TITLE: Best practices and challenges for urban tree detection, classification, and geolocation with street-level images across North American cities

RUNNING TITLE: Urban tree detection, classification, and geolocation.

AUTHORS: Thomas A. Lake[a, † *], Brit B. Laginhas[a, †], Brennen T. Farrell[a], Ross K. Meentemeyer[a], and Chris M. Jones[a*]

[a]North Carolina State University, Center for Geospatial Analytics, 5112 Jordan Hall, 2800 Faucette Dr, Raleigh, NC 27695

*Corresponding authors at: North Carolina State University, Center for Geospatial Analytics, 5112 Jordan Hall, 2800 Faucette Dr, Raleigh, NC 27695

Email addresses: talake2@ncsu.edu , cmjone25@ncsu.edu

[†]Co-first authors

ABSTRACT

Accurate and up-to-date catalogs of urban tree populations are crucial for quantifying ecosystem services and enhancing the quality of life in cities. However, identifying and mapping tree species cost-effectively remains a significant challenge. Remote sensing is an active area of research where scientists are developing general-purpose tools that can capture information about plant populations across broad spatial scales. In this study, we developed computer vision models to detect, classify, and map 100 tree genera across 23 cities in North America using Google Street View (GSV) and iNaturalist images. We validated our predictions in independent portions of each city. We then compared our predictions to existing street tree records to evaluate the spatial context of errors using generalized linear mixed-effects models. Overall, the computer vision models identified the majority of independent, ground-truthed street trees (67.1%), and performance varied across the 23 cities (67.4% ± 9.3%) and 100 genera (50.9% ± 23.0%). Matching success increased with local street-view image density, simpler local stand structure, and greater training representation, particularly from the focal city. We found that genus classification performed better in continental cities with lower relative diversity, and that seasonal changes in the appearance of trees provided visual cues that moderate classification rates. Our study demonstrates that using widely available street-level imagery is a generalizable and promising avenue for mapping tree genus distributions across urban environments.

KEYWORDS: deep learning, species distributions, convolutional neural network, phenology, urban forestry, street-level images

1. Introduction

Urban forests, comprising all trees in urban settings, influence the livability of cities and the quality of life in urban areas (Edgar et al., 2021; Nowak and Dwyer, 2007; Roy et al., 2012). These forests provide ecosystem services, including climate regulation, stormwater management, and pollution sequestration, which are valued at over $18 billion annually in the United States (Escobedo et al., 2011; Nowak and Dwyer, 2007; Nowak and Greenfield, 2018). Understanding the distribution, species composition, and health status of individual trees is important for quantifying and managing these benefits (Ma et al., 2021; McPherson et al., 2016; Nowak and Dwyer, 2007; Nowak et al., 2016). Traditionally, field surveys are used to document forest conditions (Edgar et al., 2021; Nielsen et al., 2014; Nowak et al., 2008), but collecting data on individual trees is labor-intensive and quickly becomes outdated due to disturbances from urban development and pest or disease spread (Pretzsch et al., 2017; Raum et al., 2023; Tubby and Webber, 2010). As such, there is a growing need for efficient and scalable alternatives to monitor urban forests.

Tree inventories are the primary means by which cities document the location and status of individual trees (Nielsen et al., 2014; Nowak et al., 2008; Roman et al., 2013). Inventories are typically conducted by professional arborists, and their accuracy is crucial for quantifying ecosystem services, targeting new planting efforts, and forecasting pest or pathogen infestations (Hargrave et al., 2022; Jones et al., 2021; Jones et al., 2022; Ma et al., 2021; Miller et al., 2015). However, few communities in the United States maintain updated inventories (Hargrave et al., 2022). National-level surveys, such as the USDA Urban Forest Inventory and Analysis program, use systematic plot sampling to estimate urban forest characteristics (Edgar et al., 2021), but they do not conduct a full census. This leaves cities relying on opportunistic surveys and observations

from community scientists to supplement existing records and report forest conditions (Bloniarz, 1996; Crown et al., 2018; Roman et al., 2017). To complement traditional and opportunistic field surveys, researchers are turning to remote sensing and machine learning techniques to enable more detailed and unbiased assessments of urban trees.

