et al., 2015). Wagner et al. (2019) used U-net to map the spatial distribution of *Cecropia hololeuca*, an indicator species of tropical forest disturbance, in WorldView-2 images with an overall accuracy of 97%. Kattenborn et al. (2019) performed fine-grained segmentation of vegetation species and communities on UAV imagery with U-net and obtained at least 84% of accuracy. Morales et al. (2018) used a CNN with the DeepLabv3+ architecture to perform semantic level segmentation of *Mauritia flexuosa* in UAV images and achieved an accuracy of 98%. It is important to note that these studies did not perform species identification at the ITC level; that is, they mapped tree patches belonging to a single species.

Recently, a novel approach called Mask R-CNN combined object detection and semantic segmentation to perform instance segmentation tasks (He et al., 2017). Instance segmentation aims to classify an object at the pixel level and outline its exact shape. The Mask R-CNN first uses a CNN to fit a bounding box that encapsulates the object and then a fully convolutional network to determine the pixels belonging to it. Because the Mask R-CNN is composed of two deep neural networks, it is more complex than semantic segmentation architectures. The extra complexity leads to longer computational time for training and inference (Kulikov et al., 2018).

Moreover, the Mask R-CNN is designed to outline the precise contour of objects, which is challenging in tropical forests. The crown size and shape of a tropical tree can vary widely, and depends, among other factors, on the position of the tree in the forest canopy and neighboring individuals (Hallé et al., 2012). Thus, even trees of the same species can feature distinct crown architectures, and it is usually unlikely to define a crown shape that characterizes a single species.

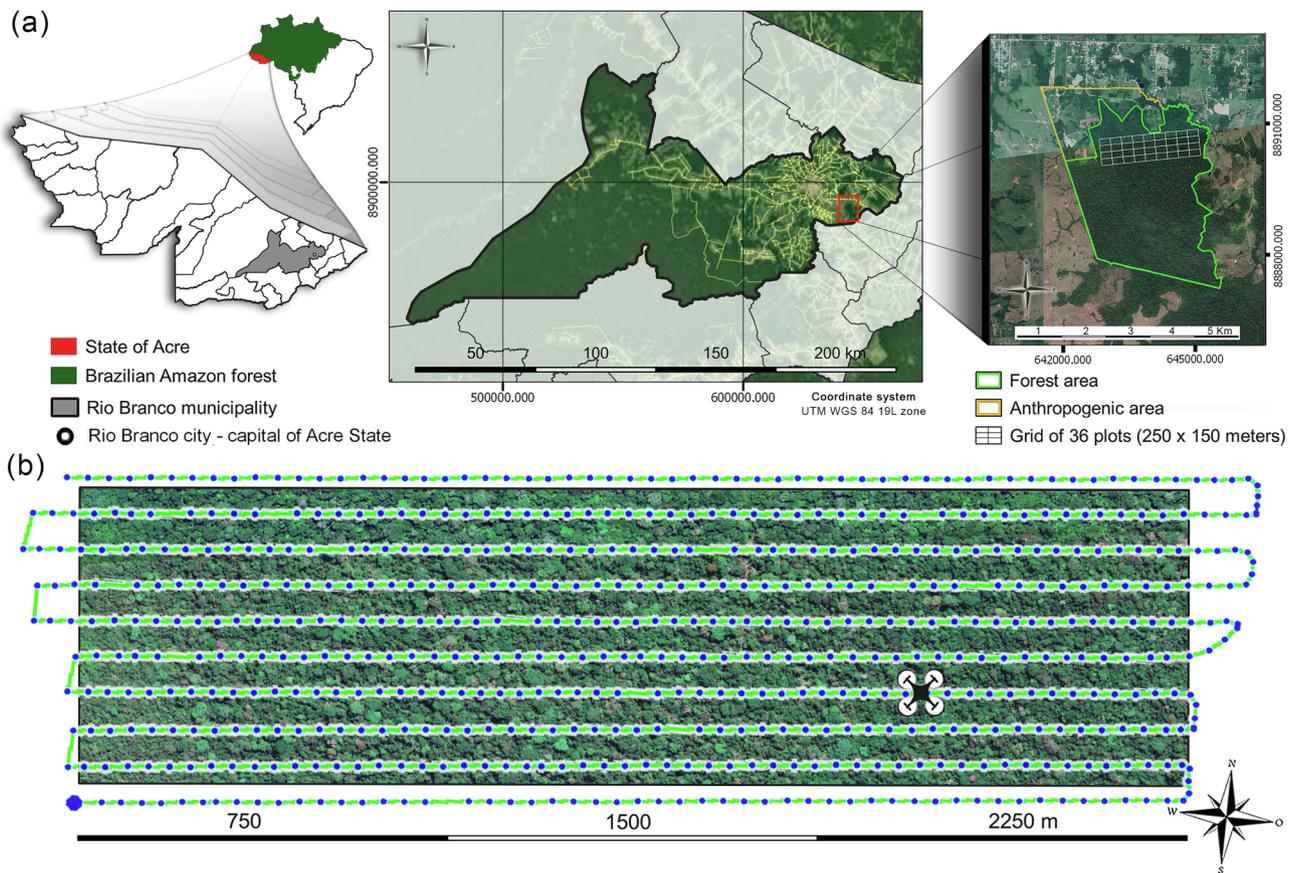
Fully convolutional networks for semantic segmentation provide the opportunity to simultaneously detect and classify ITCs in a more reliable manner and at less computational cost than instance segmentation methods, such as Mask R-CNN. However, these networks tend to blur object boundaries, when applied to high-resolution remote sensing images (Marmanis et al., 2018). The blurring of boundaries is particularly problematic for the detection of ITCs. For example, if two or more trees are close to each other, they will likely be classified as a single tree covered by one segment, limiting our ability to estimate the number of ITCs over an image.

In this study, to avoid the blurring of boundaries effect that is common to semantic segmentation networks, we developed a new method to detect and retrieve the species of ITCs. More specifically, our method is based on morphological operations that are performed on score maps derived from a fully convolutional neural network model. We tested our approach to mapping palm tree species on RGB images acquired by a low-cost UAV over a Brazilian Amazon forest.

## 2. Materials

### 2.1. Study area

The study area is an experimental forest of the Brazilian Agricultural Research Company (Embrapa) located in the municipality of Rio Branco, Acre state, southeast Amazon (10°01'22"S, 67°40'3"W) (Fig. 1a). It is a highly diverse old-growth rain forest area of 1600 ha about 200 m a.s.l. The ortho-image mosaic used in this study comprises 135 ha in the shape of a rectangle of size 600 m  $\times$  2250 m (Fig. 1b). The area annually receives 1950 mm of rain, and the annual average temperature is 24.8( $\pm$  0.8)°C (Ramos et al., 2009). The vegetation of the domain is classified as open rain forest with palms and bamboos; thus



**Fig. 1.** (a) Location of the study area in the Acre state, northwestern Brazil. A Grid of 36 plots of 250 m × 150 m were established within an old-growth rain forest with palms and bamboos. (b) Unmanned aerial vehicle (UAV) flight lines for collecting 1423 RGB images that covered the study area.

highly representative of southeast Amazonian forests (Veloso et al., 1991).

