PalmProbNet: A Probabilistic Approach to Understanding Palm Distributions in Ecuadorian Tropical Forest via Transfer Learning

Kangning Cui*[1], Zishan Shao[2], Gregory Larsen[2], Victor Pauca[2], Sarra Alqahtani[2], David Segurado[2], João Pinheiro[2], Manqi Wang[2], David Lutz[3], Robert Plemmons[2], and Miles Silman[2]

[1] City University of Hong Kong, Kowloon, Hong Kong
[2] Wake Forest University, Winston-Salem, NC, USA
[3] Dartmouth College, Hanover, NH, USA

Abstract

Palms play an outsized role in tropical forests and are important resources for humans and wildlife. A central question in tropical ecosystems is understanding palm distribution and abundance. However, accurately identifying and localizing palms in geospatial imagery presents significant challenges due to dense vegetation, overlapping canopies, and variable lighting conditions in mixed-forest landscapes. Addressing this, we introduce PalmProbNet, a probabilistic approach utilizing transfer learning to analyze high-resolution UAV-derived orthomosaic imagery, enabling the detection of palm trees within the dense canopy of the Ecuadorian Rainforest. This approach represents a substantial advancement in automated palm detection, effectively pinpointing palm presence and locality in mixed tropical rainforests. Our process begins by generating an orthomosaic image from UAV images, from which we extract and label palm and non-palm image patches in two distinct sizes. These patches are then used to train models with an identical architecture, consisting of an unaltered pre-trained ResNet-18 and a Multilayer Perceptron (MLP) with specifically trained parameters. Subsequently, PalmProbNet employs a sliding window technique on the landscape orthomosaic, using both small and large window sizes to generate a probability heatmap. This heatmap effectively visualizes the distribution of palms, showcasing the scalability and adaptability of our approach in various forest densities. Despite the challenging terrain, our method demonstrated remarkable performance, achieving an accuracy of 97.32% and a Cohen's κ of 94.59% in testing.

Index Terms: Transfer learning, Palm detection, Density map, Remote sensing.

1 Introduction

Detecting palm trees in tropical forests is important in both biological and computational domains. Biologically, palms serve as important ecological indicators, providing insights into biodiversity, soil quality, and overall health of the forest ecosystem, support human livelihoods in indigenous and rural communities, and are keystone resources for tropical wildlife [1], [2]. Computationally, particularly within image processing and machine learning, palm detection is challenging due to issues such as noise and artifacts in image orthomosaics, variable illumination within forests, lack of labeled data, and data imbalances [1]–[3]. Addressing these challenges is essential not only for ecological studies that rely on identification and localization of forest resources but also for advancing computer vision techniques for remote sensing in general.

In response to the outlined challenges, this study introduces PalmProbNet, a deep learning approach tailored for palm tree detection in Ecuador's tropical forests, capitalizing on UAV-derived orthomosaic imagery.

---

*Corresponding Author: kangnicui2-c@my.cityu.edu.hk

We employ a strategy of training two models, both with a consistent architecture, on image patches of varying sizes, resulting in probability maps that depict palm distributions. Experimental results highlight the robustness and efficacy of PalmProbNet in classifying palm trees. Our contributions are threefold: a data labeling process that categorizes two sets of image patches containing palm and non-palm characteristics in different sizes, feature extraction via transfer learning integrated with deep learning-based classification, and the application to the full UAV-derived orthomosaic imagery. The results emphasize the promise of integrating UAV technology with deep learning for efficient palm tree detection within dense forest canopies.

The subsequent sections of the paper are structured as follows: Section 2 provides a comprehensive review of related work in palm detection and deep learning. Section 3 details our dataset, including data collection, preprocessing, and labeling. Section 4 introduces PalmProbNet, our approach covering feature extraction, classification, and application to the landscape orthomosaic image. Section 5 showcases the experimental setup and its findings, succeeded by a comprehensive discussion on these outcomes. Finally, Section 6 concludes the paper and outlines directions for future research.