Advances in remote sensing have significantly enhanced our ability to map and monitor urban forests (Fassnacht et al., 2016; Ward and Johnson, 2007). Tree species in the canopy are identifiable from aerial and satellite imagery due to their unique morphological, phenological, and spectral characteristics (Fassnacht et al., 2016; Nagendra, 2001). Numerous studies have combined high-resolution (sub-meter) multispectral images with light detection and ranging (LiDAR) to inventory urban tree species (Alonzo et al., 2014; Hartling et al., 2019; Jensen et al., 2012; Pu et al., 2018; Pu and Landry, 2020; Zhang and Hu, 2012). Despite these successful cases, the broader use of this technology beyond urban areas is constrained by the high costs of acquiring and processing fine-resolution data, and further complicated by inconsistencies in image quality, acquisition timing, and sensor availability across cities (Velasquez-Camacho et al., 2021). As a result, these approaches are typically applied at local spatial scales (e.g., individual parks, neighborhoods, or cities) and are rarely extended to multiple urban areas for broader regional comparisons (Beery et al., 2022; Li et al., 2019; Velasquez-Camacho et al., 2021).

Street-level imagery, which provides terrestrial views of trees, is an emerging alternative to satellite and aerial surveys of urban forests (Berland and Lange, 2017; Berland et al., 2019; Roberts et al., 2019; Roman et al., 2017; Seiferling et al., 2017). For example, Google Street View (GSV) captures high-resolution panoramic images every 10-20 meters along roads worldwide (Anguelov et al., 2010; Biljecki and Ito, 2021). Similar platforms, such as Mapillary

(https://www.mapillary.com/), provide analogous views with crowdsourced imagery. These images capture individual tree and lower canopy characteristics, including leaf arrangement, branching structure, and trunk appearance that are typically obscured in aerial or satellite imagery (Berland and Lange, 2017; Branson et al., 2018). Recent studies have shown that street-level imagery can be used to inventory urban street trees with accuracies comparable to aerial or satellite approaches (Branson et al., 2018; Capecchi et al., 2023; Choi et al., 2022; Liu et al., 2023; Lumnitz et al., 2021; Velasquez-Camacho et al., 2024; Wegner et al., 2016). Yet, most studies examine individual neighborhoods or cities, and it remains unclear where and to what extent surveys with street-level images can generalize across larger geographic extents and into novel environments (Beery et al., 2022).

Machine learning models are increasingly used to detect and classify plant species from remotely-sensed images, including aerial, satellite, and street-level sources (Hartling et al., 2019; Lake et al., 2022; Lumnitz et al., 2021; Ventura et al., 2024). These models can achieve high accuracy when trained and tested within the same geographic area, but their performance can decline substantially when applied to new regions (Beery et al., 2022; Marconi et al., 2022). Several factors can limit model generalizability, including variation in image quality, vegetation types, seasonal timing, and local environmental conditions. The resolution and availability of images can introduce biases that affect the consistency of species detection (Kim and Jang, 2023; Ventura et al., 2024). Even when imagery is available, structural occlusions such as buildings or overlapping vegetation can lead to missed detections (Choi et al., 2022; Ventura et al., 2024). Classifying trees to the correct taxon is also error-prone, especially for congeneric species with subtle morphological differences. Machine learning models tend to perform better for species with distinctive features like unique growth habits or spectral and textural characteristics (

Capecchi et al., 2023; Choi et al., 2022; Velasquez-Camacho et al., 2024). However, these traits also vary with geography and seasonality; for instance, leaf-off conditions in winter may obscure key features and increase misclassification rates (Liu et al., 2023). Additionally, urban areas support a higher diversity of tree species (Augustinus et al., 2024; Galle et al., 2021; Jha et al., 2019), raising the likelihood of confusion among visually similar taxa. As a result, developing machine learning models that generalize reliably across regions remains a significant challenge (Arevalo-Ramirez et al., 2024; Beery et al., 2022; Van Horn and Perona, 2017).

In this study, we developed computer vision models to detect, classify, and map urban street trees across 23 cities in the United States and Canada using Google Street View (GSV) and iNaturalist images. We trained the models on a large dataset of street trees and evaluated their performance using independent test splits that captured both cross-city and within-city generalization. Our objectives were to (1) examine how variation in urban tree diversity across cities influences image classification accuracy, (2) test whether geography and the month of GSV image capture explain differences in classification rates, and (3) determine how locally varying ecological conditions and street-view image coverage affect model errors and limit generalization across cities. To address these objectives, we used model predictions to create city-level tree inventories and fitted generalized linear mixed-effect models to evaluate when and where predictions aligned with known street trees.