## 2.2. UAV images

The UAV images were collected in March 2017 with the UAV DJI Phantom 4 Professional, which is equipped with an RGB camera of 20-megapixel resolution and a 24 mm autofocus lens. The camera is attached to a three-axis electronic gimbal stabilization system to ensure a nadir view during image collection. The UAV flew 120 m above the forest canopy with a cruising speed of  $13\text{ms}^{-1}$ . A total of 1423 images with endlap of 86% and sidelap of 86.36% were taken in eight consecutive flights (Fig. 1b) that lasted about two hours and were performed between 10:00–12:00 am. Before the flights, three ground control points (GCPs) that form a triangular shape were established in the edges of the forest reserve. At each GCP, a dual-frequency GNSS receiver was installed and collected GPS and GLONASS data for 241 min. After post-processing procedures, the average horizontal and vertical precision of the GCPs was 10 cm and 3 cm. Finally, the scale-invariant features algorithm (SIFT, Lowe (2004)) that is available in the PiX4D software was used to generate an orthomosaic with a ground sampling distance (GSD) of 4 cm of the study area (Fig. 2a).

## 3. Methods

### 3.1. Individual tree crown dataset

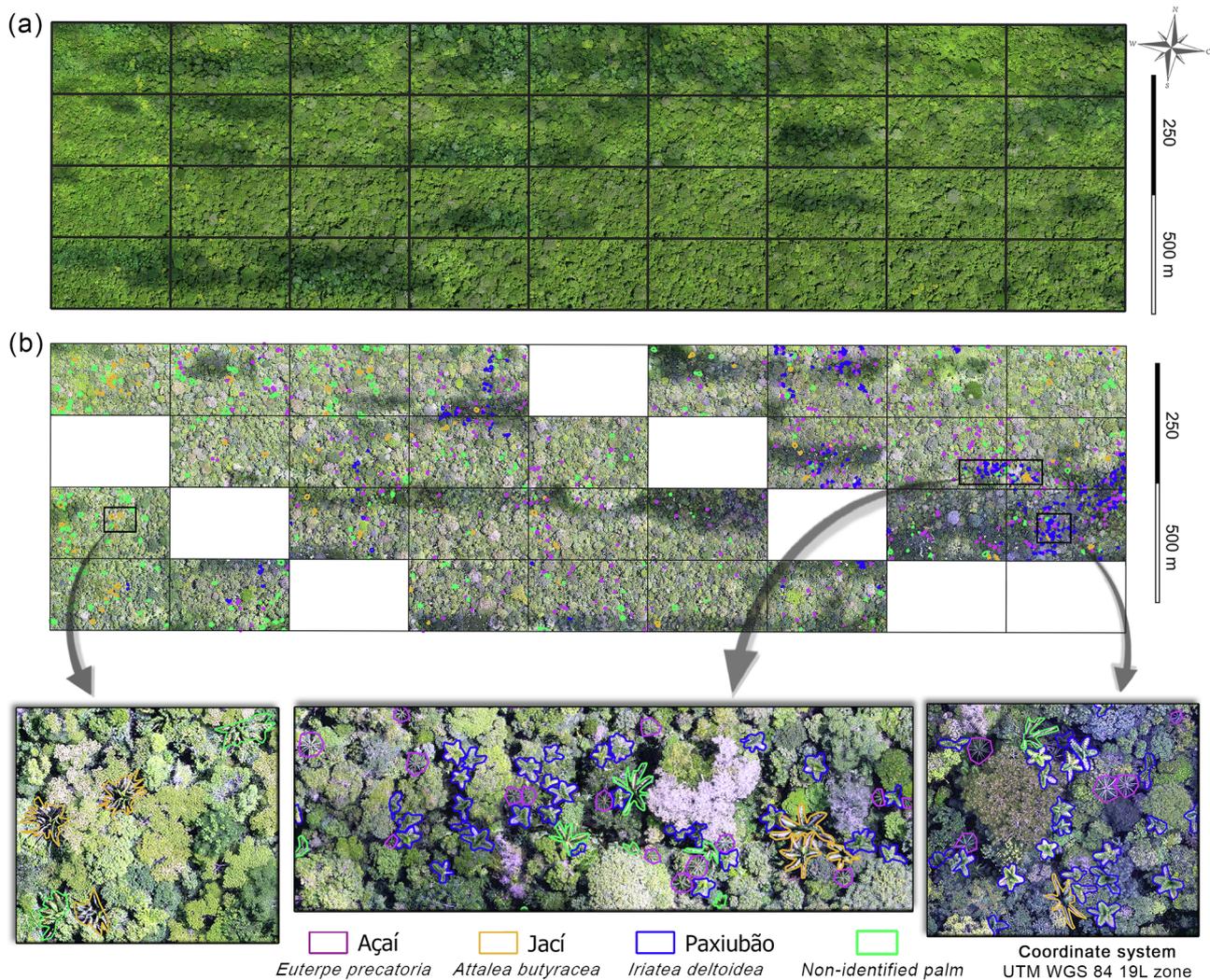
We divided the study area into a grid of 36 plots of approximately 250 m × 150 m (Fig. 2a). An initial visual inspection revealed that 28 plots had a significant number of palms (Fig. 2b). Within each plot, using a true color composition of the UAV orthomosaic on a scale of

1:150, we outlined all palm trees using the QGIS software (QGIS Development Team, 2019) and generated a shapefile of ITCs (Fig. 2b). Each ITC was analyzed by five photo-interpreters with expert knowledge in recognizing palms. Before photointerpretation, a fieldwork was performed to establish a botanical identification key that permitted recognition of species-specific characteristics in the UAV images. We retrieve the species of each tree after a consensus of the photo-interpreters. Their decision regarding the species was based mainly on the characteristics of the crown leaves (Fig. 3). Usually, we followed the arrangement of the branches to outline palm ITCs throughout the image. However, for *E. precatória*, we described a circular shape around the trunk. A precise delineation of *E. precatória* ITCs is very time-consuming because this species features a small crown (<5.1 m in diameter on average, Table 1) and a high number of leaves (>10).

Three species were identified: *Attalea butyracea*, *Euterpe precatória* and *Iriartea deltoidea*. Palms within the plots that did not belong to the species mentioned above were labeled as the class “non-identified palm”. Thus this class contains only trees other than the three target species. Table 1 shows the number of ITCs and the crown diameter of each species, which was estimated by computing the smallest circle that encloses the ITC.

### 3.2. Pixel label dataset

To obtain a pixel label dataset of palm trees, we first performed a vector-to-raster conversion of the shapefile of ITCs. This dataset contains information regarding the class of each UAV image pixel, that is, the three palm species in addition to “non-identified palm” and the background with other types of trees, totaling five classes. Second, both the orthomosaic and the labeled images were cropped to the extent of the plots, thus generating 28 images with their respective labels. Fig. 4



**Fig. 2.** True color compositions of the UAV orthomosaic (pixel = 4cm). (a) Original orthomosaic overlaid with a grid of 36 plots with dimensions 250 m × 150 m. (b) Twenty-eight plots selected for manual individual tree crown (ITC) delineation. All palm ITCs within each plot were manually outlined.

shows an example of the orthomosaic and corresponding labeled image used to train the fully convolutional network model.

