

Counting Trees from Satellite Imagery with Noisy Supervision

Dimitri Gominski¹, Maurice Mugabowindekwe^{1,2}, Qiue Xu³, Xiaowei Tong³,
Martin Brandt¹, Hieu Le⁴, Rasmus Fensholt¹, Dimitris Samaras⁵, and Loic
Landrieu⁶

¹ University of Copenhagen

² University of Rwanda

³ University of Chinese Academy of Sciences

⁴ University of North Carolina at Charlotte

⁵ Stony Brook University

⁶ LIGM, CNRS, Univ Gustave Eiffel, ENPC, Institut Polytechnique de Paris, France

Abstract. Counting individual trees is a fundamental task for environmental monitoring, yet remains largely unexplored with satellite imagery. At these resolutions, isolated trees may still be identifiable, but crown boundaries become ambiguous in dense forests, making the notion of an individual tree inherently ill-defined. Moreover, large-scale manual annotations of individual trees are prohibitively expensive. While scalable supervision can be derived from airborne LiDAR, the resulting annotations are noisy and difficult to exploit effectively. We address these challenges by formulating tree counting as a spatial density matching problem supervised through Unbalanced Optimal Transport. This formulation naturally accommodates both precise localization of isolate trees and robust density estimation in dense forests. We further introduce a self-correction mechanism that leverages transport residuals to progressively refine noisy supervision during training. We evaluate our approach on TINYTREES, a new benchmark spanning three continents and three satellite sensors, comprising over 216 million tree annotations (including 639k manually verified instances) across 25 890 km². Our method consistently outperforms detection-based, regression-based, and transport-based distribution-matching baselines, demonstrating the effectiveness of unbalanced transport and reliability-aware supervision for large-scale tree counting from satellite imagery. Code, data and models are available at github.com/dgominski/treematch.

1 Introduction

Most large-scale tree monitoring efforts focus on forest cover [15], biomass [38], or canopy height [12, 35, 42], which provide aggregate metrics and do not capture tree-level structure. However, estimating the number and spatial distribution of individual trees is central to environmental monitoring and land management. Indeed, many key applications require tree-level analysis: including reforestation tracking [7], biomass and carbon estimation [20], biodiversity assessment [5], as

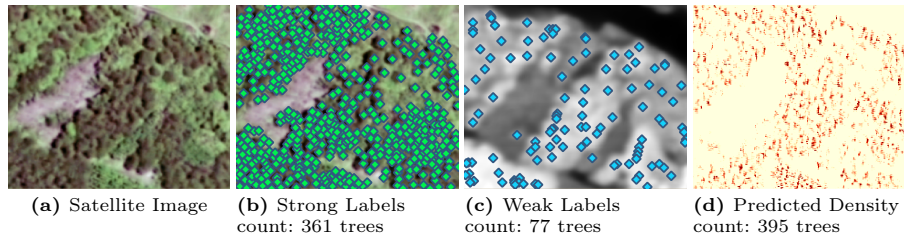


Fig. 1: Counting Trees From Space. We introduce TREEMATCH, a model to estimate the number of trees in a VHR satellite image (a). Training relies on two supervision sources: *strong labels* consisting of manually annotated tree centers (b), and *weak labels* automatically derived from airborne laser scanning (c). The model predicts a per-pixel tree density map (d), enabling precise localization of isolated trees while accommodating dense forest regions where individual boundaries are ambiguous.

well as the monitoring of anthropogenic tree structures such as hedgerows [33] and urban trees [4, 46].

Despite this, large-scale tree counting from satellite imagery remains largely unexplored, with existing datasets restricted to small-scale aerial imagery and narrow geographic coverage [8, 54], limiting both systematic evaluation and the development of dedicated methods.

Challenges. Counting trees from satellite imagery raises several unique technical challenges. Tree crowns vary widely in size, from less than 1 m^2 to several hundred square meters, and do not always admit a clear object center. They occur in heterogeneous spatial configurations, ranging from isolated individuals to dense forest canopies. At satellite resolution, individual crowns often span only a few pixels, and sensor blur can merge neighboring trees. As a result, delineation becomes ill-defined in dense regions while remaining feasible for isolated trees. Because real-world scenes combine separable instances, dense clusters, and intermediate regimes, tree counting is difficult to address with a single globally consistent formulation.

Precise manual annotations for tree counting are costly and typically limited in scale. However, large-scale supervision can be derived automatically from Airborne Laser Scanning (ALS), albeit with substantially lower reliability. Effective learning at the satellite scale therefore requires formulations that explicitly accommodate noisy and heterogeneous annotations.

Contributions. To address these challenges, we make three contributions. First, we introduce TINYTREES, the first large-scale benchmark for tree counting from satellite imagery. The dataset spans three continents and three satellite sensors, covers $25\,890 \text{ km}^2$, and contains more than 216 million tree annotations, including 639k manually verified instances. Second, we propose TREEMATCH, a density-matching framework based on unbalanced optimal transport. This formulation naturally handles heterogeneous counting regimes, enabling precise localization of isolated trees while remaining robust in dense canopies where in-

Table 1: Tree-Counting Datasets. TINYTREES substantially expands geographic coverage, annotation scale, and sensor diversity compared to existing datasets. GSD denotes Ground Sampling Distance (spatial resolution). † data not accessible.

	Location	sensor GSD (m)	area (km ²)	labels	
				manual	automatic
Yosemite Tree [8]	California	aerial (0.1m)	~10	~100k	-
† VHRTrees [43]	Turkey	Google earth (0.5m)	23	26k	-
KCL-London	London, UK	aerial (0.2m)	38	95k	-
† GF-II [54]	China	Gaofen2 (0.8m)	56	≥100k	-
DRIFT [22]	France	aerial, SkySat (0.2-0.5m)	62	-	526k
NEON Benchmark [51]	USA	aerial (0.1m)	≤100	12k	9.3M
TINYTREES (ours)	China, Africa, France	Gaofen2, Planet, SPOT6 (0.8-4.2m)	25.9k	639k	215M

stance boundaries are ambiguous. Third, we leverage marginal residuals of the unbalanced transport solution to automatically correct weak annotations during training. This mechanism enables effective learning from large-scale noisy supervision alongside sparse high-quality labels, providing a principled solution to heterogeneous annotation quality.

