Efficient Localization and Spatial Distribution Modeling of Canopy Palms Using UAV Imagery

Kangning Cui, Wei Tang, Rongkun Zhu, Manqi Wang, Gregory D. Larsen, Victor P. Pauca, Sarra Alqahtani, Fan Yang, David Segurado, Paul Fine, Jordan Karubian, Raymond H. Chan, Robert J. Plemmons, Jean-Michel Morel, Miles R. Silman

Abstract—**Understanding the spatial distribution of palms in tropical forests is essential for ecological monitoring, conservation strategies, and the sustainable integration of natural forest products into local and global supply chains. However, the analysis of remotely sensed data are challenged by overlapping palm and tree crowns, uneven shading across the canopy surface, and the heterogeneous nature of the forest landscapes, which often affect the performance of palm detection and segmentation algorithms. To overcome these issues, we introduce PalmDSNet, a deep learning framework for efficient detection, segmentation, and counting of canopy palms. To model spatial patterns, we introduce a bimodal reproduction algorithm that simulates palm propagation based on PalmDSNet outputs. We used UAV-captured imagery to create orthomosaics from 21 sites across western Ecuadorian tropical forests, covering a gradient from the everwet Chocó forests near Colombia to the drier forests of southwestern Ecuador. Expert annotations were used to create a comprehensive dataset, including 7,356 bounding boxes on image patches and 7,603 palm centers across five orthomosaics, encompassing a total area of 449 hectares. By integrating detection and spatial modeling, we effectively simulate the spatial distribution of palms in diverse and dense tropical environments, validating its utility for advanced applications in tropical forest monitoring and remote sensing analysis. The dataset can be accessed at 10.5281/zenodo.13822508, and the code to replicate the study is available at github.com/ckn3/palm-ds-sp.**

Manuscript received Month xx, 2024; revised Month xx, 2024; accepted Month xx, 2024. Date of publication Month xx, 2024; date of current version Month xx, 2024. This work was supported in part by US National Science Foundation BEE 2039850, and in part by HKRGC Grants No. CityU11301120, C1013-21GF, CityU11309922, ITF Grant No. MHP/054/22, and LU BGR 105824. *(Kangning Cui and Wei Tang contributed equally to this work.) (Corresponding author: Kangning Cui.)*

Kangning Cui and Wei Tang are with the Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong, and Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong (e-mail: kangnicui2-c@my.cityu.edu.hk; wetang7-c@my.cityu.edu.hk).

Rongkun Zhu is with the Department of Computer Science, Xidian University, Xi'an, Shaanxi, China (e-mail: zhurongkun@stu.xidian.edu.cn).

Manchi Wang, Gregory D. Larsen, David Segurado and Miles R. Silman are with the Department of Biology, Wake Forest University, Winston-Salem, NC, USA (e-mail: wangm21@wfu.edu; gl7176@gmail.com; dsegurado@wfu.edu; silmanmr@wfu.edu).

Victor P. Pauca, Sarra Alqahtani, Fan Yang and Robert J. Plemmons are with the Department of Computer Science, Wake Forest University, Winston-Salem, NC, USA (e-mail: paucavp@wfu.edu; sarra-alqahtani@wfu.edu; yangfan@wfu.edu; plemmons@wfu.edu).

Paul Fine is with the Department of Integrative Biology, University of California, Berkeley, CA, USA (email: paulfine@berkeley.edu).

Jordan Karubian is with the Department of Ecology and Evolutionary Biology, Tulane University, LA, USA, (email: jordankarubian@gmail.com).

Raymond H. Chan is with the Department of Operations and Risk Management and School of Data Science, Lingnan University, Tuen Mun, Hong Kong, and Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong (e-mail: raymond.chan@ln.edu.hk).

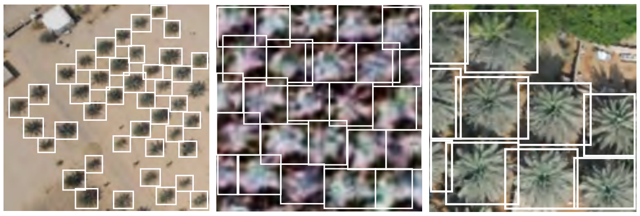
Jean-Michel Morel is with the Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong (e-mail: jeamorel@cityu.edu.hk).

Index Terms—**Environmental Sustainability, Spatial Point Pattern, Object Detection, Instance Segmentation.**

I. INTRODUCTION

PALMS (family *Arecaceae*) include many ecologically and economically important species whose spatial distributions crucially inform tropical forest ecology and conservation research. They are also central to local economies and regional to global efforts to incorporate natural forest products into sustainable livelihoods and local to international forest product supply chains [1], [2]. Tropical forests host a significant portion of global biodiversity and are increasingly threatened by deforestation and degradation [3], [4], and palms, with their distinctive ecological importance, can serve as vital indicators of both forest health and human impact, offering insights into biodiversity, soil quality, and the overall health of forest ecosystems [5]. They play a central role in shelter, food, and fiber, and are an emerging resource in the development of non-timber forest product markets, supporting human communities in indigenous and rural areas. Palms also constitute essential and often keystone resources for tropical wildlife [6]–[9]. Here, we are concerned with identifying, locating, and quantifying palms occurring naturally within tropical forests, with particular attention to their spatial distribution and natural reproduction (see Figure 1b), as opposed to palm plantations (see Figure 1a). Knowledge of the spatial distribution and abundance of palms can inform sustainable use and management, leading to economic benefits to local communities. Thus, these tasks are crucial for effective management, economic development, and conservation, as well as for understanding basic ecological questions of palm distribution and abundance.

Palms can be detected in high-resolution remotely sensed imagery by their distinctive leaves and crowns. Automated techniques have been widely used in remote sensing tasks such as land cover mapping [10]–[12], agricultural assessment [13], [14], and ecological monitoring [15], [16]. While object detection and spatial pattern analysis are well-established in computer vision and statistics, their application to complex environments – such as identifying naturally occurring palms and analyzing their spatial distribution in tropical forests – poses significant challenges [17]–[20]. First, palm species exhibit extreme class imbalances, with only two or three species typically represented in sufficient numbers for reliable detection [21], [22]. Detecting both common and rare species is essential for comprehensive forest ecological analysis.



(a) Cases from existing studies

(b) Cases from our dataset

Figure 1: Comparative Samples of Manual Labels.

Second, UAV optical imagery of tropical forests exhibits considerable variability in lighting, occlusion, and background clutter, which complicates automated analysis [17], [23]. Palms are often obscured by neighboring tree canopies, and the dense vegetation adds background noise. Inconsistent lighting across different forest areas, shadows cast by dense canopies, and varying sun angles throughout the day further complicates feature detection and localization [24]–[26].

Third, high-quality labeled datasets for tropical forests are notably scarce. Their collection and annotation require extensive fieldwork by trained experts, which is particularly demanding in Amazon [27]–[29]. Creating such data typically involves assembling raw imagery into orthomosaics – spatially referenced data products derived from UAV captures. However, this process is prone to challenges, including noise and artifacts due to image stitching (alignment, merging, and rectification) errors, varying lighting conditions, sensor discrepancies, and environmental factors such as wind and cloud cover that introduce movement and shading [30].

Moreover, efficient object detection and segmentation methods remain underutilized for UAV-based remote sensing in large, densely forested regions where substantial computational resources may be limited [31]–[33]. This shortfall highlights the need for robust and scalable algorithms capable of processing such imagery to support field applications.

Finally, large-scale spatial analysis of palms in tropical forests remains limited, constrained by localization challenges and the lack of extensive, high-quality datasets. Effective monitoring and conservation require robust statistical models that accurately represent the spatial distribution of palms across vast tropical forest areas [5]. Understanding the ecological mechanisms driving plant distribution is fundamental, especially given the urgent need to address the rapid degradation of wilderness areas in recent decades [3], [34].

Addressing the challenges of localizing and analyzing the spatial distribution of palms in tropical forests using UAV imagery, this work presents the following main contributions:

1) We develop a dataset through extensive fieldwork across 21 sites in western Ecuador, spanning a rainfall gradient from the Choco's wettest forests to the edge of the tropical dry forest at the limit of the Sechura desert (5800mm to 1400mm precipitation). This gradient drives distinct canopy palm compositions. We annotate 1,500 image patches with 7,356 bounding-boxed palm instances at two sites, and mark the landscape center points of 7,603 palms for counting across five sites.
2) We introduce a flexible framework for evaluating models in efficient palm detection, segmentation, and counting, effective even with limited computational resources. We further enhance the model's interpretability using saliency maps to spotlight critical decision areas. We also examine the balance between label volume and model performance, finding that detection is sensitive to training set size, while counting remains fairly robust.
3) We present a Poisson-Gaussian reproduction algorithm that simulates the spatial distribution of palms by combining a Poisson process with a local Gaussian distribution. The model effectively reproduces observed spatial patterns and provides insights into palm population structure and ecological dynamics across forest types.

The paper is organized as follows: Section II reviews research on object detection, segment anything models (SAMs), palm identification, and spatial point pattern analysis. Section III describes our dataset and methodology, including the study area, data collection, preprocessing, and annotation procedures, and the PalmDSNet framework for palm detection, segmentation, and counting, along with the Poisson-Gaussian model for simulating palm distributions. Section IV presents our experimental design, numerical results, and analysis of palm localization and spatial distribution. Section V summarizes our findings and proposes future research directions.

II. RELATED WORK

A. Object Detection and Segmentation

1) Object Detection: Object detection is a computer vision task that involves both classifying and localizing objects within images using bounding boxes [35]. It underpins more advanced applications in image segmentation and object tracking [36]–[41]. Two main approaches that have advanced the field are Detection Transformer (DETR) [42], [43] and You Only Look Once (YOLO) [44]–[46].

DETR [42] formulates object detection as a set prediction task, which eliminates the need on non-maximum suppression (NMS). It employs a transformer-based architecture, combined with a set-based global loss and bipartite matching. DETR utilizes a fixed set of learned object queries to capture and reason about the relationships between objects and their global context within the image. Real-Time Detection Transformer (RT-DETR) [43] enhances DETR by incorporating a convolutional backbone and an efficient hybrid encoder. This adaptation optimizes the processing of multi-scale features through a combination of intra-scale interactions and cross-scale fusion. RT-DETR achieves real-time performance while preserving high accuracy. It also provides flexibility in adjusting inference speed through modifications to decoder layers, without requiring retraining [43].

YOLO regards detection as a single regression task, directly predicting bounding boxes and class probabilities from images in real time. Its strong balance between accuracy and efficiency makes it well-suited for real-world applications such as autonomous driving, robotic navigation, and pedestrian tracking [47], [48]. YOLOv8 [44] introduces a cross-stage partial network for efficient feature extraction, an enhanced path aggregation network for multi-scale feature fusion, and an optimized detection head for multi-scale object localization and classification [49]. YOLOv9 [45] advances this by incorporating gradient enhanced learning and augmentation network and programmable gradient information to improve training efficiency. YOLOv10 [46] eliminates the need for NMS through consistent dual assignments. It features a lightweight head, spatial-channel decoupled down-sampling, and a compact inverted block to minimize inference latency.