### 3.2.1. CNN architecture

Research on deep networks (e.g., [Simonyan and Zisserman, 2014](#); [Shimodaira, 2000](#)) revealed that the number of layers (depth) is directly related to the ability of the network to learn features from data, which reduces the classification errors. This fact motivated researchers to increase network depth in the hope of improving results. However, some experiments (e.g., [Simonyan and Zisserman, 2014](#)) showed that adding more layers to deep networks leads to an increase in the training error.

To overcome the problem, [He et al. \(2016\)](#) proposed a deep residual learning framework called Residual Network (ResNet). A ResNet architecture is composed of several residual blocks that perform skip connections; that is, they forward the activations (feature maps) of a given layer to a deeper layer. For more implementation details, the reader is referred to as [He et al. \(2016\)](#). Common variations of ResNet include ResNet-18, ResNet-50, and ResNet-101 that differ from each other by the number of residual layers. In this study, we used the ResNet-18 architecture ([Fig. 5](#)) that provided a reasonable trade-off between processing time and accuracy in preliminary tests.

For semantic segmentation, we incorporated ResNet-18 in the DeepLabv3+ architecture, which is considered a state-of-the-art deep

learning model for semantic image segmentation. DeepLabv3+ was proposed by [Chen et al. \(2018\)](#) and consists basically of an encoder and a decoder block, as shown in [Fig. 5](#). The encoder module gradually reduces the spatial dimension of the input patch and captures high-level semantic information. The decoder module recovers the size of the patch, retrieving spatial information to produce sharp segmentation results. DeepLabv3+ uses the encoder-decoder structure with atrous convolution ([Chen et al., 2017](#)). This powerful technique allows capturing multi-level features of the input image by controlling the resolution of the output of convolutional layers. Atrous convolutions are applied in a parallel fashion with different rates forming a block called Atrous Spatial Pyramid Pooling (ASPP, [Fig. 5](#)). More information regarding the DeepLabv3+ architecture, including the stride rate of the pooling layers or implementation details, can be found in [Chen et al. \(2018\)](#).

The output generated by ResNet-18 with DeepLabv3+ are score maps for each class. The transposed convolutional layer performs up-sampling with five filters (one per category, [Section 3.2](#)) of size 8×8 and stride 4. Then, the softmax classifier is applied to produce score maps in which each pixel contains class membership probabilities.

### 3.2.2. ITC detection method

Information regarding the number of palm trees within a given forest area is crucial for management purposes. The architecture of the

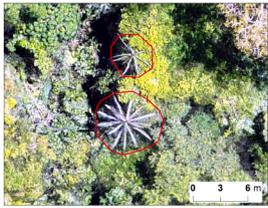
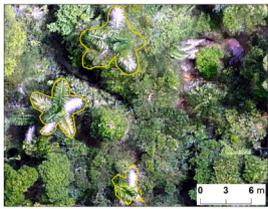
Species (popular name)	Description	Aerial view	Ground view
<i>Attalea butyracea</i> (Jací)	Palm tree with a solitary and robust trunk (DBH* > 30 cm). The leaves are usually more than 15 in number. The arrangement of the leaves at the top of the stem is spiraled. In aerial images, the leaves appear twisted in such a way that the leaflets are not easily visible. The leaf petiole does not exist, and the rachis can easily reach 7 m in length. The canopy radius can vary greatly: when the palm tree is at or above the canopy level, the crown has a radius of 10 m or more. If the palm tree is below the canopy, only the ascending pines will be visible; thus, the canopy radius is smaller. The leaflets are usually more than 150 in number, regularly arranged on each side of the leaf rachis, linear in shape, and arranged in the same plane.		
<i>Euterpe precatoria</i> (Açaí)	Palm tree with a solitary, thin trunk (DBH < 20 cm), usually with about 10 leaves concentrated at the top of the stem. The leaf petiole is 0.12-0.57 m long, and the leaf rachis is 1.6-3.6 m long. For this reason, the crown radius (considering petiole and leaf rachis) rarely reaches 7 m. The leaflets are drooping, linear in shape, regularly spaced along the rachis, and approximately 40 in number. The drooping leaflets expose the rachis and the petiole, which gives this species a peculiar starry pattern in aerial images.		
<i>Iriarteia deltoidea</i> (Paxiubão)	Palm tree with a trunk that is robust, solitary, and ventricous that easily reaches DBH of 30 cm. The leaves are 4 to 7 in number and arranged at the top of the stem. The leaf petioles are short (up to 0.13 m), and the leaf rachis is between 2 to 3 m long, which makes the crown radius of this species rarely exceed 6 m. The leaflets are broad and divided from the base into more than 15 segments arranged in various angles, giving a typical feathery appearance to the leaves. The arrangement of the leaves around the stem is regular, and in aerial photographs, it is possible to see 2 to 4 leaves from the top of the crown. The relatively small canopy radius (< 6 m), combined with the feathery appearance and the small number of visible leaves in aerial photographs, are critical aspects for the identification of this species.		

Fig. 3. Description of the palm species characteristics used to recognize them in UAV images. Ground view photos: Evandro Ferreira. \*DBH = Diameter at breast height.

network employed in this study is designed to perform patch level semantic segmentation, meaning that if two palm trees are close to each other, they will likely be classified as a single palm tree covered by one segment. To overcoming this issue, we propose a new method to detect palm ITCs and identify their corresponding species. The process is based

on morphological operations that are performed at the score maps of each species.

In a morphological operation, the neighborhood of each pixel is considered to adjust its value. Let be a score map of size  $x \times y$  in which the pixels contain class membership probabilities for the species  $\omega_i$ .

Table 1

List of palm trees with crowns manually delineated in the UAV orthomosaic and identified to the species level: species names, number of individual tree crowns (ITCs), mean crown diameter (meters), minimum crown diameter (meters) and maximum crown diameter (meters).

Palm species (popular names)	ITCs	Mean crown diameter	Min crown diameter	Max crown diameter
<i>Attalea butyracea</i> (Jací)	84	10.5	4.8	16.7
<i>Euterpe precatoria</i> (Açaí)	403	5.1	2.6	7.7
<i>Iriarteia deltoidea</i> (Paxiubão)	263	6.4	2.2	9.9
Non-identified palm	265	8.4	2.6	17.7

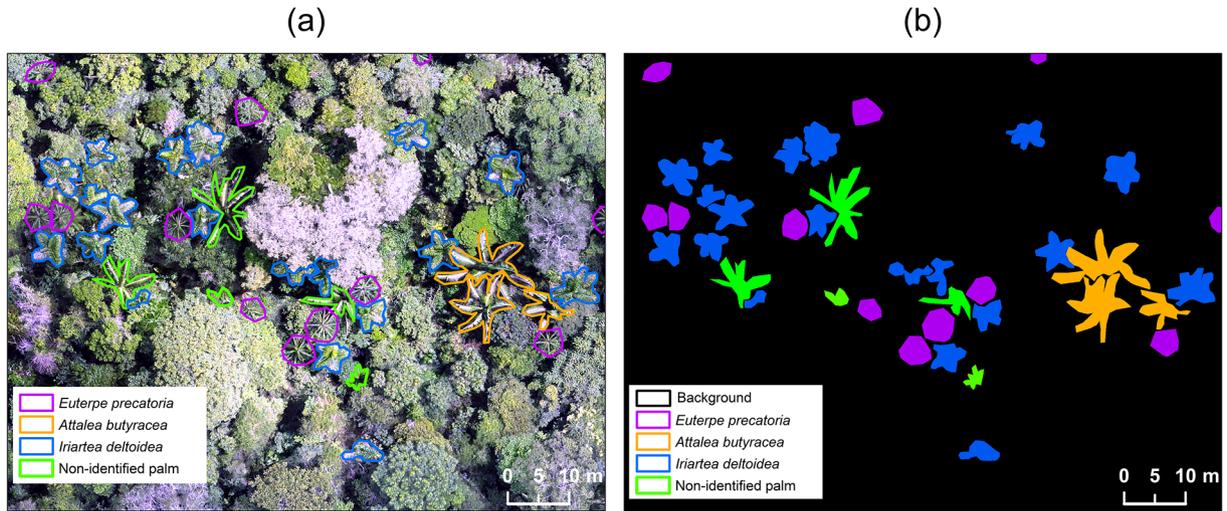


Fig. 4. (a) A true color composition of the UAV orthomosaic overlaid with manual outlined individual tree crowns (ITCs) (b) labeled image obtained after a vector-to-raster conversion of (a).

