



OPEN

Mauritia flexuosa palm trees airborne mapping with deep convolutional neural network

Luciene Sales Dagher Arce¹, Lucas Prado Osco², Mauro dos Santos de Arruda¹, Danielle Elis Garcia Furuya³, Ana Paula Marques Ramos^{3,5}, Camila Aoki¹, Arnildo Pott¹, Sarah Fatholahí⁴, Jonathan Li⁴, Fábio Fernando de Araújo^{5✉}, Wesley Nunes Gonçalves¹ & José Marcato Junior¹

Accurately mapping individual tree species in densely forested environments is crucial to forest inventory. When considering only RGB images, this is a challenging task for many automatic photogrammetry processes. The main reason for that is the spectral similarity between species in RGB scenes, which can be a hindrance for most automatic methods. This paper presents a deep learning-based approach to detect an important multi-use species of palm trees (*Mauritia flexuosa*; i.e., Buriti) on aerial RGB imagery. In South-America, this palm tree is essential for many indigenous and local communities because of its characteristics. The species is also a valuable indicator of water resources, which comes as a benefit for mapping its location. The method is based on a Convolutional Neural Network (CNN) to identify and geolocate singular tree species in a high-complexity forest environment. The results returned a mean absolute error (MAE) of 0.75 trees and an F1-measure of 86.9%. These results are better than Faster R-CNN and RetinaNet methods considering equal experiment conditions. In conclusion, the method presented is efficient to deal with a high-density forest scenario and can accurately map the location of single species like the *M. flexuosa* palm tree and may be useful for future frameworks.

The unplanned development and land occupation in both urban and rural areas are the main reasons behind deforestation, contributing to environmental degradation in riparian zones and modifying the natural landscape. Multidisciplinary research is necessary to ascertain the population of vegetative species, identifying their locations and distribution patterns. Such information is essential for the management and conservation of vulnerable ecosystems, and mapping these environments may help governmental entities to control or mitigate environmental damage properly. In the last decade, remote sensing data have been widely applied for monitoring vegetation health¹, biomass², forest management³, biodiversity⁴, among others^{5–8}. Thus, remote sensing approaches have been used to investigate areas with difficult terrain access, demonstrating great potential for the classification and detection of forest vegetation.

Remote sensing platforms can be embedded with different sensors such as RGB (Red–Green–Blue), multi-spectral and hyperspectral, LiDAR (Light Detection and Ranging), and others⁹. The identification of arboreous vegetation with remote sensing data depends on the spatial and spectral resolutions¹⁰. LiDAR sensors can produce accurate data on the height of the trees, which is an excellent variable to be adopted by automatic extraction methods in forest environments^{11,12}. Multispectral and hyperspectral sensors have the advantage of recording the spectral divergence of the vegetation, which is important for enhancing differences between species configurations, health status, etc.^{8,13}. Still, in recent years, high spatial resolution images acquired by RGB sensors

¹Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul (UFMS), Avenida Costa e Silva, Campo Grande, Mato Grosso do Sul 79070-900, Brazil. ²Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, Presidente Prudente, São Paulo 19067-175, Brazil. ³Post-Graduate Program in Environment and Regional Development, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, Presidente Prudente, São Paulo 19067-175, Brazil. ⁴Department of Geography and Environmental Management, University of Waterloo (UW), Waterloo, ON N2L 3G1, Canada. ⁵Post-Graduate Program in Agronomy, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, Presidente Prudente, São Paulo 19067-175, Brazil. ✉email: fabio@unoeste.br

have been used in many studies for vegetation identification^{7,14–17}. These sensors have a relatively low cost in comparison with others but offer limited information regarding the spectral range.

For single-tree species mapping, the literature already investigated different methods by evaluating multi-spectral and hyperspectral data^{18–20}, airborne LiDAR point clouds²¹, and multi-sensory data fusion^{22,23}. Immitzer et al.²⁴ were able to classify tree species in a temperate forest using satellite multispectral imagery. Franklin and Ahmed²⁵ evaluated UAV (Unmanned Aerial Vehicle)-based multispectral image to map deciduous vegetation. Dalponte et al.²⁶ used hyperspectral data to detect boreal tree species at pixel-level, achieving high accuracy for forest mapping. Most of these studies were conducted with hyperspectral sensors and LiDAR sensors. However, both hyperspectral and LiDAR data cost and process demand are non-attractive for rapid decision models. This is different from RGB sensors, which have a lower cost and are highly available. Guimarães et al.²⁷ demonstrated that the majority of recent applications are implementing RGB imagery data in the vegetation detection scenario.

The visual inspection of remote sensing imagery is a time-consuming, labor-intensive, and biased task. Therefore, various studies have developed multiple methods regarding the automated extraction of the vegetation features^{8,28,29}. Accurately mapping individual tree species in densely forested environments is still a challenging task, even for more robust methods. The increase in quality and quantity in remote sensing data, alongside the rapid improvement of technological resources, allowed for the development of intelligent methods in the computational vision community. By combining remote sensing data with artificial intelligence techniques, it is possible to properly map tree species and improve accuracy in applications regarding vegetation monitoring. In recent years, multiple frameworks have been implemented to assess the performance of such algorithms to accomplish this task^{2,30,31}.

During the past years, the detection and extraction of trees in high-resolution imagery were performed with more traditional machine learning algorithms, like support vector machine (SVM), random forest (RF), artificial neural networks (ANN), and others^{32–35}. They returned interesting outcomes in plenty of studies regarding vegetation analysis^{8,36–40}. However, these learners (known as shallow learners) are limited due to data complexity and may return lower accuracy in comparison with deep learning methods. When considering adverse conditions in a given forest dataset, deeper methods are required.

Identifying individual species in a scene can be a challenging task. However, state-of-the-art deep learning-based methods should be capable of identifying single tree-species with an attractive accuracy and computational load even in RGB images. Recently, deep learning-based methods have been implemented in multiple remote sensing, specifically for image segmentation, classification, and object detection approaches^{31,41,42}. Deep learning techniques are among the most recently adopted approaches to process remote sensing data^{9,43}. In a general sense, deep learning can return better performance than shallow learners, especially in the presence of large quantities of data or if the input data is highly complex^{44,45}.

In heavy-dense forested environments, the identification of single-tree species can become a challenge even for robust methods like current state-of-the-art deep networks. This motivated several studies to investigate the performance of deep learning methods to evaluate their performance on this task. A recently published research tested the performance of object detection using deep networks like YOLOv3⁴⁶, RetinaNet⁴⁷, and Faster-RCNN⁴⁸ to detect tree canopy in RGB imagery covering an urban area⁷. Another study modified the VGG16⁴⁹ to monitor the health conditions of vegetation⁵⁰. A combination of LiDAR and RGB images was also used with the RetinaNet to identify tree-crowns in UAV images¹⁶. The DenseNet⁵¹ was also implemented multispectral data to classify tree species.

The spatial and spectral divergences between the tree and non-tree are essential for automatic methods¹³. In highly-dense scenarios like heavily forested areas, the individual detection of trees could be difficult. RGB sensors are not capable of providing the same amount of spectral data as multispectral or hyperspectral sensors, which offers a potential hindrance for automatic extraction methods. Nonetheless, state-of-the-art deep learning methods based on confidence maps, instead of object detection approaches, could be capable of identifying single tree-species in highly dense areas using RGB images. Methods that could accurately identify a species, among others, may help optimize several applications in environmental planning and forest management.

In the remote sensing field, the identification of palm trees with deep neural networks is a recent topic. A study performed by⁵² investigated the performance of the RetinaNet⁴⁷ to conduct a region-wide spatial inventory of palm trees in an urban environment with high-resolution aerial RGB imagery. In this scenario, the object detection method was appropriate to singularly count palm trees. Another study conducted by⁵³ evaluated the performance of another traditional deep learning method, YOLOv3⁴⁶, to map individual palm trees using aerial imagery acquired with UAVs. As such, dealing with this type of object detection in remote sensing data has demonstrated potential, but it lacks further investigations in highly-dense environments, such as natural forests. The traditional object detection approach may not be entirely appropriate for this environment. Because of that we propose an approach that, while being able to deal with this environment, can return improved evaluation metrics.

