BiomeAzuero2022: A Fine-Grained Dataset and Baselines For Tree Species Classification with Ground Images

Ziwei Gu [*1], Gauri Jain [*1], Hongwen Song [*1], Isak Diaz [*2], Margaux Masson-Forsythe [*2], Jorge Valdes [2], Milind Tambe [1]

[1] Harvard University
[2] Earthshot Labs Biome
{ziweigu, gaurigain, hongwensong}@g.harvard.edu,
{margaux, isak}@earthshot.eco, jjvaldes1698@gmail.com, milind_tambe@harvard.edu

Abstract

It is becoming increasingly popular for organizations to invest in carbon credits programs involving planting trees to offset their carbon footprint. However, in order to correctly measure carbon sequestration, it is important to know which tree species exist in a given region. This identification task requires very well-trained local botanists who are not only an expensive resource, but also not accurate enough. AI-assisted tree classification has become an promising way to solve the problem, but low data quality has been the limiting factor. Our work focuses on determining how to build a high quality dataset of tree parts that can lead to the most accurate tree classification. We contribute **BiomeAzuero2022**, a publicly available image dataset of 9071 tree images captured in Azuero, Panama with ground truth species labels, organized by different parts of trees. We further provide a methodology for estimating feature importance via machine learning and interpretability methods with the goal of gathering higher-quality image data through a case study on **BiomeAzuero2022**.

Introduction

There is enormous potential for tree restoration efforts in forests to help mitigate climate change (Bastin 2019). In order to successfully implement tree planting initiatives, we need accurate metrics for measuring tree carbon sequestration, which is done with some baseline allometric models. These models require wood density as a parameter, which differs for every tree species, so in order to properly measure tree carbon sequestration we need a strong understanding of which species exist in any given region ((Chave J 2014)).

Our work is based out of the tropical dry forest in Azuero, Panama, which is one of the most biodiverse regions in the world. With this many species, identifying and classifying trees requires highly-trained local botanists, and even for them the task takes a lot of time and effort and is error-prone. In contrast, machine learning models have the ability to find subtle differences between thousands of inputs and classify images correctly (O'Shea 2015). AI-assisted tree classification can greatly reduce the work field ecologists need to do to identify trees on the ground.

[*]These authors contributed equally.

The first step to trying to build up good classification models is to create a dataset of labeled tree images. The process of photographing different features of a tree requires substantial manual effort and time (which we found through two months of photography field work in the region). This led us to focus our efforts on understanding what is needed from a dataset of trees in order to produce the best classification. Understanding which part of a tree and what kind of photos are most necessary for classification can greatly reduce the labor needed to photograph each tree.

We also choose to focus on strengthening our dataset as a result of recent critiques of training machine learning models that are based on low quality data. Data issues can cause significantly negative downstream effects in AI models (Sambasivan et al. 2021).

Through this work, we contribute a public dataset of 10 tree species in Azuero, Panama separated into different tree parts. We also analyze the strengths and weaknesses of our current Panama dataset using (1) saliency and class activation maps to understand the qualities needed in a good image and (2) accuracy in existing classification models when using different tree parts as inputs. This analysis will inform and improve future field photography in these forest regions.

Related Work

Tree Species Image Datasets and Classification Models

Most existing approaches for tree species classification utilize the leaf pictures as the major input source as leave often are the most accessible and informative feature of trees. (Lee et al. 2015) was one of the first to propose the state-of-the-art machine learning approach for plant species classification on 44 species from the Royal Botanic Gardens in England. Alternatively, using the LeafSnap dataset, (Barré et al. 2017) developed the "LeafSnap" system that extracts features representing the curvature of the leaf's contour over multiple scales and achieves acceptable accuracy when identifying the species from LeafSnap dataset (Kumar et al. 2012) , which consists of 184 tree species in the Northeastern United States. However leaves are far less accurate when it comes to identifying trees.