2 Related Work

2.1 Object Detection

Object detection, a pivotal task in computer vision, identifies and classifies objects within images. Broadly, the methodologies can be categorized into traditional and deep learning-based approaches [4]–[8]. Traditional approaches follow a structured pipeline: selecting informative regions, extracting hand-crafted features, and classifying these features. Specifically, regions of interest (RoIs) within the image, which contains the targeted objects, are identified using methods like sliding window or selective search [7], [9]. The selected RoIs are then subjected to feature extraction approaches to describe specific features within the region, capturing object appearance, shape, and texture information [6], [7], [10]. The features are then fed into classifiers to categorize the RoIs and finalize the object detection [5], [7]. Traditional classifiers tend to be computationally efficient due to their fewer tunable parameters, in contrast to their deep learning counterparts [6], [7].

Deep learning-based techniques, on the other hand, offer an end-to-end training and prediction process by integrating feature extraction and classification into a unified framework [1], [4], [11], [12]. Notable deep learning algorithms include You Only Look Once (YOLO) and Region-Based Convolutional Neural Networks (R-CNNs) [6]. YOLO adopts a single-shot architecture that enables it to sufficiently perform object detection in real-time, while R-CNNs leverage a two-step process that first identifies regions of interest and then classifies those regions, providing a more accurate but computationally intensive approach [4], [6], [12], [13]. Despite their capabilities, both traditional and deep learning methods encounter challenges in real-world scenarios, especially in detecting palm trees. Factors like dense canopies and diverse species in forests introduce complexities such as overlapping objects and intricate backgrounds [14], [15].

2.2 Palm Detection

Modern remote sensing technologies have significantly advanced palm tree detection by increasing the abundance, availability and resolutions of both satellite and aerial imagery. Remote sensing often leverages specialized sensors to capture distinct spectral and structural characteristics [1], [14]. For example, thermal sensors monitor surface temperature and can aid in irrigation design for palm plantations, multispectral sensors produce NDVI images which can help distinguish the presence, types, and organismal qualities of vegetation, hyperspectral sensors yield detailed profiles of spectral reflectance across large segments of the electromagnetic spectrum, and LiDAR sensors provide topographic and structural data [3], [9]. These sensors are deployed on spaceborne platforms (satellites), airborne platforms (drones or occupied aircrafts) or ground-based platforms, with trade-offs of cost, scalability, and observation scale depending on the platform and its deployment [1], [9]. In this context, unoccupied aerial vehicles (UAVs) provide a distinctly scalable, cost-effective opportunity for canopy monitoring with ultra-high image resolutions, achievable due to their typically high-resolution sensors and ability to operate safely at very low altitudes [3], [13]. UAVs enable efficient monitoring of vast extents of terrain, catering especially to needs of small-scale farmers monitoring diverse palm species [7], [9]. The integration of automated palm detection techniques, whether conventional

(a) Map of the Study Area

(b) The Original Orthomosaic

(c) The Cleaned Orthomosaic

Figure 1: Visualizations of the Orthomosaic Image Before and After Cleaning

or deep learning-based, further accentuates the potential of UAV applications for mapping and monitoring vegetation land use and land cover.

UAVs and machine learning approaches rapidly enhance vegetation analysis for agricultural and forestry management, with methods like sliding window, histogram of oriented gradients, and SVM effectively detecting palm trees in UAV imagery from Malaysia and Saudi Arabia [3], [7], [10]. However, while effective, machine learning classifiers depend on feature quality. Introducing a new dataset necessitates fresh feature extraction, which complicates the training phase [16]. In contrast, deep learning approaches, particularly CNNs, inherently and adaptively extract relevant features during training, remarkably simplifying the process. CNNs have been applied to canopy monitoring with UAVs, effectively addressing complex object detection challenges [12], [14], [15], [17]. R-CNNs can effectively identify Mauritia flexuosa palms in Amazonian forests and oil palm trees in Malaysian plantations, demonstrating CNNs' adaptability in both forest and agricultural settings [12], [13]. Moreover, other architectures like YOLO have been successfully utilized to detect loose fruits of oil palms from UAV images, contributing a major advance to precision agriculture [17].