In the presented context, this paper presents a deep learning approach to detect individual fruit species of palm trees (*Mauritia flexuosa*; L.f. Buriti) with only aerial RGB orthoimages. As contribution of this approach, a framework to identify and geolocate single species in a high-complex forested environment is demonstrated. The study compares the performance of the proposed method with other state-of-the-art object detection deep neural networks to test its robustness. The palm tree *M. flexuosa* is a valuable source of food, remedy, fiber, and light wood for both indigenous communities and local populations^{54,55}. It is also considered a native species of the Brazilian flora with both current and potential high economic values^{54,56}. Besides, this species has its ecological importance, constituting a food source, nest site, and habitat to a wide variety of species and provides multiple ecosystem services⁵⁵, which highlights the need to accurately map this species.

Stages (T)	MAE	Precision (%)	Recall (%)	F1-measure (%)
1	0.933	85.1	86.4	83.8
2	0.943	93.5	83.6	86.9
4	0.852	91.5	85.5	87.1
5	0.966	93.9	83.1	86.6

Table 1. Influence of the number of stages (T) in counting and detection of *M. flexuosa* palms-trees ($\sigma_{min} = 1$ and $\sigma_{max} = 4$ were adopted).

σ_{max}	MAE	Precision (%)	Recall (%)	F1-measure (%)
3	0.931	86.7	88.7	85.8
4	0.852	91.5	85.5	87.1
5	1.611	91.4	69.6	76.8

Table 2. Influence of the σ_{max} in counting and detection of *M. flexuosa* palms-trees ($\sigma_{min} = 1$ and stages T = 4 were adopted).

σ_{min}	MAE	Precision (%)	Recall (%)	F1-measure (%)
0.75	1.879	90.7	64.9	73.3
1	0.852	91.5	85.5	87.1
1.5	1.012	83.9	88.9	84.3
2	1.671	72.4	91	77.6

Table 3. Influence of the σ_{min} in counting and detection of *M. flexuosa* palms-trees ($\sigma_{max} = 4$ and stages T = 4 were used).

Results

Validation of the parameters. The proposed approach parameters σ_{min} , σ_{max} , and the number of stages T, are responsible for refining the prediction map. Initially, the influence of these parameters was evaluated on the *M. flexuosa* palm trees validation set. Table 1 shows the evaluation of the number of stages T used in the MSM refinement phase. In this experiment, the values of $\sigma_{min} = 1$, $\sigma_{max} = 4$ and ranges T from 1 to 5 were set, and it was discovered that T = 4 achieved the best performance among the number of analyzed stages, reaching an MAE of 0.852 trees and an F1-measure of 87.1%.

The values of σ_{min} and σ_{max} applied in the refinement stage were also evaluated. For this, the number of stages T = 4 was adopted in the subsequent steps since it obtained the best results in the previous experiment (see Table 1). Since the σ_{min} values represent the dispersion of the density maps around the center of the trees, it was found that smaller values do not correctly cover the trees and, therefore, can impair the detection. On the other hand, higher σ_{min} values are also harmful as they cover more than one tree per area. Thus, the best results were obtained with $\sigma_{max} = 4$, indicating that it fits better with the *M. flexuosa* palms-trees characteristics, and generates a more accurate refinement map.

Table 2 presents the evaluation of different values of σ_{min} responsible for the last stage of the MSM. For this, $\sigma_{max} = 4$ and T = 4 were adopted since they obtained better results in the previous experiments (Tables 2 and 3). When $\sigma_{min} = 1$, the proposed approach returned the best performance among the analyzed values. Therefore, the refinement step implemented with values of $\sigma_{min} = 1$, $\sigma_{max} = 4$, and T = 4 generated a more accurate refinement to the validation set.

Comparative results between object detection methods. The proposed method returned better performance when compared with different object detection methods like Faster R-CNN and RetinaNet. The MAE, precision, recall, and F1-measure metrics were calculated for each of them, and the results are displayed in Table 4. The proposed approach achieved high precision and good F-measure values but returned a slight-lower recall value when confronted with them. Nonetheless, it is essential to consider the tradeoff in recall difference (– 6.58% from the Faster R-CNN and – 12.35% from the RetinaNet) with the precision difference (+ 14.52 from the Faster R-CNN and + 35.49% from the RetinaNet).

Since the F1-measure uses both the precision and recall values to compute the test results, it can be assumed that the proposed approach performs better and returns a better balance between true-positive predicted and true-positive rates concerning the identification of palm trees. Nonetheless, the results are consistent with recent literature where object detection applications were applied for the identification of single tree-species^{6,7,57,58}, but performed in the non-RGB image domain. The low precision values for the bounding box method may be

Method	MAE	Precision (%)	Recall (%)	F1-measure (%)
Faster R-CNN	0.984	79	90.8	81.7
RetinaNet	3.761	58	96.6	69.8
Proposed method	0.758	93.5	84.2	86.9

Table 4. Comparative results between our method and Faster R-CNN and RetinaNet.

Method	CPU		GPU	
	Average time (s)	SD	Average time (s)	SD
Faster R-CNN	1.57	0.031	0.05	0.001
RetinaNet	1.93	0.028	0.06	0.002
Proposed Method	1.26	0.051	0.07	0.002

Table 5. Processing time evaluation of the analyzed approaches.

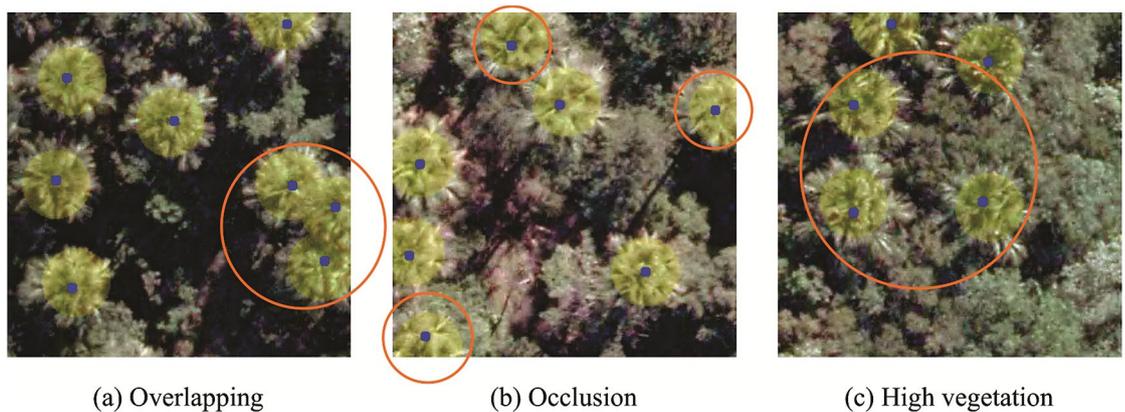


Figure 1. Qualitative results of the proposed method in three scenes: (a) an example of the detected nearby trees with overlapping, (b) detected trees with parts of the canopy occluded at the edge of the image, and (c) demons detected trees in areas of high vegetation. The orange circles highlight the detections.

explained by a high density of objects (i.e., *M. flexuosa* palm trees). This condition is described as problematic for deep networks based on these characteristics, especially when the boxes have high intersections with similar objects⁵⁹.

To verify the potential of the proposed approach in real-time processing, a comparison of its performance with other state-of-the-art methods was conducted. Table 5 shows the average processing time and standard deviation for 100 images of the test set. The values of $\sigma_{min} = 1$, $\sigma_{max} = 4$ and $T = 4$ were used to obtain the best performance in previous experiments. The results show that the approach can achieve real-time processing, delivering image detection in 0.073 seconds with a standard deviation of 0.002 using a GPU. Similarly, RetinaNet and Faster R-CNN methods obtained an average detection time and standard deviation of 0.057, 0.046, and 0.002, 0.001, respectively.