2) Segment Anything Models: Segmentation partitions an image into meaningful regions by labeling each pixel, which provides detailed analysis in fields such as medical diagnostics and ecological monitoring [50]–[53]. However, traditional segmentation methods require extensive labeled datasets, which are often costly and labor-intensive to obtain. To address this challenge, semi-supervised learning methods [54]–[57] and zero-shot approaches, such as SAMs [58]–[60], have emerged as effective alternatives. SAMs delineate all object within an image irrespective of its type and generalize beyond predefined classes, thus broadening their applicability [58]–[60]. These models can autonomously segment entire images or use prompts like points, boxes, or text to direct segmentation.

The foundational SAM [58] leverages the SA-1B dataset, which includes over 1 billion masks from 11 million images, to facilitate prompt-based and zero-shot segmentation. SAM integrates a Vision Transformer (ViT) [61] for image encoding, a prompt encoder for processing input prompts, and a dynamic mask decoder for generating segmentation maps. SAM 2 [59] extends SAM to video segmentation, allowing for object tracking across frames. Trained on both the SA-1B and SA-V datasets (includes 50.9 thousand videos and 642.6 thousand masklets), SAM 2 enhances performance by incorporating a masked autoencoder (MAE) pre-trained Hiera encoder [62], [63]. This upgrade enables the use of multiscale features, significantly improving the model's overall effectiveness.

Efforts to adapt SAM for resource-constrained environments have led to lightweight variants such as Mobile SAM [60] and FastSAM [64]. Mobile SAM employs a compact ViT-based encoder and a knowledge-distilled student model to reduce computational demands while preserving accuracy. It is trained by minimizing a distillation loss [65]: $L_d = \alpha \cdot T^2 \cdot \mathrm{MSE}\left(\mathbf{p_t}, \mathbf{p_s}\right) + (1-\alpha) \cdot \mathrm{MSE}\left(\mathbf{y}, \mathbf{p_s}\right)$, where MSE denotes the mean squared error, $\mathbf{p_t}$ (at time T) and $\mathbf{p_s}$ are teacher and student outputs, $\mathbf{y}$ is the ground truth, α balances the losses between soft targets and ground truth, and T adjusts the influence of the soft targets. This compression enables efficient deployment in mobile and low-power scenarios.

FastSAM [64] replaces the ViT backbone with a lightweight CNN to achieve real-time segmentation. It adopts a two-stage pipeline: First, all-instance segmentation using YOLOv8-seg [44] to generate object masks; second, prompt-guided refinement using point prompts that match query points to instance masks, box prompts that apply IoU matching with detection boxes, and text prompts that leverage CLIP [66] embeddings. Trained on only 2% of SA-1B, FastSAM achieves approximately 50× faster inference while improving zero-shot performance on the COCO dataset. Its real-time capabilities make it well-suited for UAV-based ecological monitoring.

3) Unified Detection and Segmentation Framework: Recent advances have led to the development of unified frameworks that jointly perform object detection and segmentation for diverse tasks. Traditional instance segmentation models, such as those based on YOLO and R-CNNs [35], integrate mask branches to predict pixel-level outputs. However, these methods rely heavily on extensive mask annotations for effective training. FastSAM [64] addresses this limitation by prompt-adapted instance segmentation, which enables flexible and efficient segmentation without relying solely on exhaustive annotations. Additionally, SAM-RSIS [67] adapts SAM for remote sensing by fine-tuning its ViT backbone and mask decoder, and uses automatic box prompting to eliminate the need for manual input. These approaches reflect a broader trend of using detection outputs (e.g., bounding boxes) as segmentation prompts. When paired with robust segmentation methods, such prompts enable precise foreground-background separation, yielding high-quality masks even in complex scenes.

Several works have integrated fine-tuned object detectors with zero-shot SAM for remote sensing tasks. For example, YOLO and SAM are combined to perform instance segmentation for building extraction and classification in [68], and for cotton boll segmentation and yield prediction using UAV imagery in [69]. Although these works demonstrate the potential of combining trained detectors with zero-shot segmentation models, they are primarily limited to structured, homogeneous environments such as urban and agricultural areas. Moreover, they do not address challenges inherent to more complex, heterogeneous landscapes, nor do they incorporate geospatial referencing essential for ecological analysis [68], [69]. In contrast, our work extends this paradigm to tropical forest environments, addressing spatial heterogeneity and producing georeferenced outputs for landscape-level analysis.

B. Palm Identification

Recent advancements in palm identification employ segmentation, classification, and object detection techniques. Segmentation divides images into polygonal segments that represent objects or classes; for palm identification, it focuses on isolating palm crowns and leaves from the background. These methods typically identify palm centers and create disc-shaped masks around them as ground truth. For instance, a U-Net model with residual networks is ultilized to segment date palms in UAV images of plantations with regular palm distributions against bare soil in [24]. Similarly, two U-Nets of varying complexities is applied to high-resolution satellite imagery of plantations in [22]. While these studies achieved precision and recall rates of 88% to 94%, they were limited by clear backgrounds and regular palm spacing.

Classification techniques assign labels to images, pixels, or objects, such as identifying palm crowns based on visual characteristics. A common approach is the sliding window technique, where a fixed-size window moves across the image to classify each sub-region, though this approach is computationally intensive. CNNs were employed with sliding windows to detect and count oil palms in QuickBird satellite images of plantations with homogeneous backgrounds in [21]. In [17], a probabilistic approach was developed for detecting palms in dense Ecuadorian rainforests, using varying sliding window sizes to create probability maps for palm presence [17]. These methods consistently achieve over 95% accuracy in patch classification but require complex postprocessing for individual palm detection.

Object detection directly localizes objects with bounding boxes, which demands high-quality annotations. YOLO is particularly favored for its balance of speed and accuracy [47]. Ag-YOLO, a YOLOv3-Tiny variant with focal loss for detecting smaller areca palms in UAV images, was introduced in [31]. Similarly, YOLOv5 was used to detect date palms in the United Arab Emirates, where palms are clearly delineated against the background [32]. These methods generally achieve precision and recall rates up to 92%, though often tested on smaller datasets with simple backgrounds.

This study contrasts previous work by using direct object detection with expert-verified bounding box annotations, specifically tailored to the complexities of dense tropical rainforests. Our method effectively handles the spatial heterogeneity and diverse vegetation structures inherent to these ecosystems, which advances wild palm identification.

C. Spatial Point Pattern Analysis

In forest ecosystems, dynamic ecological processes such as seed dispersal, competition, and mortality often give rise to characteristic spatial patterns among trees [70]. A robust analysis of these spatial distributions is crucial for understanding forest structure and its ecological and evolutionary dynamics [71], [72]. One natural approach to spatial distribution analysis is representing each tree as a point, and then investigating both (1) the characteristics of the spatial pattern – what the pattern looks like, and (2) the underlying point process responsible for its formation – how the pattern arises.

Table I: Ripley's G, F, and J Functions. Here, $\hat{d}_i$ represents the nearest neighbor distance for each observed point i, and $\tilde{d}_j$ is the nearest neighbor distance from a simulated point j to the observed points. $\mathbb{1}(\cdot)$ is an indicator function.

Function	**Formulation**
G	$G(d) = \frac{1}{N_o} \sum_{i=1}^{N_o} \mathbb{1}(\hat{d}_i < d)$, $N_o =$ observed events
F	$F(d) = \frac{1}{N_r} \sum_{j=1}^{N_r} \mathbb{1}(\tilde{d}_j < d)$, $N_r =$ simulated events
J	$J(d) = (1 - G(d))/(1 - F(d))$

To address the first question – characterizing point patterns – Ripley's statistical functions, often termed Ripley's alphabet, are widely employed [73], [74]. Among them, three key functions, G, F, and J, stand out for their simplicity and effectiveness. The function $G(d)$ calculates the proportion of nearest neighbor distances within the sample that are less than or equal to a given distance d, thereby reflecting clustering or dispersion tendencies. The F function, in contrast, mirrors the G function but measures the nearest neighbor distances from randomly generated points to the observed pattern, effectively serving as a measure for spatial randomness. The J function combines the information from both the G and F functions, offering a comprehensive view of both intra-pattern and inter-pattern spatial relationships. These functions provide concise summary statistics that effectively describe point pattern characteristics. Table I presents their mathematical definitions.

Beyond nearest-neighbor approaches, multi-scale functions such as Ripley's K function [73], [75] and its derivative, the L function [76], are frequently used to detect spatial patterns across scales. Although highly informative, these methods are computationally intensive and require extensive spatial data collection across large areas [77], [78]. Given the localized focus of our study on palm distributions, we prioritize single-scale analyses to maintain computational efficiency while capturing essential spatial patterns.

To simulate the process that generates the observed spatial patterns, stochastic point process models are employed. The homogeneous Poisson process, widely used in ecological studies [79], [80], assumes points are independently and uniformly distributed within the observation window, leading to point patterns that exhibit complete spatial randomness with no spatial trends or associations among points. However, this model often oversimplifies real ecological conditions. The heterogeneous Poisson process employs a spatially varying intensity function to better capture environmental variability in real-world senarios [80], [81]. Further, the Thomas cluster process introduces Poisson cluster processes across multiple scales, thereby enhancing the model's capacity to depict the intricate structures observed in natural forests [82], [83].

While classical models provide important foundations, they exhibit key limitations for modeling palm spatial distributions. The homogeneous Poisson process fails to capture spatial structure [79], [80], the heterogeneous Poisson process accounts for environmental heterogeneity but not local clustering [80], [81], and the Thomas cluster process models local aggregation but neglects global dispersion [82], [83]. These shortcomings limit their applicability to patterns where both clustering and uniformity coexist, as seen in tropical palms. In contrast, our method balances clustering and dispersion through a probabilistic mixture, directly fits spatial summary functions, and offers ecologically interpretable parameters for modeling complex palm patterns.

III. MATERIALS AND METHODS

A. Dataset

1) Study Sites and Data Collection: The data for this study, with spatial locations shown in Figure 2, come from tropical forest in western Ecuador Chocó centered on the Fundación para la Conservación de los Andes Tropicales Reserve and adjacent Reserva Ecológica Mache-Chindul park (FCAT; 00°23'28" N, 79°41'05" W), and the Jama-Coaque Ecological Reserve (00°06'57" S, 80°07'29" W). FCAT consists of high

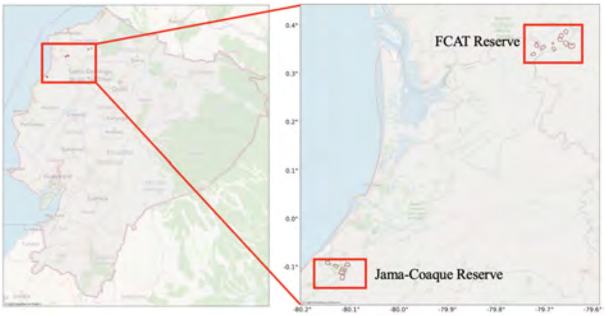
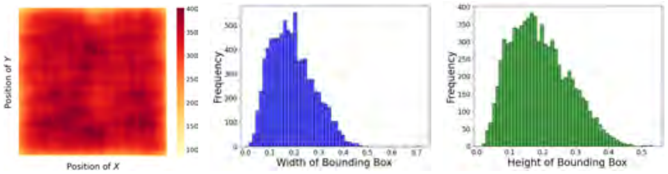


Figure 2: Map of the Study Sites Locations in Ecuador.



