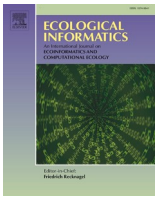
Ecological Informatics 71 (2022) 101815

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Fusing multi-season UAS images with convolutional neural networks to map tree species in Amazonian forests

Hudson Franklin Pessoa Veras [a,*], Matheus Pinheiro Ferreira [b], Ernandes Macedo da Cunha Neto [a], Evandro Orfanó Figueiredo [c], Ana Paula Dalla Corte [a], Carlos Roberto Sanquetta [a]

[a] Departament of Forest Science, Federal University of Paraná (UFPR), Pref. Lothário Meissner Avenue 900, 80210-170 Curitiba, PR, Brazil
[b] Cartographic Engineering Department, Military Institute of Engineering (IME), Praça Gen. Tibúrcio 80, 22290-270 Rio de Janeiro, RJ, Brazil
[c] Embrapa Acre, Rodovia BR-364, km 14, 69900-056 Rio Branco, AC, Brazil

ARTICLE INFO

Keywords:
Individual tree crowns
Deep learning
Multi-temporal images
Phenology
Tropical forest

ABSTRACT

Remote sensing images obtained by unoccupied aircraft systems (UAS) across different seasons enabled capturing of species-specific phenological patterns of tropical trees. The application of UAS multi-season images to classify tropical tree species is still poorly understood. In this study, we used RGB images from different seasons obtained by a low-cost UAS and convolutional neural networks (CNNs) to map tree species in an Amazonian forest. Individual tree crowns (ITC) were outlined in the UAS images and identified to the species level using forest inventory data. The CNN model was trained with images obtained in February, May, August, and November. The classification accuracy in the rainy season (November and February) was higher than in the dry season (May and August). Fusing images from multiple seasons improved the average accuracy of tree species classification by up to 21.1 percentage points, reaching 90.5%. The CNN model can learn species-specific phenological characteristics that impact the classification accuracy, such as leaf fall in the dry season, which highlights its potential to discriminate species in various conditions. We produced high-quality individual tree crown maps of the species using a post-processing procedure. The combination of multi-season UAS images and CNNs has the potential to map tree species in the Amazon, providing valuable insights for forest management and conservation initiatives.

1. Introduction

Information regarding characteristics of tropical forests, such as aboveground biomass and tree species composition, is crucial for their sustainable management. Forest inventories are the most common approach used to retrieve such characteristics. However, due to prohibitive costs, they are usually limited to small spatial extents (< 1% of the sampling area). The combination of remote sensing and forest inventory data has provided accurate information on forest characteristics over large areas (hundreds of hectares) and proved helpful in reducing field effort (Dalla Corte et al., 2020; de Almeida Papa et al., 2020).

Remote sensing images acquired by UAS platforms have been employed to retrieve characteristics of tropical forests, particularly the spatial distribution of tree species. UAS-based data acquisition usually results in hundreds of ultra-high spatial resolution (ground sampling distance, GSD < 10 cm) RGB images, which enable retrieving fine-scale features of individual tree crowns, such as the arrangement of leaves and branches. Previous investigations showed that deep learning methods can automatically learn these features on ultra-high resolution images acquired by UAS and successfully discriminate species (Ferreira et al., 2020; Kattenborn et al., 2021; Morales et al., 2018).

The most common deep learning method is convolutional neural networks (CNNs). CNNs rely on convolutions (local linear operations) to automatically extract object features (e.g., shapes, edges, texture) without user intervention (Zhang et al., 2016). CNNs were initially designed to perform scene classification, that is, classifying the image as a whole into one of several different classes. CNNs for object detection can locate the object within the image by constructing a bounding box that encloses it. Semantic segmentation networks aim to assign a label to each image pixel, while instance segmentation networks combine object detection and semantic segmentation to precisely outline the object's contour. Semantic and instance segmentation methods can

* Corresponding author.
E-mail address: hudsonveras@gmail.com (H.F.P. Veras).

https://doi.org/10.1016/j.ecoinf.2022.101815
Received 24 April 2022; Received in revised form 11 September 2022; Accepted 11 September 2022
Available online 20 September 2022