2. Materials and methods

2.1 Study cities

We examined public street trees in 23 cities in the United States and Canada. These cities have existing tree inventories, available GSV imagery, and span a wide range of geographic,

ecological, and socioeconomic contexts. Cities with smaller populations in northern, continental climates (e.g., Sioux Falls, South Dakota, and Kitchener, Ontario) have a lower diversity of trees, mostly consisting of native deciduous and coniferous species (Table 1). In contrast, metropolitan cities such as Los Angeles, California, and New York City, New York, exhibit a higher diversity of trees with a greater proportion of non-native and ornamental species (Table 1) (Love et al., 2022; Ma et al., 2020; McCoy et al., 2022). Variation in urban populations, environmental conditions, and management practices shapes the species composition and diversity of urban forests (Roman et al., 2018), presenting both opportunities and challenges for tree inventories with street-level imagery.

2.2 Data collection and image filtering

2.2.1 Selecting focal genera

We selected 100 tree genera to detect, classify, and geolocate across 23 cities (Table S1). We examined trees at the genus level to simplify the complexity of distinguishing between numerous varieties, hybrids, and subspecies commonly found in urban areas. Surveys of urban tree diversity in the United States show that the relative abundance of both species and genera follows a long-tailed distribution, where a few taxa are abundant while the majority are infrequent or rare (Love et al., 2022; Ma et al., 2020; McCoy et al., 2022). Our 100 selected genera account for over 95% of all genera reported in the 23 city inventories. These include taxa that are widely distributed (e.g., *Acer* [maple], *Fraxinus* [ash], *Quercus* [oak]), regionally abundant (e.g., *Magnolia* [magnolia], *Picea* [spruce]), and occasionally found as non-native or ornamental species (e.g., *Ailanthus* [tree of heaven], *Lagerstroemia* [crape myrtle]).

2.2.2 Images of tree genera from AutoArborist

We accessed a comprehensive dataset of GSV images to develop computer vision models that detect and classify urban tree genera (Figure 1). Beery et al. (2022) created the AutoArborist v0.16 dataset by co-locating GSV images with tree inventory records across 23 cities in North America. This dataset includes approximately 1.2 million GSV images of over 300 unique genera. The images (dimensions: 768 x 1152 pixels) also contain metadata, including capture location (i.e., latitude and longitude), capture date (i.e., month and year), and annotated bounding boxes of individual trees.

2.2.3 Images of tree genera from iNaturalist

We expanded the AutoArborist dataset with images of tree genera sourced from iNaturalist, a platform hosting millions of user-annotated plant images from around the world. We downloaded all available 'research-grade' images of our selected 100 genera in North America from the iNaturalist Licensed Observation Images webserver (https://registry.opendata.aws/inaturalist-open-data; accessed August 1st, 2024), yielding approximately 3.6 million images.

2.2.4 Filtering iNaturalist images

iNaturalist records are contributed by community scientists and contain images of trees taken from a wide range of perspectives. Many of these photos are captured from a short distance (i.e., within one meter) and show individual leaves, twigs, flowers, and fruits. While these features are important for taxonomic identification, they do not align with the street-level perspective of GSV images, where trunk and canopy structures are visible. Recent studies have

shown that mixing close-up iNaturalist images with broader views (e.g., UAV or satellite images) introduces inconsistent visual features that can confuse computer vision models and potentially degrade their performance (Beery et al., 2022; Soltani et al., 2022; Soltani et al., 2024). To address this issue, we filtered iNaturalist images to prioritize those that better matched the perspective of GSV images.

We applied a Contrastive Language-Image Pre-Training (CLIP) deep learning model to filter iNaturalist images (Radford et al., 2021). This model was jointly trained on millions of paired images and text labels and allows querying of images based on how well they align with a text prompt (Radford et al., 2021). We used the prompt "a photo of a large mature tree in an urban landscape" to retain iNaturalist images that better aligned with the perspective of GSV imagery (Figure S1). We set an upper limit on the number of filtered images per genus to avoid introducing data imbalances between common and rare genera. For genera with more than 10,000 available iNaturalist images (e.g., *Acer* [maple], with 303,000 images), we retained the 1,000 images with the closest agreement to the prompt. For genera with fewer than 10,000 iNaturalist images, we kept the top 10% of images that best matched the prompt. This strategy enabled iNaturalist imagery to serve as a supplementary source rather than a replacement for AutoArborist images. After filtering, the curated subset consisted of 59,132 images, which we then incorporated into our training dataset for the tree classification model (see below: *2.3.2. Tree classification model*; Figure S1; Table S1).