Identifying palm trees from high-resolution aerial or satellite imagery presents several challenges. Conventional classifiers may not capture hierarchical feature representations adequately, whereas deep-learning methods demand extensive training datasets [6], [9]. Integrating CNN-based methods with conventional classifiers can harness the strengths of both, offering a more robust solution for palm tree detection [5], [6]. This has been demonstrated, using CNNs to detect oil palm trees in satellite images using sliding window and post-processing techniques [18].

(a) Isolated Palm   (b) Hidden Palm   (c) Overlapped Palms

Figure 2: Palms as Captured in the Orthomosaic

3 Dataset

3.1 Raw and Orthomosaic Data

The raw data used in this study pertains to a natural reserve in the Northwest region of Ecuador, known as the Ecuadorian Choco forest (00°23′28″ N, 79°41′05″ W), see Figure 1. The reserve is a high diversity humid tropical forest at 500m elevation, receiving $\sim$ 3000 mm precipitation per year. Precipitation is seasonal, with a 5-month minimum from July-November accompanied by persistent fog (see [19] for detailed description). The forests contain 16 species of palms that can have exposed crown [19], [20].

The data was collected in collaboration with the Fundación para la Conservación de los Andes Tropicales (FCAT). The UAV imagery was collected in two campaigns. The first in June 2022 covered 95 ha using a DJI Phantom 4 RTK drone with a 1" CMOS sensor and controlled using UgCS mission planning software (CITS). Missions were flown at 90 m/s with 70% sidelap and 80% endlap. A total of 387 photos were taken and an orthomosaic and 3D point cloud were generated using Structure from Motion (SfM) in Agisoft Metashape 2.0 (Agisoft 2022). Flight plans, raw imagery, and orthomosaics are available upon request.

An orthomosaic is a large photogrammetrically orthorectified image product created from an image collection of a particular geographical region. Orthomosaics generated from SfM can contain considerable distortion due to the orthorectification process, especially near the edges of the reconstruction. Typically, distorted regions are removed to improve the visual consistency of the imagery, as shown in Figure 1.

3.2 Dual-Scale Manual Labels

A ground survey provided the exact location of 2,929 palms in the study region. Upon observation, we found that many of these palms are not visible in the orthomosaic image, due to severe occlusion or overlap with surrounding trees, see e.g. Figure 2. The resulting number of sufficiently visible palms in the orthomosaic was deemed too small for supervised classification. Thus, the following strategy was adopted to generate labeled data.

Fine-scale labeled data. We used the location of visible palms in the orthomosaic to extract 6000 small image patches of 40 by 40 pixels, capturing as much palm leaf feature diversity as possible. We endeavored to ensure that the extracted patches contain at least 90% of palm data within them. Similarly, we extracted 6000 40-by-40-pixel patches of non-palm features, such as other types of tree crowns, ground soil, roads, tree trunks, and other land cover types.

Sample images for the resulting palm and non-palm classes are shown in Figure 3. The images on the left are patches containing palm features, capturing the distinctive characteristics of a palm tree. The images on the right, on the other hand, showcase patches without palm features, highlighting variations that may help the model to differentiate between palms and non-palms.

Coarse-scale labeled data. The second part of our labeling process involved larger 100x100 pixel patches labeled as having high/low probability of containing palms. This scheme not only captures intricate leaf details but also broader tree crown contexts, crucial for distinguishing partially occluded palms from other trees.