First, we define regions of interests (ROIs) in the  $P_{oi}(x, y)$  image by selecting only the pixels with class membership probabilities that are higher than 0.8. The threshold of 0.8 was effective in removing spurious pixels while preserving pixels of the target palm species. Second, within the ROIs, we perform an erosion operation using a disk-shaped structuring element  $S(p, q)$  of radius  $r$ , which depended on the mean crown diameter of the species (Table 1). For *E. precatorea* and *I. deltoidea*  $r$  equals to seven pixels, while for *A. butyracea* and the class non-identified palm  $r$  was set to 20 pixels. The main objective of erosion is to identify high-class membership values that are related to treetops. It is defined as:

$$P \ominus S = \min_{p,q} \{P(x+p, y+q) - S(p, q) | (p, q) \in D_S\} \quad (1)$$

in which  $D_S$  is the domain of the structuring element  $S$ . Third, to identify connected blobs of pixels within the ROIs, we perform the morphological opening, which is an erosion followed by dilation with  $S(p, q)$  for both operations. Dilation is the opposite of erosion:

$$P \oplus S = \max_{p,q} \{P(x+p, y+q) - S(p, q) | (p, q) \in D_S\} \quad (2)$$

Mathematically, the morphological opening is defined as:

$$P \circ S = (P \ominus S) \oplus S \quad (3)$$

Fourth, we performed another dilation operation with the structuring element  $S(p, q)$  to increase the size of the objects detected in the previous step. Finally, we identify regional maxima locations, that is, a set of pixels with similar values, surrounded by pixels with lower values. For the sake of clarity, Fig. 6 shows the morphological operations described above applied in a synthetic image. The score map of each species is processed individually and merged to obtain a species map of the testing images.

### 3.3. Experimental set-up

First, we partitioned the plots following the proportion of 80% (22 plots) for training and 20% (6 plots) for testing. The partition procedure was repeated 30 times. At each realization, different plots were randomly chosen to train and test the fully convolutional network model, which allowed us to compute the variability in the classification

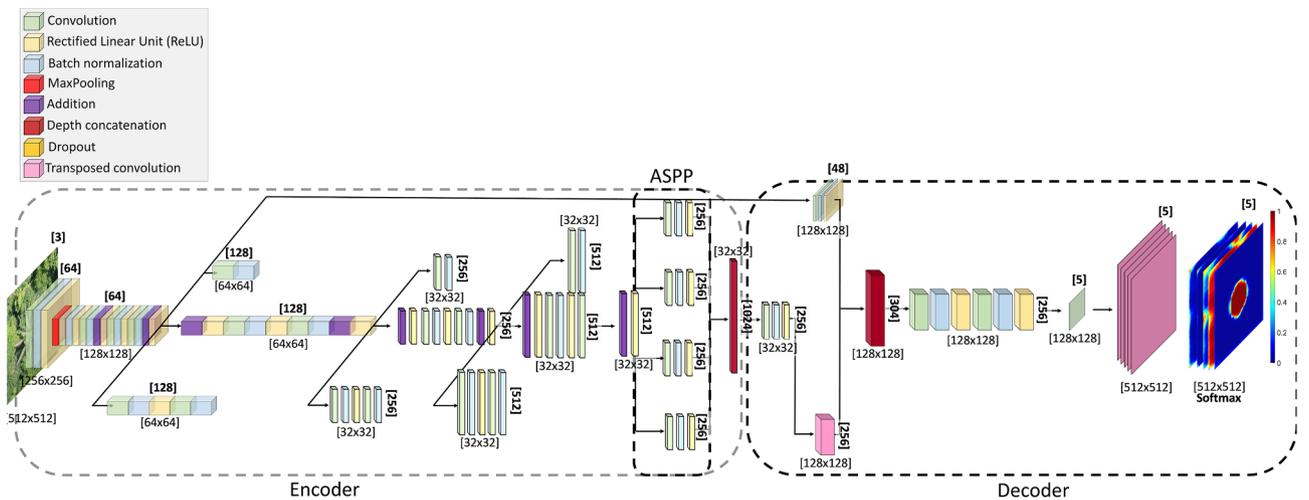
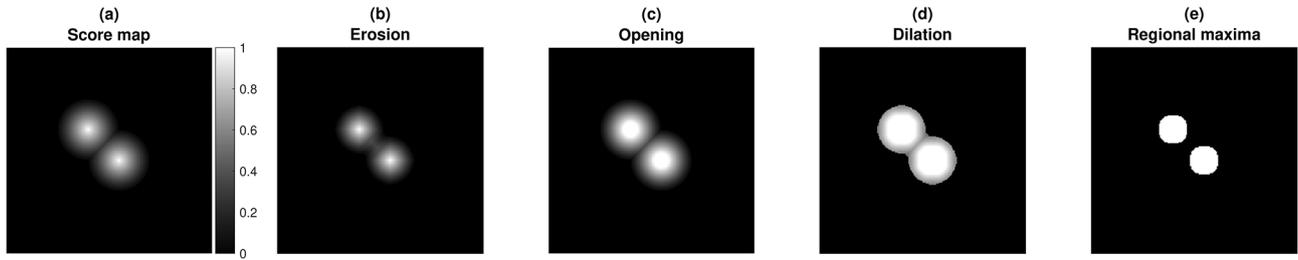


Fig. 5. ResNet-18 incorporated into the DeepLabv3+ (Chen et al., 2017) architecture. The encoder module gradually reduces the spatial dimension of the input image patch by up to 16, while the decoder module recovers the spatial dimension. The numbers in brackets highlighted in bold refer to the number of feature maps (filters) of each convolutional block. Atrous Spatial Pyramid Pooling (ASPP) is applied to capture multi-level features of the input image. A transposed convolution operation upsamples the original spatial dimension of the patch. Then, the softmax classifier is applied to generate score maps for the classes. These maps contain class membership probabilities for each pixel of the patch.