Figure 1 presents the qualitative results of the proposed method where the annotations of *M. flexuosa* palm trees are marked with yellow circles, and the blue dots indicate the correctly detected positions. This approach correctly detects the *M. flexuosa* palm trees in different capture conditions, such as overlapping trees (Fig. 1a), partial occlusion of the treetops (Fig. 1b), and highly dense vegetation areas (Fig. 1c), highlighted by orange circles. Moreover, the predicted positions have a satisfactory level of accuracy, generating detection (blue dots) close to the annotations (center of the yellow circles).

Although the method obtained good results in the detection of *M. flexuosa* palm trees, it faces some challenges. Figure 2 presents areas where the incorrect detections are shown by the red circles. The main challenge is the detection of trees with a high level of occlusion at the image boundary or by overlapping of trees (highlighted by the orange circles). However, even in these few cases, the proposed approach can correctly detect most of the palm trees.

The visual comparison of the palm tree detection approaches is shown in Fig. 3. Column (a) displays the detections obtained by the proposed method, while Columns (b) and (c) are related to the compared methods: Faster R-CNN and RetinaNet, respectively. The approach that obtained the worst result was RetinaNet (Fig. 3c), generating many false-positives (red dots) close to the *M. flexuosa* palm trees detections. On the other hand, Faster R-CNN (Fig. 3b), despite having fewer false-positives, did not properly learn the characteristics of the



Figure 2. Examples of the challenges faced by our method in the *M. flexuosa* palm tree detection task. The orange circles indicate challenging detections.

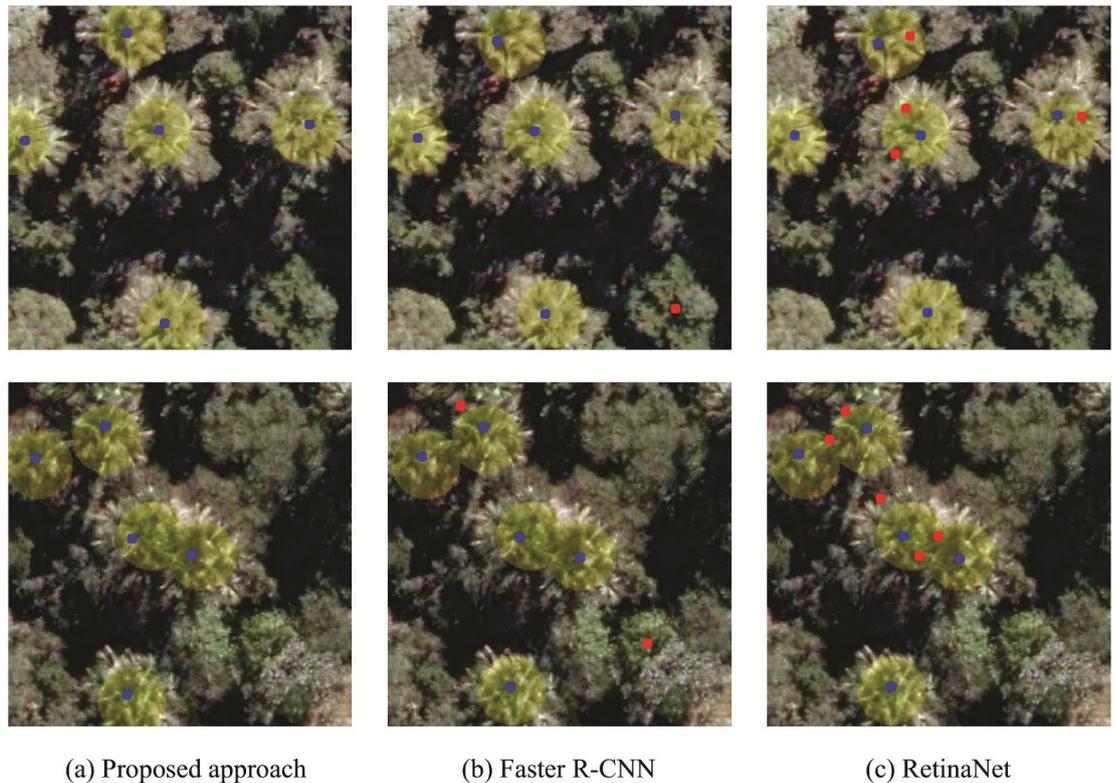


Figure 3. Visual comparison of the analyzed methods. (a) shows the detections obtained by the proposed approach; (b) indicates the detections of the Faster R-CNN and; (c) demonstrates the detections of the RetinaNet. Blue and red points correspond to correct and incorrect detection positions, respectively, and the yellow circle to *M. flexuosa* palms-trees annotation.

palm trees and incorrectly detected other tree species among them. Following the quantitative results shown in Table 4, the proposed approach has the greater precision in detecting *M. flexuosa* palm trees (Fig. 3a), while having the least number of incorrect detections (false-positives).

Discussion

This study demonstrated a feasible method to automatically map single palm tree species of the *M. flexuosa* plant genus using an RGB imagery dataset. *Mauritia flexuosa* frequently occurs at low elevations, with high density on river banks and lake margins, around water sources, and in inundated or humid areas⁵⁶. This is one of the most widely distributed palm trees in South America, Brazil. This species occurs in the Amazon region, Caatinga, Cerrado, and Pantanal, and is one of the palm trees mostly used by humans, being an important item in the diet of many indigenous groups and rural communities⁵⁶.

Mapping *M. flexuosa* palm trees is an important practice for multiple regions of South America, like Brazil, where this plant is viewed as a valuable resource. This palm is widely used for several purposes, is considered

a species of multiple use⁵⁴, occurs in areas of “Veredas”, considered protected by the Brazilian forest code, but there is still a great lack of characterization of the habitats of this species in this country. Mapping and identifying populations of palm *M. flexuosa* is relevant because it is a reliable indicator of water resources, such as streams inside dense gallery forests, slow-flowing swamp surface water, and shallow groundwater in the Cerrado region, vital for humans and wildlife, besides being a valuable source of several non-timber forest products. The approach provides useful information for sustainable economic use and conservation.

As described, single tree species identification is a challenging task even for state-of-the-art deep neural networks when considering only RGB imagery. Mainly because forest environments are constituted by multiple spectral spatial information, overlapping canopies, leaves and branches, size, growth stages, and density, among others. In this manner, studies considered different data to help solve this issue like density point information, canopy height, digital terrain and surface models, spectral divergence, etc.^{4,25,34,45}. Regardless, in this paper, it is proposed a simplification of this process by adopting little input information (i.e., label features such as points and RGB imagery) and a robust method that once trained, can rapidly perform and resolve the said task even in a real-time context.

The results of the present approach achieved satisfactory precision (93.5%), recall (84.2%), and F1-measure (86.9%) values, respectively, and a small MAE (0.758 tree). Studies that applied deep neural networks for detecting other types of arboreal vegetation returned approximated metrics. For the identification of citrus-tree, a CNN method was able to provide 96.2% accuracy¹³, and in oil palm tree detection, a deep neural network implementation returned an accuracy of 96.0% (Li et al., 2019). One different kind of palm trees than the ones evaluated in our dataset was investigated with a modification of the AlexNet CNN architecture and returned high prediction values ($R = 0.99$, with the relative error between 2.6 and 9.2%)⁵⁷. A study⁷ achieved an accuracy higher than 90% to detect single tree-species using the RetinaNet and RGB images. However, in the aforementioned papers, the tree density patterns are differently from ours, and the evaluated individual trees are more spaced from each other, which makes a simpler object detection problem.

In the described manner, the proposed method may help in mapping the *M. flexuosa* palm tree with little computational load and high accuracy. Since this approach can compute point features as labeled objects, it reduces the amount of label-work required from the human counterpart. Additionally, the method provided a fast solution to detect the palm tree’s location with a delivering image detection of 0.073 seconds and a standard deviation of 0.002 using a GPU. This information is essential for properly calculating the cost and effectiveness of the method. The presented approach may help new research while providing primary information for exploring environmental management practices in the experiment context (i.e., evaluating a keystone tree species). The proposed method could also be incorporated into areas and regions to help detect the *M. flexuosa* palm tree and contribute to decision-making conservation measures of the said species.