(a) Palms   (b) Non-Palms

Figure 3: Comparison of Small Patches

(a) Palms   **(b) Non-Palms**

Figure 4: Comparison of Large Patchs

A semi-manual approach was employed for this purpose. We used a portion of our fine-scale labeled data to train a network and identify small palm patches throughout the entire orthomosaic image. The locations of these patches reveals the presence of a palm tree, even in areas where the palm is mostly occluded. We then explore 100x100 pixel areas around these patches, labeling them as either palms or non-palms. This approach resulted in about 3367 palm and 4091 non-palm samples, as shown in Figure 4, with the left four images showing palm patches and the right four showing non-palm patches, including roads and other tree crowns. Notably, some palms emerged within the low probability patches and some non-palms within the high probability ones, underscoring the effectiveness of the large patch datasets in tree crown identification.

4 Methodology

In this section, we introduce the proposed PalmProbNet that focuses on generating a heatmap to show the probability distribution of palm presence, aiding in biological applications. PalmProbNet encompasses three pivotal stages: feature extraction, classification, and application to the landscape orthomosaic image. The subsequent sections provide a detailed introduction of these stages and Figure 5 shows the workflow of the proposed PalmProbNet.

4.1 Feature Extraction

Transfer learning is a potent technique that leverages pre-trained models, reducing the need for extensive data collection and computational resources [21]. Customization of these models is often achieved by adding specialized layers tailored to the specific task. Multi-Layer Perceptrons (MLPs), feedforward neural networks with multiple layers, are commonly used for this purpose due to their capability to learn intricate patterns [22]. Residual Networks (ResNet) address challenges in training deep neural networks by introducing short-cut or "skip" connections, making them suitable for complex tasks like palm tree detection [23]. In our study, given the limited labeled data, we utilized the pre-trained ResNet-18 model, a variant with 18 layers. Originally trained on the ImageNet dataset [24], ResNet-18 offers a robust feature hierarchy. We fine-tuned its final layers to our dataset, ensuring a robust foundation for the PalmProbNet. Additionally, we incorporated data augmentation techniques to enhance the diversity of our training data, addressing issues such as variable illumination and occlusions [13].

Figure 5: Workflow of The Proposed PalmProbNet

4.2 Classification

To classify the palm images based on the extracted features from ResNet layers, PalmProbNet employs an MLP that can capture complex patterns in data by introducing non-linearity [22]. Our MLP architecture, shown in Figure 5, incorporates a batch normalization layer, enhancing the model's ability to learn from ResNet's feature vectors and subsequently map them to the appropriate class labels. The overall workflow is illustrated in Figure 5.

4.3 Application to the Orthomosaic Image

Applying the model trained by small/large patches to large landscape orthomosaic images in PalmProbNet involves several steps. Depending on the training set, the model uses either 40×40 or 100×100 pixel patches. The process begins by loading the image, preprocessing, and capturing patches using a sliding window with a stride of 10. Patches with 25% or more missing pixels are immediately categorized as 'non-palm'. The ResNet layers then extracts features from each patch, which are predicted by the trained MLP layers. Probabilities from these predictions are aggregated into an accumulator array, with overlapping regions being averaged to ensure accurate probability estimates. Following the predictions, the averaged probabilities are resized to generate a heatmap, which is then overlaid onto the orthomosaic image to emphasize the areas containing palms. The final results, including the highlighted image indicating the palm regions and the heatmap of the landscape orthomosaic, are saved for further analysis, capturing all computed probabilities.

5 Experiment

5.1 Data Division and Augmentation

To ensure a fair evaluation of PalmProbNet, we divided our dataset (for both small and large patches datasets) into two parts: 80% for training and 20% for testing. The training set was further subjected to 5-fold cross-validation, providing a more reliable assessment of the model's performance.

Data augmentation enhances model robustness and performance, especially in image-based tasks with limited labeled datasets. For the training set, we implemented several preprocessing and augmentation techniques. The images were resized to 224×224 pixels for compatibility with the pre-trained ResNet layers. Augmentations included random horizontal and vertical flips, which mirror the image in different orientations, and color jittering to simulate various lighting conditions. Lastly, normalization was applied to scale pixel values to a standardized range, facilitating faster convergence during training. For the test set, we avoided augmentations that might alter the intrinsic characteristics of the images, preserving consistency with the model's expectations.