**Fig. 6.** Illustration of the morphological operations of the individual tree crown (ITC) detection method performed in a synthetic image. (a) score map of a given species showing two ITCs that are close to each other. In this image, each pixel contains the probability of species occurrence. (b) result of an erosion operation (Eq. 1) performed within a region of interest (ROI) that contains only the pixels with class membership probabilities that are higher than 0.8. (c) result of the morphological opening (Eq. 2) of (b). (d) Dilation (Eq. 2) of (c). (e) Regional maxima of (d) showing two ITCs of a given species apart from each other.

accuracy, depending on the data used in the classification process. After selecting the training plots, we concatenated them into a single image. We extracted randomly positioned patches of size 512×512 pixels from both the orthomosaic and the labeled images to feed the network. The images did not contain border pixels (zero value pixels). Thus, the model was trained with five classes (Fig. 4), that is, four palm classes and the background. The inclusion of the background class in the training process was necessary for the network to learn how to distinguish palms from other types of trees.

The mini-batch size was eight, and the maximum number of epochs was 15. A mini-batch is a subset of the training data that is used by the stochastic gradient descent with momentum (SGDM) algorithm (Murphy, 2012) to update the network parameters (weights and biases). An epoch is a full pass over the entire training set, composed of 2000 random patches. At each epoch, a different set of 2000 patches were extracted from the images. Random reflection and rotation augmentation of the patches were used during training. Data augmentation is a common practice to prevent the network from overfitting and consists of random rotations and flips of the input images.

To avoid biased training of the network due to class imbalance, we used class weighting to balance the classes. We calculated the class weights by the ratio between the median frequency of all classes to the frequency of each class. The weights of the ResNet-18 model were initialized with pre-trained values of the ImageNet database (Deng et al., 2009), and the learning rate was 0.001. The score maps of all species in the testing phase were produced patches-wise on patches of size 2000×2000 pixels.

Training and inference were performed on a desktop workstation with an Intel Core i7-8700 3.2 GHz CPU, 24 GB of main memory, and an NVIDIA® GeForce Titan V GPU with 12 GB of dedicated memory and 5120 CUDA® cores. All image processing procedures were performed in MATLAB® environment.

### 3.4. Accuracy assessment

The accuracy assessment was performed with the test images, which were not used for training, and their respective labels. The results are presented by confusion matrices that show the percentage of correctly

**Table 2**

Example of a confusion matrix showing the percentage of correctly classified pixels in the diagonal cells (highlighted in bold) and the misclassification rate between the species in the off-diagonal cells.

Predicted class	True class				Producer's accuracy
	<i>E. precatória (E)</i>	<i>A. butyracea (A)</i>	<i>I. deltoidea (I)</i>	Non-identified palm (N)	
<i>E. precatória (E)</i>	<b><math>TP_{EE}</math></b>	$M_{EA}$	$M_{EI}$	$M_{EN}$	$TP_{EE} + M_{EA} + M_{EI} + M_{EN}$
<i>A. butyracea (A)</i>	$M_{AE}$	<b><math>TP_{AA}</math></b>	$M_{AI}$	$M_{AN}$	$TP_{AA} + M_{AE} + M_{AI} + M_{AN}$
<i>I. deltoidea (I)</i>	$M_{IE}$	$M_{IA}$	<b><math>TP_{II}</math></b>	$M_{IN}$	$TP_{II} + M_{IE} + M_{IA} + M_{IN}$
Non-identified palm (N)	$M_{NE}$	$M_{NA}$	$M_{NI}$	<b><math>TP_{NN}</math></b>	$TP_{NN} + M_{NE} + M_{NA} + M_{NI}$
User's accuracy	$TP_{EE} + M_{AE} + M_{IE} + M_{NE}$	$TP_{AA} + M_{EA} + M_{IA} + M_{NA}$	$TP_{II} + M_{EI} + M_{AI} + M_{NI}$	$TP_{NN} + M_{EN} + M_{AN} + M_{IN}$	

classified pixels per species and the misclassification rate between all pairwise combinations of species. An example of a confusion matrix for a 4-class classification problem is shown in (Table 2).

In the confusion matrix,  $TP_{ii}$  is the percentage of true positives of species (i), and  $M_{ij}$  is the number of pixels that truly belong to species (i) but were classified as species (j). The producer's accuracy (PA), also known as recall, represents the probability that the pixels of a particular species in the reference are correctly classified. PA is computed as the ratio between the number of correctly identified pixels of a given species divided by the number of pixels of that species:

$$PA_i = \frac{TP_i}{TP_i + \sum_j M_{ij}} \tag{4}$$

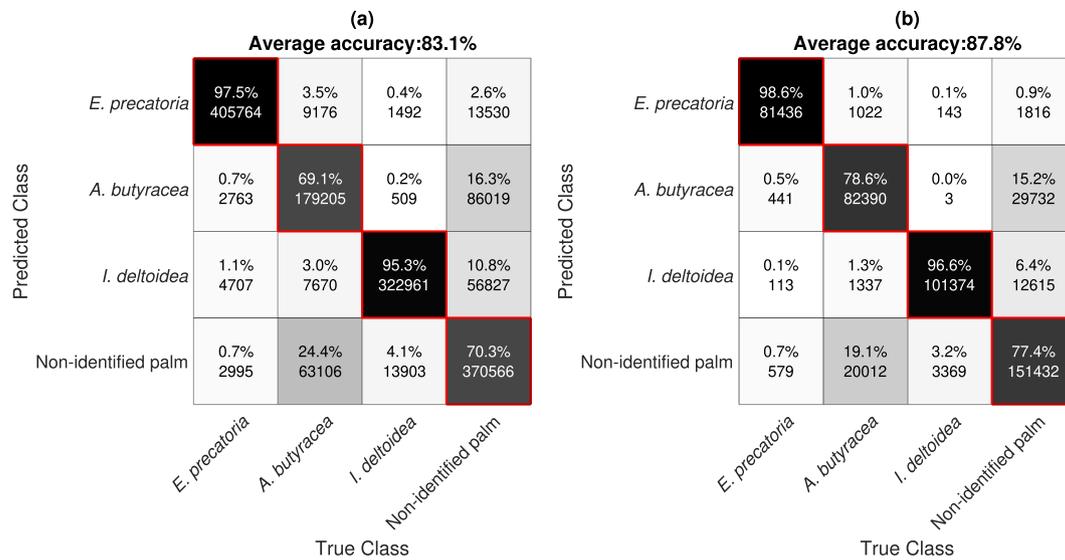
The user's accuracy (UA), also known as precision, represents the probability that a pixel classified as a given species represents that species in the reference. UA is computed as the ratio between the number of correctly identified pixels of a given species to the number of pixels that have been predicted by the classifier as a that species:

$$UA_i = \frac{TP_i}{TP_i + \sum_j M_{ij}} \tag{5}$$

Finally, we computed the percentage of correctly detected ITCs per species at each realization, which allowed us to assess the performance of the ITC detection method (Section 3.2.2). The ITC of a given species was considered to be correctly detected if it intersects a single reference ITC. For the sake of comparison, we computed the classification accuracy, and the ITC detection rate obtained using conventional patch-wise semantic segmentation (from now on referred to as CSS).