Results. Across three continents and three satellite sensors, TREEMATCH consistently outperforms detection-based, regression-based, and distribution-matching baselines. Our formulation demonstrates improved robustness to weak supervision and maintains strong performance across diverse forest types and imaging conditions.

2 Related Work

Object Counting. Object counting is a well-established problem in computer vision. A common approach converts object annotations into density maps by convolving point locations with Gaussian kernels and training models using pixel-wise regression losses [23, 34]. Subsequent work adapts kernel sizes based on local density [49]. Alternatively, detection-based methods localize instances via bounding boxes [25] or points [40]. While effective when object boundaries are distinct, detection-based approaches degrade under heavy overlap or clutter, where separability assumptions fail—conditions typical of dense forest canopies.

Tree Mapping. Recent advances in remote sensing have enabled large-scale tree analysis, including isolated tree detection [44], canopy height estimation [21, 42], and species mapping [31]. These efforts are supported by public datasets providing crown delineations, point annotations [45], LiDAR point clouds [17], and species labels [1, 13]. However, tree counting itself is typically treated as

a by-product of detection [6, 8], and no dedicated large-scale benchmark exists for satellite imagery. Moreover, trees in continuous canopy exhibit overlapping and irregular crowns [48], making precise delineation difficult. Although high-resolution data and specialized models can improve segmentation [11, 52], such approaches are costly and difficult to scale. Consequently, recent work increasingly relies on airborne LiDAR to derive large-scale supervision [12, 14]. Yet LiDAR-derived labels introduce structured and correlated errors, motivating noise-robust formulations.

Distribution Matching for Counting. Crowd counting faces similar challenges of density variation, overlap, and annotation uncertainty. This has motivated probabilistic formulations that replace strict pixel-wise regression with distribution-matching objectives [29, 50], improving robustness to spatial ambiguity. Closest to our work, Ma *et al.* [30] propose to use UOT for crowd counting. Despite these parallels, distribution-matching approaches have not been systematically explored for large-scale tree counting from satellite imagery.

3 Method

We first review unbalanced optimal transport (Sec. 3.1), then describe how it can be used for tree density regression (Sec. 3.2). We then present a method to bootstrap the results of marginal transport to correct weak annotations (Sec. 3.3). Finally, we present our approach to leverage both strong and weak tree annotations to train our model (Sec. 3.4).

3.1 Primer on Unbalanced Optimal Transport

We briefly review optimal transport and its unbalanced variant, focusing on the elements required for our formulation. Readers already familiar with optimal transport may skip this subsection.

Optimal Transport. Optimal Transport (OT) [47] measures the cost of transforming one probability distribution into another while accounting for the geometry of the underlying space. Let $\mu \in \mathbb{R}_+^n$ and $\nu \in \mathbb{R}_+^n$ be two discrete probability distributions satisfying $\mathbf{1}^\top \mu = \mathbf{1}^\top \nu = 1$, where $\mathbf{1}$ denotes a vector of ones. Let $C \in \mathbb{R}_+^{n \times n}$ be a cost matrix, where C_{ij} represents the cost of transporting a unit mass from location i to location j . OT seeks a transport plan γ that matches the marginals of μ and ν : $\Gamma(\mu, \nu) = \{\gamma \in \mathbb{R}_+^{n \times n} \mid \gamma \mathbf{1} = \mu, \gamma^\top \mathbf{1} = \nu\}$. The Monge–Kantorovich OT problem is then defined as

$$W_C^{\text{OT}}(\mu, \nu) = \min_{\gamma \in \Gamma(\mu, \nu)} \langle C, \gamma \rangle, \quad (1)$$

where $\langle C, \gamma \rangle = \sum_{i,j} C_{ij} \gamma_{ij}$. Intuitively, OT computes the minimum effort required to reshape one distribution into the other.

Entropy-Regularized Optimal Transport. Solving (1) exactly is computationally expensive. Cuturi *et al.* [10] introduce a negative entropy penalty, leading to the relaxed objective

$$W_{C,\varepsilon}^{\text{ROT}}(\mu, \nu) = \min_{\gamma \in \Gamma(\mu, \nu)} \langle C, \gamma \rangle + \varepsilon H(\gamma), \quad H(\gamma) = \sum_{i,j} \gamma_{ij} (\log \gamma_{ij} - 1), \quad (2)$$

where $\varepsilon > 0$ controls the smoothness of the transport plan. The solution of (2) can be efficiently approximated using Sinkhorn iterations. This procedure is fast, GPU-friendly, and differentiable, which makes entropy-regularized OT widely used as a loss in deep learning.

Unbalanced Optimal Transport. In many applications, including density regression, the two distributions may not have the same total mass. Enforcing strict marginal constraints is therefore unrealistic and can destabilize optimization. Unbalanced Optimal Transport (UOT) [9] relaxes these constraints by allowing mass creation or destruction at a controlled cost. Given two nonnegative measures μ and ν with arbitrary total mass, UOT solves

$$W_{C,\varepsilon,\tau}^{\text{UOT}}(\mu, \nu) = \min_{\gamma \geq 0} \langle C, \gamma \rangle + \varepsilon H(\gamma) + \tau \text{KL}(\gamma \mathbf{1} \parallel \mu) + \tau \text{KL}(\gamma^\top \mathbf{1} \parallel \nu). \quad (3)$$

where $\tau > 0$ controls the penalty for deviating from mass conservation with respect to the source distribution μ and the target distribution ν . and $\text{KL}(\cdot \parallel \cdot)$ denotes the generalized Kullback–Leibler divergence between nonnegative measures [9].

3.2 Learning to Count Trees with Optimal Transport

Optimal transport offers a natural way to compare spatial measures by jointly accounting for geometry and mass distribution. The unbalanced formulation is especially appropriate in our setting, as it decouples geometric alignment from exact mass conservation, thereby accommodating counting errors and annotation noise.

Setting. We consider an input image $x \in \mathbb{R}^{H \times W \times C}$ defined on a pixel grid of size $N = H \times W$. Ground-truth annotations are represented as a nonnegative density map $y \in \mathbb{R}_+^N$. Depending on the supervision source, y may either encode a dense tree-cover map or a sparse binary map, where $y_i = 1$ indicates the annotated center of a tree crown at pixel i . This representation defines a discrete measure over the image domain, rather than a probability distribution. The total number of annotated trees is given by $\sum_i y_i$.

We train a model ϕ that maps x to a nonnegative density map $z \in \mathbb{R}_+^N$, where z_i represents the predicted mass at pixel i . As for y , the prediction is not normalized. The predicted tree count is obtained by summation, $\sum_i z_i$.