5.2 Model Training and Evaluation

In the training of PalmProbNet, we fine-tuned the MLP layer parameters for task-specific adaptations, while keeping most ResNet parameters frozen to guarantee robust feature extraction. The final layer of ResNet was specifically fine-tuned to better align the network's output with the unique characteristics of our dataset. During each fold of the cross-validation, the training data was augmented using the described techniques, while the validation data remained unchanged. We submitted SLURM jobs on Wake Forest University's cluster to accelerate the computation. The training employed the Negative Log-Likelihood loss function to evaluate prediction accuracy.

The core training phase consisted of 500 epochs for small patches and 200 epoches for large patches, employing a batch size of 64. We monitored the validation loss to select the best model, characterized by the lowest loss across epochs and folds. This model was then retrained on the entire training dataset with the same augmentation techniques and evaluated on a non-augmented test set. This systematic training approach, integrating cross-validation and validation loss monitoring, provided a robust evaluation framework, ensuring optimal model performance and reliability.

We utilized several metrics to evaluate the performance of PalmProbNet. Accuracy, given by Acc $= \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}}$, measures the proportion of correct classifications. Average Accuracy (AA) is the mean of individual class accuracies, offering a balanced view, especially for imbalanced datasets. Cohen's κ coefficient gauges prediction reliability by assessing agreement between predicted and actual classes beyond chance. For binary classification, Precision, defined as Precision $= \frac{\text{True Positives}}{\text{True Positives + False Positives}}$, evaluates the model's exactness, while Recall, given by Recall $= \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$, measures its completeness. The ROC AUC represents the model's discriminative power between classes.

5.3 Numerical Results and Discussions

In evaluating the numerical results, the trained models tailored for both large and small patch sizes were tested on the testing patches and deployed on the landscape orthomosaic image. A stride of 10 was employed during this process, which allowed for a comprehensive and efficient scan of the entire image.

Our comparative analysis of patch classification, encompassing both small and large patches, involved three configurations of hidden layer nodes (`nnodes`): 64, 128, and 256. The performances, shown in Table 1, reveal a uniform level of effectiveness across the configurations for both patch sizes. For small patches, the 256-node configuration showed a slight edge, although the overall stable metrics across all configurations indicate that additional fine-tuning may be necessary for significant performance improvements. In the realm of large patches, the 256-node setup similarly displayed a marginal but noticeable enhancement in performance measures, such as accuracy, average accuracy, and Cohen's κ coefficient. These parallel findings across different patch sizes underscore the robustness of the classification approach and suggest that the 256-node configuration provides a subtle yet consistent advantage.

Figure 6(a-c) presents the probability heatmaps generated from the landscape orthomosaic image using the PalmProbNet trained on small patches. These heatmaps visually encode the model's assessed probability of palm presence, with brighter areas corresponding to higher probabilities of palm features. The intensity within each 10×10 segment of the heatmap is indicative of the model's confidence in identifying palm elements within that specific area.

Considering three different hidden layer configurations (`nnodes` $= 64, 128, 256$), the results, while broadly similar, exhibit subtle differences upon closer inspection. Each configuration efficiently detects a majority of the palm trees, though there are occasions where non-palm tree crowns are erroneously classified as

Patch Size	nnodes	Acc	ROC AUC	AA	κ	Precision	Recall
	64	0.9101	0.9642	0.9105	0.8203	0.9126	0.9101
Small	128	0.9041	0.9653	0.9047	0.8084	0.9099	0.9041
	256	**0.9110**	**0.9674**	**0.9114**	**0.8221**	**0.9132**	**0.9110**
	64	0.9698	**0.9963**	0.9694	0.9391	0.9698	0.9698
Large	128	0.9712	0.9949	0.9705	0.9418	0.9712	0.9712
	256	**0.9732**	0.9959	**0.9735**	**0.9459**	**0.9733**	**0.9733**

Table 1: Classification Results for Different Configurations

palms. Such instances emphasize the necessity for more refined model adjustments to reduce false positives. A visual evaluation suggests that configurations with `nnodes` = 64 and `nnodes` = 256 marginally surpass the performance of `nnodes` = 128. The consistent detection patterns across diverse node configurations underline the reliability of the feature extraction and classification phases, simultaneously indicating areas where further tuning could elevate model accuracy.