To expand the variety of plant images from merely leaves, the citizen science website iNaturalist published an state-of-the-art dataset iNat2017 for image classification and detection for "in the wild data" featuring large numbers of imbalanced, fine-grained, categories (Van Horn et al. 2018). iNaturalist also embeds their own species identification model into their official website, which is trained on all species (animals, insects, plants, etc.) from around the world, while we would like to focus specifically on tree classification in Azuero. What's more, the model proposed by iNaturalist is unfortunately not available for commercial applications and to not perform well for all parts of a tree (only achieving around 0.1 accuracy using tree profile or leaf pictures from BiomeAzuero2022).

Extensive studies have focused on utilizing Light Detection and Ranging (LiDAR) data (Brandtberg 2007) (Jones, Coops, and Sharma 2010) (Zhou et al. 2017) for tree species classification tasks in recent years, which allows for the acquisition of three-dimensional point clouds of scanned areas. While LiDAR-based methods sound promising in large scale applications, they usually suffer from the issues of data sparsity and the difficulty in tree separation. Alternative remote-sensing methods, such as those based on multispectral/hyperspectral data (Clark and Roberts 2012) (Ghosh et al. 2014) and Synthetic Aperture Radar (SAR) (Schmitt, Shahzad, and Zhu 2015) (Ranson et al. 2001) data, have similar limitations in individual tree separation and information de-noising. By contrast, we are photographing high quality tree images on the ground which eliminates the tree separation problem and maximize the data quality.

Interpretable Image Classification

In order for machine learning models to move towards their integration into the real world for greater impact, they have to be transparent and produce explainable results. The most common way to interpret image classification is perhaps through saliency maps, which highlight the most important pixels of an image for its classification (Simonyan, Vedaldi, and Zisserman 2013). Saliency is usually measured by computing the gradient of the activation function for a particular class with respect to every pixel in an image. Another promising class of visualization methods is class activation maps (CAMs), which are generated by calculating the gradients at the last layer in a deep neural network that contains spatial information, instead of the last layer (Zhou et al. 2016).

While a lot of work is done on optimizing the interpretability methods above, the focus is mainly on model behavior validation; exploratory analysis is possible, but clearly expensive (Balayn et al. 2021). We propose a procedure that generates insights from saliency maps and class activation maps which ultimately sheds light on the data gathering process, providing guidelines for image capturing based on where attention is paid by the models.

Dataset

In this section we describe the details of the dataset, including the data collection process, data features, and challenges encountered while collecting the dataset, which future researchers might find helpful when constructing their own datasets. The dataset is publicly available at https://www.kaggle.com/datasets/earthshot/azuero-trees-1024.

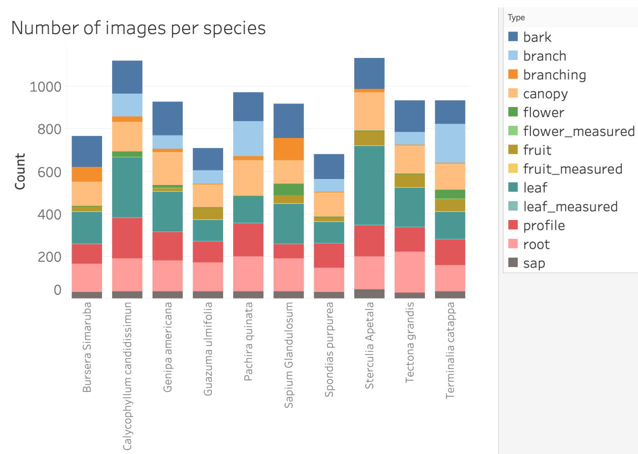

Figure 1: Dataset structure of the 10 most common tree species in Panama

Data collection

BiomeAzuero2022 contains 9071 pictures collected by the Biome team from Earthshot Labs with the aim of streamlining tree inventory work for improved biomass calculation and verification. Biome collaborated with local Smithsonian botanist , Jorge Valdes, in the tropical dry forests at Azuero, Panama for data collection and the detailed tree image collection process is shown in Figure 2. Clear instructions and guidelines for photo gathering are given to local botanist. Guidelines include number of pictures to take for each tree, specification of parts of trees (such as canopy, bark, etc.), and the direction of certain types of pictures (e.g. the picture of leaves should be taken horizontally). Figure 3 shows two examples of the guidelines.