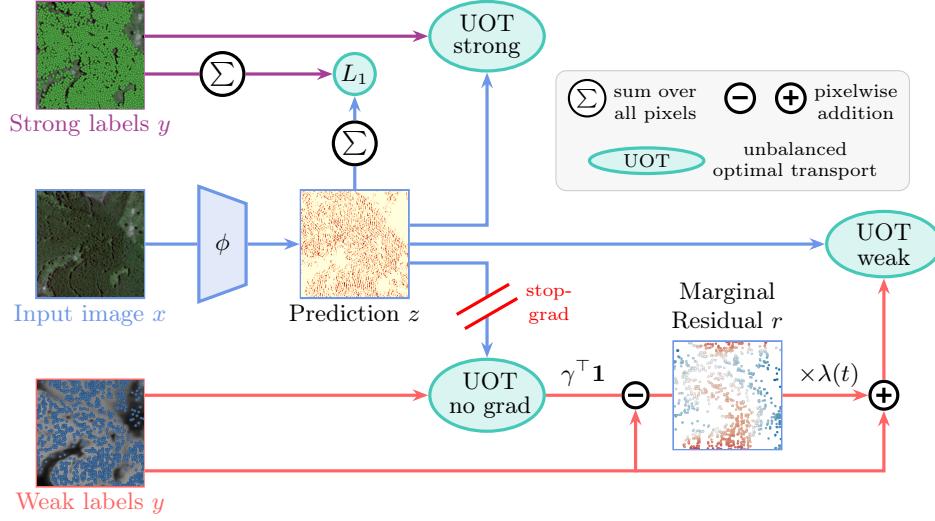


Fig. 2: TREEMATCH. Given a VHR satellite image, the network ϕ predicts a continuous tree density map. Training uses two supervision sources: — strong labels (manual point annotations) and — weak labels (automatically derived tree-cover masks). Strong labels supervise both a global count loss and a UOT-based alignment loss. For weak labels, we first solve a UOT problem (without backpropagation) to compute marginal residuals. These residuals are used to correct the weak annotations. A second UOT loss is then applied between the corrected labels and the prediction, from which gradients are propagated to update the model.

Count Supervision. A first supervision signal enforces consistency between predicted and ground-truth counts:

$$\mathcal{L}^{\text{count}}(z, y) = \left| \sum_i z_i - \sum_i y_i \right|. \quad (4)$$

While this loss enforces global count consistency, it provides no spatial supervision.

Density Matching with Unbalanced Optimal Transport. Directly comparing the predicted density z and the annotated measure y using a KL divergence is inappropriate in our setting. Indeed, KL heavily penalizes spatial misalignment: assigning mass to a nearby but unannotated pixel can incur arbitrarily large loss. In our setting, the precise location of tree centers is inherently ambiguous, especially in dense canopies. Instead, we align z and y using Optimal Transport, which compares spatial measures while explicitly accounting for geometric proximity. We define the cost matrix C using squared Euclidean

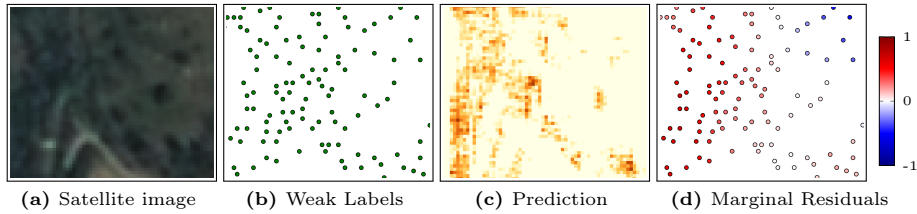


Fig. 3: Marginal Residuals. A satellite image (a) is annotated with weak labels (b) and processed by the model to produce a density prediction (c). The resulting marginal residuals (d) highlight disagreements between annotation and prediction, revealing potential missed trees (in red) and spurious annotations (in blue).

distances between pixel locations:

$$C_{ij} = \|\text{pos}(i) - \text{pos}(j)\|_2^2, \quad (5)$$

where $\text{pos}(i)$ denotes the 2D coordinates of pixel i . Transporting mass over long distances is penalized, encouraging the predicted density to concentrate near annotated locations while tolerating small localization shifts.

DM-Count [50] applies classical (balanced) OT to crowd counting by normalizing both the prediction and the ground truth into probability distributions. This normalization implicitly enforces equal total mass between the two measures. In our setting, however, annotations may be noisy or incomplete, and predicted counts may differ substantially from ground truth. We instead adopt *Unbalanced* Optimal Transport (UOT), which relaxes the mass conservation constraint. UOT allows mass to be locally created or removed at a controlled penalty, decoupling geometric alignment from exact mass equality. The loss is defined as

$$\mathcal{L}^{\text{UOT}} = W_{C, \varepsilon, \tau}^{\text{UOT}}(z, y), \quad (6)$$

where ε controls entropic regularization and τ governs tolerance to mass discrepancies. This formulation preserves the geometric alignment properties of OT while explicitly accommodating counting uncertainty.

3.3 Bootstrapping Marginal Residuals

Tree annotations derived from automated pipelines are inherently noisy, often containing spurious detections or missing trees in dense forest regions. These errors occur at the level of individual trees, making them difficult to address with a single global—or even image-level—relaxation parameter. Rather than introducing ad-hoc filtering rules, we exploit the local mass discrepancy that naturally arises when solving the Unbalanced Optimal Transport (UOT) problem.

Marginal Residuals from Transport. When solving UOT between the predicted density z and the annotated measure y , mass may be both discarded and created at a controlled cost. Let γ denote the optimal transport plan. The column marginals $\gamma^\top \mathbf{1}$ represent the amount of mass effectively matched to each annotation under the transport geometry. We define the *marginal residual* as

$$r = \gamma^\top \mathbf{1} - y. \quad (7)$$

As illustrated in Fig. 3, positive values indicate that the transport assigns more mass to a location than originally annotated, suggesting missed detections. Negative values indicate that annotated mass is not supported by the prediction, suggesting over-annotation.