Conclusion

This paper presents an approach based on deep networks to map single species of fruit palm trees (*Mauritia flexuosa*) in aerial RGB imagery. According to the performance assessment, the method returned an MAE of 0.75 trees and F1-measure of 86.9%. A comparative study also shows that the proposed method returned better accuracy than state-of-the-art methods like Faster R-CNN and RetinaNet under the same experimental conditions. Besides, this approach took a shorter time to detect the palm trees with 0.073 seconds for delivering image detection and achieved a standard deviation of 0.002 using the GPU. In future implementations, it should be possible to add new strategies in this CNN architecture to overcome challenges regarding other tree patterns. Still, the identification of individual species can help to assist in both monitoring and mapping important singular species. As such, the proposed method may assist in new research for the forest remote sensing community that includes data obtained with RGB sensors. As a future study, different takes on the detection approach could be implemented to enhance the precision of the method, one of which being the investigation of different loss functions and approaches to detect each tree.

Methods

The method proposed in this paper is composed of three main phases (see Fig. 4): (1) the dataset was composed of aerial RGB orthoimages obtained from a riparian zone of a well-known populated region of *M. flexuosa* palm trees. With specialist assistance, the palm trees in the RGB images were visually identified and labeled in a Geographical Information System (GIS). The image and labeled data were split into groups of training, validation, and testing subsets; (2) the object detection approach was applied in a computational environment; (3) the performance of the proposed method was compared with other networks.

Study area and mapped species. The riparian zone of the upper-stream of the Imbiruçu brook, located near the city of Campo Grande, in the state of Mato Grosso do Sul, Brazil was selected for the study (Fig. 5). This stream, formed by a dendritic drainage system, is inserted in the hydrographical basin of the Paraguay River and covered by the Cerrado (Brazilian Savanna) biome. This area is composed of a highly complex forest patch containing a widespread of palm tree species *Mauritia flexuosa* (popular name Buriti). The Arecaeae is a dioecious⁶⁰ long-living species and grows naturally in flooded areas, providing water balance for rivers and other water bodies. In highly dense, monodominant stands in flooded areas, mature *M. flexuosa* palm trees reach 20 m high⁶⁰. The evaluated site in our experiment, in specific, is one of the well-known locations where a large number of samples of this species is sufficient to train a deep neural network.

The aerial RGB orthoimages were provided by the city hall of Campo Grande, State of Mato Grosso do Sul, Brazil. The ground sample distance (GSD) of the orthoimages is 10 cm. A total of 43 orthoimages with dimensions

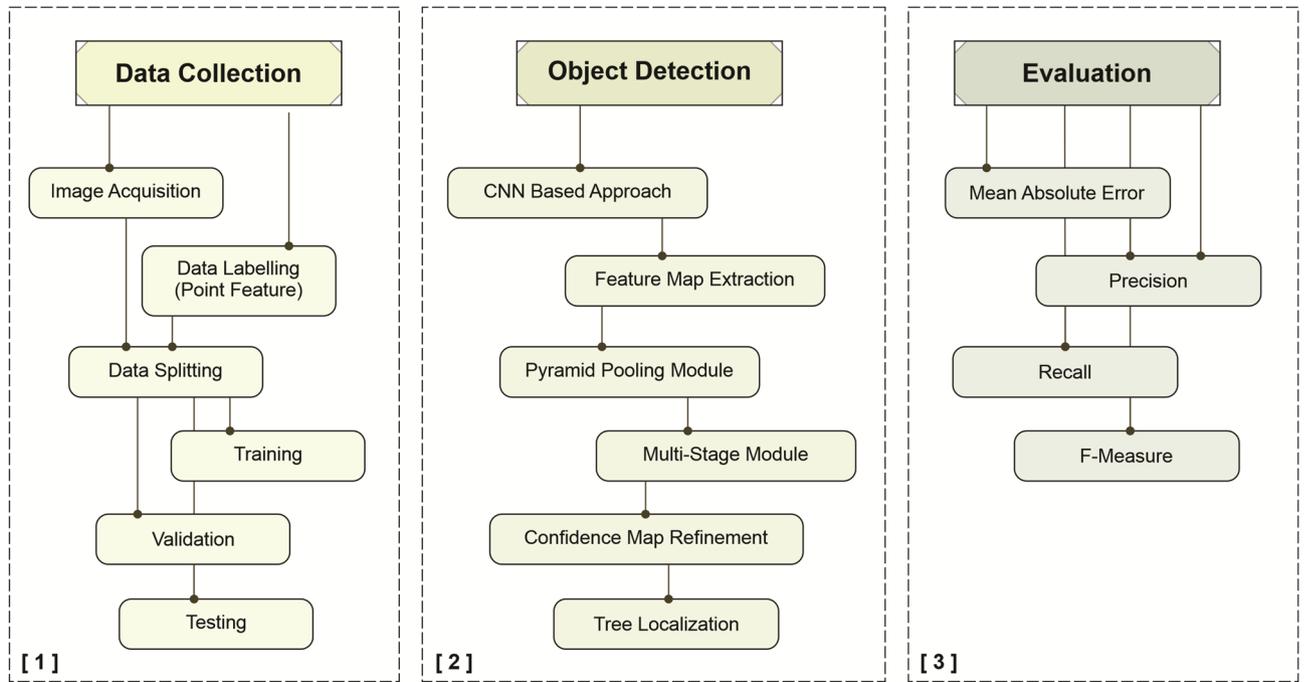


Figure 4. Summarized phases of the proposed approach.

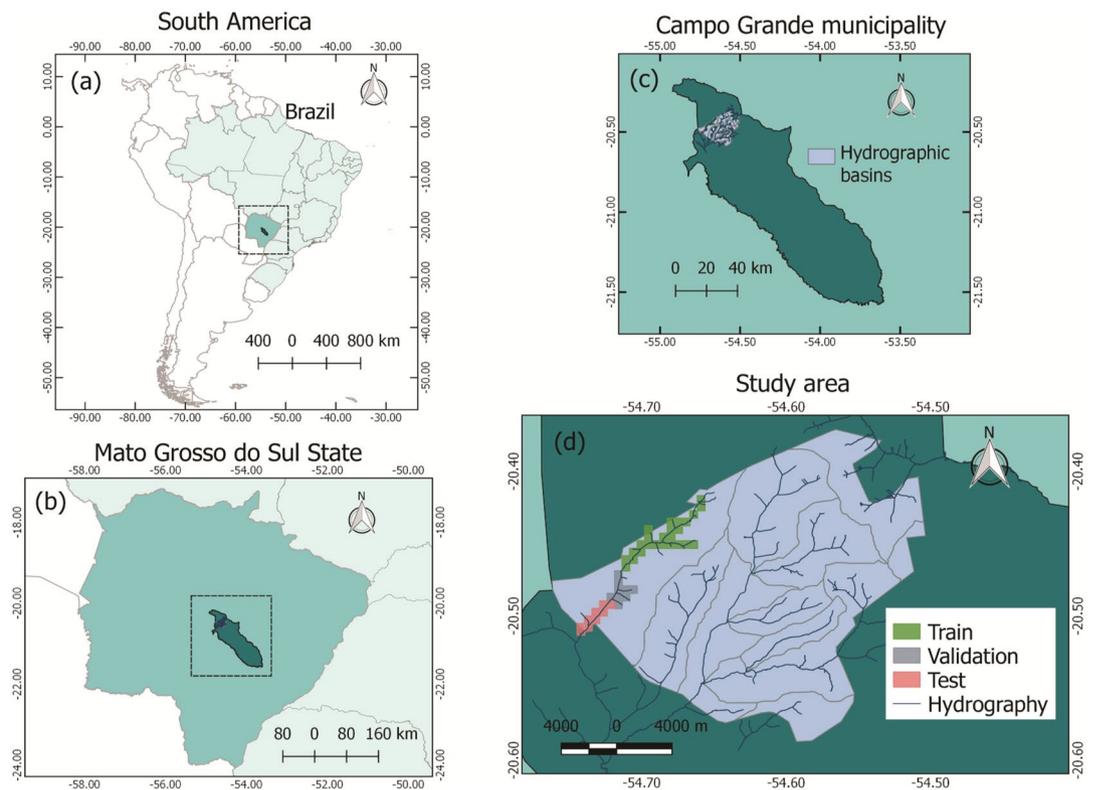


Figure 5. Location map of the study area in (a) South America and Brazil, (b) Mato Grosso do Sul, (c) Campo Grande, and (d) study area. Map created with QGIS 3.16.9: <https://qgis.org/en/site/>.

