Detection and Geographic Localization of Natural Objects in the Wild: A Case Study on Palms

Kangning Cui[1,2,†] Rongkun Zhu[3,†] Manqi Wang[1] Wei Tang[2]
Gregory Larsen[1] Victor Pauca[1] Sarra Alqahtani[1] Fan Yang[1]
David Segurado[1] David Lutz[4] Jean-Michel Morel[2] Miles Silman[1]

[1]Wake Forest University
[2]City University of Hong Kong
[3]Xidian University
[4]Colby-Sawyer College

IJCAI-25 AI and Social Good Track

[†]Denotes equal contribution.

Outline

The Challenge: Finding Palms in Natural Forests

Figure: Detection is easy in plantations (left), but hard in natural forests (right).

Ecological & Economic Significance of Palms:

- Vital to tropical forest ecology and biodiversity.
- Support local livelihoods and are key resources for tropical wildlife.
- Act as bioindicators of forest health and environmental impact.

The Problem:

- Most research focuses on plantations (ordered, sparse).
- Natural forests are chaotic with: irregular spacing, overlapping crowns, complex backgrounds, uneven lighting.

Our Contributions

1. **The PALMS Dataset**
   - A Large-scale UAV imagery dataset for **PA**lm **L**ocalization in **M**ulti-**S**cale from **21 ecologically diverse sites** in western Ecuador.
   - Annotated with **8,830** bounding boxes and **5,026** georeferenced ground-truth center points for palms.

Our Contributions

1. **The PALMS Dataset**
   - A Large-scale UAV imagery dataset for **PA**lm **L**ocalization in **M**ulti-**S**cale from **21 ecologically diverse sites** in western Ecuador.
   - Annotated with **8,830** bounding boxes and **5,026** georeferenced ground-truth center points for palms.

2. **The PRISM Pipeline**
   - An end-to-end efficient framework for **P**rocessing, **I**nference, **S**egmentation, and **M**apping.
   - Ensures trustworthiness with interpretability via saliency maps and confidence calibration.

Our Contributions

1. **The PALMS Dataset**
   - A Large-scale UAV imagery dataset for **PA**lm **L**ocalization in **M**ulti-**S**cale from **21 ecologically diverse sites** in western Ecuador.
   - Annotated with **8,830** bounding boxes and **5,026** georeferenced ground-truth center points for palms.
2. **The PRISM Pipeline**
   - An end-to-end efficient framework for **P**rocessing, **I**nference, **S**egmentation, and **M**apping.
   - Ensures trustworthiness with interpretability via saliency maps and confidence calibration.
3. **Comprehensive Validation**
   - Demonstrated strong generalization across four distinct reserves.
   - Achieved high performance, locating palm centers with a median error of less than 1.5 meters.

The PALMS Dataset: Data from the Field

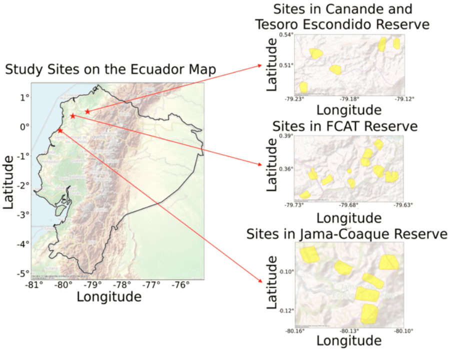


Figure: Study sites across a rainfall and ecological gradient in western Ecuador.

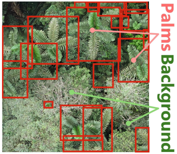


Figure: Bounding box annotations.

Figure: Georeferenced palm centers.

- Data collected from four reserves spanning wet to dry tropical forests.
- Captures high variation in palm species, density, and canopy structure.
- High-resolution orthomosaics created from thousands of UAV images.

The PRISM Pipeline at a Glance

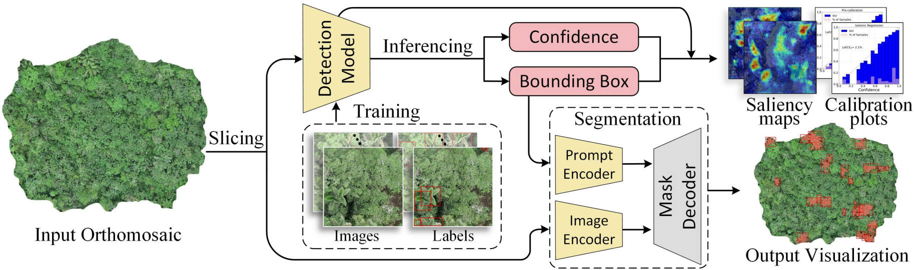


Figure: Our modular pipeline: from orthomosaic to georeferenced coordinates.