(a) Box Distribution (b) Width Distribution (c) Height Distribution

Figure 3: Statistics of Bounding Boxes. Panel (a) shows the bounding box heatmap of the labels, where the color intensity represents the occurrence frequency of palm instances across image patches. Panels (b) and (c) display the width and height distributions of all bounding boxes, respectively. The horizontal axes in (b) and (c) represent bounding box dimensions scaled relative to the image size (unitless).

diversity humid tropical forest at $\sim$500 m elevation, receiving $\sim$3000 mm precipitation yr^{-1} accompanied by persistent fog during drier months. Jama-Coaque spans from the boundary of tropical moist deciduous/tropical moist evergreen forest at the lower elevations ($\sim$1000 mm precipitation yr^{-1}, $\sim$250 m asl) to fog-inundated wet evergreen forests above 580m to 800m. These forests host several regularly occurring canopy-exposed palm species [84]–[86], with *Iriartea deltoidea* and *Socratea exorrhiza* dominant at FCAT, and *Astrocaryum standleyanum* and *Socratea exorrhiza* at Jama-Coaque.

To capture detailed spatial data, UAV surveys were conducted in two phases across 96 plots in 21 areas, covering a total of 1995 hectares with a ground spacing distance (GSD) $<$6 cm. The first phase, in June 2022, involved flying over 95 hectares using a DJI Phantom 4 RTK drone equipped with a 1” CMOS sensor. Mission planning was done using GS RTK, with flights conducted at an altitude of 90 meters above ground level (AGL) with 70% sidelap and 80% frontlap. The 387 resulting images were processed with Agisoft Metashape 2.0 to create orthomosaics – georeferenced image created through photogrammetry by stitching together multiple overlapping images of a geographic area. These images are orthorectified to standardize perspective and spatial scale while embedding spatial metadata, although localized distortions may still occur, particularly along the edges. The second phase, conducted in February 2023, expanded coverage to 1900 hectares, flying at 150 meters AGL and capturing 8458 photos using the same drone and software. Subsequent data processing involved noise reduction, edge trimming of orthomosaics, and the generation of Digital Surface Models and Digital Terrain Models.

(a) Bounding Boxes (b) Crown Centers

Figure 4: Annotation examples. Panel (a) shows manually annotated bounding boxes for palm crowns in the image patch. Panel (b) displays georeferenced crown centers labeled on an orthomosaic. Note that the red segments in (b) are generated by SAM 2 and are not part of the manual annotations.

2) Manual Labels: To create a high-quality training and validation dataset, we manually annotated drone imagery from two sites within the FCAT Reserve. A total of 1,500 patches (800×800 pixels each) were extracted from orthomosaics, representing various image qualities and palm densities. Precise bounding boxes were annotated around palm crowns and isolated leaves. Figure 3 provides a detailed analysis of the bounding box annotations, including their spatial distribution and size variability. For counting tasks, we also manually labeled five orthomosaics using ArcGIS Pro 3.3.1 by placing georeferenced points at the centers of visible palm crowns. Figure 4 illustrates representative examples of our annotations, including both bounding boxes and crown centers.

Unlike studies focused on structured plantation palms [21], [24], [32], our research examines natural forest scenes (see Figure 1). Manual annotation of bounding boxes was particularly hard due to crown overlap, occlusion by other vegetation, and orthomosaic distortions Similarly, labeling the center points of palms across five sites was labor-intensive but essential for conducting landscape-level analysis and validating counting accuracy against expert assessments.

B. Palm Detection and Segmentation Network

This section introduces the Palm Detection and Segmentation Network (PalmDSNet), a framework designed to detect, segment, and count palms in the dense, heterogeneous tropical forests (see Figure 5). The detection target is the individual palm crown, with bounding box annotations used for training and evaluation, while manually labeled crown centers support landscape-level validation. Segmentation is also applied to visualize palm crowns. Unlike prior works that focus on structured environments or omit geospatial analysis, PalmDSNet addresses the spatial heterogeneity of natural forests, incorporates georeferenced outputs, and uses expert-verified labels for accurate detection in complex ecological settings.

PalmDSNet is modular, which supports interchangeable detection and segmentation backbones tailored to specific contexts. Detection results are produced by trained backbones and used to guide zero-shot segmentation. Furthermore, saliency maps are incorporated to highlight focal areas influencing model decisions that enhance interpretability.

1) Palm Detection: The detection networks aim to locate palm crowns in orthomosaic images, which are partitioned into overlapping patches for detailed analysis. These networks

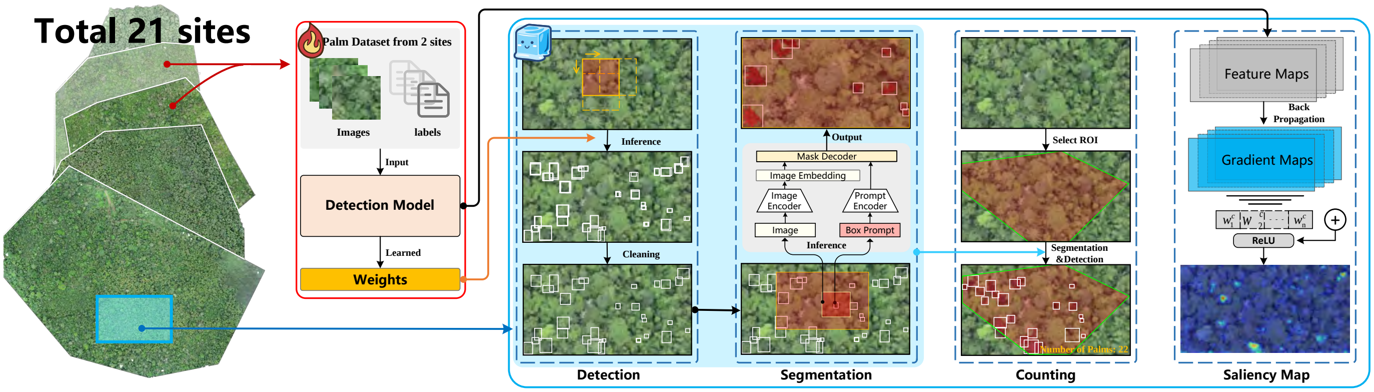


Figure 5: Workflow of PalmDSNet. The model is trained on labeled patches for accurate palm detection, with the refined weights then used to infer bounding boxes on new images. These boxes guide the segmentation and counting processes across selected regions or landscape scales. Saliency maps are also produced to enhance the interpretability of palm localization.

are designed to be flexible, allowing for the substitution of alternative models based on specific analytical needs [42]–[46]. Training involves various image augmentations, such as adjustments in hue, saturation, and brightness, as well as rotations, scaling, translations, and flips, to enhance generalization across diverse visual conditions. During inference, detections from adjacent patches are merged using NMS to avoid duplication. The output is a set of georeferenced bounding boxes representing individual detected palms.

2) Palm Segmentation: Segmentation is performed during inference due to the absence of segmentation masks in the training dataset. Detected palms serve as centers, with their surrounding areas included for contextual information. The bounding box from the detection phase acts as a prompt for a chosen SAM, allowing it to focus on and segment the palm crown. This targeted approach efficiently delineates palms without requiring full-scene processing.

3) Palm Counting: Palm counting is conducted either at the landscape scale or within a region of interest (ROI). For ROIs, a user-drawn polygon is enclosed within a rectangular area to ensure full coverage during detection and segmentation. The detection module identifies individual palms, and logs their coordinates followed by NMS. The segmentation branch then generates masks for these palms based on the bounding box outputs. These masks, along with bounding boxes, are visualized to enable thorough analysis and verification within the ROI. Center point annotations are used to validate model predictions across landscape-scale orthomosaics in this task.

4) Saliency Map: To enhance interpretability, we employ Grad-CAM [87], [88], which creates coarse localization maps by using gradients flowing into the final convolutional layer. These maps highlight the the image regions that most influence the model's predictions. By providing visual explanations of the model's focus, Grad-CAM helps verify that the network emphasizes relevant features for accurate palm localization.

C. Poisson-Gaussian Palm Reproduction Algorithm

This section introduces the proposed Poisson-Gaussian model to simulate palm spatial reproduction that optimizes the reproduction parameters to minimize the discrepancy between observed and simulated point patterns (see Algorithm 1). This model provides insights into both global and local aspects of palm reproduction, including long-range and short-range reproductive dynamics. Understanding the spatial distribution of palms can further reveal how different palm species and individuals of the same species interact in tropical forests, specifically whether they coexist or compete for resources.

The Poisson-Gaussian model combines a Poisson point process with a local Gaussian distribution to simulate palm dispersal dynamics [89]. The Poisson process represents long-range, random dispersal like animal-mediated spread, yielding a uniform background distribution; while the Gaussian distribution models the tendency of seeds to fall and germinate near parent trees, creating localized clusters. This bimodal approach better captures ecological processes such as negative density dependence [90], where increased density reduces seed germination and survival rates, promoting species coexistence and realistic spatial distribution patterns. Unlike the Student's t distribution [91] which favors tight clustering, the proposed model balances both local aggregation and global dispersion.

The optimization process begins with initializing p^*, σ^*, and d_{min}. Ripley's functions $\mathbf{g}$ and $\mathbf{f}$ are computed for the observed set X and serve as benchmarks for comparison. The `Simulate` function generates synthetic point patterns based on candidate values of p and σ, starting with a randomly placed point and probabilistically adding new points: with probability p, offspring are drawn from a Gaussian distribution centered around a randomly selected parent, and with probability $1-p$, offspring are uniformly distributed. This process continues until the synthetic set $\hat{X}$ matches the size of X. For each parameter combination, N simulations are performed to compute the total discrepancy. Simulated patterns are compared to the observed data using Ripley's functions. The discrepancy d_i for each trial is measured as the integral of absolute differences between observed and simulated functions. The total discrepancy d is the sum of discrepancies across all trials, with the optimal parameters being those that minimize d.