5619 × 5946 pixels were used in the study. This aerial image dataset was composed of 1394 scenes, where 5334 palm trees were manually labeled and used as ground-truth (Fig. 6).

Proposed method. This study proposes a CNN method that uses the RGB image as an input and, throughout a confidence map refinement, returns a prediction map with tree locations (Fig. 7). The objects' position is calculated after a 2D confidence map estimation, based on previous works⁵⁸. The first step of the architecture

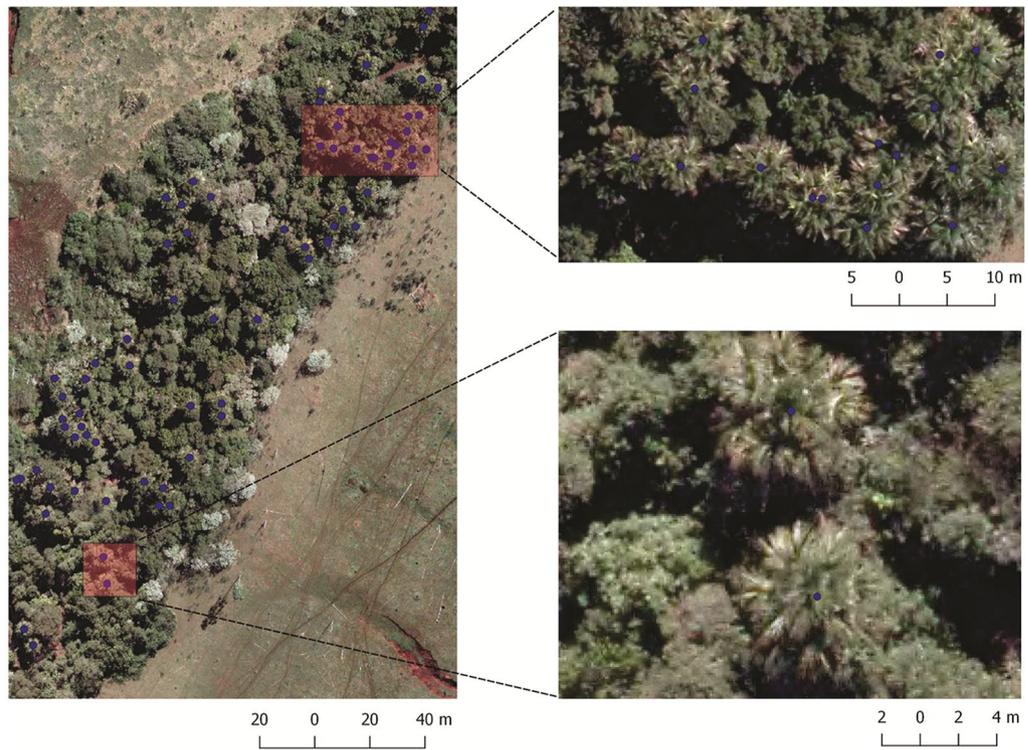


Figure 6. Examples of the labeled dataset. *M. flexuosa* palm trees are represented with blue dots.

extracts the feature map (Fig. 7a). In a sequential step, the feature map goes through the Pyramid Pooling Module (PPM)⁶¹. The last step of the architecture produces a confidence map in a Multi-Stage Module (MSM)⁵⁸ that enhances the position of the tree by adjusting the prediction to its center.

Feature map extraction and PPM. For the feature map extraction (Fig. 7b), the proposed CNN is based on the VGG-19⁴⁹. Here, the network is composed of 8 convolutional layers with 64, 128, and 256 filters with a 3×3 size window, with Rectified Linear Units (ReLU) functions in all layers. The spatial volume size was reduced in half after the second and fourth layers by a max-pooling layer (2×2 window). The PPM⁶¹ was used (Fig. 7c) to extract global and local information, which helps the CNN to be less variant to tree scale differences. The extracted features are upsampled to size equivalent to the input feature map and concatenated with it to create an enhanced version of the image.

Tree localization. The MSM step (Fig. 7d) estimates the confidence map from the feature map extracted in the previous module. The MSM is composed of T refinement stages, where the first stage contains 3 layers of 128 filters with 3×3 size, 1 layer with 512 filters of 1 1 size, and one final layer with 1 filter that generates the confidence map C_1 from the first stage. The position of the trees predicted in the first stage is refined in the $T - 1$ stages. In each stage $t \in [2, 3, \dots, T]$, the prediction (C) is returned from a previous stage ($t - 1$) and the feature map from the PPM module is concatenated. The final layer in this step has a sigmoid activation function since the method considers the probability of occurrence of a tree to exist or not $[0,1]$. The concatenation process allows for both global and local context information to be incorporated in it. At the end of each stage, a loss function (1) is adopted to avoid the vanishing gradient problem. The general loss function is calculated by the following Eq. (2).

$$f_t = \sum_p \|\hat{C}_t(p) - C_t(p)\|_2^2 \quad (1)$$

where $\hat{C}_t(p)$ is the ground-truth confidence map of location (p) in stage (t).

$$f = \sum_{t=1}^T f_t \quad (2)$$

The confidence map is generated by a 2D Gaussian kernel at the center of the labeled tree. A standard deviation σ_t controls the spread of the peak for each Gaussian kernel (Fig. 8). Different values of σ_t were used to refine the predictions. The value of σ_1 in the MSM is set to maximum (σ_{max}) while the σ_t in the final stage is set to minimum (σ_{min}). In the early phases of the experiment, different values for t were adopted to improve its robustness. Finally, the tree location is estimated by the peaks of the confidence map (Fig. 8). These peaks are