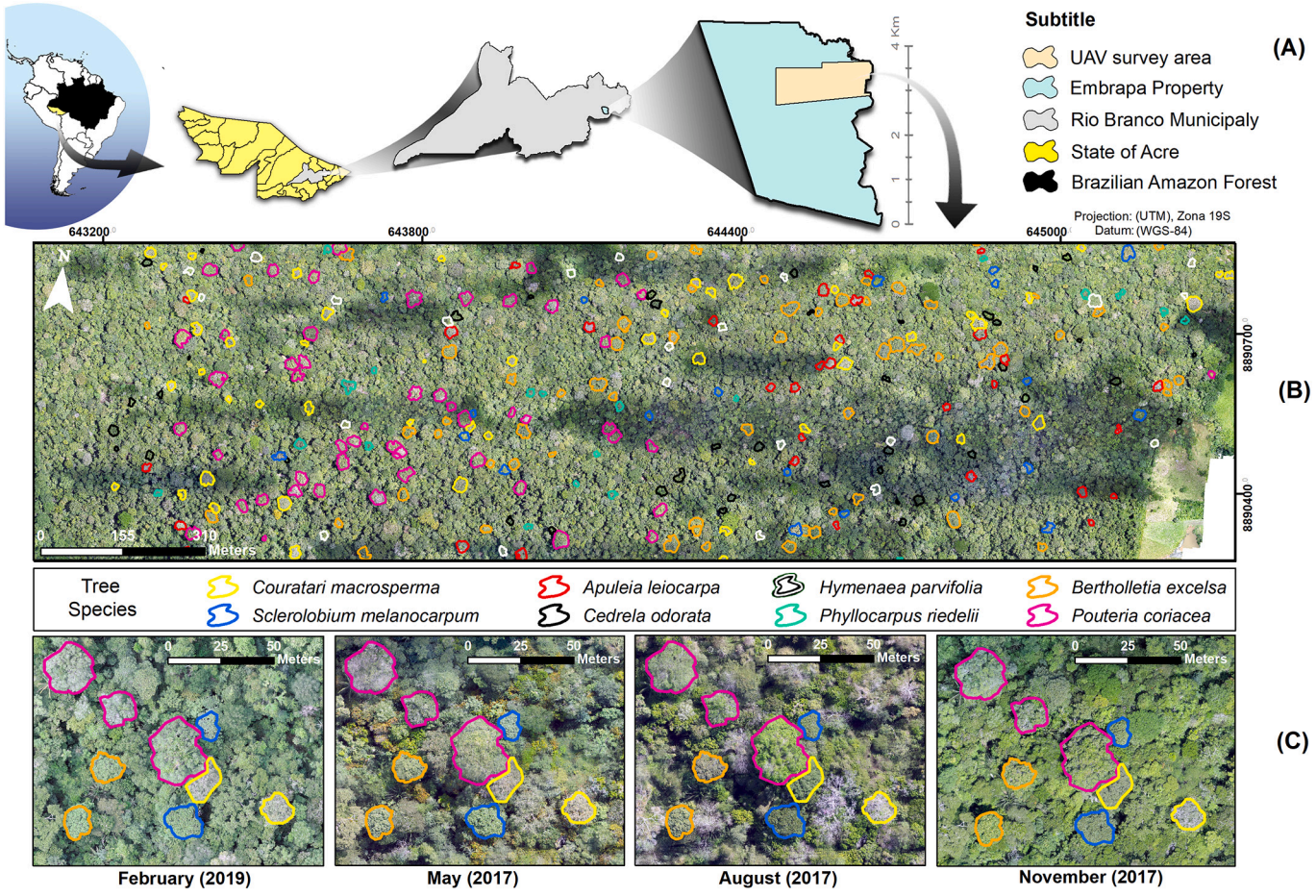


Fig. 1. (A) Location of the study area in the Acre state, southwestern Amazon, Brazil. (B) True color (RGB) composition of the orthomosaic with manually delineated individual tree crowns (ITCs) overlaid. (C) Examples of UAS images acquired in February, May, August, and November.

simultaneously classify and detect ITCs, making them an alternative to object-based approaches that require an image segmentation step before classification (Blaschke, 2010). Image segmentation algorithms are computationally intensive and require empirical parameter tuning. Moreover, applying these algorithms to delineate ITCs in tropical forests usually does not produce reliable results because of the structural complexity of the canopy (Tochon et al., 2015; Wagner et al., 2018).

Although ITC delineation provides valuable insights for forest management, such as crown size and shape, retrieving the exact contour of ITCs to map tree species is not essential. ITC delineation can be performed using instance segmentation approaches such as Mask R-CNN (He et al., 2017) or improved YOLO variants (e.g., (Hurtik et al., 2022)). However, these methods have prohibitive computational costs because they are composed of two deep neural networks, one to detect the objects and another to label the pixels that belong to them. In turn, semantic segmentation networks provide a more straightforward and fast way to map tree species at the ITC level (Ferreira et al., 2020).

Recent works, motivated by the great flexibility in data acquisition provided by UAS, used remote sensing images from multiple seasons to map tree species. Most studies were performed in temperate forests (e.g., Natesan et al. (2020); Grybas and Congalton (2021); Belcore et al. (2021); Liu (2022)) in which seasons are well defined and tree species diversity is limited. In tropical forests, tree species usually have complex phenological patterns, and selecting the most relevant time frame for image acquisition to map a set of tree species is challenging.

Images acquired by RGB sensors onboard UAS enable capturing and quantifying fine-scale features of ITCs. CNNs can automatically learn species-specific features in ultra-high spatial resolution RGB images. Previous studies showed the capacity of CNNs to differentiate among palm species in Amazonian forests by detecting crown architectural patterns (e.g., Ferreira et al. (2020); Morales et al. (2018). To improve tree species mapping, it is worth exploring other features, such as seasonal variations in plant characteristics caused by phenology. In this regard, multi-season images can be used to track phenological events such as flowering, fruiting, and leaf falling. The combined use of CNNs with multi-season UAS images has been poorly investigated, particularly in humid tropical forests where trees show highly diverse phenological patterns.

This study examines the utility of using multi-season RGB images acquired by low-cost UAS and CNNs to map tree species at the ITC level in Brazilian Amazonian forests. We tested the following hypothesis: CNNs can learn species-specific phenological characteristics; thus, fusing multi-season images for model training improves classification accuracy. To test our hypothesis, we defined two fundamental objectives (i) acquire UAS images along the year, encompassing the two main seasons in Amazonia; (ii) assess the classification accuracy of a CNN model trained with single-date and multi-date images.

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, Brazil (10°01′22”S, 67°40′3”W) (Fig. 1a). It is a highly diverse native rainforest area of 1600 ha about 200 m above the sea level (a.s.l).

Table 1
List of tree species with crowns manually delineated in the UAS orthomosaic and identified to the species level: species names, number of individual tree crowns (ITCs) and average number of pixels per ITC.

Scientific name	Popular name	No of ITCs	No of pixels	Average no of pixels per ITC	Average crown diameter (pixels)
Pouteria coriacea (Pierre) Pierre	Abiorana-rosa	74	16,371,007	221,230	616
Bertholletia excelsa Bonpl.	Castanheira	85	15,653,626	184,160	550
Cedrela odorata L.	Cedro	52	3,906,764	75,130	359
Apuleia leiocarpa (Vogel) J.F. Macbr.	Garapeira	45	5,210,028	115,778	446
Phyllocarpus riedelii Tul.	Guaribeiro	24	2,014,050	83,919	398
Hymenaea parvifolia Huber	Jutaí	32	3,685,441	115,170	443
Sclerolobium melanocarpu Ducke	Tachi	27	3,352,427	115,170	432
Couratari macrosperma http://A.C. Sm.	Tauari	67	9,967,430	124,164	505

The orthomosaic used in this study comprises 260 ha. The experimental forest area receives 1950 mm of rain annually, and the average temperature is 24.8 (±0.8)° C (Ramos et al., 2009). The vegetation of the area is classified as an open rainforest with the presence of palms and bamboos (Veloso et al., 1991).

2.2. UAS images

We collected aerial images with the UAS DJI Phantom 4 Pro, equipped with an RGB (red, green, blue) CMOS sensor of 20-megapixel resolution, a 24 mm autofocus lens, and a manual shutter with autonomy up to 30 min. The UAS flew 120 m above the forest canopy with a cruising speed of 10 m/s, resulting in images with a GSD of 4 cm. We took a total of 1585 images with a forward overlap of 86.0% and side lap of 86.36% in eight consecutive flights. Before the flights, we established three ground control points (GCPs) on the edges of the forest reserve. A dual-frequency GNSS receiver was installed at each GCP and collected GPS and GLONASS data for 241 min. The average horizontal and vertical precision of the GCPs after post-processing were 10 cm and 3 cm, respectively. Finally, we used the PiX4D software program (Pix4D Inc.) to generate orthomosaics of the study area.