The model's novelty lies in its iterative, mixed-generation mechanism, which dynamically selects parent points and controls clustering-dispersion balance through parameter p. Unlike

Algorithm 1: Poisson-Gaussian Palm Reproduction

```
Input: $\mathbf{p}$ (list of candidate $p$), $\boldsymbol{\sigma}$ (list of candidate $\sigma$),
       $X$ (set of observed palm coordinates), $N$
       (number of Simulations)
Output: $p^*, \sigma^*$ (optimal $p$ and $\sigma$ that minimizes $d$)
Initialize $p^* = 0$, $\sigma^* = 0$, $d_{min} = \infty$, $\hat{X} = \emptyset$;
Compute Ripley's functions $\mathbf{g} = G(X)$ and $\mathbf{f} = F(X)$;
Procedure Simulate($p, \sigma$):
    Initialize $\hat{X}$ with a random 2D point generated
      uniformly across the spatial extent;
    while $|\hat{X}| < |X|$ do
        Select a random parent palm $\mathbf{x}$ from $\hat{X}$ and
          generate a random number $p_r$ from $[0, 1]$;
        if $p_r < p$ then
            Generate offspring palm using a Gaussian
              distribution $\mathcal{N}(\mathbf{x}, \mathbf{\Sigma})$ around the parent $\mathbf{x}$
              with $\mathbf{\Sigma} = [\sigma^2, 0; 0, \sigma^2]$;
        else
            Generate offspring palm from a 2D uniform
              distribution;
        end
        Append offspring palm to $\hat{X}$;
    end
    return $\hat{X}$;
foreach $p$ in $\mathbf{p}$ do
    foreach $\sigma$ in $\boldsymbol{\sigma}$ do
        Initialize $d = 0$;
        for $i = 1$ to $N$ do
            $\hat{X}_i$ = Simulate($p, \sigma$);
            Compute $\mathbf{g_s}_i = G(\hat{X}_i)$ and $\mathbf{f_s}_i = F(\hat{X}_i)$;
            Integrate the difference for the $i$-th trial:
              $d_i = \int_x |\mathbf{g} - \mathbf{g_s}_i| \, dx + \int_x |\mathbf{f} - \mathbf{f_s}_i| \, dx$;
            $d = d + d_i$;
        end
        if $d < d_{min}$ then
            Update $d_{min} = d, p^* = p$, $\sigma^* = \sigma$;
        end
    end
end
return $p^*, \sigma^*$;
```

the Thomas cluster process [82], [83], which uses a strictly two-stage generation process with fixed parent-child relationships and requires three parameters (parent intensity ρ_p, mean offspring μ, and cluster spread σ), our approach grows clusters incrementally from observed patterns. Moreover, our data-driven optimization of Ripley's G and F functions provides more flexible pattern matching. This model contrasts with the Thomas process [82], which typically relies on maximum likelihood estimation and theoretical assumptions, limiting its flexibility in handling complex, non-ideal clustering patterns. This yields a more practical framework for empirical ecological patterns with fewer parameters (just p and σ) and lower computational complexity.

IV. EXPERIMENTAL RESULTS

This section outlines the performance evaluation of our PalmDSNet framework, systematically addressing several research objectives across diverse ecological sites. We begin by evaluating the performance of various detection models, including RT-DETR [43], YOLOv8 [44], YOLOv9 [45], and YOLOv10 [46] under different training conditions to assess their accuracy and efficiency. Saliency maps [87] are used to visualize the focus regions of these models for palm detection, particularly in densely populated tree crowns. The bounding box annotations, which represent the target palm crowns, are used during both training and evaluation for detection tasks.

Following detection, we assess the segmentation capibilities of SAM [58], SAM 2 [59], and Mobile SAM [60], focusing on inference time and the visual quality of segmentation results. We then explore how training data volume affects the accuracy of these models and evaluate their counting performance, which is crucial for ecological analysis. Finally, we apply the models to orthomosaics for large-scale landscape analysis. The annotated crown centers, which are georeferenced, are used to validate the counting performance.

We also evaluate our bimodal palm reproduction model, which combines a Poisson point process with a local Gaussian distribution to simulate palm spatial distribution, accounting for both random spread and localized clustering. By comparing simulated and PalmDSNet derived patterns, we assess its ability to replicate real-world palm distributions, which provides insights into the ecological dynamics of palm populations, including interspecies interactions and habitat coexistence.

A. Experimental Setting

To evaluate PalmDSNet and the proposed Poisson-Gaussian reproduction model, we conducted experiments to test both robustness and reliability. For PalmDSNet, we adopted two data allocation strategies to quantitatively analyze the difference between limited and extensive training datasets: (1) *10% Training Set:* We used 10% of the available data for training over 100 epochs. This setup was designed to simulate data-scarce conditions, which is a common challenge in deep learning applications, particularly in remote sensing. By limiting the training data, we aimed to evaluate the model's ability to generalize and perform well under week supervision. (2) *90% Training Set:* We used 90% of the data for training over 300 epochs. This setup allowed us to maximize data utilization and assess the model's performance with abundant labeled data. The extended training epochs ensured that the model could fully utilize the larger dataset to optimize its parameters. Both trained models were then rigorously validated to assess their detection and counting performance. Images were cropped to 800×800 pixels with a stride of 400 to balance spatial coverage and computational efficiency. The model was configured to detect a single class to simplify the architecture and reduce computational requirements. While four RTX 3090 GPUs were utilized for training, only one was used during validation and testing to simulate resource-limited operational settings.

For the Poisson-Gaussian palm reproduction model, simulations were conducted using an Intel Xeon Silver 4210 CPU

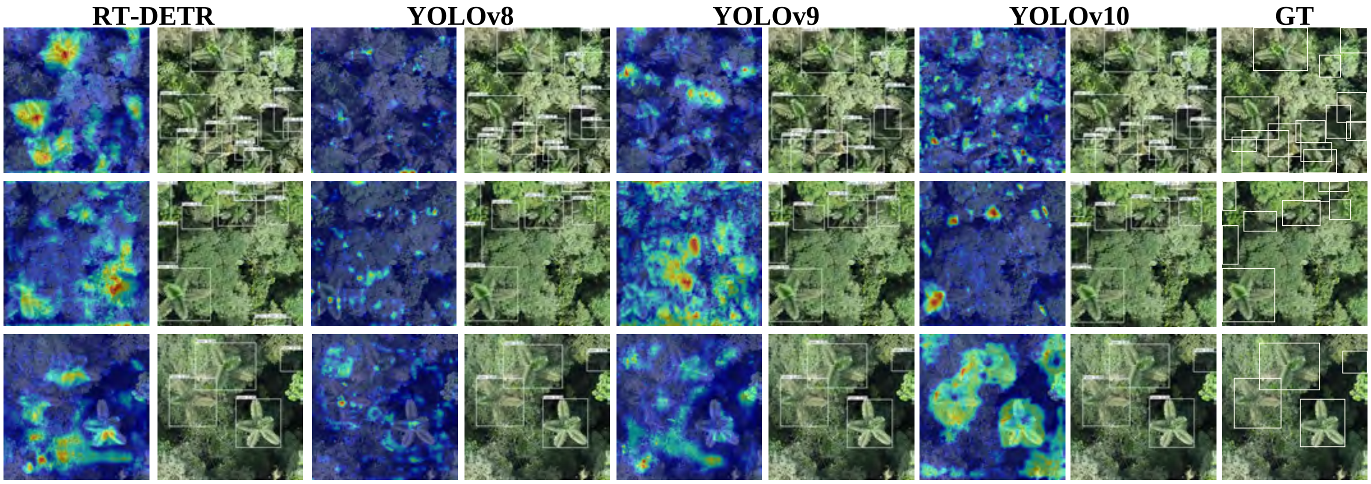


Figure 6: Comparative Visualization of Detection Outcomes and Saliency Maps Across Different Models in PalmDSNet. Each pair of columns displays detection results with bounding boxes and confidence levels alongside saliency maps derived from detection model weights. The final column showcases the labeled ground truth (GT).

Table II: Performance Comparison of Detection Models for PalmDSNet.

Model	GFLOPS	Params (M)	Latency (ms)	Validation (10% Training)					Validation (90% Training)				
				Precision	Recall	AP_{50}	AP_{75}	mAP	Precision	Recall	AP_{50}	AP_{75}	mAP
RT-DETR	222.5	65.5	22.4	0.845	0.778	0.843	0.560	0.532	**0.923**	0.905	**0.955**	**0.822**	**0.734**
YOLOv8	257.4	68.1	22.9	0.859	0.773	0.849	0.569	0.532	0.902	**0.919**	0.944	0.725	0.672
YOLOv9	189.1	57.4	24.3	**0.869**	**0.795**	**0.878**	**0.640**	**0.590**	0.897	0.914	0.950	0.769	0.680
YOLOv10	**169.8**	**31.6**	**21.2**	0.838	0.781	0.846	0.562	0.529	0.912	0.891	0.938	0.735	0.662

(40 cores, 2.20 GHz). Each simulation was repeated 10 times ($N = 10$) to mitigate random variance. The Ripley's G and F functions were computed at sampled values and compared to observed data, with their absolute differences integrated using the trapezoidal rule [92] to identify optimal parameters.

Our evaluation utilized a series of standard metrics for a thorough assessment of the model's performance. Precision and Recall measure the model's accuracy and error rates in detecting objects. Average Precision (AP) at various IoU thresholds quantifies precision across different recall levels, and Mean Average Precision (mAP) aggregates these measures for a comprehensive performance overview. We also examine computational demands by comparing Giga Floating Point Operations Per Second (GFLOPS), the number of parameters (Params), and latency to evaluate efficiency during training and inference. Model counting performance was validated by comparing predictions against expert annotations across five sites. For the bimodal reproduction model, optimal parameters p^* and σ^* were selected based on the minimum integrated difference between observed and simulated point patterns.

B. Performance Evaluation of PalmDSNet

1) Detection: Table II summarizes the performance of detection models in PalmDSNet under two training regimes: 10% and 90% of the data used for training. The evaluated models vary in computational costs and inference efficiencies. YOLOv10 is the lightest model (31.6M parameters, 169.8 GFLOPS), while YOLOv8 is the heaviest (68.1M parameters, 257.4 GFLOPS). Nevertheless, all models maintain consistent inference times under 25ms per frame, thus showing suitability for real-time applications and on-board deployment.

Detection performance improves consistently with increased training data. With only 10% of the data for training, YOLOv9 outperforms others, achieving mAP (0.59) and AP_{75} (0.64) – both over 6% higher than the nearest competitor. Its precision (0.869) and recall (0.795) are also the highest under this setting, which indicates strong capability in both correctly detecting palms and retrieving the majority of true palm instances. This suggests YOLOv9's generalizes well from limited supervision, making it a practical choice for scenarios with scarce annotations. In contrast, RT-DETR under 10% training exhibits lower precision (0.845), recall (0.778), and mAP (0.532), indicating challenges in accurately localizing palms with limited samples. YOLOv10, while being the most lightweight and fastest model, shows the lowest mAP (0.529) under this setting, which suggests a trade-off between efficiency and localization accuracy when data are limited.

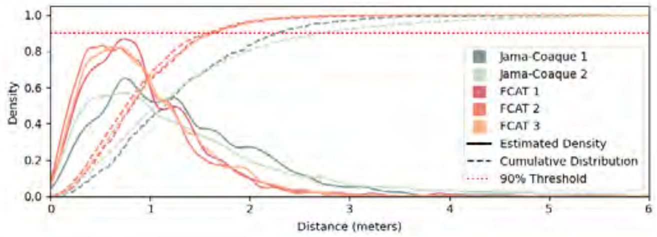
With 90% training data, all models show improved detection accuracy and bounding box quality. RT-DETR leads in mAP (0.734) and AP_{75} (0.822), outperforming the next best by about 5%, with strong precision (0.923) and recall (0.905). This indicates its effectivenes in both detection and localization when data are abundant. The gains over its 10% training performance (mAP: +0.202, AP_{75}: +0.262) further illustrates the data efficiency of ViT-based methods. In contrast, YOLOv10, while efficient, still lags in mAP (0.662) and AP_{75} (0.735), indicating a prioritization of efficiency over detection accuracy with ample data. These trends affirm the importance

Table III: Counting Performance Comparison of PalmDSNet Across 5 Sites. Sites marked with ⋆ are training locations.