2.3 Tree detection, classification, and geolocation

2.3.1 Tree detection model

We developed an object detection model (YOLO; You Only Look Once) (Redmon et al., 2016) to identify trees in GSV images (Figure 1). We selected the YOLOv5 model (Jocher et al., 2022; Redmon et al., 2016) over comparable object detection models, such as R-CNN (Girshick et al., 2014) and Faster R-CNN (Ren et al., 2017), due to its efficiency, accuracy, and ability to detect multi-scale objects (Terven and Cordova-Esparza, 2023). Previous studies have also used similar versions of YOLO models (e.g., YOLOv3) to detect trees in street-level images (Choi et al., 2022; Liu et al., 2023).

To train the object detection model, we used annotated images of trees from the AutoArborist dataset (Beery et al., 2022). Because our goal was to build an object detection model that identified all cases of a single class (i.e., 'tree') across all cities, we randomly sampled images across all genera to capture variability in tree appearance and image conditions. We sampled 750,000 images annotated with bounding boxes from the AutoArborist dataset for model training. An additional 75,000 annotated images were withheld for validation to fine-tune model parameters during training, and a separate 75,000-image test set was reserved for final evaluation. To improve generalization to unseen images, we applied image augmentations (e.g., random horizontal flips, rotations, and color shifts) during training to vary the appearance and scale of trees.

We trained the model for 25 epochs and evaluated performance on the test dataset using precision and recall, calculated at an Intersection over Union (IoU) threshold of 0.5. IoU measures the overlap between the predicted and annotated bounding boxes. A detection was considered a true positive if the IoU with the annotated box is at least 0.5; otherwise, it was classified as a false positive. We selected this threshold to account for potential inaccuracies in

the AutoArborist data; for example, defining a bounding box for complex objects (i.e., trees) is somewhat subjective (Everingham et al., 2010).

We calculated precision as the ratio of true positive detections to total predicted detections, which reflects the ability to avoid false positives. We calculated recall as the ratio of true positives to the total number of annotated trees, measuring the ability to correctly detect trees. These metrics are commonly used to evaluate the performance of object detection models (Terven and Cordova-Esparza, 2023). For additional details on the object detection model architecture, training, and evaluation, see Supplementary Methods 1.1.

2.3.2 Tree classification model

We developed a convolutional neural network (CNN) to classify tree genera detected in GSV images (Figure 1). CNNs have shown state-of-the-art performance in image classification tasks across disciplines (Kattenborn et al., 2021; Krizhevsky et al., 2012). We selected the EfficientNetV2-S model for its balance of network complexity and computational efficiency (Tan and Le, 2021).

To train and evaluate the image classification model, we used images of 100 tree genera sourced from AutoArborist and iNaturalist (range: 793 images for *Larix* [larch] to 45,743 images for *Acer* [maple]; Table S1). To assess model generalization within and across cities, we applied a two-stage data splitting strategy. For each city, we first split the dataset into two spatially distinct subsets based on geographic boundaries (Figure S2). Approximately 70% of images (N = 789,820) were allocated to model development, and the remaining 30% (N = 356,341) were withheld as a geographically independent test set for final evaluation. This spatial partitioning ensured that training and testing images did not overlap geographically.

Within the 70% development set, we further partitioned images into model training (80%), validation (10%), and internal test (10%) sets using random sampling. The validation set was used to monitor performance and tune model parameters during training, while the internal test set provided a benchmark to evaluate classification performance across cities (see below: *2.4.1 Effect of diversity on genus classification*).

We trained the classification model for up to 100 epochs and terminated training when the validation loss did not decrease after 10 consecutive epochs to limit overfitting. To improve generalization to unseen images, we applied image augmentations (e.g., random horizontal flips, rotations, and color shifts) during training. After training, we evaluated model performance by calculating recall (i.e., true positive rate), precision (i.e., proportion of predicted positives that are correct), and F1 score (a measure of performance that accounts for both model recall and precision) for each genus separately using the 30% geographic testing split. For additional details on the image classification model architecture, training, and evaluation, see Supplementary Methods 1.2.

2.3.3. Tree classification model interpretation

Image classification performance tends to increase with additional training images (Goodfellow et al., 2016; Hestness et al., 2017). To examine this effect across genera, we calculated the Spearman rank correlation coefficient between the number of training images per genus and test-time performance (measured as F1 score calculated from the 30% geographic testing split).

To understand which visual features lead to correct classifications, we applied gradient-weighted class activation maps (Grad-CAMs) (Selvaraju et al., 2017). Class activation maps

highlight regions of an image that contribute strongly to model predictions (Zhou et al. 2016). Previous work using Grad-CAM has shown that bark and trunk features are important for tree species classification (Kim et al., 2022). We calculated the activation maps from the penultimate convolutional layer in the image classification model. We randomly selected 100 images from each genus from the AutoArborist 30% geographic testing split and calculated Grad-CAMs for each image separately. We visually inspected these maps to verify that the model focused on biologically meaningful features, such as leaves, branches, fruit, and crown structure, during the prediction process.