We acquired UAS images in different months to capture species-specific phenological characteristics such as leaf fall, flowering, changes in leaf color, and fruiting patterns. The aim was to encompass two main seasons in Amazonia: the dry and wet seasons. Thus, we used images from February (peak of the wet season), May (end of the wet season), August (peak of the dry season), and November (end of the dry season). Examples of multi-season UAS images are shown in Fig. 1c.

3. Methods

3.1. Individual tree crown (ITC) dataset

In the experimental forest area, we measured and identified to the species level all trees with a diameter at breast height (DBH) >50 cm. For this study, we selected the most economically important species for nut and timber production and delineated their ITCs in the UAS images (GSD = 4 cm). We identified each ITC to the species level using information from the forest inventory and expert knowledge from botanical identification. Manual ITC delineation was carefully performed so that a single ITC polygon encompasses the same tree in all four images (Fig. 1c). We outlined a total of 406 ITCs, corresponding to eight species. The number of ITCs per species, as well as the number of pixels, are shown in Table 1.

3.1.1. Tree species classification method

We performed tree species classification using the ResNet-18 (He et al., 2016) CNN incorporated into the DeepLabv3+ architecture (Chen et al., 2018), which is considered a state-of-the-art deep learning model for semantic segmentation and was successfully used for tree species classification from remotely sensed images (Morales et al., 2018; Ferreira et al., 2021, 2020). We used ResNet-18 because it has the best trade-off between computational time and classification accuracy compared to other ResNet variants in preliminary tests. The ResNet-18 comprises residual blocks that forward the feature maps (activations) from shallow to deeper layers using the so-called skip connections, which allows the training of much deeper models. DeepLabv3+ is an encoder-decoder architecture designed to capture multi-level features. The encoder module performs feature extraction by gradually reducing the spatial dimension of the input image. The decoder recovers the original size of the input image and performs semantic segmentation using features extracted by the encoder. (Fig. 2). More details regarding ResNet-18 and DeepLabv3+ can be found in He et al. (2016) and Chen et al. (2018), respectively.

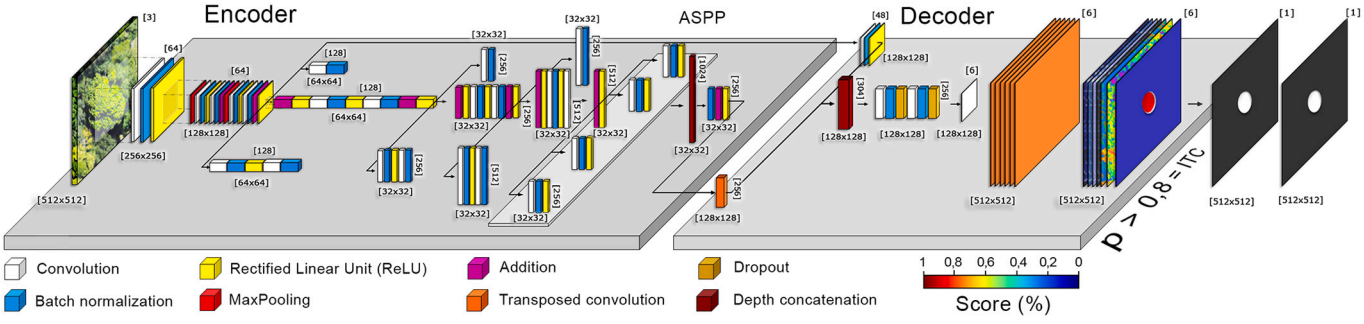


Fig. 2. DeepLabv3+ architecture with the ResNet-18 backbone network. The encoder module uses convolution operations and activation functions (ReLU) to extract features of tree species. The spatial dimension of the image is reduced by a factor of 16 at the end of the convolution process. The decoder module recovers the spatial dimension using transposed convolution operations. The softmax classifier is applied to generate score maps for each class. Finally, a probability rule is applied to detect the ITCs of target tree species within the input image patch.