Site	Size (ha)	No. of Counts	Counting Accuracy (10% Training)				Counting Accuracy (90% Training)			
			RT-DETR	YOLOv8	YOLOv9	YOLOv10	RT-DETR	YOLOv8	YOLOv9	YOLOv10
FCAT 1⋆	21.7	284	**0.997**	0.979	**0.997**	0.993	0.989	**0.997**	**0.997**	0.993
FCAT 2⋆	119.4	2,569	**0.998**	0.988	0.992	0.994	**0.999**	0.998	0.998	0.997
FCAT 3	103.1	2,206	**0.998**	0.995	0.993	0.995	**0.999**	0.998	0.998	0.995
Jama-Coaque 1	112.2	732	**0.969**	0.876	0.872	0.862	**0.975**	0.870	0.909	0.833
Jama-Coaque 2	92.3	1,596	**0.938**	0.776	0.812	0.775	**0.940**	0.743	0.824	0.791

(a) Sample Region (b) SAM (c) SAM 2 (d) Mobile SAM (e) FastSAM

Figure 7: Comparison of SAM Segmentation Results. While panels (b–d) demonstrate comparable segmentation performance, panel (e) reveals FastSAM's failure to segment palms effectively.



Statistics	FCAT 1	FCAT 2	FCAT 3	Jama-Coaque 1	Jama-Coaque 2
Mean (m)	0.90	0.84	0.84	1.20	1.30
Median (m)	0.77	0.73	0.74	1.06	1.05
Standard Deviation (m)	0.60	0.54	0.51	0.78	1.02

Figure 8: Distribution of Distance Shifts Across Sites.

Table IV: Latency of Different Configurations for PalmDSNet Across 5 Sites. Segmentation latency is measured based on the detection results of RT-DETR.

Model	FCAT 1	FCAT 2	FCAT 3	Jama-Coaque 1	Jama-Coaque 2
RT-DETR	42.9	378.0	273.5	177.1	143.0
YOLOv8	34.1	297.6	213.1	138.9	112.0
YOLOv9	34.7	297.2	216.4	140.4	113.8
YOLOv10	31.3	273.8	197.5	134.3	102.3
SAM	175.9	1454.9	1219.5	496.6	590.4
SAM 2	150.4	1265.4	1052.8	445.7	509.1
Mobile SAM	107.2	835.8	654.0	341.7	334.2

of data volume in optimizing model performance and guide model choice based on resource availability.

Figure 6 illustrates detection performance across varied scenes using models trained with 90% of the data. In the first row, YOLOv9 struggles with densely overlapped palm crowns, misdirecting attention toward non-palm regions and missing detections. In the second row, RT-DETR shows slightly higher confidence but yields two false positives in the lower right, likely due to a low threshold aimed at boosting recall, as supported by its saliency map. Similarly, YOLOv8 produces a false positive in the upper right, while YOLOv10, though more precise, fails to tightly bound a palm leaf there. In contrast, all models perform well in row three, correctly detecting palms in a simpler setting. Saliency maps reveal prioritization of palm leaves and centers, with centers often highlighted more intensely; however, they also display activities in non-palm zones, illustrating that while saliency maps guide model focus, the final detection head – layers responsible for final decision-making – ultimately confirms the presence of palms, which shows the disparity between model focus and actual output.

2) Segmentation: Figure 7 compares the segmentation performance of SAM, SAM 2, Mobile SAM, and FastSAM at Jama-Coaque 1 – an ecologically distinct site from the training environment, differing in rainfall, palm species, and forest structure. The results demonstrate the models' adaptability to out-of-sample data, despite some expected false alerts due to the intentionally low confidence threshold.

SAM provides precise boundary alignment with palm leaves but tends to fragment palm crowns into multiple segments where they overlap with other tree crowns. SAM 2 improves upon this by delivering better boundary segmentation while reducing the occurrence of small, broken segments, therefore providing a more cohesive representation of palm crowns. In contrast, Mobile SAM includes broader areas as foreground, resulting in more unified palm segmentation but with less precise boundaries. FastSAM, though efficient, struggles with dense tropical canopies, often failing to segment palm crowns and instead detecting irrelevant areas, even with prompts (Figure 7(e)). This stems from its lightweight CNN backbone, sparse training (2% of SA-1B), and its approach of segmenting images into pieces before matching prompts, which reduces its precision in complex environments.

3) Counting: Table III evaluates counting accuracy across five sites and examines the impact of training data quantity. Counting accuracy is defined as the proportion of detected centers located within a 5-meter radius of the labeled centers. FCAT 1 and FCAT 2, marked with ⋆, used for training, achieve near-perfect accuracy across all models and training ratios. FCAT 3, though not used in training, achieves high accuracy, especially for RT-DETR (0.998 with 10% data, 0.999 with 90%), indicating strong generalizability within the same ecological reserve. This suggests that PalmDSNet effectively captures structural and visual features that are consistent across

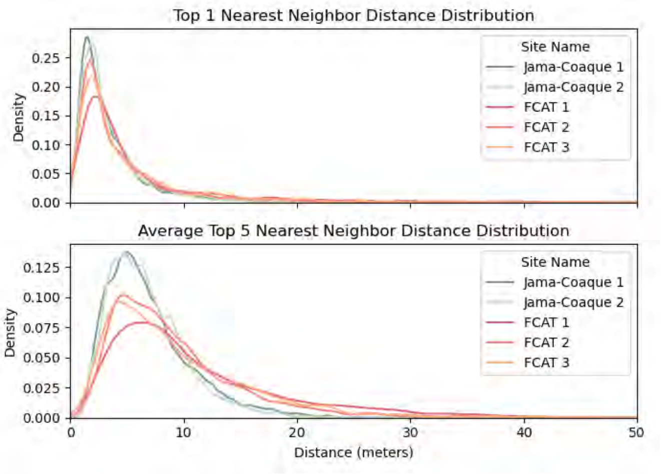


Figure 9: Distribution of Distances to Nearest Neighbors Among Detected Palm Centers Across Sites.

spatially separated but ecologically similar areas.

Performance declines at geographically distinct reserves. At Jama-Coaque 1, RT-DETR achieves the highest accuracy (0.975 with 90% data), only slightly improving over its 10% result (0.969). In contrast, YOLOv10's performance drops (0.833 with 90% data vs. 0.862 with 10% data), indicating potential overfitting or limited adaptability. A similar trend appears at Jama-Coaque 2, where RT-DETR remains robust (0.938 with 10% data, 0.940 with 90% data), but YOLOv8's accuracy drops sharply (0.776 to 0.743) despite increased training data. This shows that not all models benefit equally from larger training sets in cross-domain settings and underscores the need for models capable of handling domain shift.

Importantly, increasing training data from 10% to 90% yields minimal accuracy gains across most models and sites (e.g., RT-DETR: -0.8% to +0.6%). This implies additional data primarily refines localization rather than counting accuracy. It also emphasizes the need to balance data volume with model choice, particularly for cross-domain adaptation in ecologically diverse environments like tropical forests.

Figure 8 further quantifies localization performance using estimated density and cumulative distributions of distance shifts between predicted and true centers, which occur typically when only part of a palm is visible, causing the bounding box center to skew towards visible leaves. In FCAT sites, these shifts are minimal, with mean distances below 0.90 meters and medians under 0.77 meters. Jama-Coaque sites show greater shifts (mean $<$ 1.30 meters; median $<$ 1.06 meters), yet still within the 5-meter radius but indicating less precise localization in new environments.

Moreover, FCAT sites display lower variability in distance shifts (standard deviation 0.51–0.60 meters) compared to Jama-Coaque (0.78–1.02 meters). This indicates more consistent palm localization in FCAT regions, likely due to more homogeneous environmental conditions or better palm visibility. The increased variance in Jama-Coaque reflects the model's greater difficulty in adapting to variability in forest composition and palm species. Nonetheless, the model demonstrates robust adaptability across diverse environments.

Table V: Parameters p^* and σ^* for different sites.

Parameters	FCAT 1	FCAT 2	FCAT 3	Jama-Coaque 1	Jama-Coaque 2
p^*	0.49	0.52	0.46	0.64	0.51
σ^*	50	70	70	80	60

4) Landscape-level Processing: The efficiency of PalmDSNet presented in Table IV substantiates the its capability for landscape-level processing, which is essential for rapid ecological assessments. Detection (first 4 rows) involves identifying palms and recording georeferenced outputs in a CSV, while segmentation (last 3 rows) generates and stores the combined segmentation mask overlaid on the image. Models maintain a high throughput, processing each frame within 25 ms on an RTX 3090, which shows their potential for real-time applications and suitability for on-board UAV deployment. Detection times vary from 31.3 to 378 seconds depending on the area of the site, while segmentation durations vary from 107.2 to 1454.9 seconds, influenced by detected palm counts. Overall, full-site processing times for detection, segmentation, and counting span from about 2.5 to 30 minutes, confirming the method's potential real-time use in varied environments.

To reduce missed detections, we use a sliding window with half-patch strides during inference, ensuring each region of the orthomosaic is processed four times. This overlap improves detection of partially visible palms, while redundant outputs are filtered using NMS. With precision and recall per patch just over 90%, this strategy ensures high overall coverage and efficient large-scale palm extraction, as evidenced by consistent counting accuracy across different ecosystems.

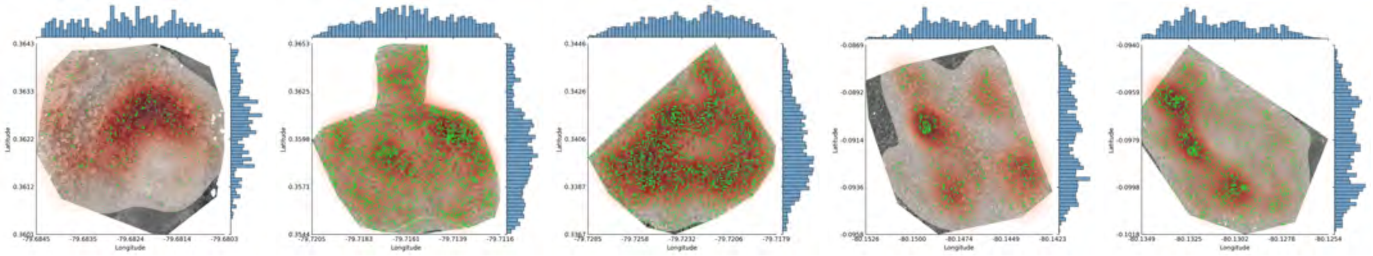
C. Spatial Distribution Analysis

The analysis of spatial distribution in palm populations offers critical insights into their ecological dynamics and interspecies relationships. By examining spatial patterns, we can better understand species coexistence and competition, which are essential for guiding sustainable forest management and utilization strategies [78], [93]. This section explores palm distribution through nearest neighbor analysis, distribution randomness analysis, and spatial point pattern simulation.