2.3.4. Tree geolocation with depth estimation

We determined the location of each detected and classified tree through geolocation. We used the latitude and longitude of a GSV image, the angle of each tree detected in the image, and an estimated distance between the tree and the image to geolocate predictions.

To estimate tree-to-image distances, we applied the monocular depth estimation model developed by Bhat et al. (2023), referred to as ZoeDepth. Monocular depth estimation models infer object distances from single-color images (Godard et al., 2018) and have been applied to approximate distances between GSV images and urban objects, including street trees (Krylov et al., 2018; Liu et al., 2023; Lumnitz et al., 2021). We selected the pre-trained ‘ZoeD-NK’ model, which was trained to estimate depth from street-level images in outdoor scenes (Bhat et al., 2023). We calculated tree-to-image distances as the median depth within the predicted bounding box for each tree. To account for distortions caused by the projection of panoramic GSV images, we scaled each depth estimate by a factor of three (Lumnitz et al., 2021).

2.4 Impact of diversity, image timing, and geography on genus classification

2.4.1 Effect of diversity on genus classification

We examined how variation in urban tree diversity across cities affects image classification rates. Within each city, genus diversity follows a long-tailed distribution, where a few genera are common and many are rare (Love et al. 2022; Ma et al. 2020). This class imbalance can bias model predictions, often resulting in higher accuracy for common genera than for rare ones (Beery et al., 2022; Van Horn and Perona, 2017).

To quantify urban tree diversity, we calculated three metrics for each city, including richness (the total number of genera), evenness (the relative abundance of genera, ranging from 0 - 1), and the Shannon diversity index (which integrates both richness and evenness; Hill, 1973) using the full AutoArborist inventory dataset (Table 1). Cities with high richness but low evenness may have many unique genera with a few that overwhelmingly dominate, whereas cities with high richness and high evenness have a more balanced distribution across genera. Shannon diversity values typically range from 1.5 to 4.0, with higher values indicating both greater richness and a more even distribution among taxa (Love et al., 2022). All diversity metrics were computed using the ‘diversity’ function in the R package ‘vegan’ version 2.6-4 (Oksanen et al., 2013).

To evaluate model performance across cities, we used the internal 10% test set (N = 74,631 images) created during the initial model development split (see *Section 2.3.2*). We used this test set to ensure the richness, evenness, and diversity measures were representative samples of each city. We then calculated the weighted F1 score across genera for each city. This metric is the harmonic mean of precision and recall, averaged across genera and weighted by the number of samples in each genus. This weighting ensures that performance estimates reflect the class

imbalance in each city, giving proportionally more influence to common genera while still incorporating rare ones.

To test whether classification performance was associated with urban tree diversity, we fit separate beta regression models for each diversity metric using the R package 'betareg' (Cribari-Neto and Zeileis, 2010). Model diagnostics confirmed that assumptions were met. One city, Kitchener, Ontario, was excluded from all regressions as an outlier due to anomalously low classification performance (Figures S3 - S5).

2.4.2. Effects of image timing and geography on genus classification

Seasonal changes in the appearance of trees and regional differences in genus composition (e.g., differing species) may influence genus classification rates. We analyzed image classification performance in relation to the month of GSV image capture and geography.

We grouped cities into four geographic regions based on climate similarity (Figure S6). These regions were defined using a principal component analysis (PCA) of 19 bioclimatic variables from the WorldClim v2.1 dataset (https://worldclim.org/), which summarize temperature and precipitation patterns relevant to tree distributions and phenology (Boucher-Lalonde et al., 2012; Fick and Hijmans, 2017). We sampled climate variables at a 2.5' (~4.5 km x 4.5 km) resolution centered on each city using the 'raster' R package (Hijmans et al., 2015). The first two PCA axes explained 77.6% of climate variation across cities. We then used these axes to assign each city to one of four geographic regions (Table 1, Figure S6).

We calculated the F1 score for each genus by month and region combination using the 30% geographically independent test set (Figure S2). Because most genera lacked sufficient imagery across all month–region combinations, we restricted the analysis to genera with at least

50 such combinations (N = 17 genera). This threshold ensured adequate temporal and spatial coverage for meaningful trend detection while avoiding sparse cases where ANOVA models failed due to missing data.