Figure 6(d-f) displays the probability heatmaps derived from the landscape orthomosaic image, utilizing PalmProbNet trained on large patches. Each map is the result of a different hidden layer node configuration (`nnodes` = 64, 128, 256), with the intensity of the highlights indicating the likelihood of palm presence. The three configurations yield very similar results, making it challenging to determine a clear superior performer. However, a common issue across all configurations is the misclassification of certain areas. Notably, patches adjacent to blank areas—those that have been either deleted due to poor image quality or are missing— are prone to be inaccurately labeled as containing palms. Addressing these misclassifications may require incorporating training samples that include these edge cases or applying post-processing steps to exclude them, thereby refining the model's ability to discern true palm features from artifacts introduced by image processing.

6 Conclusion and Future Work

Identification of ecologically and economically important species within tropical forests canopies expands the frontiers of research and conservation. The paper details a thorough investigation into identifying palm trees in Neotropical rain forest using PalmProbNet on UAV-captured images. In particular, it tackles challenges commonly encountered in detection tropical palm canopies like noise and uneven lighting, providing precise palm localization and landscape spatial distributions. PalmProbNet employed two models with the same architecture but trained on different-sized image patches. The models, particularly with a 256-node configuration, yielding robust results: for small patches, an accuracy of 0.9110, ROC AUC of 0.9674, and a Cohen's κ coefficient of 0.8221; for large patches, an accuracy of 0.9732, ROC AUC of 0.9959, and a Cohen's κ coefficient of 0.9459. Probability heatmaps were introduced as a new way to illustrate palm distributions, useful for both environmental and economic research and computational tasks. The integration of UAVs and deep learning has been validated for palm detection in dense forests through high classification accuracy across various model configurations.

Future enhancements to PalmProbNet will focus on refining the heatmap by including training samples with edge cases and employing post-processing to filter out patches near blank areas. We plan to include a fairer comparison of PalmProbNet with other classifiers, and explore segmentation networks using the produced heatmap. We also intend to involve superpixel-based methods for the fast and precise localization of individual palms in orthomosaic images, enhancing our understanding of their distribution [25]. Additionally, we aim to augment our approach by integrating UAV imagery with satellite data from platforms like Planet or WorldView, capitalizing on the complementary strengths of local detail and broad-scale context to improve detection accuracy [26].

References

[1] K. Kipli, S. Osman, A. Joseph, *et al.*, "Deep learning applications for oil palm tree detection and counting," *Smart Agricultural Technology*, vol. 5, p. 100 241, 2023.

[2] K. L. Chong, K. D. Kanniah, C. Pohl, and K. P. Tan, "A review of remote sensing applications for oil palm studies," *Geo-spatial Information Science*, vol. 20, no. 2, pp. 184–200, 2017.

[3] N. Khan, M. A. Kamaruddin, U. U. Sheikh, Y. Yusup, and M. P. Bakht, "Oil palm and machine learning: Reviewing one decade of ideas, innovations, applications, and gaps," *Agriculture*, vol. 11, no. 9, p. 832, 2021.

[4] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[5] M. R. S. Devi, V. V. Kumar, and P. Sivakumar, "A review of image classification and object detection on machine learning and deep learning techniques," in *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India: IEEE, 2021, pp. 1–8.

[6] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[7] Y. Wang, X. Zhu, and B. Wu, "Automatic detection of individual oil palm trees from uav images using hog features and an svm classifier," *International Journal of Remote Sensing*, vol. 40, no. 19, pp. 7356–7370, 2019.