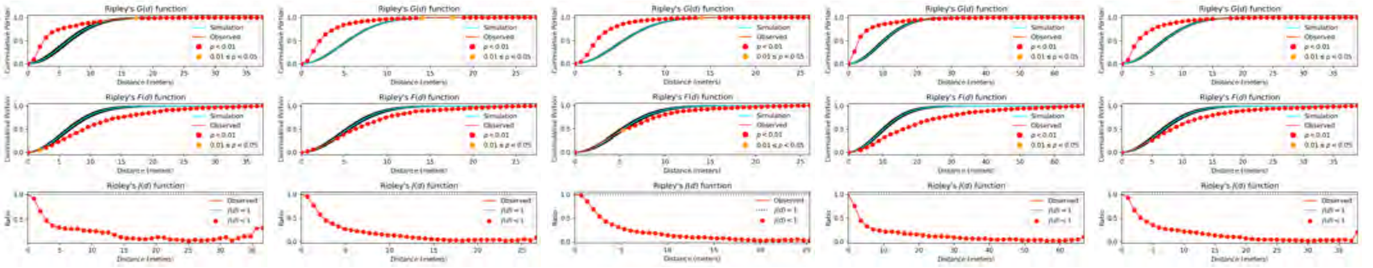
1) Nearest Neighbor Analysis: Figure 9 depicts the distance distributions from each detected palm to its nearest neighbors. The results reveal compact clustering, which suggests biological interactions or shared habitat preferences of palms. The average top five nearest neighbors' distance distribution shows that while palms generally form tight clusters, they also participate in larger, slightly more dispersed groups. This pattern is more distinct in Jama-Coaque than FCAT, reflecting ecological heterogeneity and differences in demographic or environmental drivers – an area warranting further study.

2) Distribution Randomness Analysis: This analysis evaluates whether the spatial distribution of palms across various sites deviates from a random pattern by comparing observed patterns with those generated by a Poisson point process. We aim to identify underlying ecological patterns that may indicate clustering or dispersion within these environments.

Figure 10 presents the observed spatial distribution of palms across five study sites. These regions display heterogeneous



(a) FCAT 1 (b) FCAT 2 (c) FCAT 3 (d) Jama-Coaque 1 (e) Jama-Coaque 2



(f) FCAT 1 (g) FCAT 2 (h) FCAT 3 (i) Jama-Coaque 1 (j) Jama-Coaque 2

Figure 10: Visualization and Randomness Analysis of Palm Spatial Distribution. The first row (a-e) presents kernel density estimates with histograms illustrating palm distribution across regions. Green dots indicate detected palm locations, with darker KDE regions showing higher clustering. The second row (f-j) compares Ripley's functions of the detected palms to those generated by a Poisson process, highlighting deviations from randomness. Blue-shaded areas represent the 95% confidence intervals with the black line indicating the mean for a Poisson process, while the red lines represent Ripley's functions of the observed data. Red and orange points denote regions with p-values below 1% and 5%, respectively, indicating significant differences from randomness. These indices quantify palm clustering or dispersion relative to a random distribution.

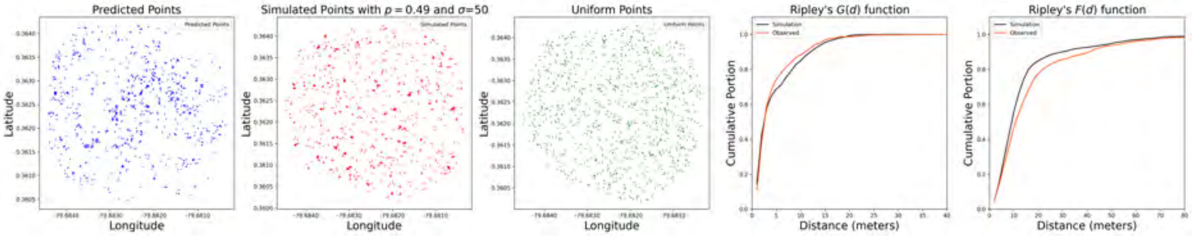
densities, with certain areas exhibiting higher concentrations. The first row visually indicates non-random structures with clusters and dispersed regions are clearly visible. The second row quantitatively compares observed point patterns of palm distributions with those generated by a Poisson point process using Ripley's functions. For the G and F plots, the central black line represents the expected pattern if palms were randomly scattered throughout the forest. The blue-shaded area around this line indicates the 95% confidence interval within which random data points would typically fall, with different confidence levels $(1-p)$ for the sampled points being random, depicted with varying colors for $p < 0.01$ and $p < 0.05$. The J function plot includes a dashed line at $J(d) = 1$, with values below this line colored red to indicate clustering.

For the G function, the observed line ascends more steeply than the random case, indicating a higher degree of clustering, with palms closer to their nearest neighbors than random placement would suggest. Conversely, in the F function, the observed line falls below the random case, implying that random points must travel a greater distance to find a nearby palm, indicating larger gaps or empty spaces in the actual palm distribution compared to a random arrangement. The J function, with values below 1, further supports these observations by indicating that the palms are more clustered than would be expected under randomness. Together, these Ripley's functions reveal significant deviations from randomness and suggest complex ecological interactions shaping palm distributions. The observed pattern's clustering is more prominent than randomness yet less extreme than one would expect with a Student's t distribution. This sets the stage for future work to explore the underlying ecological drivers of the distribution.

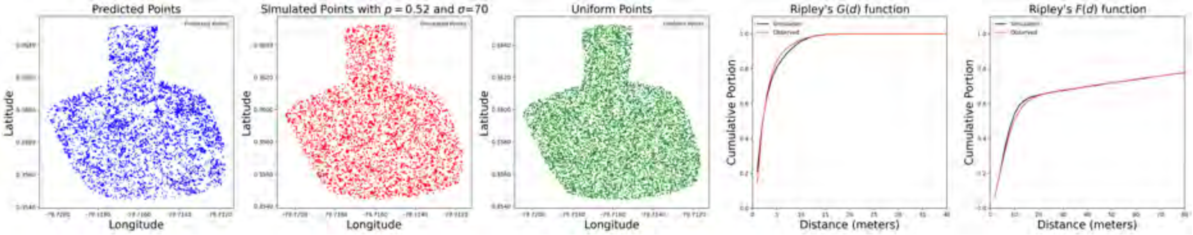
3) Simulation of Spatial Point Pattern: Given the statistically significant evidence that the distribution of palms is far from random, as anticipated due to the influence of various factors such as habitat suitability, as well as human and animal activities, we seek to employ a statistical yet straightforward model that can simulate and explain the spatial point patterns of palm distribution. To this end, we utilize the proposed Poisson-Gaussian palm reproduction model, which integrates the inherent stochastic nature of palm reproduction with spatially dependent factors, thereby generating distributions that can be compared with those predicted by PalmDSNet.

To measure clustering and dispersion in the simulated palm distributions, we use Ripley's G and F functions, with G indicating clustering and F showing dispersion between palms. By comparing these functions from our simulations to those derived from PalmDSNet, we aim to align the simulated distributions with the predicted ones. Ecologically, the parameter p indicates the degree of global randomness, with higher values denoting greater randomness and lower values indicating local clustering with a Gaussian distribution. The parameter σ controls the local range of palm reproduction, reflecting the degree of clustering around individual palms.

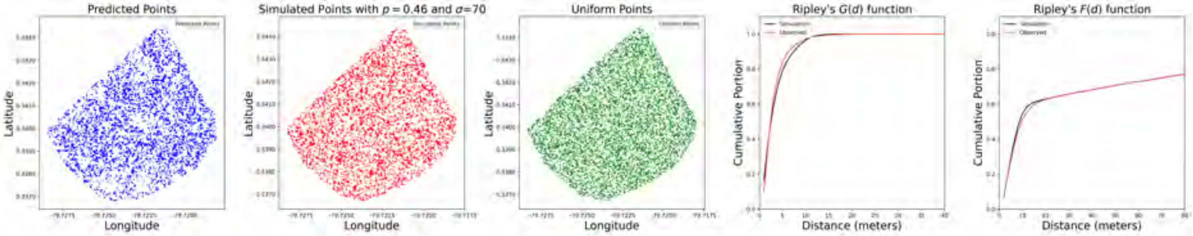
Table V lists the grid search results of the optimal (p^*, σ^*) pairs for the five study sites. Four of the sites yield stable results, with p^* ranging from 0.46 to 0.52 and σ^* between 50 to 80 ($\sim$2$-$4 m). Jama-Coaque 1 stands out with a higher p^* of 0.64, indicating more dispersed clusters, as also reflected by the slightly larger σ^* for this site. This greater dispersion is evident in the estimated density shown in Figure 10d. The



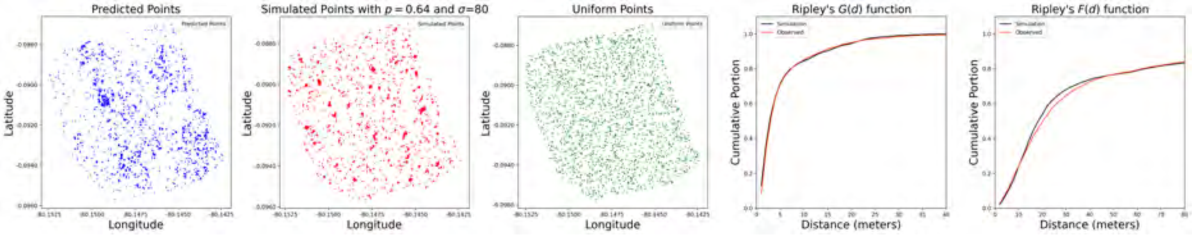
(a) FCAT 1



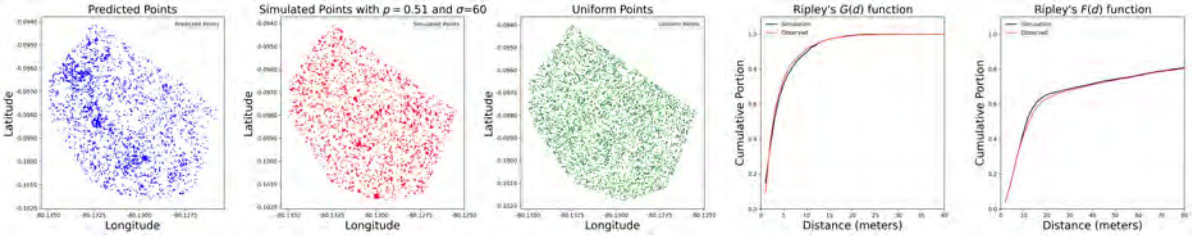
(b) FCAT 2



(c) FCAT 3



(d) Jama-Coaque 1



(e) Jama-Coaque 2

Figure 11: Comparison of Simulated and Predicted Point Patterns. The first column shows PalmDSNet's predicted palm distribution, the second column shows the simulated distribution, and the third column depicts a uniform random pattern. The fourth and fifth columns compare Ripley's G and F functions, respectively, between the simulated and predicted patterns.

robustness of p^* and σ^* across sites demonstrates the model's capability to replicate typical palm spatial patterns with only two parameters. This simplicity helps in understanding the spread of palm species in forests under various scenarios.

Figure 11 compares our simulated point patterns to those predicted across sites. The first column presents the predicted spatial distribution of detected palms by PalmDSNet, the second column shows the simulated distribution, and the third column depicts random points with a uniform distribution. The fourth and fifth columns compare Ripley's G and F functions, respectively, between the simulated and predicted distributions. This last comparison shows an excellent fit for all sites. With the exception of FCAT 1, the simulated point patterns in the second column are visually more homogeneous than the predicted ones. This is easily explained by the fact that the distribution parameters were estimated globally. This observation should lead us to perform slightly more localized parameter estimation in the future.