For the remaining 17 genera, we tested whether classification performance varied by month, region, and their interaction using ANOVA (Type II sum of squares). We also included the number of training images for each genus–month–region combination as a covariate to evaluate how training data availability influenced model performance. Model assumptions were assessed via diagnostic checks before analysis. Pairwise comparisons were evaluated using Tukey's post hoc tests with the 'emmeans' R package (Lenth, 2025).

2.5 Generalizability of computer vision-based tree inventories

To evaluate model generalizability, we applied our computer vision models across cities (Figure 1) and compared predicted trees to AutoArborist inventory tree data. Predicted and inventory trees were paired using a greedy matching approach, and we modeled the probability of a matched pair with binomial generalized linear mixed effects models (GLMMs, logit link).

2.5.1 Sampling inventory trees

We restricted this analysis to AutoArborist inventory trees located within the 30% geographic test split of each city (Figure S2). To capture local variability, we divided each city into 100-meter grid cells and retained only those containing at least one inventory tree. Next, we stratified the grids by inventory tree density quantiles (0.25, 0.5, 0.75) and randomly selected 5% of cells within each category. This sampling design yielded 103,428 inventory trees across 6,306 grids in 23 cities (Table S2).

For each sampled grid, we downloaded all available Google Street View images within and 100 meters beyond the grid boundary to minimize edge effects from missing imagery during predictions (Supplementary Methods 1.3). We then applied the computer vision models to the downloaded GSV images to detect, classify, and geolocate trees.

2.5.2 Matching computer vision-based predictions to inventory trees

Predicted trees were linked to inventory trees using a greedy, one-to-one genus-matching approach. For each inventory tree, we computed a 20-meter circular buffer and identified all congeneric predicted trees within this area. Geodesic (great-circle) distances between inventory and predicted trees were calculated in meters from their geographic coordinates (WGS84), and the closest pair was iteratively selected and removed from further consideration, yielding a one-to-one assignment. Inventory trees were left unmatched if no congeneric predicted trees fell within their buffer or if all congeneric predicted trees were already matched. We tested multiple buffer distances (5-meter intervals from 5 to 30 meters) and selected a 20-meter buffer to balance matching performance while maintaining a local scope of analysis (Figure S12).

2.5.3 Statistical analysis with generalized linear mixed models

We modeled the probability that an inventory tree matched with a congeneric predicted tree using binomial GLMMs. The binary response was one for a correct match and zero otherwise. The full model was specified as:

$$\square\square\square\square\square(\square_{\square\square\square\square}) = \square_{0} + \sum_{\square=1}^{11} \square_{\square}\square_{\square,\square\square\square\square} + \square_{\square} + \square_{\square(\square)} + \square_{\square} \quad (1)$$

Where $\rho_{\square\square\square\square}$ is the probability that an inventory tree $\square$ in a grid $\square$ of a city $\square$ belonging to a genus $\square$ is correctly matched to a congeneric predicted tree. The intercept is denoted as β_0, $\square_{\square,\square\square\square\square}$ are the 11 predictors (described in section 2.5.4), $\square_{\square} \sim \square\left(0, \square^2_{\square\square\square\square}\right)$ is the random intercept for city, $\square_{\square(\square)} \sim \square\left(0, \square^2_{\square\square\square\square}\right)$ is the random intercept for sampling grid IDs nested within city, and $\square_{\square} \sim \square\left(0, \square^2_{\square\square\square\square\square}\right)$ is the random intercept for genus.

2.5.4 Fixed and random effects

We evaluated 11 fixed effects hypothesized to influence model generalizability. These included five ecological predictors, including tree abundance, congeneric tree abundance, genus richness, nearest-neighbor (NN) tree distance in meters, and temperature seasonality. All predictors except seasonality were measured within a 20-meter buffer around each inventory tree. These variables capture local density and spatial structure of nearby trees, which may influence tree matching outcomes by varying occlusion (tree abundance, NN tree distance), misclassification risk (genus richness), or match likelihood (congeneric tree abundance). Temperature seasonality (BIO4) was extracted for each grid from the WorldClim v2.1 dataset (30” resolution, ~1 km x 1 km; https://worldclim.org/) using the ‘terra’ R package (Hijmans, 2022).

We modeled four street-level image coverage predictors within a 20-meter buffer around each inventory tree, including the number of street-level images, the distance to the nearest street-level image (in meters), the modal year of capture of the street-level images, and the season of capture of the street-level images. Street-level image counts and distances represent the availability and diversity of perspectives. The capture year reflects recency, which influences image resolution. The capture season indicates whether the date the image was taken falls in the

growing season or dormant season (Supplementary Methods 1.4), depicting cyclical variation in phenological traits that affect tree appearance.