Bootstrapped Target. We use this residual to progressively refine the supervision signal:

$$\tilde{y}(t) = y + \lambda(t) r, \quad (8)$$

where $\lambda(t) \in [0, \alpha]$ increases from 0 to $\alpha \in [0, 1]$ during training. Early in training, $\lambda(t) \approx 0$ and supervision relies on the raw annotations. As the model becomes better calibrated, $\lambda(t)$ increases and the target progressively incorporates transport-consistent corrections. Importantly, the transport plan γ is always computed with respect to the original annotations y , ensuring that the correction does not depend on previously reweighted targets.

3.4 Learning with Weak and Strong Tree Labels

Manually annotating individual trees at scale is prohibitively expensive. Instead, approximate tree-center or tree-cover annotations can be derived automatically from airborne laser scanning (ALS) data (see Sec. 4.1). While these weak annotations are informative, they become noisy in dense forests and may contain missed or spurious detections. In contrast, manual annotations are accurate but sparse. We therefore consider two training sets: a strong set $\mathcal{D}^{\text{strong}}$ with reliable spatial localization and counts, and a larger weak set $\mathcal{D}^{\text{weak}}$ with potentially inaccurate annotations.

Given a prediction $\phi(x)$, strong labels supervise both the global count and spatial alignment using the counting loss $\mathcal{L}^{\text{count}}$ and the UOT-based distribution matching loss $\mathcal{L}^{\text{UOT}}$. For weak labels, we first compute marginal residuals (Sec. 3.3) and use them to form corrected annotations $\tilde{y}$. This correction is computed without backpropagation. We then apply the UOT loss between the prediction and $\tilde{y}$, with backpropagation. The overall objective is

$$\mathcal{L} = \sum_{(x,y) \in \mathcal{D}^{\text{strong}}} \left(\mathcal{L}^{\text{count}}(\phi(x), y) + \mathcal{L}^{\text{UOT}}(\phi(x), y) \right) \quad (9)$$

$$+ \sum_{(x,y) \in \mathcal{D}^{\text{weak}}} \mathcal{L}^{\text{UOT}}(\phi(x), \text{stop_grad}(\tilde{y})), \quad (10)$$

Table 2: Overview of the TINYTREES datasets. Dataset statistics by region and split. Weak annotations are obtained from height/crown predictions derived from aerial LiDAR scans, while strong labels are manual annotations obtained via photointerpretation ($\odot$) or field surveys (A). $\star$: use medium density LiDAR, likely undercounted.

Location	Sensor /GSD	Split	area (km ²)	#trees	label source
China	Gaofen-2 0.8m	Train-weak	50.3	7.68M $\star$	height reg.
		Train-strong	5.9	54.6k	manual $\odot$
		Test	7.02	58.4k	manual $\odot$
Rwanda	PlanetScope 3.4-4.2 m	Train-weak	209.3	3.40M	crown seg.
		Train-strong	48.4	309k	manual $\odot$
		Test	33.9	195k	manual $\odot$
France	Spot6 1.5m	Train-weak	23.3k	204M	high-res ALS
		Train-strong	0.35	11.1k	manual A
		Test	0.35	10.8k	manual A

where `stop_grad` denotes a stop-gradient operator preventing gradients from flowing through the correction step. This formulation enables large-scale learning from noisy supervision while preserving reliable count and spatial constraints where clean annotations are available. Although weak labels are typically far more abundant than strong ones, we sample $\mathcal{D}^{\text{weak}}$ and $\mathcal{D}^{\text{strong}}$ to have comparable sizes. This ensures accurate count supervision from strong annotations while stabilizing training and preventing circular feedback from the correction mechanism applied to noisy labels.

Implementation Details. The transport plan smoothness ε is set to 0.005, inducing an effective interaction radius of approximately $\sqrt{\varepsilon} \approx 0.07$ in normalized coordinates. The mass deviation penalty τ is set between 0.1 and 0.5 depending on the dataset, balancing the cost of creating unit mass against the cost of transporting unit mass across one unit spatial distance (in normalized coordinates). Smaller values of τ reduce mass flexibility, while larger values allow moderate shifts in density. The scheduler $\lambda(t)$ increases from 0 to $\alpha = 0.8$ over the first 400 optimization steps using a tempered sigmoid schedule. For all data splits, we use AdamW [26] with default parameters. We train for 500 epochs with a cosine annealing learning rate scheduler with an initial value of 8e-5. Weak and strong training samples are drawn in stratified 1:1 ratio.

4 Experiments

We first introduce the TINYTREES datasets (Sec. 4.1), and present our baselines and metrics (Sec. 4.2). We then detail quantitative results and analysis (Sec. 4.3), followed by an ablation study (Sec. 4.4).

4.1 The TINYTREES Dataset

We introduce TINYTREES, the first large-scale dataset for tree counting from satellite imagery. The dataset spans three continents and three satellite sensors with ground sampling distances (GSD) ranging from 0.8 m to 4.2 m. In total, it covers over 25k km² and contains more than 216 million tree annotations. An overview is provided in Tab. 2.

Areas of Interest. TINYTREES covers three geographically and ecologically distinct regions.

- **China.** Very-high-resolution Gaofen-2 imagery (0.8 m GSD) covering temperate forests. Manual annotations were produced through expert photo-interpretation, supported by higher-resolution reference imagery. Canopy height maps from supervised regression trained with medium-density ALS are available ($\sim 2pt/m^2$).
- **Rwanda.** PlanetScope imagery (3–4 m GSD) covering heterogeneous tropical landscapes. Manual annotations were created through expert photo-interpretation. Large-scale tree crown predictions derived from aerial imagery are available.
- **France.** SPOT-6 imagery (1.5 m GSD, pansharpened from 6 m) covering temperate forests. We have access to high-density LiDAR-HD ALS data ($\geq 20pt/m^2$) to derive canopy height maps [17], and forest inventory data on 15 m radius plots from the French National Forest Office are available.

Annotation Types. Each region contains two forms of supervision: strong and weak. *Strong annotations* are obtained through manual photo-interpretation or *in situ* field measurements, providing high precision but limited geographic coverage.

Weak annotations are automatically derived at large scale but are inherently noisy, particularly in dense canopy where tree separation is imperfect. They are obtained from dedicated canopy height models or large-scale crown prediction pipelines. For France, weak labels are generated from ALS-derived canopy height maps using a fully automated peak detection procedure [14]. In Rwanda, weak labels originate from a semi-automatic national-scale crown segmentation product [32]. For the China split, we derive pseudo-density targets from canopy height maps: canopy pixels above a 3 m threshold are assigned the average tree density estimated from the strongly annotated subset.