## 4. Results

The mean of the producer's accuracy and misclassification rates of each palm species, obtained after training and testing the model 30 times with different plots, are shown in Fig. 7. At each realization, the computation time to train the network was 45±3 min, while the production of score maps of the species and application of the ITC



**Fig. 7.** Confusion matrices that show the classification accuracy of Amazonian palm species in UAV images. The producer’s accuracy of each species is distributed along with the diagonal cells (highlighted in red). The percentage of misclassification between a pairwise combination of species is shown in the off-diagonal cells. Each cell contains absolute values (pixels) and relative percentages. (a) Refers to the conventional patch-wise semantic segmentation and (b) to the ITC detection method (Section 3.2.2). The classification was performed using the ResNet-18 model adapted for semantic segmentation by the DeepLabv3+ architecture (Fig. 5). The model was trained and tested with UAV images and their respective labeled images (Fig. 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

detection method lasted 15±3 min. The mean of the producer’s accuracy was 83.1% and 87.8% for the CSS and the proposed ITC detection method. The main difference between the two approaches was related to *A. butyracea* that was classified with 69.1% of accuracy by CSS and 78.6% by the proposed method. Non-identified palms showed the lowest classification success, given the misclassification rate concerning *A. butyracea* (16.3% for CSS and 15.2% for the ITC detection method). *E. precatorea* and *I. deltoidea* were classified with more than 97% of accuracy for both approaches.

On average, the pixel-wise classification accuracy, for both the CSS and the proposed approach, did not differ significantly. For example, the difference between the average producer’s accuracy of CSS and the proposed method was 4.7%, and in terms of average user’s accuracy, this difference reached 5.3% (Table 3). However, the proposed method was able to detect, on average correctly, 34.7 percentage points more ITCs than CSS. Specifically, the ITC detection rate for *A. butyracea*, *E. precatorea*, *I. deltoidea* and non-identified palms were, respectively, 37.3, 31, 43.5, and 27.1 percentage points superior to the proposed method.

A comparison between the CSS and the proposed approach is depicted in Fig. 8. In the first example (Fig. 8a), it is shown the ITC detection and species classification of *E. precatorea* and *A. butyracea*. The steps of the proposed approach are illustrated in the central portion of the figure. One can note how the method gradually detects palm trees through morphological operations performed in areas of high-class

membership (>0.8) within the score maps of each species. The last column of Fig. 8a shows that the proposed approach successfully separated each palm tree present in the image, while the CSS merged the ITC of *A. butyracea* with those from non-identified palms (depicted in green). Another example is presented in Fig. 8b that shows three individuals of *E. precatorea* that are close to each other. It is noteworthy that the CSS merged the three ITCs into one segment, while the proposed approach separated them.

### 5. Discussion

We developed a new method based on fully convolutional neural networks to simultaneously detect and classify Amazonian palm species using RGB images acquired by a low-cost UAV platform. Previous works demonstrated the potential of deep learning methods to identify palm trees in monoculture plantations (Li et al., 2019; Zheng et al., 2019) or oligarchic tropical forest formations (Morales et al., 2018). However, to the best of our knowledge, this is the first study demonstrating the potential of UAV-borne images and deep learning to classify multiple palm species at the ITC level in Amazonian forests.

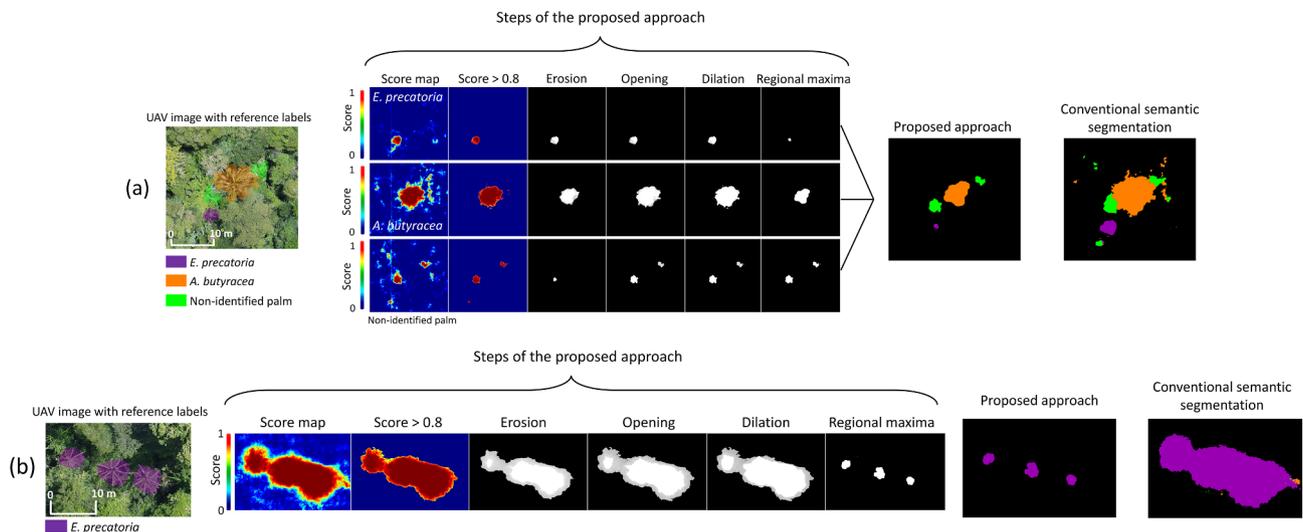
Our results show that *E. precatorea* and *I. deltoidea* can be mapped with high accuracy rates (>97%). This classification success is attributed to species-specific crown architectures -the leaflets of *E. precatorea* are dropping, and the leaf rachis is usually long (>3 m), attributing to the crown a peculiar starry pattern in UAV images (Fig. 3, Fig. 2b). The

**Table 3**

Producer’s accuracy, user’s accuracy, and individual tree crown (ITC) detection rate for the conventional patch-wise semantic segmentation (CSS) and the proposed approach (Section 3.2.2).

Species (popular names)	Producer’s accuracy (mean±SD,%)		User’s accuracy (mean±SD,%)		ITC detection rate (mean±SD,%)	
	CSS*	Proposed approach	CSS*	Proposed approach	CSS*	Proposed approach
<i>Attalea butyracea</i> (Jací)	69.1±7.7	78.6±5.5	66.7±6.9	73.2±5.3	34.1±8	71.4±6.1
<i>Euterpe precatorea</i> (Açaí)	97.5±1.4	98.6±1.4	94.4±4.1	96.5±3.5	64.6±7.8	95.6±4.4
<i>Iriartea deltoidea</i> (Paxiubão)	95.3±4.3	96.6±3.4	82.3±5.6	87.8±4.7	44.2±8.9	87.7±4.3
Non-identified palm	70.3±6	77.4±6.2	82.2±5.8	86.3±4.3	41.3±7.5	68.5±5.1
Average	83.1±4.9	87.8±4.4	81.4±5.6	86.0±4.5	46.1±8.1	80.8±5.0

[\*]Conventional patch-wise semantic segmentation



**Fig. 8.** Two illustration examples of individual tree crown (ITC) detection and species classification of Amazonian palms using conventional patch-wise semantic segmentation (CSS) and the proposed approach (Section 3.2.2). Both approaches are based on the score maps of the palm species (Table 1) obtained by the ResNet-18 model incorporated in the DeepLabv3+ architecture (Fig. 5). The results of the steps of the proposed approach are illustrated in the central portion of (a) and (b). One can note how the method gradually detects the palm trees through morphological operations performed in areas of high-class membership (score >0.8) within the score maps of each species. The last columns of (a) and (b) show a comparison between CSS results and the proposed approach. It is worth noting that CSS merged into single segments the trees that were close to each other, while the proposed method separated them.

crown of *I. deltoidea* are likewise very characteristic. The leaflets are broad and divided into several segments that confer a feathery appearance to the leaves. Moreover, the crown radius is small (< 6 m), and the leaves are regularly arranged at the top of the stem (Fig. 3). The CNN model was able to learn the above-mentioned species-specific characteristics in UAV images. Such ability of CNNs is explained by the fact that they are inspired by the human visual cortex (Cadieu et al., 2014). It is worth noting that the photo-interpreters easily identified ITCs of *E. precatória* and *I. deltoidea* during the construction of the ITC dataset (Section 3.1). However, both CNNs and photo-interpreters are prone to errors. For example, we observed relatively high rates (>18%) of misclassification between *A. butyracea* and non-identified palms. The classification errors between these classes may arise from the similarity of their canopy structures. Moreover, the number *A. butyracea* individuals in our dataset were small if compared to the other species (Table 1), which limited the extraction of its species-specific characteristics by the network.