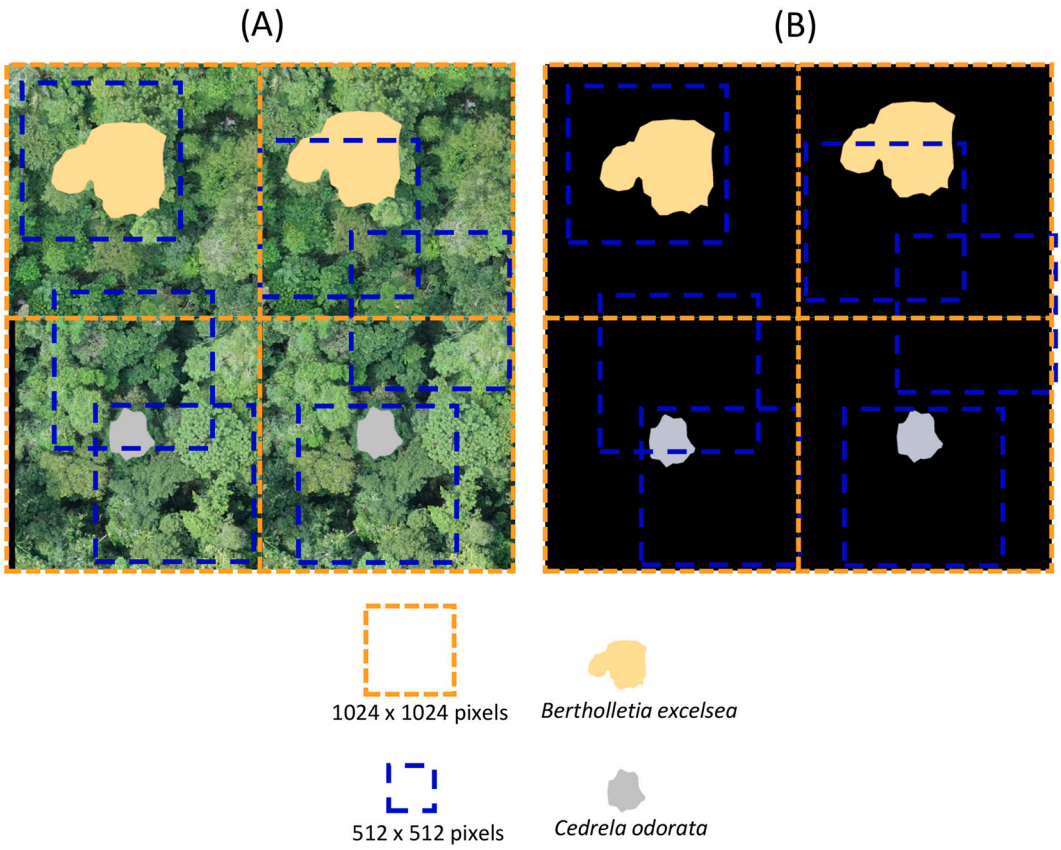


Fig. 3. Illustration of the model training strategy. (A) Synthetic image created from the original orthomosaic. (B) Labeled image. Image patches of size 1024 × 1024 (orange dotted squares) are created from randomly selected pixels of individual tree crowns of the target species (Table 1). These patches are concatenated into one rectangular tiled image. Then, patches of 512 × 512 (blue dotted squares) are randomly extracted from the synthetic and labeled image to feed the network with training data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Semantic segmentation networks are usually trained with densely labeled datasets, in which all pixels from a training image patch are labeled, meaning that they belong to a given class. This further means that it is expected that all classes present in the image patch should be known beforehand. In the case of map tree species in tropical forests, and given the highly diverse nature of these ecosystems, knowing all the tree species contained in an image patch is very unlikely. Thus, we trained our model using only the pixels from the target tree species (Table 1). To do this, we used a modified cost function that propagates the errors of desired classes only, as proposed by Martins et al. (2021).

3.1.2. Post-processing procedures

A semantic segmentation model will classify all pixels from an input image patch according to the number of classes used to train the model. In our case, the image pixels are labeled into eight species. However, the study area contains >200 tree species (Section 2.1), and labeling all of them as one of the eight target species represents an oversimplification. To avoid labeling all image pixels, we decided to classify them based on class membership probabilities, which indicate the confidence level of the predictions. A pixel is labeled in the scores maps of each species if its probability of belonging to that species is higher than 0.8. Otherwise, the pixel is considered an unknown species. We verified that this procedure produced realistic species maps in which ITCs of the target species are detected.

3.2. Experimental set-up

First, we separated the ITCs into 60% for training and 40% for testing. Then, we performed a vector-to-raster conversion of the ITC polygons to produce a densely labeled dataset containing class information of each image pixel. Notably, pixels from non-target species were not considered during network training (see Section 3.1.1).

We produced a synthetic image using the training ITCs. We randomly chose five pixels for each ITC, from which we grew patches of size 1024 × 1024 pixels. We then extracted these patches from the UAS and the labeled image and concatenated them into one rectangular tiled image (Fig. 3). Finally, we randomly extracted 2500 patches of size 512 × 512 pixels from the synthetic image to feed the network with training data. We extracted a new set of 2500 patches for each epoch. We used the stochastic gradient descent with momentum (SGDM) algorithm to update the network parameters (weights and biases) (Murphy, 2012). The total number of epochs was 25, as the classification accuracy did not change significantly after this number of epochs. The mini-batch size was 12, according to the available memory in the GPU. We used random reflection and rotation operations to augment the training data and to prevent the network from overfitting.

We performed two experiments to evaluate the use of multi-temporal images to discriminate among the species: (i) we used the images from each month (February, May, August, and November) individually, and (ii) we combined the images from each month into a single 12-bands image to train our model.

Training and inference were performed on a desktop workstation with an Intel Core i7–8700 3.2GHz CPU, 64GB of main memory, and an NVIDIA® GeForce Titan V GPU with 12GB of dedicated memory and 5120 CUDA® cores.

3.3. Accuracy assessment

The accuracy assessment was performed with the testing ITCs, which were not used for training. We computed the confusion matrices with