Dataset Splits. All images are extracted as 64×64 RGB+NIR patches at nominal satellite resolution. Each region is partitioned into three subsets:

- **Train-weak:** large geographic areas with automatic annotations,
- **Train-strong:** smaller areas with manual annotations,
- **Test:** held-out areas with manual annotations only.

To prevent spatial leakage, a 1 km buffer separates training and test regions. Because manual annotations cover non-square geometries, a small fraction ($< 10\%$) of annotated trees may appear in overlapping patches.

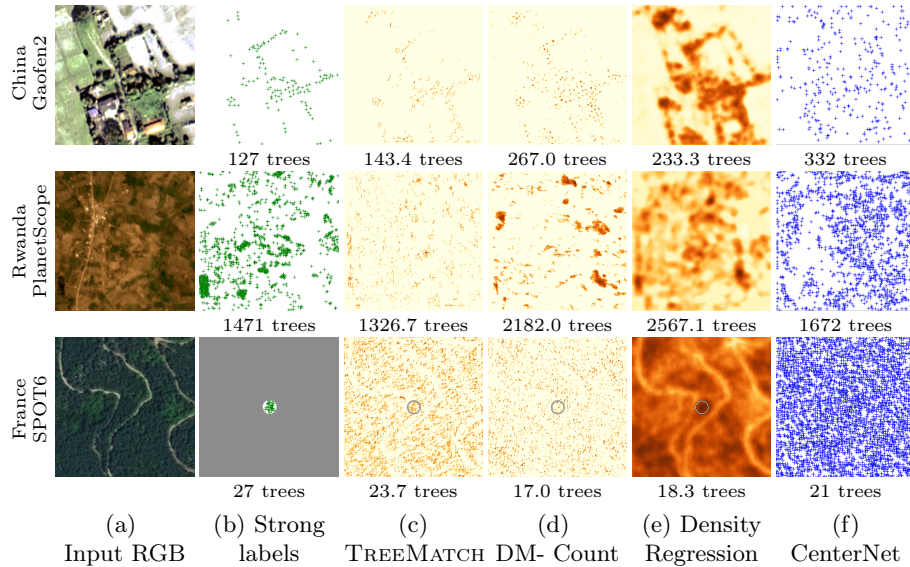


Fig. 4: Qualitative results across regions and sensors. Each row corresponds to a different dataset: China (Gaofen2), Rwanda (PlanetScope), and France (SPOT6). In the France split, ground-truth annotations are only available on 15m radius inventory plots. From left to right: input RGB image (a), strong manual annotations (b), TREEMATCH’s predictions (c), DM-Count (c), density regression baseline (d), and CenterNet detector (f). TREEMATCH produces spatially coherent density estimates across both isolated and dense tree regimes, while baselines either oversmooth dense regions or miss trees in cluttered areas.

4.2 Baselines and Metrics

Baselines and Competing Methods. Following common practice in forest analysis, all models use a UNet architecture [37] with a ResNet50 encoder [16]. We chose this design for its stability and ease of training, and evaluate other backbones in the ablation. We evaluate the following training strategies:

1. **Density Regression.** Tree annotations are converted into density maps using fixed-width Gaussian kernels, and the model is trained with an ℓ_2 loss [36, 39, 49].
2. **Count Regression.** Global sum pooling is applied to the 1-channel output of UNet to directly regress the total tree count for each image.
3. **Detection.** We evaluate both box-based and point-based detectors, including YOLOv8 [18], P2P [40], TreeFormer [2], and CenterNet [56].
4. **Distribution Matching.** We evaluate VRSNet [28], and DM-Count [50], a crowd-counting method based on balanced optimal transport.
5. **Bayesian Approach.** We evaluate a Bayesian approach designed for crowd counting [29]. In addition, we reimplemented an uncertainty-aware dense

Table 3: Quantitative Evaluation on TINYTREES. We compare regression baselines (top), detection-based vision models (middle), and distribution-matching methods (bottom). We indicate the type of supervision used during training for the best configuration: strong manual annotations and/or weak ALS-derived labels. TREEMATCH consistently outperforms competing approaches.

- *did not converge* n/a *official implementation not compatible with partial annotation*

Method	strong weak	China / Gaofen2			Rwanda / PlanetScope			France / Spot6		
		RMSE ↓	R^2 ↑	nMAE ↓	RMSE ↓	R^2 ↑	nMAE ↓	RMSE ↓	R^2 ↑	nMAE ↓
Density regression	✓	64.6	0.55	37.4	79.3	0.36	54.0	163.4	0.20	40.2
Count regression	✓✓	67.9	0.50	43.7	75.8	0.42	53.8	149.2	0.33	39.1
Uncertainty-aware reg.	✓	63.3	0.57	37.9	80.5	0.34	55.3	163.0	0.20	40.6
YOLOv8 [18]	✓	-	-	-	448.7	-0.50	85.0	n/a	n/a	n/a
CenterNet [56]	✓✓	82.9	0.26	51.8	98.9	0.01	69.8	181.5	0.01	45.6
P2P [40]	✓✓✓	98.7	-0.05	60.4	-	-	-	177.2	0.06	45.6
TreeFormer [2]	✓	68.7	0.49	41.2	86.0	0.25	59.8	179.7	0.03	45.1
Bayes. Crowd Count. [29]	✓	168.8	-2.06	134.1	80.5	0.34	60.9	162.0	0.21	43.7
VRSNet [27]	✓	67.5	0.51	41.2	81.7	0.32	56.9	165.1	0.18	40.3
DM-Count [50]	✓✓	65.0	0.55	39.6	76.8	0.40	53.4	158.0	0.25	40.9
TREEMATCH (ours)	✓✓	60.6	0.60	36.6	72.4	0.47	51.1	147.2	0.35	37.4

regression baseline [21], in which the model predicts both the mean and variance of the density at each pixel.

Evaluation Metrics. We assess counting accuracy using the following metrics:

1. **RMSE.** Image-level root mean squared error on tree counts, in *trees per hectare*.
2. **nMAE.** Dataset-level normalized mean absolute error.
3. **R^2 .** Coefficient of determination between predicted and ground-truth counts, computed at the image-level.