Core Components:

- **Detection:** Fine-tuned models locate palms in orthomosaic patches.
- **Segmentation:** Detections are used as prompts for a zero-shot Segment Anything Model (SAM) to generate precise masks.
- **Mapping:** Outputs are georeferenced for landscape-scale analysis.
- **Interpretability:** Grad-CAM and confident calibration.

Detection Performance: Fast and Accurate

Table: Detection model performance comparison.

Model	GFLOPS ↓	Params (M) ↓	FPS ↑	Precision ↑	Recall ↑	AP_{50} ↑	AP_{75} ↑	mAP ↑
DINO	1920.3	218.2	18.98 ± 0.95	0.7629 ± 0.0177	0.8494 ± 0.0071	0.8169 ± 0.0166	0.5455 ± 0.0150	0.5102 ± 0.0101
DDQ	1232.6	218.6	19.18 ± 0.96	0.7825 ± 0.0124	$\mathbf{0.8566 \pm 0.0123}$	0.8541 ± 0.0129	0.6354 ± 0.0137	0.5736 ± 0.0130
RT-DETR	222.5	65.5	151.49 ± 0.70	$\mathbf{0.8869 \pm 0.0230}$	0.7598 ± 0.0310	0.8416 ± 0.0181	0.6198 ± 0.0181	0.5769 ± 0.0145
YOLOv8	226.7	61.6	174.92 ± 0.86	0.8729 ± 0.0165	0.7997 ± 0.0203	0.8667 ± 0.0141	0.6777 ± 0.0137	0.6148 ± 0.0128
YOLOv9	**169.5**	53.2	114.96 ± 0.30	0.8763 ± 0.0176	0.7976 ± 0.0209	$\mathbf{0.8741 \pm 0.0109}$	0.6762 ± 0.0146	0.6162 ± 0.0122
YOLOv10	169.8	**31.6**	$\mathbf{177.04 \pm 1.14}$	0.8716 ± 0.0121	0.7968 ± 0.0089	0.8626 ± 0.0129	$\mathbf{0.6794 \pm 0.0112}$	$\mathbf{0.6173 \pm 0.0090}$
YOLO11	194.4	56.8	170.40 ± 0.95	0.8721 ± 0.0095	0.7896 ± 0.0127	0.8684 ± 0.0108	0.6677 ± 0.0180	0.6115 ± 0.0109

Key Findings:

- **YOLOv10 (Selected):** Best overall trade-off, achieving the highest mAP, AP_{75} and inference speed with the fewest parameters.
- **DDQ:** Highest recall, ideal when finding all instances is prioritized.
- **RT-DETR:** Highest precision, but misses more palms (lower recall).

Zero-Shot Segmentation: Generalizing Across Ecosystems

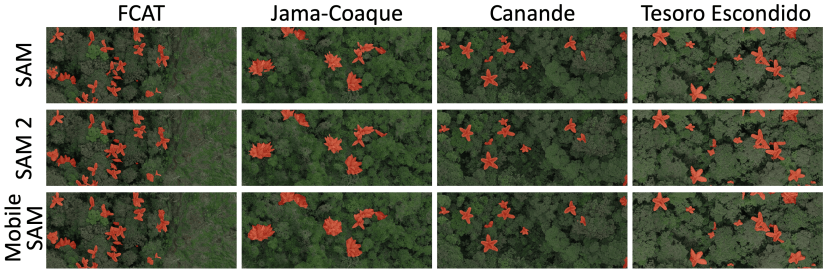
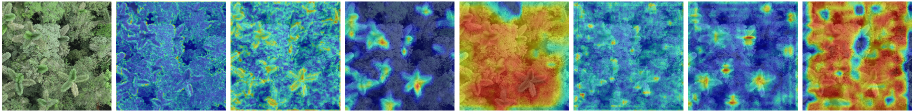


Figure: SAM 2 provides the most robust segmentation on unseen data.

We use boxes as prompts for **zero-shot segmentation** on new ecosystems.

- Original SAM: Sometimes produces incomplete masks.
- MobileSAM: Tends to include background areas (over-segments).
- **SAM 2 (Selected):** Most balanced and accurate segmentation.