V. CONCLUSION AND FUTURE WORK

The PalmDSNet framework has proven effective in detecting, segmenting, and counting palms across diverse forest environments. We found that increasing the number of training labels enhances the precision of bounding boxes, though it has a minimal impact on counting accuracy. PalmDSNet supports efficient, large-scale landscape analysis even under resource-constrained scenarios, including on-board UAV processing and integration into autonomous flight control systems, despite using weak supervision, where labels for bounding boxes and crown centers only sparsely overlap. By delivering the framework to function under these conditions, we ensure its generalizability to new geographic regions without retraining, while still extracting useful ecological information.

Additionally, our spatial analysis across multiple forest reserves shows that palms tend to form tightly-knit clusters, yet their spatial arrangements vary significantly across ecosystems. The Poisson-Gaussian reproduction model, which employs a brute-force method to identify the optimal parameters that align with observed point patterns using Ripley's G and F functions, has also shown satisfying performance. This model reliably captures the spatial patterns of palm distribution across various forest reserves, with consistent parameter ranges observed in most sites. By quantifying the level of clustering and its deviation from randomness, the model enhances our understanding of spatial processes of palms, which is crucial for conservation efforts, management strategies, economic planning, and supply chain logistics.

The integration of these two models: PalmDSNet for data-driven prediction and the Poisson-Gaussian model for ecological interpretation, is necessary given the nature of our data and the distinct goals of each stage. The two-step process separates the prediction and interpretation tasks, and reflects the reality of ecological data workflows, where stakeholders often operate with disparate annotation formats and goals. For instance, ecologists and conservation practitioners frequently rely on georeferenced crown annotations created in GIS softwares, while computer scientists typically label bounding boxes using Python tools. Aligning these annotations spatially at scale is inherently hard, especially since cropped image patches used for training may not preserve georeferenced information.

Furthermore, in practice, different stakeholders require different outputs: some applications (e.g., automated counting) benefit directly from accurate detection results, while others (e.g., ecological modeling, biodiversity assessments) may already have access to center annotations and are primarily interested in analyzing spatial structure. The modularity of our approach supports both use cases, which allows broader adoption in conservation, ecological monitoring, and land management. By addressing these diverse needs, our framework is well suited for real-world ecological applications, particularly in regions where expert-labeled, comprehensive datasets are difficult and costly to produce.

Future work will expand the dataset via semi-automated labeling, covering new regions in Ecuador and a 1000 km^2 area near Iquitos, Peru, with species-level differentiation. We will incorporate a normalized Digital Surface Model to integrate height information for refining palm localization in dense and overlapping canopies [94]. To improve accuracy metrics and better reflect real-world performance, we will also develop a symmetric counting metric that considers both palm centers captured by predictions and predictions supported by true centers. Applying the framework to multi-year imagery will further enable tracking of temporal dynamics in palm populations under environmental change [95], [96]. Deployment on edge devices, such as the NVIDIA Jetson series, will enable real-time inference in the field.

The framework is broadly adaptable to conservation tasks, such as species-level canopy mapping, land cover segmentation, and monitoring of threats such as illegal mining [55], [97]. Future extensions will incorporate more complex spatial models that account for environmental drivers, demographic processes, and anthropogenic impacts, particularly after incorporating species-level classifications.

REFERENCES

[1] W. L. Eiserhardt, J.-C. Svenning, W. D. Kissling, *et al.*, “Geographical ecology of the palms (arecaceae): Determinants of diversity and distributions across spatial scales,” *Ann. Bot.*, vol. 108, no. 8, pp. 1391–1416, 2011.

[2] N. Y. P. Zambrana, A. Byg, J.-C. Svenning, *et al.*, “Diversity of palm uses in the western amazon,” *Biol. Conserv.*, vol. 16, pp. 2771–2787, 2007.

[3] W. J. Sutherland, R. P. Freckleton, H. C. J. Godfray, *et al.*, “Identification of 100 fundamental ecological questions,” *J. Ecol.*, vol. 101, no. 1, pp. 58–67, 2013.

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

[5] F. H. Wagner, R. Dalagnol, X. Tagle Casapia, *et al.*, “Regional mapping and spatial distribution analysis of canopy palms in an amazon forest using deep learning and vhr images,” *Remote Sens.*, vol. 12, no. 14, p. 2225, 2020.

[6] N. C. Pitman, J. E. G. Andino, M. Aulestia, *et al.*, “Distribution and abundance of tree species in swamp forests of amazonian ecuador,” *Ecography*, vol. 37, no. 9, pp. 902–915, 2014.

[7] Y. Malhi, T. A. Gardner, G. R. Goldsmith, *et al.*, “Tropical forests in the anthropocene,” *Annual Review of Environment and Resources*, vol. 39, pp. 125–159, 2014.

[8] Y. van der Hoek, S. Álvarez Solas, and M. C. Peñuela, “The palm mauritia flexuosa, a keystone plant resource on multiple fronts,” *Biol. Conserv.*, vol. 28, no. 3, pp. 539–551, 2019.

[9] J. Terborgh, “Community aspects of frugivory in tropical forests,” in *Frugivores and seed dispersal*, Springer, 1986, pp. 371–384.

[10] S. L. Polk, K. Cui, A. H. Chan, *et al.*, “Unsupervised diffusion and volume maximization-based clustering of hyperspectral images,” *Remote Sens.*, vol. 15, no. 4, p. 1053, 2023.

[11] X. Ma, J. Yang, T. Hong, *et al.*, “Stnet: Spatial and temporal feature fusion network for change detection in remote sensing images,” in *Proc. ICME*, 2023, pp. 2195–2200. DOI: 10.1109/ICME55011.2023.00375.

[12] R. Guan, W. Tu, Z. Li, *et al.*, “Spatial-spectral graph contrastive clustering with hard sample mining for hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, 2024.

[13] R.-F. Wang and W.-H. Su, “The application of deep learning in the whole potato production chain: A comprehensive review,” *Agriculture*, vol. 14, no. 8, p. 1225, 2024.

[14] Z. Chang-Tao, W. Rui-Feng, T. Yu-Hao, *et al.*, “Automatic lettuce weed detection and classification based on optimized convolutional neural networks for robotic weed control,” *Agronomy*, vol. 14, no. 12, p. 2838, 2024.

[15] P. Ma, T. G. Mondal, Z. Shi, *et al.*, "Early detection of pipeline natural gas leakage from hyperspectral imaging by vegetation indicators and deep neural networks," *Environ. Sci. & Technol.*, 2024.

[16] S. L. Polk, A. H. Chan, K. Cui, *et al.*, "Unsupervised detection of ash dieback disease (hymenoscyphus fraxineus) using diffusion-based hyperspectral image clustering," in *Proc. IGARSS*, IEEE, 2022, pp. 2287–2290.

[17] K. Cui, Z. Shao, G. Larsen, *et al.*, "Palmprobnet: A probabilistic approach to understanding palm distributions in ecuadorian tropical forest via transfer learning," in *Proceedings of the 2024 ACM Southeast Conference*, 2024, pp. 272–277.

[18] X. Tagle Casapia, L. Falen, H. Bartholomeus, *et al.*, "Identifying and quantifying the abundance of economically important palms in tropical moist forest using uav imagery," *Remote Sens.*, vol. 12, no. 1, p. 9, 2019.

[19] J. Peng, X. Bu, M. Sun, *et al.*, "Large-scale object detection in the wild from imbalanced multi-labels," in *Proc. CVPR*, 2020, pp. 9709–9718.

[20] Z. Wang, B. Li, C. Wang, *et al.*, "Airshot: Efficient few-shot detection for autonomous exploration," *arXiv preprint arXiv:2404.05069*, 2024.

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

[22] M. Freudenberg, N. Nölke, A. Agostini, *et al.*, "Large scale palm tree detection in high resolution satellite images using u-net," *Remote Sens.*, vol. 11, no. 3, p. 312, 2019.

[23] J. Y. Park, H. C. Muller-Landau, J. W. Lichstein, *et al.*, "Quantifying leaf phenology of individual trees and species in a tropical forest using unmanned aerial vehicle (uav) images," *Remote Sens.*, vol. 11, no. 13, p. 1534, 2019.

[24] M. B. A. Gibril, H. Z. M. Shafri, A. Shanableh, *et al.*, "Deep convolutional neural network for large-scale date palm tree mapping from uav-based images," *Remote Sens.*, vol. 13, no. 14, p. 2787, 2021.

[25] K. Cui and R. J. Plemmons, "Unsupervised classification of aviris-ng hyperspectral images," in *Proc. WHISPERS*, IEEE, 2021, pp. 1–5.

[26] H. Tao, J. Li, Z. Hua, *et al.*, "Dudb: Deep unfolding based dual-branch feature fusion network for pan-sharpening remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, 2023.

[27] R. Muscarella, T. Emilio, O. L. Phillips, *et al.*, "The global abundance of tree palms," *Glob. Ecol. Biogeogr.*, vol. 29, no. 9, pp. 1495–1514, 2020.

[28] C. G. Hidalgo Pizango, E. N. Honorio Coronado, J. del Águila-Pasquel, *et al.*, "Sustainable palm fruit harvesting as a pathway to conserve amazon peatland forests," *Nat. Sustain.*, vol. 5, no. 6, pp. 479–487, 2022.

[29] K. Cui, R. Li, S. L. Polk, *et al.*, "Superpixel-based and spatially-regularized diffusion learning for unsupervised hyperspectral image clustering," *IEEE Trans. Geosci. Remote Sens.*, 2024.

[30] J. Zhang, S. Xu, Y. Zhao, *et al.*, "Aerial orthoimage generation for uav remote sensing," *Information Fusion*, vol. 89, pp. 91–120, 2023.

[31] Z. Qin, W. Wang, K.-H. Dammer, *et al.*, "Ag-yolo: A real-time low-cost detector for precise spraying with case study of palms," *Front. Plant Sci.*, vol. 12, p. 753 603, 2021.

[32] T. Jintasuttisak, E. Edirisinghe, and A. Elbattay, "Deep neural network based date palm tree detection in drone imagery," *Comput. Electron. Agric.*, vol. 192, p. 106 560, 2022.

[33] X. Ma, R. Che, T. Hong, *et al.*, "Sacanet: Scene-aware class attention network for semantic segmentation of remote sensing images," in *Proc. ICME*, 2023, pp. 828–833. DOI: 10.1109/ICME55011.2023.00147.

[34] M. Di Marco, S. Ferrier, T. D. Harwood, *et al.*, "Wilderness areas halve the extinction risk of terrestrial biodiversity," *Nature*, vol. 573, no. 7775, pp. 582–585, 2019.

[35] Z. Zou, K. Chen, Z. Shi, *et al.*, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[36] F. Pan, Y. Wu, K. Cui, *et al.*, "Accurate detection and instance segmentation of unstained living adherent cells in differential interference contrast images," *Comput. Biol. Med.*, vol. 182, p. 109 151, 2024.