4.3 Results and Analysis

We report the performance of all methods across all splits in Tab. 3. Overall, the task has proven to be challenging, with R^2 values remaining below 0.6 for all models and regions. This contrasts with related tasks such as crowd counting, where instance-level metrics can reach substantially higher accuracy. Across all datasets and evaluation metrics, TREEMATCH achieves the best overall performance. Improvements are consistent in terms of RMSE, R^2 , and normalized MAE, indicating robustness across continents, biomes, and imaging conditions. Qualitative results are shown in Fig. 4.

Performance Analysis. We make several observations:

- **Regression baselines.** Simple regression approaches provide surprisingly strong baselines. In particular, image-level count regression performs competitively, highlighting the importance of global count supervision. Density

regression also performs well, although its optimal Gaussian std of 4px leads to blurry outputs (see Fig. 4f). However, these methods are not well-suited to heterogeneous annotation quality. In particular, incorporating weak supervision degrades the performance of density regression.

- **Detection-based methods.** Detection models such as YOLOv8 and CenterNet perform poorly in this setting. They struggle particularly in dense forest regions where tree boundaries are ambiguous and individual crowns are not clearly separable (see Fig. 4f). In several cases, these models either fail to converge or produce unstable predictions, confirming that classical instance detection is ill-suited to satellite-based tree counting across heterogeneous forest regimes.
- **Distribution-matching methods.** DM-Count [50] performs relatively well, confirming the relevance of transport-based formulations for counting tasks. However, its reliance on balanced OT and TV losses lead to over-confident activations (see Fig. 4f) and limit robustness under partial or noisy annotations. Our unbalanced formulation is blurrier but generalizes better, consistently improving upon DM-Count across all datasets.

Efficiency. Our model is computationally efficient at inference, requiring only a single forward pass. It processes approximately 99M pixels/s, corresponding to about 887 km²/s for PlanetScope imagery. Training is also fast and stable: a full training run on a single data split completes in under one hour on a single NVIDIA 3090 GPU. These characteristics suggest that the approach is well-suited for large-scale deployment.

Spatial Generalization. In the previous experiments, we trained one model per region–sensor pair. An important question is whether a single model can generalize across biomes and satellite platforms. Table 4 reports cross-region evaluation results, as well as the performance of a jointly trained global model. We observe that some regions are substantially easier to learn from than others. In particular, the model trained on the China split achieves the best overall performance. This can likely be attributed to the higher spatial resolution of Gaofen imagery and the quality of the photo-interpreted strong labels. In contrast, the model trained on the Rwanda split generalizes poorly, likely due to the specificity of its tropical biome and the calibration of its sensor.

While a model jointly trained on all splits performs slightly worse than specialized models on their training split, it still achieves strong overall results. This suggests that cross-region generalization is feasible and could be further improved by leveraging sensor-agnostic architectures [3, 53] and larger, more diverse training datasets.

4.4 Ablation Study and Perspectives

We conduct an ablation study to quantify the contribution of each component of our framework. Results are reported in Table 5.

Table 4: Cross-Region Spatial Generalization.

Each row corresponds to a model trained on a specific region (or jointly on all regions), and each column reports performance (nMAE) on a target region.

Training	Evaluation			
	China	Rwanda	France	Overall
China	36.6	99.9	76.9	84.8
Rwanda	1066	51.1	180	287
France	154	99.9	37.4	111
Joint Training	42.4	51.7	42.9	49.3

- A) **Role of supervision sources.** Removing weak, ALS- or model-derived supervision (*No Train-weak*) increases RMSE from 72.4 to 76.1, showing that large-scale noisy labels provide a valuable complementary training signal. Removing strong, manual supervision (*No Train-strong*) leads to a larger degradation, with RMSE increasing to 82.4, confirming the importance of accurate counts for calibrating the model. Uniform sampling slightly improves performance on the Rwanda split (71.0 RMSE), although the effect was not consistent on other splits due to varying strong/weak ratios.
- B) **Loss formulation.** Training with count supervision only ($\mathcal{L}_{\text{count}}$ *only*) increases RMSE to 75.5, highlighting the importance of spatial distribution matching. Replacing unbalanced OT with balanced OT further degrades performance (78.3 RMSE), confirming that explicitly tolerating mass mismatch is critical under weak supervision. Removing residual correction (*No residuals*) increases RMSE to 79.0, demonstrating the benefit of our correction mechanism for noisy annotations.
- C) **Choice of backbone.** Using a ViT-S backbone yields performance nearly identical to the baseline (72.5 RMSE), indicating limited sensitivity to moderate capacity reduction. A hierarchical Swin-S backbone slightly improves performance (71.0 RMSE), consistent with prior findings that hierarchical encoders are well-suited to forest imagery analysis [12].

Impact of Annotation Type. For the China split, weak supervision is provided as continuous forest cover maps with constant density rather than pointwise annotations, due to the low resolution of the available canopy height data. For comparison, we derived pointwise weak annotations using the local peak detection procedure employed for France. This alternative leads to a slight performance decrease (nMAE = 39.6 vs. 36.6). These results indicate that our model can accommodate both pointwise and continuous weak labels, and that adapting the annotation form to its corresponding modality improves performance.

Limitations and Perspectives. Optimal transport losses incur a quadratic memory complexity $O(N^2)$ with respect to the number of pixels $N = H \times W$, which limits scaling to larger images with broader spatial context. Sinkhorn-based implementations alleviate part of this issue through efficient computation, but further improvements could be achieved through more memory-efficient optimal transport algorithms [41].

Table 5: Ablation Study. Effects of supervision strategy (A), loss formulation (B), and backbone architecture (C) on performance. All results are reported for the Rwanda split.

	RMSE ↓	R2 ↑	nMAE ↓
Used config.	72.4	0.47	51.1
A No Train-weak	76.1	0.41	52.9
A No Train-strong	82.4	0.31	57.0
A Uniform sampling	71.4	0.48	50.8
B $\mathcal{L}_{\text{count}}$ only	75.5	0.42	53.5
B Balanced OT	78.3	0.38	53.6
B No residuals	79.0	0.37	55.0
C ViT-S backbone	72.5	0.47	51.6
C SWIN-S backbone	71.0	0.49	50.6

Tree counting and canopy height estimation are closely related, suggesting strong potential for joint modeling: counts constrain stand density, while height maps capture vertical structure. These signals are linked through well-established ecological relationships, such as allometric scaling between tree height and volume [19] and self-thinning dynamics relating stand volume and density [55]. Combining the two could yield more consistent estimates of forest structure and biomass, and the growing availability of large-scale canopy height products [12, 24, 35] makes such modeling increasingly feasible at continental scales.