[8] K. Cui, S. Camalan, R. Li, *et al.*, "Semi-supervised change detection of small water bodies using rgb and multispectral images in peruvian rainforests," in *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Rome, Italy: IEEE, 2022, pp. 1–5.

[9] Z. Khuzaimah, N. M. Nawi, S. N. Adam, B. Kalantar, O. J. Emeka, and N. Ueda, "Application and potential of drone technology in oil palm plantation: Potential and limitations," *Journal of Sensors*, vol. 2022, pp. 1–18, 2022.

[10] S. Malek, Y. Bazi, N. Alajlan, H. AlHichri, and F. Melgani, "Efficient framework for palm tree detection in uav images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4692–4703, 2014.

[11] S. Camalan, K. Cui, V. Pauca, *et al.*, "Change detection of amazonian alluvial gold mining using deep learning and sentinel-2 imagery," *Remote Sensing*, vol. 14, no. 7, p. 1746, 2022.

[12] W. Marin, I. Mondragon, and J. Colorado, "Aerial identification of amazonian palms in high-density forest using deep learning," *Forests*, vol. 13, no. 5, p. 655, 2022.

[13] X. Liu, K. H. Ghazali, F. Han, and I. I. Mohamed, "Automatic detection of oil palm tree from uav images based on the deep learning method," *Applied Artificial Intelligence*, vol. 35, no. 1, pp. 13–24, 2021.

[14] L. P. Osco, J. M. Junior, A. P. M. Ramos, *et al.*, "A review on deep learning in uav remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, p. 102 456, 2021.

[15] Y. Diez, S. Kentsch, M. Fukuda, M. L. L. Caceres, K. Moritake, and M. Cabezas, "Deep learning in forestry using uav-acquired rgb data: A practical review," *Remote Sensing*, vol. 13, no. 14, p. 2837, 2021.

[16] F. E. Fassnacht, H. Latifi, K. Stereńczak, *et al.*, "Review of studies on tree species classification from remotely sensed data," *Remote Sensing of Environment*, vol. 186, pp. 64–87, 2016.

[17] M. H. Junos, A. S. Mohd Khairuddin, S. Thannirmalai, and M. Dahari, "Automatic detection of oil palm fruits from uav images using an improved yolo model," *The Visual Computer*, vol. 38, pp. 2341–2355, 2022.

[18] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote Sensing*, vol. 9, no. 1, p. 22, 2016.

[19] S. Lueder, K. Narasimhan, J. Olivo, *et al.*, "Functional traits, species diversity, and species composition of a neotropical palm community vary in relation to forest age," *Frontiers in Ecology and Evolution*, vol. 10, p. 678 125, 2022.

[20] L. Browne and J. Karubian, "Diversity of palm communities at different spatial scales in a recently fragmented tropical landscape," *Botanical Journal of the Linnean Society*, vol. 182, no. 2, pp. 451–464, 2016.

[21] K. Weiss, T. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[22] D. Rumelhart, G. Hinton, R. Williams, *et al.*, *Learning internal representations by error propagation*, 1985.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA: IEEE, 2016, pp. 770–778.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA: IEEE, 2009, pp. 248–255.

[25] K. Cui, R. Li, S. Polk, *et al.*, "Superpixel-based and spatially-regularized diffusion learning for unsupervised hyperspectral image clustering," *arXiv preprint arXiv:2312.15447*, 2023.

[26] S. Camalan, K. Cui, V. Pauca, *et al.*, "Detecting change due to alluvial gold mining in peruvian rainforest using recursive convolutional neural networks and contrastive learning," in *AGU Fall Meeting Abstracts*, vol. 2022, Chicago, USA: AGU, 2022, B52G–0905.

(a) With nnodes = 64, Small

(b) With nnodes = 128, Small

(c) With nnodes = 256, Small

(d) With nnodes = 64, Large

(e) With nnodes = 128, Large

(f) With nnodes = 256, Large

Figure 6: Probability Heatmaps Produced by PalmProbNet Trained on Small & Large Patches