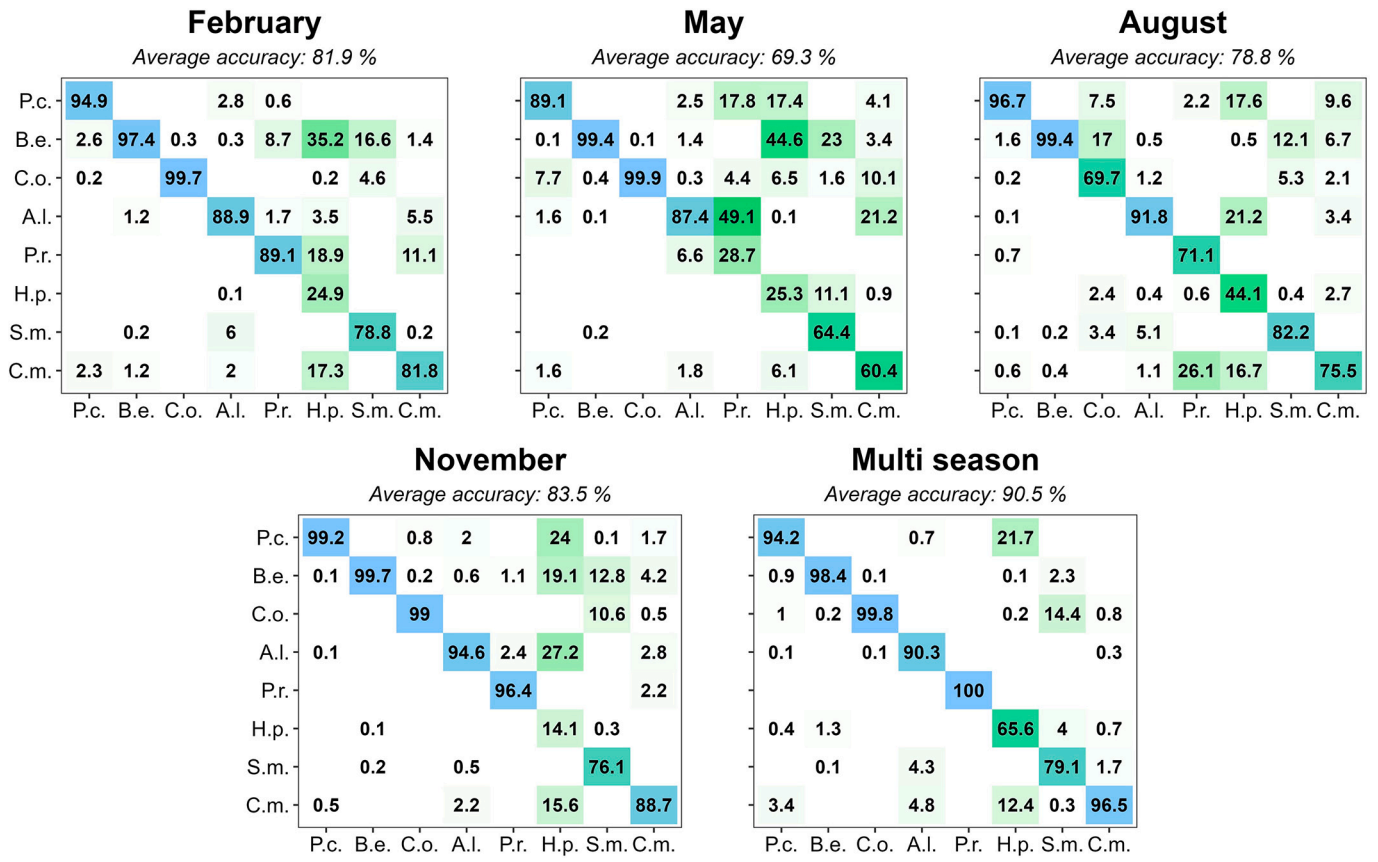


Fig. 4. Confusion matrices with the classification accuracy of tree species mapping using UAS images acquired in February, May, August, and November and a multi-season composition obtained by combining the images from each month. The producer's accuracy of each species in the diagonal cells (highlighted in blue), while the misclassification rate between species is shown in the off-diagonal cells. P.c. (*Pouteria coriacea*), B. e. (*Bertholletia excelsa*), C.o. (*Cedrela odorata*), A.p. (*Apuleia leiocarpa*), P.r. (*Phyllocarpus riedelii*), H.p. (*Hymenaea parvifolia*), S.m. (*Sclerolobium melanocarpum*), C.m. (*Couratari macrosperma*). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

intersecting samples between the reference and predicted ITCs. From the confusion matrices, we calculated the user's and producer's accuracy, the F1-score. The user's accuracy (UA), also known as precision, is the probability that a pixel classified as a particular species represents that species in the reference:

$$UA_i = \frac{TP_i}{TP_i + \sum_j M_{i'j}} \quad (1)$$

Where 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 class (j). The producer's accuracy (PA), also known as recall, is the probability that the pixels of a particular species are correctly classified in the reference, or:

$$PA_i = \frac{TP_i}{TP_i + \sum_j M_{ij'}} \quad (2)$$

The F1-score is the harmonic mean between the user's and producer's accuracies, representing their balance. F1-score is computed as:

$$2\frac{(UA^{*}PA)}{(UA + PA)} \quad (3)$$

Since we are not interested in outlining the ITC boundaries perfectly, we did not compute accuracy metrics designed for object detection, such as the intersection over union (IoU). Instead of IoU, we computed the percentage of ITCs correctly predicted in the test set. An ITC was considered correctly predicted if >50% of its pixels were correctly labeled by the model.

4. Results

The classification accuracy of tree species is shown in Fig. 4. The highest average accuracy was 90.5%, using images from all seasons. Combining images acquired in different seasons generally resulted in a significant decrease in the misclassification rate between species. For example, the misclassification between *P. riedelii* and *A. leiocarpa* was 49.1% in May and 0% in the multi-season composition. Similarly, *H. parvifolia* and *B. excelsa* showed misclassification rates above 35% in February and May but only 0.1% in the multi-season scenario. Such decreases in misclassification between species suggest that the network automatically learned how to select the most discriminative characteristics of each species in the multi-season images.

The highest average accuracies were observed in the rainy season, meaning February and November, with 81.9% and 83.5%, respectively (Fig. 4). The average accuracy did not reach 80% in the dry season (May and August). Some species showed remarkable changes in the producer's accuracy between seasons. For example, the producer's accuracy of *C. odorata* was 99.7% in February and dropped to 69.7% in August (Fig. 4). Similarly, *P. riedelii* was classified with 96.4% accuracy in November and with 28.7% in May. The drops in the producer's accuracy of these species can be explained by leaf fall in the dry season. Fig. 5 shows an ITC of *C. odorata* in UAS images acquired in August (Fig. 5A) and November (Fig. 5B). One can note that the absence of leaves significantly changes the crown structure by exposing primary and lateral branches. Conversely, the fully foliated crown in the rainy season forms a billowy pattern. UAS images acquired in August and May show differences caused by leaf fall in an ITC of *P. riedelii* (Fig. 5C,D). Unlike,