5 Conclusion

In this work, we addressed counting individual trees from satellite imagery under intrinsic ambiguity and heterogeneous supervision. We proposed an approach based on unbalanced optimal transport that naturally accommodates localization ambiguity, uncertain total counts, and annotations of varying quality. Experiments on our TINYTREES benchmark, spanning three continents and satellite sensors, show consistent improvements over detection-based, regression-based, and distribution-matching baselines. These results suggest that unbalanced optimal transport offers a principled and flexible foundation for large-scale tree counting under real-world annotation constraints.

Acknowledgement

This work was funded by the National Key Research and Development Program of China (grant agreement no. 2023YFF1305700) and the Novo Nordisk Foundation (grant number NNF21OC0069116). This work was supported by the ANR project SHARP ANR-23-PEIA-0008 in the context of the PEPR IA and Hi! PARIS (ANR-23-IACL-0005), the Danish National Research Foundation through the Center for Remote Sensing and Deep Learning of Global Tree Resources (TreeSense, DNRF192), and the U.S. National Science Foundation (award IIS-2212046). We thank the French Forestry Office (ONF) for providing the forest inventory data used in the France split.

References

1. Ahlswede, S., Schulz, C., Gava, C., Helber, P., Bischke, B., Förster, M., Arias, F., Hees, J., Demir, B., Kleinschmit, B.: TreeSatAI Benchmark Archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing. *Earth System Science Data* (2023)
2. Amirkolaee, H.A., Shi, M., Mulligan, M.: TreeFormer: A Semi-Supervised Transformer-Based Framework for Tree Counting From a Single High-Resolution Image. *TGRS* (2023)
3. Astruc, G., Gonthier, N., Mallet, C., Landrieu, L.: AnySat: An earth observation model for any resolutions, scales, and modalities. In: *CVPR* (2025)
4. Beery, S., Wu, G., Edwards, T., Pavetic, F., Majewski, B., Mukherjee, S., Chan, S., Morgan, J., Rathod, V., Huang, J.: The Auto Arborist Dataset: A Large-Scale Benchmark for Multiview Urban Forest Monitoring Under Domain Shift. In: *CVPR* (2022)
5. Beloiu, M., Heinzmann, L., Rehush, N., Gessler, A., Griess, V.C.: Individual Tree-Crown Detection and Species Identification in Heterogeneous Forests Using Aerial RGB Imagery and Deep Learning. *Remote Sensing* (2023)
6. Brandt, M., Tucker, C.J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L.V., Hiernaux, P., Diouf, A.A., Kergoat, L., Mertz, O., Igel, C., Gieseke, F., Schöning, J., Li, S., Melocik, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Montagu, A., Dendoncker, M., Fensholt, R.: An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* (2020)
7. Cao, S., Sun, G., Zhang, Z., Chen, L., Feng, Q., Fu, B., McNulty, S., Shankman, D., Tang, J., Wang, Y., et al.: Greening china naturally. *Ambio* (2011)
8. Chen, G., Shang, Y.: Transformer for Tree Counting in Aerial Images. *Remote Sensing* (2022)
9. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.: Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation* (2018)
10. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: *NeurIPS* (2013)
11. Firoze, A., Wingren, C., Yeh, R.A., Benes, B., Aliaga, D.: Tree Instance Segmentation With Temporal Contour Graph. In: *CVPR* (2023)
12. Fogel, F., Perron, Y., Besic, N., Saint-André, L., Pellissier-Tanon, A., Schwartz, M., Boudras, T., Fayad, I., d’Aspremont, A., Landrieu, L., Ciais, P.: Open-Canopy: Towards Very High Resolution Forest Monitoring. In: *CVPR* (2025)
13. Gaydon, C., Roche, F.: PureForest: A Large-Scale Aerial Lidar and Aerial Imagery Dataset for Tree Species Classification in Monospecific Forests. arxiv.org/abs/2404.12064 (2024)
14. Gominski, D., Brandt, M., Tong, X., Liu, S., Mugabowindekwe, M., Li, S., Reiner, F., Davies, A., Fensholt, R.: Trees as Gaussians: Large-Scale Individual Tree Mapping. *Journal of Remote Sensing* (2026)
15. Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., et al.: High-resolution global maps of 21st-century forest cover change. *Science* (2013)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR* (2016)
17. IGN: LIDAR HD, <https://geoservices.ign.fr/lidarhd>
18. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8 (2023), <https://github.com/ultralytics/ultralytics>

19. Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D.A.: Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Global Change Biology* (2017)
20. Kankare, V., Holopainen, M., Vastaranta, M., Puttonen, E., Yu, X., Hyypä, J., Vaaja, M., Hyypä, H., Alho, P.: Individual tree biomass estimation using terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* (2013)
21. Lang, N., Jetz, W., Schindler, K., Wegner, J.D.: A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution* (2023)
22. Li, S., Gominski, D., Brandt, M., Tong, X., Ciais, P.: Get Your Embedding Space in Order: Domain-Adaptive Regression for Forest Monitoring. In: *ECCV* (2024)
23. Lian, D., Li, J., Zheng, J., Luo, W., Gao, S.: Density Map Regression Guided Detection Network for RGB-D Crowd Counting and Localization. In: *CVPR* (2019)
24. Liu, S., Brandt, M., Nord-Larsen, T., Chave, J., Reiner, F., Lang, N., Tong, X., Ciais, P., Igel, C., Pascual, A., Guerra-Hernandez, J., Li, S., Mugabowindekwe, M., Saatchi, S., Yue, Y., Chen, Z., Fensholt, R.: The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. *Science Advances* (2023)
25. Liu, Y., Shi, M., Zhao, Q., Wang, X.: Point in, Box Out: Beyond Counting Persons in Crowds. In: *CVPR* (2019)
26. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. In: *ICLR* (2019)
27. Luo, T., Gao, W., Belotserkovsky, A., Nedzved, A., Deng, W., Ye, Q., Fu, L., Chen, Q., Ma, W., Xu, S.: VrsNet - density map prediction network for individual tree detection and counting from UAV images. *JAG* (2024)
28. Luo, T., Gao, W., Belotserkovsky, A., Nedzved, A., Deng, W., Ye, Q., Fu, L., Chen, Q., Ma, W., Xu, S.: VrsNet-density map prediction network for individual tree detection and counting from UAV images. *International Journal of Applied Earth Observation and Geoinformation* **131**, 103923 (2024)
29. Ma, Z., Wei, X., Hong, X., Gong, Y.: Bayesian Loss for Crowd Count Estimation With Point Supervision. In: *ICCV* (2019)
30. Ma, Z., Wei, X., Hong, X., Lin, H., Qiu, Y., Gong, Y.: Learning to count via unbalanced optimal transport. In: *AAAI* (2021)
31. Mu, Y., Guo, J., Shahzad, M., Zhu, X.X.: National-scale tree species mapping with deep learning reveals forest management insights in Germany. *International Journal of Applied Earth Observation and Geoinformation* (2025)
32. Mugabowindekwe, M., Brandt, M., Chave, J., Reiner, F., Skole, D.L., Kariryaa, A., Igel, C., Hiernaux, P., Ciais, P., Mertz, O., Tong, X., Li, S., Rwanyiziri, G., Dushimiyimana, T., Ndoli, A., Uwizeyimana, V., Lillesø, J.P.B., Gieseke, F., Tucker, C.J., Saatchi, S., Fensholt, R.: Nation-wide mapping of tree-level above-ground carbon stocks in Rwanda. *Nature Climate Change* (2023)
33. Muro, J., Blickensdörfer, L., Don, A., Köber, A., Asam, S., Schwieder, M., Erasmí, S.: Hedgerow mapping with high resolution satellite imagery to support policy initiatives at national level. *Remote Sensing of Environment* (2025)
34. Oñoro-Rubio, D., López-Sastre, R.J.: Towards Perspective-Free Object Counting with Deep Learning. In: *ECCV* (2016)