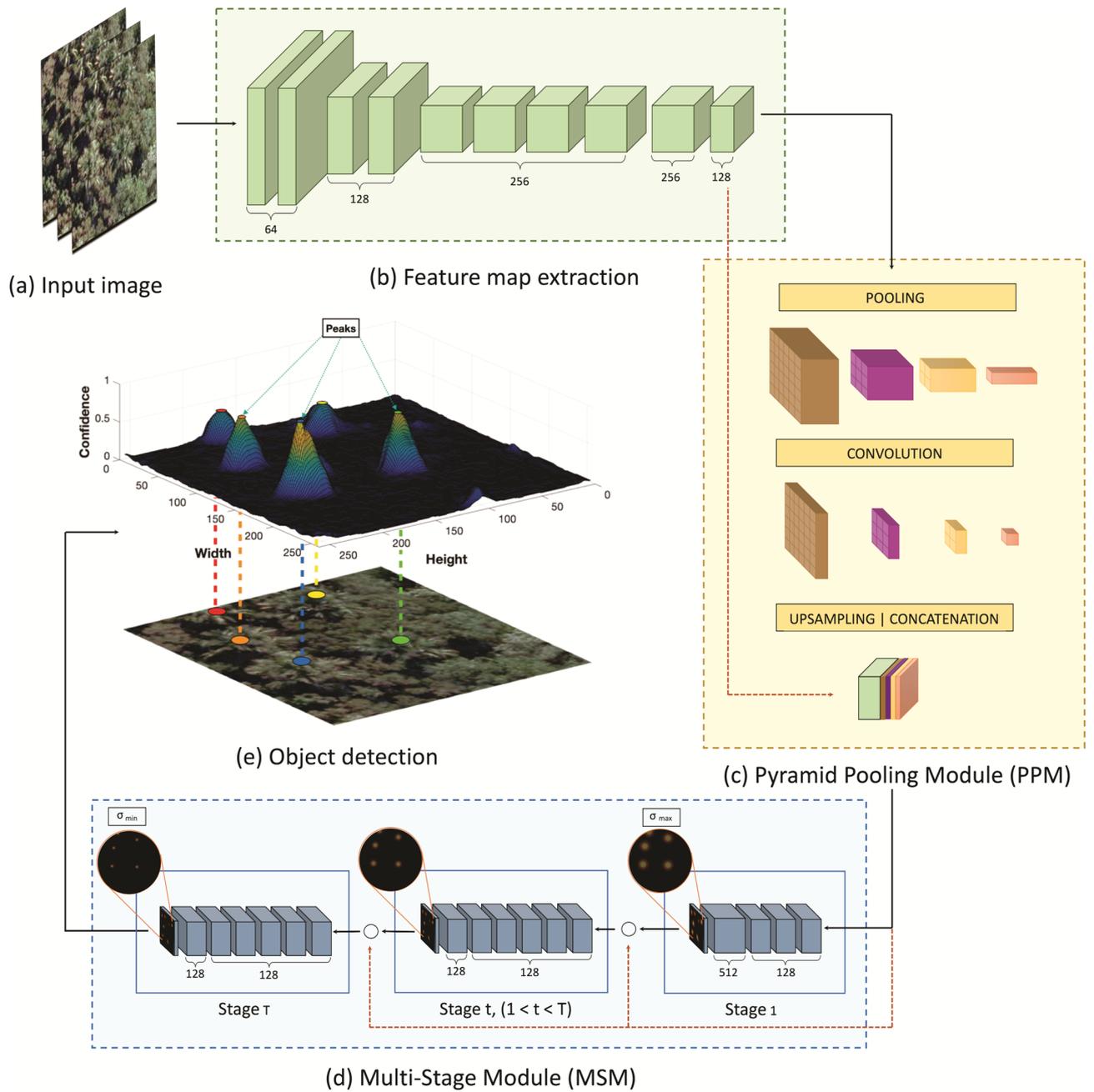


Figure 7. Proposed CNN. The feature map (b) is extracted from the input image (a) and improved by the PPM (c). The result is used as input at the MSM step (d), where T stages enhance the prediction positions (e).

considered the local maximum of the confidence map, representing a high probability of a tree occurrence. $P = (x_p, y_p)$ is considered as a local maximum if $C_T(p) > C_T(v)$ for all neighbors v . Here, v is given by $(x_p \pm 1, y_p)$ or $(x_p, y_p \pm 1)$.

A peak in the confidence map is defined as a real tree if $C_T(p) > \tau$ (Fig. 7e). To prevent the network from confusing trees in a nearby range from each other, a distance of δ is defined. For this study, τ equal to 1 pixel and δ equal to 0.35 were defined as valid metrics. These values were defined during a previous experimental phase.

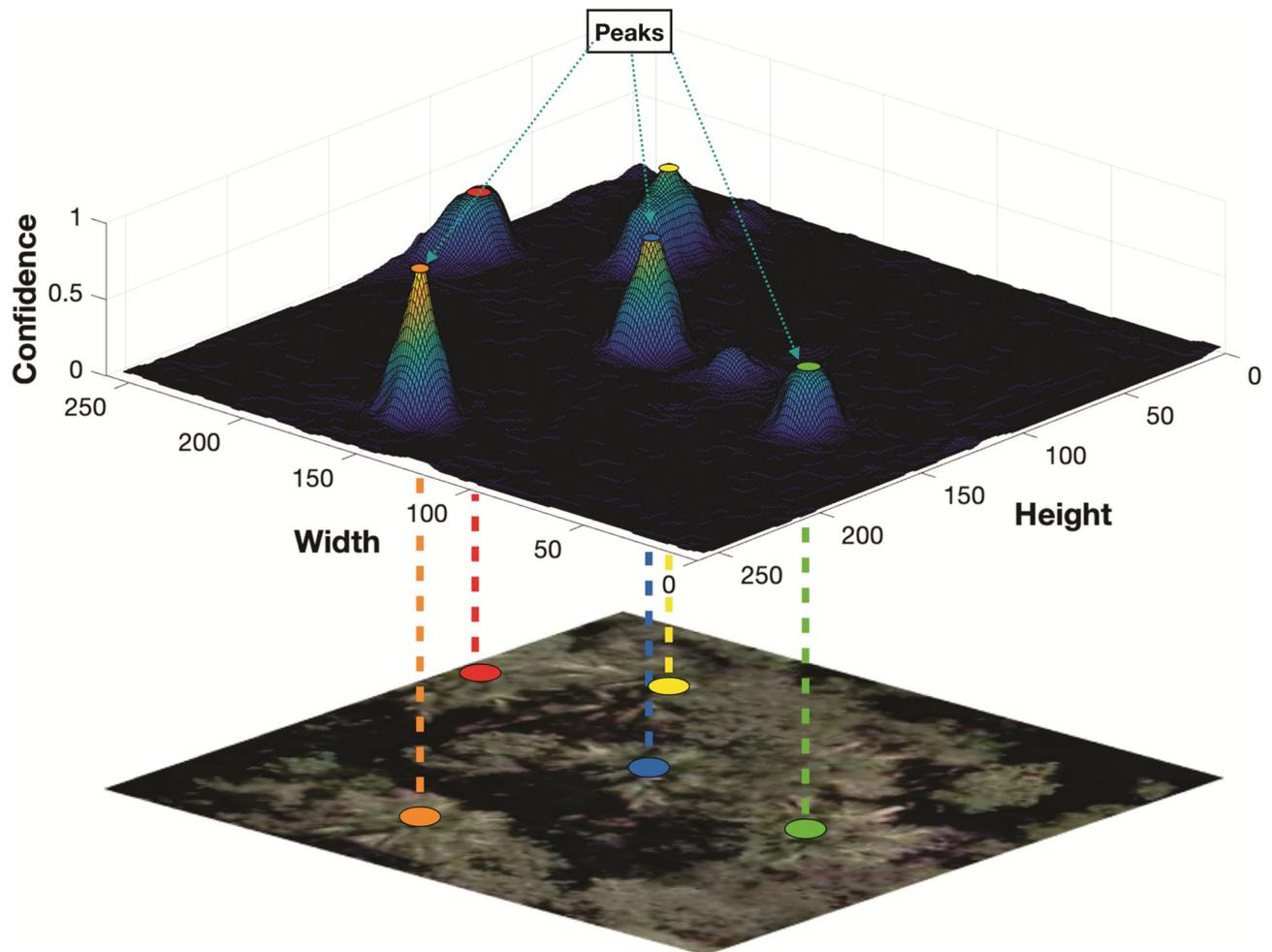


Figure 8. Tree localization example from a refined confidence map. Figure was created with Matplotlib 3.4.2: <https://matplotlib.org/>.

Experimental setup. For the experimental setup, the RGB orthomosaics were separated into training, validation, and testing, respectively (Fig. 9). They were split into nonoverlapping patches of 256×256 pixels ($25.6 \text{ m} \times 25.6 \text{ m}$). The patches were then divided into training (42.3%), validation (34.5%), and testing (23.2%) sets. Table 6 lists the number of samples (trees) and image patches, and Fig. 9 displays examples of the orthomosaics used to extract the datasets. For the training process, the CNN was initialized with pre-trained weights from ImageNet and a Stochastic Gradient Descent optimizer was applied with a moment equal to 0.9. For this, the validation set was used to adjust the learning rate and the number of epochs, which were set to 0.001 and 100, respectively.

The performance of the proposed network was assessed with the following metrics: mean absolute error (MAE); precision (P); recall (R), and; F1-measure (F1). The results were compared with Faster R-CNN and RetinaNet methods. Since these methods are based on bounding boxes, the plant position (x, y) from the labeled ground truth was used as a center of the box. The correct size of the box corresponds with the size occupied by the tree canopy. To perform this comparison, the same conjuncts of training, validation, and testing datasets were adopted for the three methods. Likely, an inverse process was applied during the test phase, where the position of the tree was obtained by the center of the point inside the predicted bounding-box of the RetinaNet and Faster R-CNN methods. This allowed applying the same metrics (MAE, P, R, and F1) for comparing the performances of each neural network.