We also considered two predictors to account for biases from prior model knowledge. These included the number of AutoArborist training images associated with the inventory tree's genus, and the proportion of those images originating from the tree's focal city (the city-level count divided by the genus-level count). These predictors quantify the model's familiarity with each genus and its geographic context, which may affect classification performance and, consequently, tree-matching outcomes.

We included random intercepts for city and genus to account for the hierarchical structure in the data and a nested random intercept for sampling grid ID within city to account for spatial autocorrelation.

2.5.5 Model selection and evaluation

All numeric predictors were centered and scaled. Skewed predictors (all except temperature seasonality and capture year/season) were log-transformed to meet the assumption of linearity on the link scale. We fitted all 2,046 submodels of the full specification in R v.4.5.1 (R Core Team, 2025) using the 'lme4' (Bates et al., 2015) and 'MuMIn' (Bartoń, 2024) R packages. Models were ranked by the Bayesian Information Criterion (BIC), which was chosen as a conservative criterion given the large sample size (Schwarz, 1978). Diagnostics for the top-ranked model (minimum BIC) included residual plots, uniformity, dispersion, and outlier checks using the 'DHARMa' R package (Hartig, 2024), variance inflation factors using the 'performance' R package (Lüdecke et al., 2021), and spatial autocorrelation tests using the 'spdep' R package (Bivand, 2022) (Supplementary Methods 1.5)

We evaluated consistency in predictor selection, effect size, and direction for the full model and all submodels with ΔBIC < 10. For the top-ranked model, we assess goodness-of-fit by reporting the marginal R^2 (variance explained by fixed effects only) and the conditional R^2 (variance explained by fixed and random effects), both calculated using the 'performance' R package (Lüdecke et al., 2021). We further decomposed the random effect variance to estimate the percent of total variance attributable to each grouping factor. We visualized the influence of fixed effects on the probability of correct tree matching with partial effect plots. Lastly, we examined significant deviations in group levels from overall means by extracting best linear unbiased predictions for city and genus.

3. Results

3.1 Tree detection and classification

3.1.1 Object detection accurately identifies trees in street-level images

The object detection model identified trees with high precision (92.2%) and recall (82.4%) when evaluated using withheld testing images (Figure S7). Both high precision and recall values indicate a low proportion of false positive detections and that the model identified most trees present in GSV images.

3.1.2 Classification performance and model interpretation

Classification performance varied widely among 100 genera (mean F1 score = 0.55; range: *Phoenix* [date palm], F1 = 0.88; *Carya* [hickory], F1 = 0.06; Table S3). In general, higher performance was correlated with a greater number of training images ($p < 0.001$; $R^2 = 0.50$; Figure S8). However, the genus with the highest performance (*Phoenix* [date palm]; N = 3,790

images, F1 = 0.88) had considerably fewer training images relative to the most common class (*Acer* [maples]*;* N = 45,743 images; F1 = 0.75).

The gradient-weighted class activation maps (Grad-CAMs) provided insight into different visual features underlying correct genus classifications (Figure 2). For example, the model correctly classified *Citrus* trees by focusing on the characteristic orange fruit (Figure 2C). For genera where fruiting or flowering features were absent, structural features such as canopy shape and branching patterns were important for classification (e.g., for deciduous genera like *Acer* and *Betula*, but also for palms like *Phoenix*) (Figure 2). Likewise, bark color and texture informed correct classifications. The reddish-brown, furrowed bark common to the conifers in the genus *Pinus* and *Thuja*, and white-brown papery bark found in the deciduous genus *Betula*, guided correct classifications (Figure 2). Across genera, Grad-CAMs revealed that classification relied on features that were biologically meaningful and visually distinctive.

3.2 Impact of diversity, image timing, and geography on genus classification

3.2.1 Urban tree diversity affects classification rates among cities

Genus diversity within cities impacted overall classification performance. Cities with a higher Shannon diversity index had significantly lower classification performance relative to less-diverse cities ($p < 0.001$; Figure 3, Table S4). Likewise, cities with a greater number of genera (i.e., genus richness) had lower performance ($p = 0.045$; Figure S9, Table S4). The relative abundance of genera within cities also affected classification accuracy ($p < 0.001$; Figure S9, Table S4). Cities where few genera are dominant (e.g., Sioux Falls, SD; evenness = 0.48) had higher performance relative to cities where no single genus dominates (e.g., Los Angeles, CA; evenness = 0.89) (Table S4).