Given the data-driven nature of deep learning, large amounts of data are required to successfully train a model. In the case of tree species, ground truth information needed to construct pixel label datasets to train a model is difficult and costly to obtain, particularly in tropical forests. Here, we used expert knowledge of photo-interpreters to outline ITCs of target palm species manually in the UAV orthomosaic. The visual interpretation was only possible given the hyperspatial resolution, which allowed clear visualization of palm crowns and identification of species-specific characteristics. The ground sampling distance (pixel = 4 cm) of the UAV orthomosaic played an essential role in the accurate identification and fine-grained segmentation of palm species. The transferability of our approach to different data sets, composed of images acquired with varying flight altitudes or other cameras, for example, requires further investigation.

Regarding the performance of the fully convolutional network model (ResNet-18 and DeepLabv3+ architecture) combined with the ITC detection method in our study proved to be effective. In this realm, our results agree with Morales et al. (2018) who used the DeepLabv3+ architecture to perform accurate semantic level segmentation of the *Mauritia flexuosa* palm in swamps of the Peruvian Amazon with 98.1% of classification accuracy. It is worth noting that Morales et al. (2018) performed conventional patch-wise semantic segmentation, thus not

detecting individual trees. Tagle Casapia et al. (2020) developed an object-based classification scheme for the identification and quantification of Amazonian palms using UAV images. From the confusion matrix provided by the authors, the producer's accuracies for *A. butyracea* and *E. precatória* were of 74.6% and 90.4%, respectively. The method developed by Tagle Casapia et al. (2020) was based on image segmentation (region growing) to outline ITCs before the classification step. The delineation of ITCs in tropical forests is challenging due to the high degree of crown overlapping between neighboring trees. The choice of the segmentation parameters is usually arbitrary, limiting the reproducibility of the approach to other areas.

The commercial exploitation and management of palm species in the Amazon region are usually based on *in situ* counting of individual trees (e.g., Lopes et al., 2019; Muñiz-Miret et al., 1996). Thus, delineating the exact border/shape of the ITCs is less critical than detecting its location in the image. For this reason, we avoided using standard instance segmentation methods such as the Mask R-CNN (He et al., 2017). Moreover, the crown shape of a given species depends on how it was manually delineated. In the majority of cases, we followed the arrangement of the branches to outline palm ITCs throughout the image, except for *E. precatória* in which we described a circular shape around the trunk. Trees of *E. precatória* feature small crowns ( $\approx 5$  m, Table 1) and usually more than ten leaves (Fig. 3), which require considerable effort and time to perform a precise ITC delineation. Thus, we decided to describe a circular shape around the trunk to outline ITCs of *E. precatória*. It is important to note that our choice did not impact model performance because *E. precatória* reached the highest classification accuracy (> 97.5%).

The training process of a CNN involves the use of image patches of predefined size. In our work, we extracted randomly positioned patches from the UAV orthomosaic and corresponding pixel label images to feed the networks with training data (see Section 3.3). The application of such a random patch extraction procedure was possible because we carefully performed the manual delineation of all palm ITCs within each 250 m  $\times$  150 m plots (Fig. 1b). The suitability of random patch extraction to train CNNs for tree species mapping in highly diverse sites such as tropical forests still requires further investigation, given the difficulty to ensure that the target species is not present in a patch that is labeled as "background".

Our study was based on UAV images acquired with only three (RGB) channels. It is worth noting that overlapping RGB images from UAVs provide the opportunity to derive point clouds for three-dimensional (3D) characterization. The inclusion of 3D information in the classification process of palm species is an exciting research opportunity, considering that each species features a unique crown architecture.

The high accuracy rates (> 95%) achieved for some palm species suggest that spectral information played a secondary role in the classification process, which contradicts results from previous studies showing that the classification success usually depends on high spectral resolutions (Fassnacht et al., 2016). However, recent technological advancements enabled the manufacture of UAV hyperspectral sensors (Sankey et al., 2017), thus allowing both hyperspatial and hyperspectral resolution. Further research is needed to assess the potential of such sensors for palm species mapping in tropical environments.

## 6. Conclusions

Our study shows the potential of RGB images acquired by a low-cost UAV and deep learning to map the spatial distribution of palm species in Amazonian forests. We developed a method based on the score maps derived from a fully convolutional network model to detect and classify ITCs of three palm species (*Attalea butyracea*, *Euterpe precatoria* and *Iriartea deltoidea*). *Euterpe precatoria* (popularly known as *Açaí*) and *Iriartea deltoidea* (commonly known as *Paxiubão*) were detected, respectively, with  $98.6 \pm 1.4\%$  and  $96.6 \pm 3.4\%$  of accuracy. Moreover, our method outperformed conventional patch-wise semantic segmentation since it successfully detected trees that were close to each other, producing accurate maps of palms occurrence in Amazonian forests. Maps of the spatial distribution of *Euterpe precatoria* can significantly assist management projects in the Amazon, providing a valuable tool to aid decision making and community-based forest monitoring programs. Finally, the approach developed in this study can be applied to other forest areas where high-resolution UAV images (with at least 4 cm pixels) are available.

## CRedit authorship contribution statement

**Matheus Pinheiro Ferreira:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Danilo Roberti Alves de Almeida:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Daniel de Almeida Papa:** Conceptualization, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Juliano Baldez Silva Minervino:** Resources, Data curation. **Hudson Franklin Pessoa Veras:** Resources, Data curation. **Arthur Formighieri:** Resources, Data curation. **Caio Alexandre Nascimento Santos:** Resources, Data curation. **Marcio Aurélio Dantas Ferreira:** Resources, Data curation. **Evandro Orfanó Figueiredo:** Resources, Data curation, Writing - original draft. **Evandro José Linhares Ferreira:** Supervision, Writing - original draft, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research. D. Almeida was supported by the São Paulo Research Foundation (#2018/21338-3 and #2019/14697-0).