Figure 2: Image collection process

Features

The tree classification dataset contains 10 species of trees photographed from tropical dry forest at Azuero, Panama. Up to 10 trees are photographed for each species. For each species, images are organized into categories, including tree profile, close-up bark, roots, canopy leaves, sap (optional), sap with white balance card (optional, only if provided), leaves bottom (optional), leaves top (optional), fruit/seed (optional), flower (optional).

Figure 1 shows the dataset structure of the 10 most common species. Time stamps and the GPS coordinates of each image are also recorded as part of the image meta data.

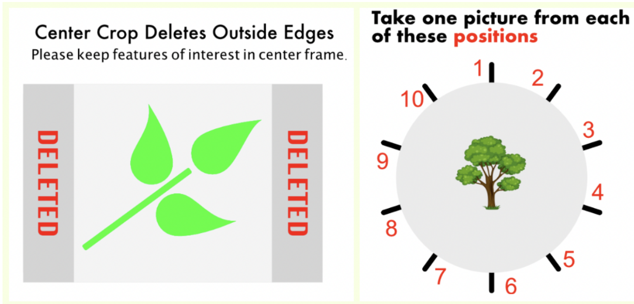

Figure 3: Example Guidelines

Model	Accuracy	F-1	Recall	Precision
MobileNet	0.35	0.44	0.35	0.65
DenseNet121	0.31	0.39	0.31	0.61
MobileNetV2	0.3	0.37	0.3	0.54
ResNet50	0.09	0.13	0.09	0.23
EfficientNetB5	0.08	0.08	0.08	0.13

Table 1: Model performance on all Panama collected images before pre-processing with train/test split 7-3 trees

Challenges

Labor limitations and data inconsistency are the main challenges. Even though the local botanist expert tried to strictly follow the guideline, mistakes (with respect to the number of pictures for each part) are inevitable especially when he has to take thousands of photos in several days. The collected dataset may not strictly follow the instructions, which led to additional efforts in downstream data reorganization and cleaning. We want to improve the efficiency and correctness by updating the photo guidelines which focus on important features tree so that local botanist expert can take less pictures for each tree.

Methods

Model

To run the feature importance analysis, we used the convolutional neural network backbone architecture called MobileNet after determining that it was the best performing model architecture (Table 1) for this dataset with 7-3 trees train/test split.

Macro Features

The first aspect of the dataset we choose to understand is which tree parts are needed for the most accurate tree classification. We call these features macro features because they refer to which images are being used instead of the content of each image. We look at feature importance in a multi-image input model and analyze which combination of images lead to highest accuracy. We expect that this process will inform which subset of images is most important to be taken while in the field.

Micro Features

We then study the micro features within each part of trees; that is, pixels in an image that contribute most to its classification. Building on the macro feature analysis, we focus our analysis on bark, branch, canopy, leaf, and profile as the other parts (fruit, flower, etc.) are harder to find. The workflow is as follows. First, all the mis-classified images from the testing set are collected, and an example is randomly selected. Then, a correctly classified image most similar to the previous image from the same part of tree is selected. After that, saliency maps and class activation maps are generated for both images and results are compared by a pair of researchers. We hypothesize that there are patterns with how images are (mis)classified and that visualizing them provides a mechanism for informing how images should be captured. We describe recurring patterns and themes from our experiments in the next section.

Experiments/Results

Macro Features

We use a modified 3 images input Mobile Net model to classify 10 tree species. Then we stress-tested the model using different combinations of tree parts to see which tree parts are most important for accurate classification. For this we used an 80-20 train test split. We try to measure how much accuracy each tree part adds on. Figure 4 shows:

$$addon_p = \frac{\sum_{\{S \subseteq U | p \in S\}} a_S - a_{S \setminus \{p\}}}{|\{S \subseteq U \mid p \in S\}|}$$

where p is a tree part, U is all possible tree parts, and a is the accuracy of the model given some subset of tree parts. We see that the profile image is the most helpful in terms of classifying the tree, followed by leaves, and bark is last. For now, we only analyze combinations of bark, leaves, and profile, but will run a larger scale analysis of different combinations of all the tree features in the future.