35. Pauls, J., Zimmer, M., Kelly, U.M., Schwartz, M., Saatchi, S., Ciais, P., Pokutta, S., Brandt, M., Gieseke, F.: Estimating Canopy Height at Scale. In: ICML (2024)
36. Ranjan, V., Le, H., Hoai, M.: Iterative Crowd Counting. In: ECCV (2018)
37. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
38. Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabile, V., Araza, A., De Bruin, S., Herold, M., Quegan, S., Rodríguez-Veiga, P., et al.: The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data* (2021)
39. Shi, Z., Mettes, P., Snoek, C.: Counting With Focus for Free. In: ICCV (2019)
40. Song, Q., Wang, C., Jiang, Z., Wang, Y., Tai, Y., Wang, C., Li, J., Huang, F., Wu, Y.: Rethinking Counting and Localization in Crowds: A Purely Point-Based Framework. In: ICCV (2021)
41. Sun, C., Hu, J., Jiang, H.: MAP-UOT: A memory-efficient approach to unbalanced optimal transport implementation. *arXiv:2412.11079* (2024)
42. Tolan, J., Yang, H.I., Nosarzewski, B., Couairon, G., Vo, H.V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., et al.: Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial LiDAR. *Remote Sensing of Environment* (2024)
43. Topg  il, S.N., Sertel, E., Aksoy, S.,   nsalan, C., Fransson, J.E.: VHRTrees: A new benchmark dataset for tree detection in satellite imagery and performance evaluation with YOLO-based models. *Frontiers in Forests and Global Change* (2025)
44. Tucker, C., Brandt, M., Hiernaux, P., Kariryaa, A., Rasmussen, K., Small, J., Igel, C., Reiner, F., Melocik, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Fitts, Y., Morin, A., Pinzon, J., McClain, D., Morin, P., Porter, C., Loeffler, S., Kergoat, L., Issoufou, B.A., Savadogo, P., Wigneron, J.P., Poulter, B., Ciais, P., Kaufmann, R., Myneni, R., Saatchi, S., Fensholt, R.: Sub-continental-scale carbon stocks of individual trees in African drylands. *Nature* (2023)
45. Veitch-Michaelis, J., Cottam, A., Schweizer, D., Broadbent, E., Dao, D., Zhang, C., Zambrano, A.A., Max, S.: OAM-TCD: A globally diverse dataset of high-resolution tree cover maps. In: *NeurIPS Datasets and Benchmarks Track* (2024)
46. Ventura, J., Pawlak, C., Honsberger, M., Gonsalves, C., Rice, J., Love, N.L.R., Han, S., Nguyen, V., Sugano, K., Doremus, J., Fricker, G.A., Yost, J., Ritter, M.: Individual tree detection in large-scale urban environments using high-resolution multispectral imagery. *International Journal of Applied Earth Observation and Geoinformation* (2024)
47. Villani, C.: Optimal Transport: Old and New. *Grundlehren der mathematischen Wissenschaften*, Springer (2009)
48. Voulgaris, G.: Bridging Classical and Modern Computer Vision: PerceptiveNet for Tree Crown Semantic Segmentation. In: *CVPR Workshops* (2025)
49. Wan, J., Chan, A.: Adaptive Density Map Generation for Crowd Counting. In: *ICCV* (2019)
50. Wang, B., Liu, H., Samaras, D., Hoai, M.: Distribution Matching for Crowd Counting. In: *NeurIPS* (2020)
51. Weinstein, B.G., Graves, S.J., Marconi, S., Singh, A., Zare, A., Stewart, D., Bohlman, S.A., White, E.P.: A benchmark dataset for canopy crown detection and delineation in co-registered airborne RGB, LiDAR and hyperspectral imagery from the National Ecological Observation Network. *PLOS Computational Biology* (2021)

52. Wielgosz, M., Puliti, S., Xiang, B., Schindler, K., Astrup, R.: SegmentAnyTree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. *Remote Sensing of Environment* (2024)
53. Xiong, Z., Wang, Y., Zhang, F., Stewart, A.J., Hanna, J., Borth, D., Papoutsis, I., Saux, B.L., Camps-Valls, G., Zhu, X.X.: Neural plasticity-inspired foundation model for observing the Earth crossing modalities. *arXiv:2403.15356* (2024)
54. Yao, L., Liu, T., Qin, J., Lu, N., Zhou, C.: Tree counting with high spatial-resolution satellite imagery based on deep neural networks. *Ecological Indicators* (2021)
55. Yoda, K., Kira, T., Ogawa, H., Hozumi, K.: Self-Thinning in Overcrowded Pure Stands under Cultivated and Natural Conditions. *Journal of Biology, Osaka City University* **14**, 107–129 (1963)
56. Zhou, X., Wang, D., Krähenbühl, P.: Objects as Points. *arxiv.org/abs/1904.07850* (2019)