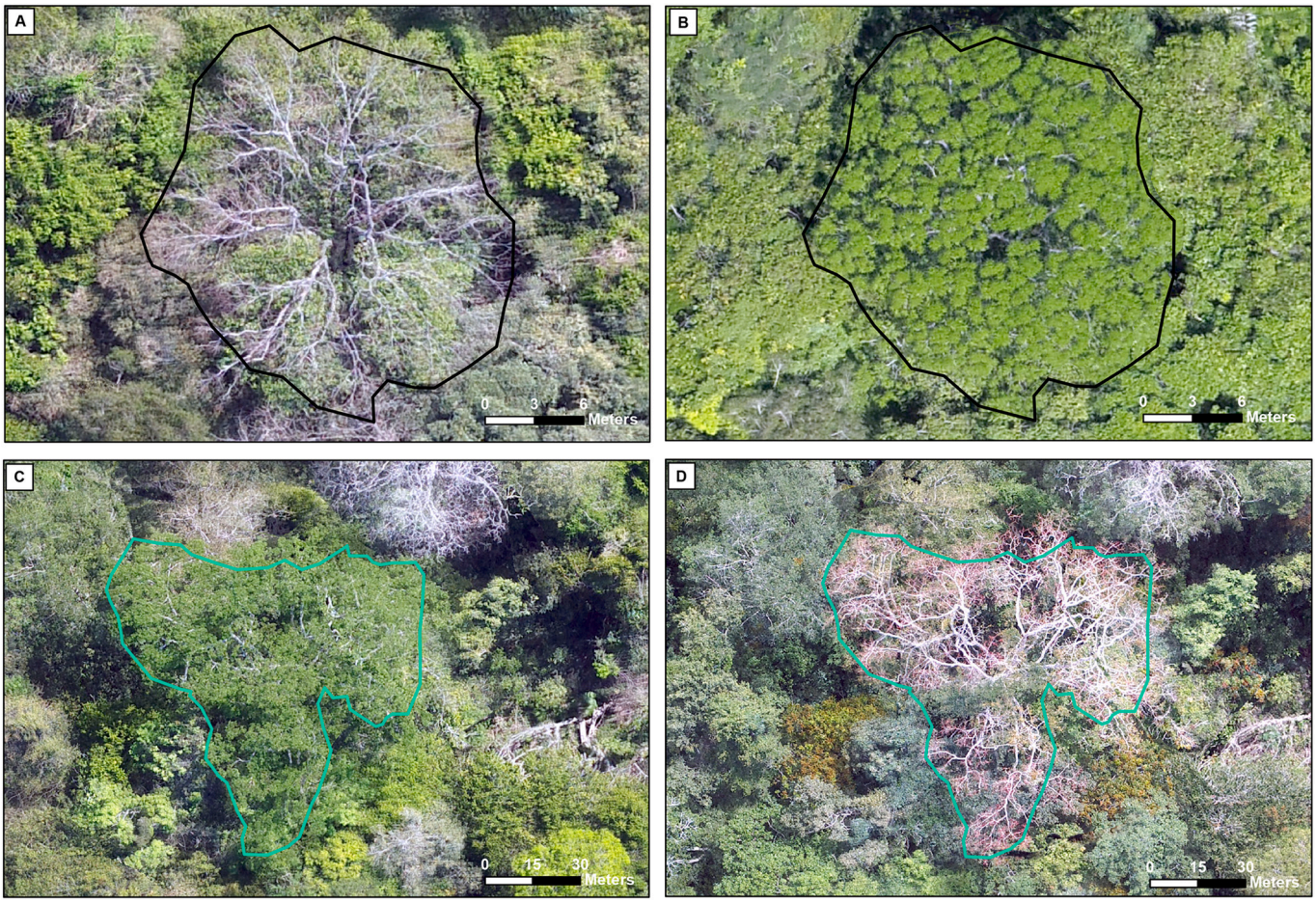


Fig. 5. UAS images of an individual tree crown of *C. odorata* obtained in August (A) and November (B) and *P. riedelii* in August (C) and May (D).

Table 2
User's accuracy (UA), producer's accuracy(PA), F1-score (F1), and percentage of individual tree crowns (ITCs) that were correctly predicted in the test set. The accuracy metrics result from classifications of UAS images acquired in February, May, August, and November and a multi-season composition of the images of all seasons.

Period	Accuracy	Species								Average
		P.c.	B.e.	C.o.	A.l.	P.r.	H.p.	S.m.	C.m.	
February	UA (%)	99.1	93.7	98.1	83.4	60.0	98.8	83.2	86.2	87.8
	PA(%)	94.9	97.4	99.7	88.9	89.1	24.9	78.8	81.8	81.9
	F1 (%)	96.9	95.5	98.9	86.0	71.7	39.8	80.9	84.0	81.7
	ITCs (%)	90.0	100.0	95.0	94.0	80.0	69.0	82.0	85.0	86.9
May	UA (%)	96.0	93.6	76.3	66.6	32.1	41.0	97.9	91.4	74.4
	PA(%)	89.1	99.4	99.9	87.4	28.7	25.3	64.4	60.4	69.3
	F1 (%)	92.4	96.4	86.5	75.6	30.3	31.3	77.7	72.7	70.4
	ITCs (%)	93.0	85.0	86.0	83.0	50.0	62.0	82.0	89.0	78.8
August	UA (%)	94.2	95.8	64.5	82.4	78.9	86.8	83.3	82.3	83.5
	PA(%)	96.7	99.4	69.7	91.8	71.1	44.1	82.2	75.5	78.8
	F1 (%)	95.4	97.5	67.0	86.9	74.8	58.5	82.7	78.8	80.2
	ITCs (%)	93.0	94.0	67.0	83.0	70.0	85.0	73.0	89.0	81.8
November	UA (%)	95.8	95.9	91.7	84.4	91.4	84.3	96.7	95.8	92.0
	PA(%)	99.2	99.7	99.0	94.6	96.4	14.1	76.1	88.7	83.5
	F1 (%)	97.5	97.8	95.2	89.2	93.8	24.2	85.1	92.1	84.4
	ITCs (%)	93.0	94.0	71.0	100.0	80.0	85.0	100.0	96.0	89.9
Multi-season	UA (%)	98.2	99.0	89.6	99.0	100.0	64.5	77.8	88.4	89.6
	PA(%)	94.2	98.4	99.8	90.3	100.0	65.6	79.1	96.5	90.5
	F1 (%)	96.1	98.7	94.4	94.4	100.0	65.1	78.4	92.3	89.9
	ITCs (%)	100.0	97.0	95.0	100.0	60.0	85.0	100.0	81.0	89.8

C. odorata, in August, the crown of *P. riedelii* does have leaves.

The combination of images from different months increased both the average classification accuracy (Fig. 4) and the F1-score of the species (Table 2). For example, the F1-score of *H. parifolia* in November, February and May did not exceed 40% but increased to 65.1% after combining multi-season images. Using images from multiple seasons also increased the percentage of correctly predicted ITCs per species (Table 2). On average, more ITCs were correctly predicted in the rainy months (February and November) than in the dry seasons (May and August). Most of the tree species had >60% of their ITCs correctly