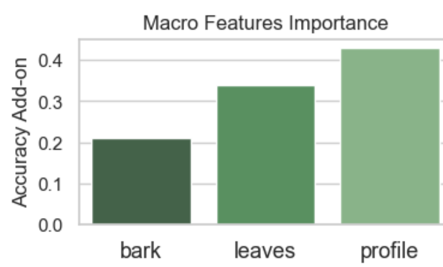

Figure 4: The plot shows how much accuracy on average is added on when each feature is added to the dataset. These results are based on an 80-20 train test split which tends to report higher accuracy, so these values are primarily used to compare features with each other

Micro Features

There seem to be two main types of issues with data that cause an image to be mis-classified, type (1) too much noise in the background, and type (2) missing key characteristics of a component.

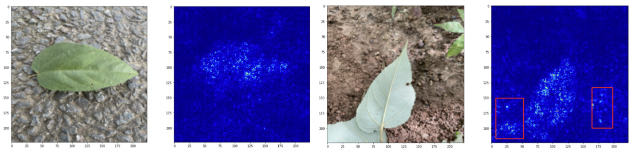

Figure 5: Correctly classified (left) and incorrectly classified (right) leaf images; Multiple leaves in the image add to noise.

In the first example, when the model predicts the class correctly, attention is paid entirely to the leaf itself, whereas in the mis-classified one the model spares attention to the other leaf in the photo, possibly treating the set of the two leaves as a single leaf. Similar examples indicate that photographers should minimize error (1) by capturing only one leaf in the photo.

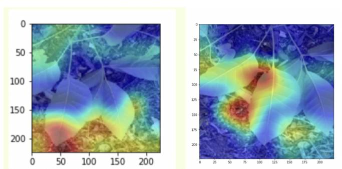

Figure 6: Correctly classified (left) and incorrectly classified (right) branch structure; Noise in the background overshadows tips of leaves.

When it comes to branch structure, the model invariably focuses on the tip of leaves as a distinguishing feature in correctly classified examples. However, in the mis-classified example shown above, the model fails to capture it because it is distracted by random noises in the background. Again, photographers could have reduced error (1) by removing salient features in the background.

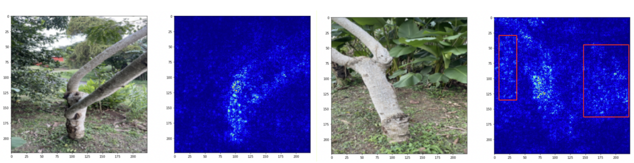

Figure 7: Correctly classified (left 2) and incorrectly classified (right 2) profile images; Other profile-like objects nearby confuse the model.

Figure 9 shows a pair of profile images, where the second one is mis-classified because other profile-like objects near the tree of interest interferes with the classification. Thus, in cases where the background cannot be cleared, photographers should try to minimize the number of similar objects in the background.

Equally problematic is issue (2). In this pair of canopy photos, the model predicts the correct label when it attends

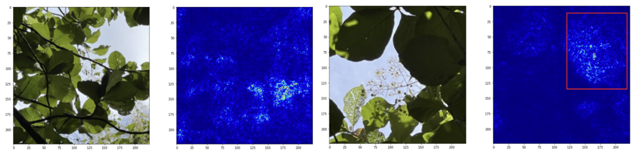

Figure 8: Correctly classified (left 2) and incorrectly classified (right 2) canopy images; The full distribution of leaves is not captured.

to many of the leaves there, thereby capturing the full distribution of leaves. By contrast, the model only focuses on 2 or 3 leaves in the mis-classified case. Therefore, We conjectured that canopy photos need to capture a larger number and a wider distribution of leaves. We noticed the same patterns in other canopy examples as well, where poorly-scoped images lead to the omission of key canopy features and incorrect classification.

Recommendation for Data Gathering Practices

From a macro feature perspective, photographers should focus more on leaf and profile pictures. Bark on the other hand does not add much and can likely be removed from the dataset collection.