## References

- 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 Brasílica* 32, 527–539. <https://doi.org/10.1590/0102-33062017abb0400>.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>.
- Baker, W.J., Dransfield, J., 2016. Beyond genera palmarum: progress and prospects in palm systematics. *Bot. J. Linn. Soc.* 182, 207–233.
- Bernal, R., Torres, C., García, N., Isaza, C., Navarro, J., Vallejo, M.I., Galeano, G., Balslev, H., 2011. Palm management in south america. *Botanical Rev.* 77, 607–646. <https://doi.org/10.1007/s12229-011-9088-6>.
- Cadieu, C.F., Hong, H., Yamins, D.L., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J., 2014. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Comput. Biol.* 10, e1003963. <https://doi.org/10.1371/journal.pcbi.1003963>.
- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Machine Intell.* 40, 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>.
- Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818.
- Cho, M.A., Malahlela, O., Ramoelo, A., 2015. Assessing the utility worldview-2 imagery for tree species mapping in south african subtropical humid forest and the conservation implications: Dukuduku forest patch as case study. *Int. J. Appl. Earth Obs. Geoinf.* 38, 349–357. <https://doi.org/10.1016/j.jag.2015.01.015>.
- Clark, M.L., Roberts, D.A., Clark, D.B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens. Environ.* 96, 375–398. <https://doi.org/10.1016/j.rse.2005.03.009>.
- 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*. Ieee, pp. 248–255.
- Fassnacht, F.E., Latifi, H., Sterenczak, K., Mdozewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* 186, 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>.
- Feret, J., Asner, G.P., 2013. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sens.* 51, 73–84. <https://doi.org/10.1109/TGRS.2012.2199323>.
- Ferreira, M.P., Wagner, F.H., Aragão, L.E., 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 J. Photogram. Remote Sens.* 149, 119–131. <https://doi.org/10.1016/j.isprsjprs.2019.01.019>.
- Ferreira, M.P., Zorzea, M., Zanotta, D.C., Shimabukuro, Y.E., de Souza Filho, C.R., 2016. Mapping tree species in tropical semi-deciduous forests with hyperspectral and multispectral data. *Remote Sens. Environ.* 179, 66–78. <https://doi.org/10.1016/j.rse.2016.03.021>.
- Hallé, F., Oldeman, R.A., Tomlinson, P.B., 2012. *Tropical Trees and Forests: An Architectural Analysis*. Springer Science & Business Media.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Kattenborn, T., Eichel, J., Fassnacht, F.E., 2019. Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery. *Sci. Rep.* 9, 2045–2322. <https://doi.org/10.1038/s41598-019-53797-9>.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Adv. Neural Informat. Process. Syst.* 1097–1105.
- Kulikov, V., Yurchenko, V., Lempitsky, V., 2018. Instance segmentation by deep coloring. arXiv preprint arXiv:1807.10007.
- Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pp. 2169–2178. doi:10.1109/CVPR.2006.68.
- 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 Sens.* 11, 11. <https://doi.org/10.3390/rs11010011>.
- Li, W., Fu, H., Yu, L., Cracknell, A., 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* 9, 22.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>.
- Lopes, E., Soares-Filho, B., Souza, F., Rajão, R., Merry, F., Ribeiro, S.C., 2019. Mapping the socio-ecology of non timber forest products (ntfp) extraction in the brazilian amazon: The case of açaí (*euterpe precatoria* mart) in acre. *Landscape Urban Plann.* 188, 110–117. <https://doi.org/10.1016/j.landurbplan.2018.08.025>. landscape dynamics of family forest owners.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60, 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- Marmaris, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018.

- Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS J. Photogramm. Remote Sens.* 135, 158–172.
- Morales, G., Kemper, G., Sevillano, G., Arteaga, D., Ortega, I., Telles, J., 2018. Automatic segmentation of *mauritia flexuosa* in unmanned aerial vehicle (uav) imagery using deep learning. *Forests* 9. <https://doi.org/10.3390/f9120736>.
- Muñiz-Miret, N., Vamos, R., Hiraoka, M., Montagnini, F., Mendelsohn, R.O., 1996. The economic value of managing the açafá palm (*euterpe oleracea* mart.) in the floodplains of the amazon estuary, Pará, Brazil. *For. Ecol. Manage.* 87, 163–173.
- Murphy, K.P., 2012. *Machine Learning: A Probabilistic Perspective*. MIT press.
- Omer, G., Mutanga, O., Abdel-Rahman, E.M., Adam, E., 2015. Performance of support vector machines and artificial neural network for mapping endangered tree species using worldview-2 data in dukuduku forest, south africa. *IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens.* 8, 4825–4840. <https://doi.org/10.1109/JSTARS.2015.2461136>.
- 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 Ecol. Manage.* 411, 35–45.
- Paneque-Gálvez, J., McCall, M.K., Napoletano, B.M., Wich, S.A., Koh, L.P., 2014. Small drones for community-based forest monitoring: An assessment of their feasibility and potential in tropical areas. *Forests* 5, 1481–1507. <https://doi.org/10.3390/f5061481>.
- 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. *Appl. Veg. Sci.* 15, 548–559. <https://doi.org/10.1111/j.1654-109X.2012.01196.x>.
- Peters, C.M., 1992. The ecology and economics of oligarchic forests. *Adv. Econ. Botany* 9, 15–22. URL <http://www.jstor.org/stable/43931385>.
- Ponti, M.A., Ribeiro, L.S.F., Nazare, T.S., Bui, T., Collomosse, J., 2017. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: 2017 30th SIBGRAP Conference on Graphics, Patterns and Images Tutorials (SIBGRAP-T), pp. 17–41. doi:10.1109/SIBGRAP-T.2017.12.
- QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation. URL <http://qgis.osgeo.org>.
- Ramos, A.M., dos Santos, L.A.R., Fortes, L.T.G., 2009. Normais climatológicas do Brasil, 1961-1990. Instituto Nacional de Meteorologia-INMET, Ministério da Agricultura.
- Rocha, E., 2004. Potencial ecológico para o manejo de frutos de açazeiro (*euterpe precatoria* mart.) em áreas extrativistas no acre, brasil. *Acta amazônica* 34, 237–250. <https://doi.org/10.1590/S0044-59672004000200012>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, pp. 234–241.
- Sankey, T., Donager, J., McVay, J., Sankey, J.B., 2017. Uav lidar and hyperspectral fusion for forest monitoring in the southwestern usa. *Remote Sens. Environ.* 195, 30–43. <https://doi.org/10.1016/j.rse.2017.04.007>.
- Shimodaira, H., 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Stat. Plann. Inference* 90, 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 preprint arXiv:1409.1556.
- Tagle Casapia, X., Falen, L., Bartholomeus, H., Cárdenas, R., Flores, G., Herold, M., Honorio Coronado, E.N., Baker, T.R., 2020. Identifying and quantifying the abundance of economically important palms in tropical moist forest using uav imagery. *Remote Sens.* 12, 9.
- Veloso, H.P., Rangel-Filho, A.L.R., Lima, J.C.A., 1991. *Classificação da vegetação brasileira, adaptada a um sistema universal*. Ibge.
- Wagner, F.H., Sanchez, A., Tarabalka, Y., Lotte, R.G., Ferreira, M.P., Aida, M.P., Gloor, E., Phillips, O.L., Aragao, L.E., 2019. Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images. *Remote Sens. Ecol. Conservat.*
- 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, 1309.
- Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* 4, 22–40. <https://doi.org/10.1109/MGRS.2016.2540798>.
- Zheng, J., Li, W., Xia, M., Dong, R., Fu, H., Yuan, S., 2019. Large-scale oil palm tree detection from high-resolution remote sensing images using faster-rcnn, in: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 1422–1425. <https://doi.org/10.1109/IGARSS.2019.8898360>.