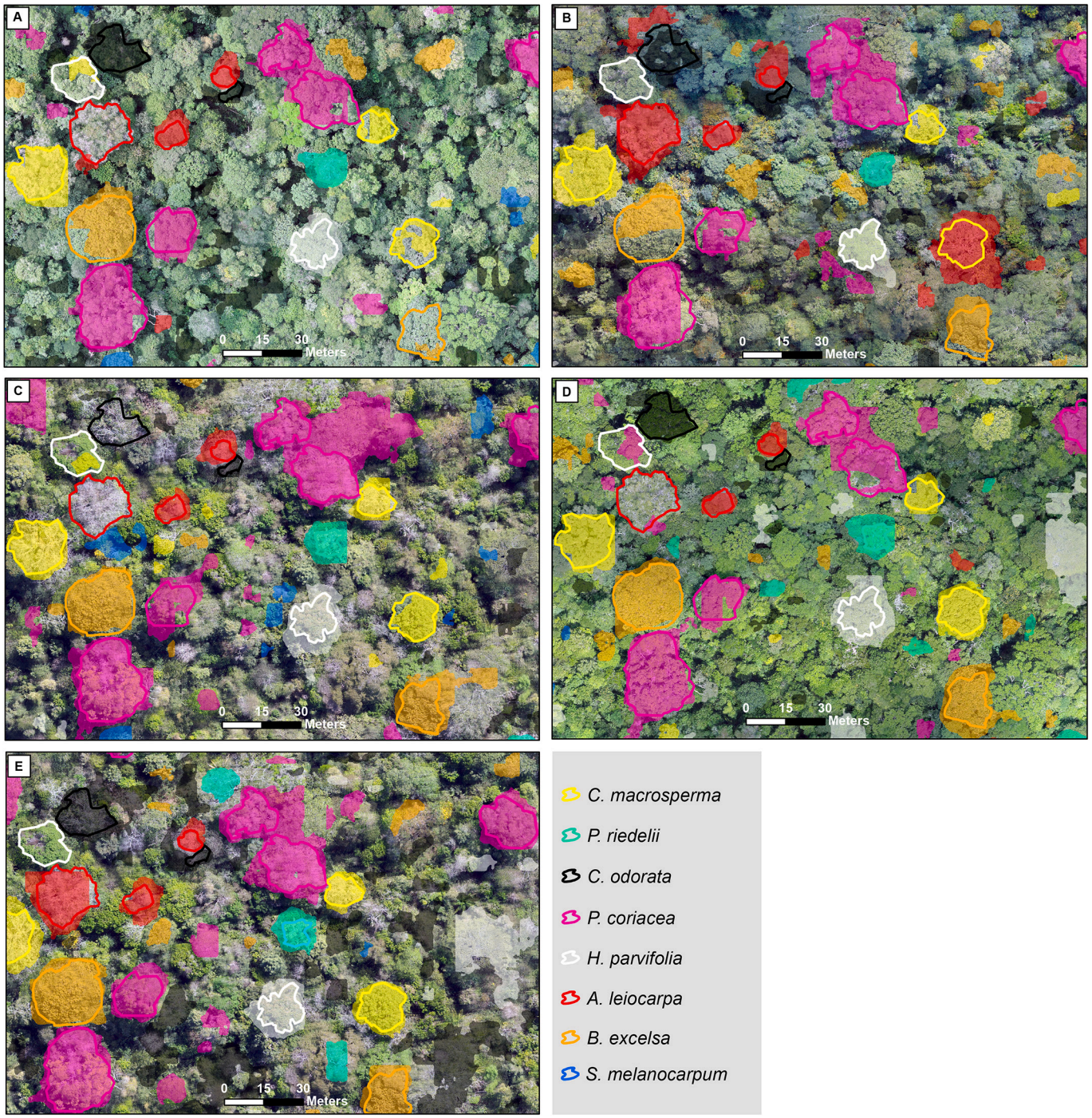


Fig. 6. Model predictions were obtained using images from (a) February, (b) May, (c) August, (d) November, and (e) a combination of all months. Reference ITCs not used to train the model are shown as unfilled polygons, while model predictions are shown as filled polygons. For clarity, the UAS image shown in (e) is from August.

predicted in the test set, except *P. riedelii*.

Examples of predictions obtained for images of each month are shown in Fig. 6. *P. coriacea*, *B. excelsa* and *A. leiocarpa* presented the homogeneous morphological characteristics during the year. The neural network model detected them in all months. Conversely, *C. odorata* showed a strong leaf fall in the dry season and was not detected by the model using the August image only (Fig. 6c). By combining images from multiple seasons *C. odorata* was successfully detected (Fig. 6e). In general, one can note a reduction in misclassification after using images from multiple seasons.

5. Discussion

Tree species mapping in Amazonian forests using UAS images and CNNs is a powerful tool for forest conservation and management. User's accuracy in November ranged from 84.4% to 95.9%. Even species with <30 ITCs (*P. riedelii* and *S. melanocarpu*) reached high levels of classification accuracy, with F1-scores of 93.8% and 85.1%, respectively. *H. parvifolia* was the most challenging species to classify. Its ITCs showed highly variable patterns with several leafless trees, as well as a high degree of overlap between neighboring crowns. More ITCs of *H. parvifolia* are needed to incorporate its canopy structural variability in model training.

Nuijten et al. (2019) examined the multi-seasonal consistency of ITC segmentation in a Canadian forest and found that the beginning of leaf fall is an adequate period to segment ITCs because the canopy is less dense. The authors argue that when leaves begin to fall, the border among ITCs is more evident. Similarly, we observed an improvement in the visualization of *H. parvifolia* ITCs when their leaves started to fall in August. However, the beginning of leaf fall varied among the species. For example, in August, all *C. odorata* trees have already lost all their leaves, which negatively impacted the classification accuracy. Thus, we believe that the total absence of leaves hampers species detection.

The low cost of acquiring RGB images with a UAS and the flexibility of data acquisition promotes a more frequent use of multi-season images for tree species mapping. Natesan et al. (2020) used an RGB camera onboard a UAS to classify pine species in Canada over three years. The combination of annual images improved the classification accuracy, which reached 90% in some cases. Modzelewska et al. (2021) used hyperspectral data acquired in July (early summer), August (late summer), and October (autumn) acquired over a forest area in Poland to discriminate among seven tree species. The classification results show that a multi-season dataset produced the best results reaching 84–94% of overall accuracy. In a temperate forest area in New Hampshire, USA, Grybas and Congalton (2021) investigated whether the classification accuracy of 14 tree species could be improved with multi-temporal UAS data. The authors collected five RGB images between April and June, encompassing the spring and summer seasons. They showed that a five-date stack of the multi-temporal images provided the best results (61.1% of overall accuracy). Studies using multi-season RGB images acquired by UAS, particularly on the Brazilian Amazon, do not exist, thereby highlighting the relevance of this study. To the best of our knowledge, this is the first study to show the potential of multi-season UAS images for classifying tree species in tropical forests. Another innovative aspect of our work is the classification method used. Most studies employed pixel- or object-based approaches along with classical machine learning algorithms such as random forests and support vector machines. In pixel-based approaches, pixels are usually labeled based on their spectral characteristics, thus neglecting the spatial relationship among neighboring pixels. Object-based procedures use a segmentation algorithm to delineate tree crows before classifying them. Image segmentation is challenging because it depends on empirical parameter settings and is computationally intensive, mainly if applied in UAS images featuring hyperspatial resolutions (GSD < 10 cm). The CNN models vanquish the limitations of both pixel-based and object-based approaches by simultaneously performing ITC delineation and classification. Moreover, CNN-based methods are robust to differences in viewing and illumination conditions (Natesan et al., 2020) and our results suggest that they learn species-specific phenological patterns.

Combining UAS images and deep learning for mapping tree species in Amazonian forests paves the way for a new era in conservation and forest resource management (Onishi and Ise, 2021). The use of UAS for mapping native forests in the Amazon contributes to more optimized planning of environmental activities, helps to reduce costs and fieldwork, enabling access to the qualitative, quantitative, and spatial explicit characteristics of the forest in a faster and more dynamic way. Data collection flexibility provided by UAS particularly enabled taking advantage of multi-season images for tree species mapping.