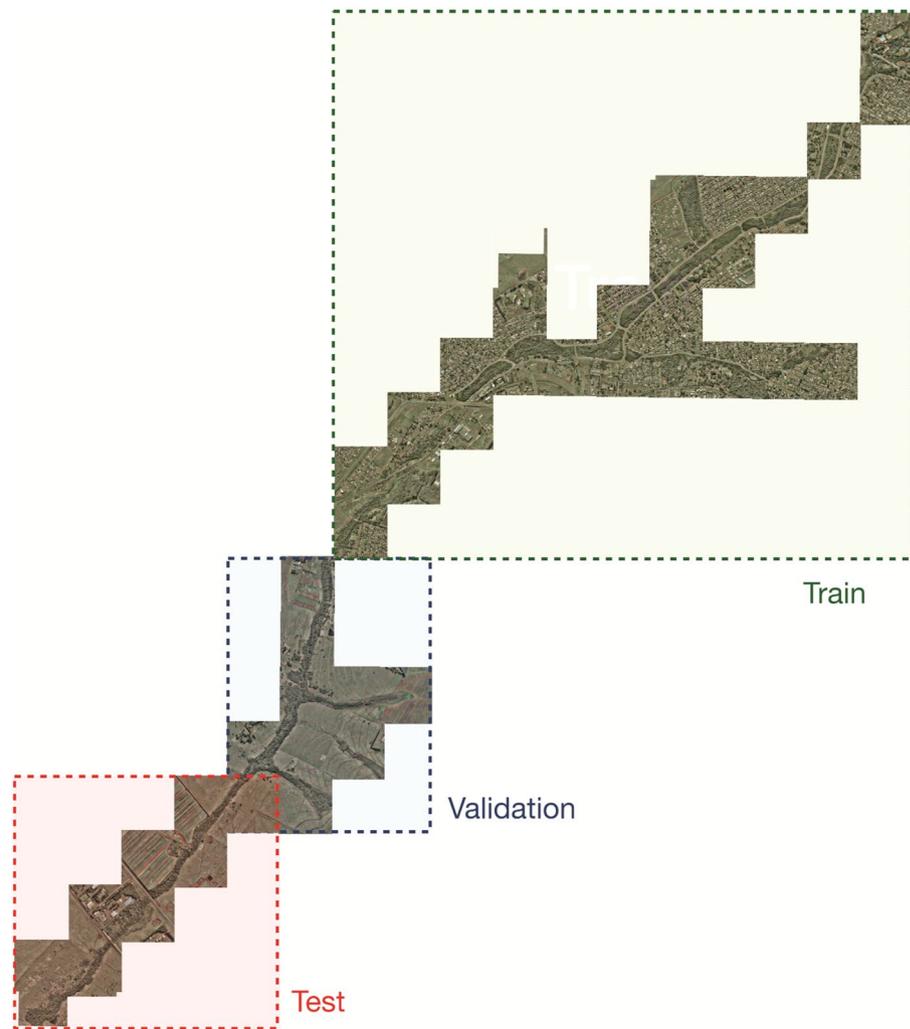


Figure 9. Training, validation, and testing datasets separated per region. Figure created with QGIS 3.16.9: <https://qgis.org/en/site/>.

Dataset	Number of patches	Number of samples (palm trees)
Training	590	1784
Validation	481	2296
Testing	323	1254

Table 6. Description of the training, validation, and testing sets of the *M. flexuosa* palms-trees dataset.

Received: 30 March 2021; Accepted: 17 August 2021
Published online: 04 October 2021

References

- Näsi, R. *et al.* Using UAV-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level. *Remote Sens.* **7**, 15467–15493. <https://doi.org/10.3390/rs71115467> (2015).
- Navarro, A. *et al.* The application of unmanned aerial vehicles (UAVs) to estimate above-ground biomass of mangrove ecosystems. *Remote Sens. Environ.* **242**, 111747. <https://doi.org/10.1016/j.rse.2020.111747> (2020).
- Reis, B. P. *et al.* Management recommendation generation for areas under forest restoration process through images obtained by UAV and LiDAR. *Remote Sens.* **11**, 1508. <https://doi.org/10.3390/rs11131508> (2019).
- Saarinen, N. *et al.* Assessing biodiversity in boreal forests with UAV-based photogrammetric point clouds and hyperspectral imaging. *Remote Sens.* **10**, 338. <https://doi.org/10.3390/rs10020338> (2018).
- Casapia, X. T. *et al.* Identifying and quantifying the abundance of economically important palms in tropical moist forest using UAV imagery. *Remote Sens.* **12**, 9. <https://doi.org/10.3390/rs12010009> (2019).
- Li, L. *et al.* Quantifying understory and overstory vegetation cover using UAV-based RGB imagery in forest plantation. *Remote Sens.* **12**, 298. <https://doi.org/10.3390/rs12020298> (2020).