Apart from identifying the prominent features, our analysis also confirms, or, in certain cases, complements, current guidelines for image capturing. More concretely, we propose the following guidelines for photographers on the ground, driven by our micro-feature analysis:

- **Bark images** should focus on areas of the bark that are clear of other foliage and avoid areas with excessive shading.
- **Branch images** should make the tips of leaves visible.
- **Canopy images** should show the full distribution of leaves, instead of just a few leaves.
- **Leaf images** should contain only 1 leaf in the photo with a cleared and smooth background.
- **Profile images** should have as few other trees and vines in the background as possible

To summarize, in all cases, photographers should aim to avoid the two types of errors mentioned previously by minimizing noise in the image while including key characteristics of the part they are capturing.

Conclusion

In this paper, we introduce **BiomeAzuero2022**, a tree image data set consisting of 9071 images taken in Azuero, Panama and the associated species labels grouped by parts of trees, in hopes of encouraging the machine learning community to take on the challenge of identifying the tropical dry forest inventory from ground images. We also present baselines for species classification and a feature-importance-based methodology for improving the data gathering processes for tree images.

Acknowledgments

This work was done in partnership with the Earthshot Labs Biome team (https://www.earthshot.eco/biome). We also want to thank Professor Milind Tambe for the great lectures on AI for Social Good.

References

Balayn, A.; Soilis, P.; Lofi, C.; Yang, J.; and Bozzon, A. 2021. What do you mean? Interpreting image classification with crowdsourced concept extraction and analysis. In *Proceedings of the Web Conference 2021*, 1937–1948.

Barré, P.; Stöver, B. C.; Müller, K. F.; and Steinhage, V. 2017. LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics*, 40: 50–56.

Bastin, Y. G. C. M. D. R. M. R. D. Z. C. M. . C. T. W., Finegold. 2019. The global tree restoration potential. In *Science (American Association for the Advancement of Science), 365(6448)*, 76–79.

Brandtberg, T. 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5): 325–340.

Chave J, B. A. C. E. C. M. D. W. D. A. E. T. F. P. G. R. H. M. M.-Y. A. M. W. M.-L. H. M. M. N. B. N. A. N. E. O.-M. E. P. R. P. P. R. C. S. J. V. G., Réjou-Méchain M. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. In *Glob Chang Bio))*, 76–79.

Clark, M. L.; and Roberts, D. A. 2012. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensing*, 4(6): 1820–1855.

Ghosh, A.; Fassnacht, F. E.; Joshi, P. K.; and Koch, B. 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, 26: 49–63.

Jones, T. G.; Coops, N. C.; and Sharma, T. 2010. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sensing of Environment*, 114(12): 2841–2852.

Kumar, N.; Belhumeur, P. N.; Biswas, A.; Jacobs, D. W.; Kress, W. J.; Lopez, I. C.; and Soares, J. V. 2012. Leafsnap: A computer vision system for automatic plant species identification. In *European conference on computer vision*, 502–516. Springer.

Lee, S. H.; Chan, C. S.; Wilkin, P.; and Remagnino, P. 2015. Deep-plant: Plant identification with convolutional neural networks. In *2015 IEEE international conference on image processing (ICIP)*, 452–456. IEEE.

O'Shea, R., Nash. 2015. An Introduction to Convolutional Neural Networks. *arXiv preprint arxiv.1511.08458*.

Ranson, K.; Sun, G.; Kharuk, V.; and Kovacs, K. 2001. Characterization of forests in Western Sayani Mountains, Siberia from SIR-C SAR data. *Remote Sensing of Environment*, 75(2): 188–200.

Sambasivan, N.; Kapania, S.; Highfill, H.; Akrong, D.; Paritosh, P.; and Aroyo, L. M. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15.

Schmitt, M.; Shahzad, M.; and Zhu, X. X. 2015. Reconstruction of individual trees from multi-aspect TomoSAR data. *Remote Sensing of Environment*, 165: 175–185.

Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Van Horn, G.; Mac Aodha, O.; Song, Y.; Cui, Y.; Sun, C.; Shepard, A.; Adam, H.; Perona, P.; and Belongie, S. 2018. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8769–8778.

Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; and Torralba, A. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2921–2929.

Zhou, T.; Popescu, S. C.; Lawing, A. M.; Eriksson, M.; Strimbu, B. M.; and Bürkner, P. C. 2017. Bayesian and classical machine learning methods: a comparison for tree species classification with LiDAR waveform signatures. *Remote Sensing*, 10(1): 39.