6. Conclusions

In this work, we show that fusing RGB images from multiple seasons with CNNs improves tree species mapping in Amazonian forests. By fusing images acquired in the rainy (November and February) and dry seasons (May and August), we verified an improvement in the classification accuracy of 27.2 percentage points compared with single-date classifications. We developed an effective model training strategy in which a synthetic image is produced based only on the reference ITCs. The use of CNNs and multi-season UAS images proved helpful in mapping economic relevant tree species in the Amazon and has the potential to contribute to forest management and conservation. Future studies will focus on expanding the ITC database by including more species. Moreover, we intend to analyze the generalization capability of our models to map tree species in other areas using low-cost RGB images.

Declaration of Competing Interest

None.

Data availability

No data was used for the research described in the article.

Acknowledgments

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research. M.P. Ferreira was supported by the Brazilian National Council for Scientific and Technological Development (CNPq) (grant 306345/2020-0) and the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ) grants #248496/2019 and #259727/2021. This study was funded in part by the Coordination for the Improvement of Higher Level Personnel (CAPES) - Finance Code 001 (A. Corte 88887.373249/2019-00), MC-371 TIC/CNPq No 28/2018 (408785/2018-7; 438875/2018-4), CNPq No 09/2018 (302891/2018-8).

References

de Almeida Papa, D., de Almeida, D.R.A., Silva, C.A., Figueiredo, E.O., Stark, S.C., Valbuena, R., Rodriguez, L.C.E., d'Oliveira, M.V.N., 2020. Evaluating tropical forest classification and field sampling stratification from lidar to reduce effort and enable landscape monitoring. For. Ecol. Manag. 457, 117634. URL: https://www.sciencedirect.com/science/article/pii/S037811271931179X https://doi.org/10.1016/j.foreco.2019.117634.

Belcore, E., Pittarello, M., Lingua, A.M., Lonati, M., 2021. Mapping riparian habitats of natura 2000 network (91e0*, 3240) at individual tree level using UAS multi-temporal and multi-spectral data. Remote Sens. 13, 1756.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16.

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.

Dalla Corte, A.P., Rex, F.E., Almeida, D.R.A.D., Sanquetta, C.R., Silva, C.A., Moura, M.M., Wilkinson, B., Zambrano, A.M.A., Cunha Neto, E.M.D., Veras, H.F.P., Moraes, A.D., Klauberg, C., Mohan, M., Cardil, A., Broadbent, E.N., 2020. Measuring individual tree diameter and height using gatoreye high-density uav-lidar in an integrated crop-livestock-forest system. Remote Sens. 12.

Ferreira, M.P., de Almeida, D.R.A., de Almeida Papa, D., Minervino, J.B.S., Veras, H.F.P., Formighieri, A., Santos, C.A.N., Ferreira, M.A.D., Figueiredo, E.O., Ferreira, E.J.L., 2020. Individual tree detection and species classification of amazonian palms using UAS images and deep learning. For. Ecol. Manag. 475, 118397 https://doi.org/10.1016/j.foreco.2020.118397.

Ferreira, M.P., Lotte, R.G., D'Elia, F.V., Stamatopoulos, C., Kim, D.H., Ribeiro, M.B.N., Benjamin, A.R., 2021. Accurate mapping of Brazil nut trees (*Bertholletia excelsa)* in amazonian forests using worldview-3 satellite images and convolutional neural networks (vol 63, 101302, 2021). Ecol. Inform. 64.

Grybas, H., Congalton, R.G., 2021. A comparison of multi-temporal RGB and multispectral UAS imagery for tree species classification in heterogeneous New Hampshire forests. Remote Sens. 13, 2631.

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.

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.

Hurtik, P., Molek, V., Hula, J., Vajgl, M., Vlasanek, P., Nejezchleba, T., 2022. Poly-yolo: higher speed, more precise detection and instance segmentation for yolov3. Neural Comput. & Applic. 34, 8275–8290.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.

Liu, H., 2022. Classification of urban tree species using multi-features derived from four-season RedEdge-MX data. Comput. Electron. Agric. 194, 106794.

Martins, G.B., La Rosa, L.E.C., Happ, P.N., Coelho Filho, L.C.T., Santos, C.J.F., Feitosa, R.Q., Ferreira, M.P., 2021. Deep learning-based tree species mapping in a highly diverse tropical urban setting. Urban For. Urban Green. 64, 127241.

Modzelewska, A., Kamińska, A., Fassnacht, F.E., Stereńczak, K., 2021. Multitemporal hyperspectral tree species classification in the białowieża forest world heritage site. Forestry 94, 464–476.

Morales, G., Kemper, G., Sevillano, G., Arteaga, D., Ortega, I., Telles, J., 2018. Automatic segmentation of *Mauritia flexuosa* in unmanned aerial vehicle (UAS) imagery using deep learning. Forests 9. https://doi.org/10.3390/f9120736.

Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT press.

Natesan, S., Armenakis, C., Vepakomma, U., 2020. Individual tree species identification using dense convolutional network (DenseNet) on multitemporal rgb images from uas. J. Unmanned Vehicle Syst. 8, 310–333.

Nuijten, R., Coops, N., Goodbody, T., Pelletier, G., 2019. Examining the multi-seasonal consistency of individual tree segmentation on deciduous stands using digital aerial photogrammetry (DAP) and unmanned aerial systems (UAS). Remote Sens. 11, 739. URL: https://doi.org/10.3390/rs11070739.

Onishi, M., Ise, T., 2021. Explainable identification and mapping of trees using UAV RGB image and deep learning. Sci. Rep. 11. URL: https://doi.org/10.1038/s41598-020-79653-9.

Ramos, A.M., Dos Santos, L.A.R., Fortes, L.T.G., 2009. Normais climatológicas do brasil, 1961-1990. Instituto Nacional de Meteorologia-INMET.

Tochon, G., Féret, J., Valero, S., Martin, R., Knapp, D., Salembier, P., Chanussot, J., Asner, G., 2015. On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images. Remote Sens. Environ. 159, 318–331. https://doi.org/10.1016/j.rse.2014.12.020.

Veloso, H.P., Rangel-Filho, A.L.R., Lima, J.C.A., 1991. Classificação da vegetação 600 brasileira, adaptada a um sistema universal. Instituto Brasileiro de Geografia e Estatística.

Wagner, F.H., Ferreira, M.P., Sanchez, A., Hirye, M.C., Zortea, M., Gloor, E., Phillips, O. L., de Souza Filho, C.R., Shimabukuro, Y.E., Aragão, L.E., 2018. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. ISPRS J. Photogramm. Remote Sens. 145, 362–377. https://doi.org/10.1016/j.isprsjprs.2018.09.013. sI: Latin America Issue.

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.