[37] K. Gao, H. Zhang, and G. Wu, "A multispectral vision-based machine learning framework for non-contact vehicle weigh-in-motion," *Measurement*, vol. 226, p. 114 162, 2024.

[38] Y. Li, H. Wang, Z. Li, *et al.*, "DAANet: Dual Attention Aggregating Network for Salient Object Detection*," in *2Proc. ROBIO*, IEEE, 2023, pp. 1–7. DOI: `10.1109/ROBIO58561.2023.10354933`.

[39] Q. Liu, C. Kaul, J. Wang, *et al.*, "Optimizing vision transformers for medical image segmentation," in *Proc. ICASSP*, IEEE, 2023, pp. 1–5.

[40] Y. Li, H. Wang, and A. Katsaggelos, "CPDR: Towards Highly-Efficient Salient Object Detection via Crossed Post-decoder Refinement," in *Proc. BMVC*, BMVA, 2024.

[41] Z. Zhu, X. Li, J. Zhai, *et al.*, "Podb: A learning-based polarimetric object detection benchmark for road scenes in adverse weather conditions," *Inf. Fusion*, vol. 108, p. 102 385, 2024.

[42] N. Carion, F. Massa, G. Synnaeve, *et al.*, "End-to-end object detection with transformers," in *Proc. ECCV*, Springer, 2020, pp. 213–229.

[43] Y. Zhao, W. Lv, S. Xu, *et al.*, *Detrs beat yolos on real-time object detection*, arXiv preprint arXiv:2304.08069, 2023. arXiv: 2304.08069.

[44] G. Jocher, A. Chaurasia, and J. Qiu, *Ultralytics yolov8*, https://github.com/ultralytics/ultralytics, version 8.0.0, Accessed: 2024-05-10, 2023.

[45] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, *Yolov9: Learning what you want to learn using programmable gradient information*, arXiv preprint arXiv:2402.13616, 2024. arXiv: 2402.13616 `[cs.CV]`.

[46] A. Wang, H. Chen, L. Liu, *et al.*, *Yolov10: Real-time end-to-end object detection*, arXiv preprint arXiv:2405.14458, 2024. arXiv: 2405.14458 `[cs.CV]`.

[47] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using yolo: Challenges, architectural successors, datasets and applications," *Multim. Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, 2023.

[48] H. Wang, B. Zhu, Y. Li, *et al.*, "Sygnet: A svd-yolo based ghostnet for real-time driving scene parsing," in *Proc. ICIP*, IEEE, 2022, pp. 2701–2705.

[49] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, *et al.*, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proc. CVPR Workshops*, 2020, pp. 390–391.

[50] W. Tang, K. Cui, and R. H. Chan, "Optimized hard exudate detection with supervised contrastive learning," in *Proc. ISBI*, IEEE, 2024, pp. 1–5.

[51] Q. Liu, X. Gu, P. Henderson, *et al.*, "Multi-scale cross contrastive learning for semi-supervised medical image segmentation," *arXiv preprint arXiv:2306.14293*, 2023.

[52] X. Ma, Z. Wu, R. Lian, *et al.*, *Rethinking remote sensing change detection with a mask view*, 2024. arXiv: 2406.15320 `[cs.CV]`. [Online]. Available: https://arxiv.org/abs/2406.15320.

[53] W. Zhang, M. Ma, Y. Jiang, *et al.*, "Center-guided classifier for semantic segmentation of remote sensing images," *arXiv preprint arXiv:2503.16963*, 2025.

[54] Q. Liu, X. Gu, P. Henderson, *et al.*, "Certainty-guided cross contrastive learning for semi-supervised medical image segmentation," *Authorea Preprints*, 2024.

[55] 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 *Proc. WHISPERS*, IEEE, 2022, pp. 1–5.

[56] Q. Liu, P. Henderson, X. Gu, *et al.*, "Learning semi-supervised medical image segmentation from spatial registration," *arXiv preprint arXiv:2409.10422*, 2024.

[57] R. Li, K. Cui, R. H. Chan, *et al.*, "Classification of hyperspectral images using svm with shape-adaptive reconstruction and smoothed total variation," in *Proc. IGARSS*, IEEE, 2022, pp. 1368–1371.

[58] A. Kirillov, E. Mintun, N. Ravi, *et al.*, "Segment anything," in *Proc. ICCV*, 2023, pp. 4015–4026.

[59] N. Ravi, V. Gabeur, Y.-T. Hu, *et al.*, "Sam 2: Segment anything in images and videos," *arXiv preprint arXiv:2408.00714*, 2024.

[60] C. Zhang, D. Han, Y. Qiao, *et al.*, *Faster segment anything: Towards lightweight sam for mobile applications*, arXiv preprint arXiv:2306.14289, 2023. arXiv: 2306.14289.

[61] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[62] K. He, X. Chen, S. Xie, *et al.*, "Masked autoencoders are scalable vision learners," in *Proc. CVPR*, 2022, pp. 16 000–16 009.

[63] C. Ryali, Y.-T. Hu, D. Bolya, *et al.*, "Hiera: A hierarchical vision transformer without the bells-and-whistles," in *Proc. ICCV*, PMLR, 2023, pp. 29 441–29 454.

[64] X. Zhao, W. Ding, Y. An, *et al.*, "Fast segment anything," *arXiv preprint arXiv:2306.12156*, 2023.

[65] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[66] A. Radford, J. W. Kim, C. Hallacy, *et al.*, "Learning transferable visual models from natural language supervision," in *Proc. ICML*, PMLR, 2021, pp. 8748–8763.

[67] M. Luo, T. Zhang, S. Wei, *et al.*, "Sam-rsis: Progressively adapting sam with box prompting to remote sensing image instance segmentation," *IEEE Trans. Geosci. Remote Sens.*, 2024.

[68] A. Khatua, A. Bhattacharya, A. K. Goswami, *et al.*, "Developing approaches in building classification and extraction with synergy of yolov8 and sam models," *Spat. Inf. Res.*, pp. 1–20, 2024.

[69] J. Reddy, H. Niu, J. L. L. Scott, *et al.*, "Cotton yield prediction via uav-based cotton boll image segmentation using yolo model and segment anything model (sam)," *Remote Sens.*, vol. 16, no. 23, p. 4346, 2024.

[70] I. Petritan, R. Marzano, A. M. Petritan, *et al.*, "Overstory succession in a mixed quercus petraea–fagus sylvatica old growth forest revealed through the spatial pattern of competition and mortality," *For. Ecol. Manag.*, vol. 326, pp. 9–17, 2014.

[71] K. Gadow, C. Zhang, C. Wehenkel, *et al.*, "Forest structure and diversity," *Continuous Cover Forestry, Book Series Managing Forest Ecosystems*, vol. 23, pp. 29–83, 2012.

[72] F. May, A. Huth, and T. Wiegand, "Moving beyond abundance distributions: Neutral theory and spatial patterns in a tropical forest," *Proc. R. Soc. B: Biol. Sci.*, vol. 282, 2015.

[73] B. D. Ripley, *Spatial statistics*. Wiley, 1981.

[74] A. Baddeley, E. Rubak, and R. Turner, *Spatial point patterns: Methodology and applications with R*. CRC Press, Taylor & Francis Group, 2016.

[75] P. Haase, "Spatial pattern analysis in ecology based on ripley's k-function: Introduction and methods of edge correction," *J. Veg. Sci.*, vol. 6, pp. 575–582, 1995.

[76] J. Besag, "Contribution to the discussion on dr. ripley's paper," *J. R. Stat. Soc., B: Stat. Methodol.*, vol. 39, no. 2, pp. 193–195, 1977.

[77] F. Goreaud, "Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes," *Dissertation*, 2000.

[78] M. Ben-Said, "Spatial point-pattern analysis as a powerful tool in identifying pattern-process relationships in plant ecology: An updated review," *Ecol. Process.*, vol. 10, pp. 1–23, 2021.

[79] E. Velázquez, I. Martinez Cano, S. Getzin, *et al.*, "An evaluation of the state of spatial point pattern analysis in ecology," *Ecography*, vol. 39, 2016.

[80] T. Wiegand and K. A. Moloney, "Rings, circles, and null-models for point pattern analysis in ecology," *Oikos*, vol. 104, pp. 209–229, 2004.

[81] I. Renner, J. Elith, A. Baddeley, *et al.*, "Point process models for presence-only analysis," *Methods Ecol. Evol.*, vol. 6, 2015.

[82] R. Law, J. Illian, D. Burslem, *et al.*, "Ecoogical information from satial patterns of plants: Insights from point process theory," *J. Ecol.*, vol. 97, pp. 616–628, 2009.

[83] M. Jácome-Flores, M. Delibes, T. Wiegand, *et al.*, "Spatial patterns of an endemic mediterranean palm recolonizing old fields," *Ecol. Evol.*, vol. 6, pp. 1–13, Nov. 2016. DOI: 10.1002/ece3.2504.

[84] L. Browne and J. Karubian, "Diversity of palm communities at different spatial scales in a recently fragmented tropical landscape," *Bot. J. Linn. Soc.*, vol. 182, no. 2, pp. 451–464, 2016.

[85] 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," *Front. Ecol. Evol.*, vol. 10, p. 678 125, 2022.

[86] K. Cui, R. Zhu, M. Wang, *et al.*, "Detection and geographic localization of natural objects in the wild: A case study on palms," *arXiv preprint arXiv:2502.13023*, 2025.

[87] R. R. Selvaraju, M. Cogswell, A. Das, *et al.*, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. ICCV*, 2017, pp. 618–626.

[88] J. Gildenblat and contributors, *Pytorch library for cam methods*, https://github.com/jacobgil/pytorch-grad-cam, 2021.

[89] J. F. C. Kingman, *Poisson processes*. Clarendon Press, 1992, vol. 3.

[90] M. R. Metz, W. P. Sousa, and R. Valencia, "Widespread density-dependent seedling mortality promotes species coexistence in a highly diverse amazonian rain forest," *Ecology*, vol. 91, no. 12, pp. 3675–3685, 2010.

[91] J. S. Clark, M. Silman, R. Kern, *et al.*, "Seed dispersal near and far: Patterns across temperate and tropical forests," *Ecology*, vol. 80, no. 5, pp. 1475–1494, 1999.

[92] F. B. Hildebrand, *Introduction to numerical analysis*. Courier Corporation, 1987.

[93] S. Tao, N. Labrière, K. Calders, *et al.*, "Mapping tropical forest trees across large areas with lightweight cost-effective terrestrial laser scanning," *Ann. For. Sci.*, vol. 78, no. 4, p. 103, 2021.

[94] D. J. Dixon, Y. Zhu, and Y. Jin, "Canopy height estimation from planetscope time series with spatio-temporal deep learning," *Remote Sens. Environ.*, vol. 318, p. 114 518, 2025.

[95] X. Qiu, J. Hu, L. Zhou, *et al.*, "Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods," *arXiv preprint arXiv:2403.20150*, 2024.

[96] X. Qiu, X. Wu, Y. Lin, *et al.*, "Duet: Dual clustering enhanced multivariate time series forecasting," *arXiv preprint arXiv:2412.10859*, 2024.

[97] S. Camalan, K. Cui, P. 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, 2022, B52G–0905.