7. dos Santos, A. A. *et al.* Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors* **19**, 3595. <https://doi.org/10.3390/s19163595> (2019).
8. Miyoshi, G. T., Imai, N. N., Tommaselli, A. M. G., de Moraes, M. V. A. & Honkavaara, E. Evaluation of hyperspectral multitemporal information to improve tree species identification in the highly diverse Atlantic forest. *Remote Sens.* **12**, 244. <https://doi.org/10.3390/rs12020244> (2020).
9. Morales, G. *et al.* Automatic segmentation of *Mauritia flexuosa* in unmanned aerial vehicle (UAV) imagery using deep learning. *Forests* **9**, 736. <https://doi.org/10.3390/f9120736> (2018).
10. Voss, M. & Sugumaran, R. Seasonal effect on tree species classification in an urban environment using hyperspectral data, LiDAR, and an object-oriented approach. *Sensors* **8**, 3020–3036. <https://doi.org/10.3390/s8053020> (2008).
11. Andersen, H.-E., Reutebuch, S. E. & McGaughey, R. J. A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Can. J. Remote Sens.* **32**, 355–366. <https://doi.org/10.5589/m06-030> (2006).
12. Ganz, S., Käber, Y. & Adler, P. Measuring tree height with remote sensing—A comparison of photogrammetric and LiDAR data with different field measurements. *Forests* **10**, 694. <https://doi.org/10.3390/f10080694> (2019).
13. Csillik, O., Cherbini, J., Johnson, R., Lyons, A. & Kelly, M. Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones* **2**, 39. <https://doi.org/10.3390/drones2040039> (2018).
14. Berveglieri, A., Imai, N. N., Tommaselli, A. M., Casagrande, B. & Honkavaara, E. Successional stages and their evolution in tropical forests using multi-temporal photogrammetric surface models and superpixels. *ISPRS J. Photogram. Remote Sens.* **146**, 548–558. <https://doi.org/10.1016/j.isprsjprs.2018.11.002> (2018).
15. Cao, J. *et al.* Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models. *Remote Sens.* **10**, 89. <https://doi.org/10.3390/rs10010089> (2018).
16. Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A. & White, E. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sens.* **11**, 1309. <https://doi.org/10.3390/rs11111309> (2019).
17. Torres, D. L. *et al.* Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. *Sensors* **20**, 563. <https://doi.org/10.3390/s20020563> (2020).
18. Liu, L., Song, B., Zhang, S. & Liu, X. A novel principal component analysis method for the reconstruction of leaf reflectance spectra and retrieval of leaf biochemical contents. *Remote Sens.* **9**, 1113. <https://doi.org/10.3390/rs9111113> (2017).
19. Maschler, J., Atzberger, C. & Immitzer, M. Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. *Remote Sens.* **10**, 1218. <https://doi.org/10.3390/rs10081218> (2018).
20. Hennessy, A., Clarke, K. & Lewis, M. Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sens.* **12**, 113. <https://doi.org/10.3390/rs12010113> (2020).
21. Hamraz, H., Contreras, M. A. & Zhang, J. Forest understory trees can be segmented accurately within sufficiently dense airborne laser scanning point clouds. *Sci. Rep.* **7**, 1–9. <https://doi.org/10.1038/s41598-017-07200-0> (2017).
22. Cho, M. A. *et al.* Mapping tree species composition in south African savannas using an integrated airborne spectral and LiDAR system. *Remote Sens. Environ.* **125**, 214–226. <https://doi.org/10.1016/j.rse.2012.07.010> (2012).
23. Apostol, B. *et al.* Species discrimination and individual tree detection for predicting main dendrometric characteristics in mixed temperate forests by use of airborne laser scanning and ultra-high-resolution imagery. *Sci. Total Environ.* **698**, 134074. <https://doi.org/10.1016/j.scitotenv.2019.134074> (2020).
24. Immitzer, M., Atzberger, C. & Koukal, T. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sens.* **4**, 2661–2693. <https://doi.org/10.3390/rs4092661> (2012).
25. Franklin, S. E. & Ahmed, O. S. Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data. *Int. J. Remote Sens.* **39**, 5236–5245. <https://doi.org/10.1080/01431161.2017.1363442> (2017).
26. Dalponte, M., Orka, H. O., Gobakken, T., Gianelle, D. & Naeset, E. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **51**, 2632–2645. <https://doi.org/10.1109/tgrs.2012.2216272> (2013).
27. Guimarães, N. *et al.* Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities. *Remote Sens.* **12**, 1046. <https://doi.org/10.3390/rs12061046> (2020).
28. Kattenborn, T., Eichel, J. & Fassnacht, F. E. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Sci. Rep.* **9**, 1–9. <https://doi.org/10.1038/s41598-019-53797-9> (2019).
29. Onishi, M. & Ise, T. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **11**, 1–15. <https://doi.org/10.1038/s41598-020-79653-9> (2021).
30. Näsi, R. *et al.* Remote sensing of bark beetle damage in urban forests at individual tree level using a novel hyperspectral camera from UAV and aircraft. *Urban For. Urban Green.* **30**, 72–83. <https://doi.org/10.1016/j.ufug.2018.01.010> (2018).
31. Nezami, S., Khoramshahi, E., Nevalainen, O., Pölönen, I. & Honkavaara, E. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. *Remote Sens.* **12**, 1070. <https://doi.org/10.3390/rs12071070> (2020).
32. Nevalainen, O. *et al.* Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. *Remote Sens.* **9**, 185. <https://doi.org/10.3390/rs9030185> (2017).
33. Raczko, E. & Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur. J. Remote Sens.* **50**, 144–154. <https://doi.org/10.1080/22797254.2017.1299557> (2017).
34. Tuominen, S. *et al.* Assessment of classifiers and remote sensing features of hyperspectral imagery and stereo-photogrammetric point clouds for recognition of tree species in a forest area of high species diversity. *Remote Sens.* **10**, 714. <https://doi.org/10.3390/rs10050714> (2018).
35. Xie, Z., Chen, Y., Lu, D., Li, G. & Chen, E. Classification of land cover, forest, and tree species classes with ZiYuan-3 multispectral and stereo data. *Remote Sens.* **11**, 164. <https://doi.org/10.3390/rs11020164> (2019).
36. Maxwell, A. E., Warner, T. A. & Fang, F. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* **39**, 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343> (2018).
37. Osco, L. P. *et al.* Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sens.* **11**, 2925. <https://doi.org/10.3390/rs11242925> (2019).
38. Marrs, J. & Ni-Meister, W. Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. *Remote Sens.* **11**, 819. <https://doi.org/10.3390/rs11070819> (2019).
39. Imangholiloo, M. *et al.* Characterizing seedling stands using leaf-off and leaf-on photogrammetric point clouds and hyperspectral imagery acquired from unmanned aerial vehicle. *Forests* **10**, 415. <https://doi.org/10.3390/f10050415> (2019).
40. Pham, T., Yokoya, N., Bui, D., Yoshino, K. & Friess, D. Remote sensing approaches for monitoring mangrove species, structure, and biomass: Opportunities and challenges. *Remote Sens.* **11**, 230. <https://doi.org/10.3390/rs11030230> (2019).
41. Ma, L. *et al.* Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogram. Remote Sens.* **152**, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015> (2019).
42. Safonova, A. *et al.* Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens.* **11**, 643. <https://doi.org/10.3390/rs11060643> (2019).
43. Kamilaris, A. & Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **147**, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016> (2018).

44. Khamparia, A. & Singh, K. M. A systematic review on deep learning architectures and applications. *Exp. Syst.* **36**, e12400. <https://doi.org/10.1111/exsy.12400> (2019).
45. Sothe, C. *et al.* Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GISci. Remote Sens.* **57**, 369–394. <https://doi.org/10.1080/15481603.2020.1712102> (2020).
46. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement (2018). [arXiv:1804.02767](https://arxiv.org/abs/1804.02767).
47. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection (2018). [arXiv:1708.02002](https://arxiv.org/abs/1708.02002)
48. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks (2016). [arXiv:1506.01497](https://arxiv.org/abs/1506.01497)
49. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition (2015). [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
50. Sylvain, J.-D., Drolet, G. & Brown, N. Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS J. Photogram. Remote Sens.* **156**, 14–26. <https://doi.org/10.1016/j.isprsjprs.2019.07.010> (2019).
51. Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M. & Carron, J. Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning. *Sensors* **19**, 1284. <https://doi.org/10.3390/s19061284> (2019).
52. Culman, M., Delalieux, S. & Tricht, K. V. Individual palm tree detection using deep learning on RGB imagery to support tree inventory. *Remote Sens.* **12**, 3476. <https://doi.org/10.3390/rs12213476> (2020).
53. Aburasain, R. Y., Edirisinghe, E. A. & Albatay, A. Palm tree detection in drone images using deep convolutional neural networks: Investigating the effective use of YOLO v3. In *Digital Interaction and Machine Intelligence*, 21–36, https://doi.org/10.1007/978-3-030-74728-2_3 (Springer International Publishing, 2021).
54. Bortolotto, I. M., Damasceno-Junior, G. A. & Pott, A. Preliminary list of native food plants from mato grosso do sul, brazil. *Iheringia, Série Botânica* **73**, 101–116 (2018). <https://doi.org/10.21826/2446-8231201873s101>
55. van der Hoek, Y., Solas, S. Á. & Peñuela, M. C. The palm *Mauritia flexuosa*, a keystone plant resource on multiple fronts. *Biodiver. Conserv.* **28**, 539–551. <https://doi.org/10.1007/s10531-018-01686-4> (2019).
56. Agostini-Costa, T. d. S., Faria, J. P., Naves, R. V. & Vieira, R. F. *Espécies Nativas da Flora Brasileira de Valor Econômico Atual ou Potencial Plantas para o Futuro - Região Centro-Oeste* (Ministério do Meio Ambiente - MMA, 2016).
57. Djerriri, K., Ghabi, M., Karoui, M. S. & Adjoudj, R. Palm trees counting in remote sensing imagery using regression convolutional neural network. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2627–2630 (2018). <https://doi.org/10.1109/IGARSS.2018.8519188>
58. Osco, L. P. *et al.* A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogram. Remote Sens.* **160**, 97–106. <https://doi.org/10.1016/j.isprsjprs.2019.12.010> (2020).
59. Goldman, E. *et al.* Precise detection in densely packed scenes (2019). [arXiv:1904.00853](https://arxiv.org/abs/1904.00853)
60. Holm, J. A., Miller, C. J. & Cropper, W. P. Population dynamics of the dioecious amazonian palm *Mauritia flexuosa*: Simulation analysis of sustainable harvesting. *Biotropica* **40**, 550–558. <https://doi.org/10.1111/j.1744-7429.2008.00412.x> (2008).
61. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. Pyramid scene parsing network (2017). [arXiv:1612.01105](https://arxiv.org/abs/1612.01105)

Acknowledgements

The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance Code 001).

Author contributions

All authors reviewed the manuscript.

Funding

This research was partially funded by CNPq (p: 433783/2018-4, 303559/2019-5, 304052/2019-1, and 310517/2020-6) and CAPES Print (p: 88881.311850/2018-01). The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and CAPES (Finance Code 001), and NVIDIA® for the donation of the Titan X graphics card used in the experiments.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.F.d.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021