Visualizing What the Model "Sees" with Grad-CAM



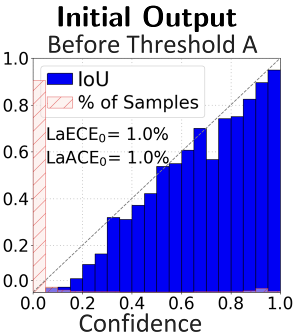
Input | Layer 1 | Layer 4 | Layer 7 | Layer 10 | Layer 16 | Layer 19 | Layer 22

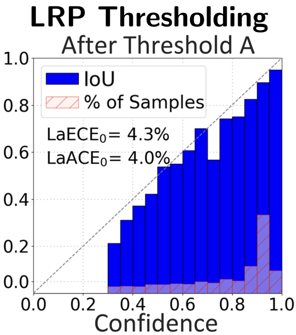
Figure: Hierarchical Feature Learning in YOLOv10 through Grad-CAM Plots.

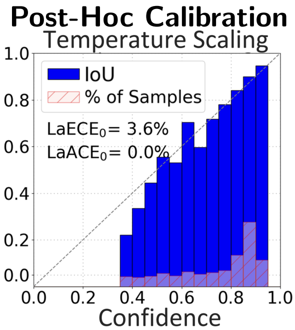
Hierarchical Feature Learning

The analysis confirms the model learns a meaningful progression: early layers focus on low-level edges and textures; intermediate layers integrate spatial context; and deep layers exhibit focused activation over entire palm crowns.

Model Interpretability: A Step-by-Step Guide to Calibration







The Problem

The uncalibrated model is unreliable; its confidence scores are poorly correlated with true accuracy.

The First Fix

LRP-based thresholding is applied to prune the large number of unreliable predictions with low confidence.

The Final Result

A post-hoc method, e.g., Temperature Scaling, is then applied to align confidence with accuracy.

Counting Performance: How Well Does It Generalize?

Table: Bidirectional counting performance across sites.

Site	Area (ha)	Counts	Pred2GT		GT2Pred	
			Ratio	Median (m)	Ratio	Median (m)
FCAT	21.62	471	0.9361	1.10	0.8854	0.77
Jama-Coaque	111.93	952	0.9348	1.50	0.8151	1.14
Canande	101.20	1,273	0.8956	0.82	0.7667	0.72
Tesoro Escondido	86.76	2,330	0.8981	1.09	0.9253	0.91

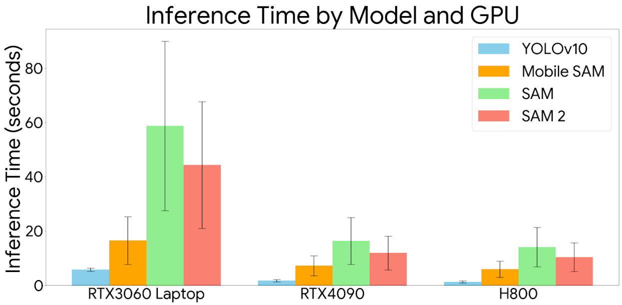
Key Metrics:

- **Pred2GT (Precision-like):** Proportion of our predictions that are correct.
- **GT2Pred (Recall-like):** Proportion of real palms that we found.

Summary

- High precision (90%) across all sites. PRISM rarely makes up palms.
- Recall is more variable (77-93%). Some sites are harder than others.
- Localization is excellent, with a median error of < 1.5 **meters**.

Feasibility for Real-Time Analysis



Key Findings:

- Detection is Real-Time Ready: YOLOv10 is fast enough (1.2–5.7s / image) for live processing on a UAV, even with mid-range hardware.
- Segmentation is Costly: Segmentation speed varies greatly, making it an optional step for time-critical missions.
- Conclusion: The core detection pipeline is efficient and stable, meeting the requirements for field deployment.

Conclusion and Future Work

Summary & Key Achievements

We introduced **PRISM**, a robust and efficient pipeline for detecting natural objects from UAV imagery, validated on our new, large-scale **PALMS** dataset. Key achievements include:

- High accuracy and strong generalization to new environments.
- Proven potential for real-time processing on UAVs.
- A trustworthy design that incorporates calibration and interpretability.

Future Work:

- Onboard deployment on UAVs for in-field validation.
- Adaptation to other ecologically critical species (e.g., pines).
- Application to lower-resolution satellite data for scalable monitoring.

Thank You

Questions?

Link to Code

Link to Data

Acknowledgments: We thank the IRSC Lab at WFU, Centro de Innovación Científica Amazónica, and Andrew Sabin Family Center for Environment and Sustainability. This research was supported by the US NSF BEE 2039850, and CityU Starting Grant 9380162.