3.2.2 Genus classification varies by geography and month

Genus classification varied by geographic region and, to a lesser extent, by the month of GSV image capture (Table S5). We detected regional differences in F1 score for the majority of tested genera, including *Acer* (maple), *Fraxinus* (ash), *Quercus* (oak), *Ulmus* (elm), *Prunus* (cherry/plum), *Tilia* (basswood), *Malus* (apple), *Platanus* (sycamore), *Magnolia* (magnolia), and *Crataegus* (hawthorn) (all $p < 0.01$; Table S5, Figure 4, Figure S10).

The month of GSV image capture significantly influenced classification for the most common genus (*Acer;* $p < 0.001$; Table S5, Figure 4). For *Acer*, images captured from summer months (e.g., July and August) experienced fewer errors from false positives and false negatives relative to images captured during winter periods (e.g., January and February) (Figure 4). This effect was independent of the number of monthly training images and was consistent across regions (Table S5). For all other tested genera, monthly differences were not significant ($p > 0.05$; Table S5, Figure S11).

3.3 Generalizability of computer vision-based tree inventories

3.3.1 Baseline matching rates across genera and cities

The computer vision-based tree inventory approach matched 67.1% of the AutoArborist inventory trees to predicted congeneric trees across all cities and genera. Matching performance varied more strongly across genera (mean = 50.9%, SD = 23.0%; Table S6) compared to cities (mean = 67.4%, SD = 9.4%; Figure S12).

3.3.2 Selection of ecological, image coverage, and prior-knowledge predictors

Among the 2,046 candidate models evaluating how ecological factors, street-level image coverage, and prior knowledge predictors influence generalizability, the best-supported model (with a BIC weight of 0.702) retained eight predictors from the full model. These included local ecological predictors (tree abundance, genus richness, congeneric tree abundance, nearest-neighbor tree distance), local image coverage predictors (nearest street-level image distance, number of street-level images), and prior knowledge predictors (number of AutoArborist genus-level training images, and the proportion of AutoArborist genus-level training images originating from the focal city).

Two additional models (ΔBIC < 10) received limited support. These differed from the top-ranked model by either excluding genus richness or including the modal season of street-level image capture as a fixed effect. The full model was strongly over-parameterized (ΔBIC = 30.77) (Figure 5A).

Effect sizes for shared predictors across supported models (ΔBIC < 10) were highly consistent in both magnitude and direction (mean SD of fixed-effect odds ratio estimates = 0.008; Figure 5B). Predictors unique to specific models exerted negligible influence on matching performance. Given the stability of effect sizes across supported models and the strong support for the top-ranked model, all subsequent model evaluations were based on the top-ranked model.

3.3.3 Model diagnostics and variance explained by fixed and random effects

Model diagnostics for the top-ranked model confirmed that all assumptions were met, with no evidence of overdispersion, non-uniform residuals, influential outliers, or spatial autocorrelation (Figures S13–S14). Fixed effects explained 20.5% of the variance in tree matching (marginal R^2), increasing to 46.7% (conditional R^2) when random effects were also

included. Variance attributable to the random effects was dominated by genus (Var = 0.78, SD = 0.89), with a comparatively smaller contribution from city (Var = 0.46, SD = 0.46). These values align with the variance partitioning estimates, with 12.7% of the total variance attributed to genus and 3.5% to city (Table 2).

3.3.4 Influence of ecological, image coverage, and prior-knowledge predictors

All fixed effects retained in the top-ranked model had a significant impact on inventory tree-matching probability (Table 2). Prior-knowledge predictors (model familiarity with training images) exerted the strongest effects. A multiplicative increase in the number of AutoArborist genus-level training images (log-transformed) more than doubled the odds of matching (OR = 2.01, 95% CI: 1.71–2.35, $p < 0.001$). The proportion of AutoArborist genus-level training images originating from the focal city also had a substantial positive effect (OR = 1.60, 95% CI: 1.53–1.67, $p < 0.001$). Local street-level image coverage also contributed comparably strong effects. A multiplicative increase in the number of street-level images (log-transformed) was associated with higher matching success (OR = 1.55, 95% CI: 1.51-1.58, $p < 0.001$), while increased distance to the nearest street-level image (log-transformed) reduced matching success (OR = 0.67, 95% CI: 0.67–0.68, $p < 0.001$). Local ecological context showed mixed effects. Greater nearest-neighbor tree distances (log-transformed; OR = 1.44, 95% CI: 1.41–1.48, $p < 0.001$) and higher congeneric tree abundance (log-transformed; OR = 1.42, 95% CI 1.37–1.46, $p < 0.001$) increased matching success, whereas higher tree abundance (log-transformed; OR = 0.74, 95% CI: 0.71–0.77, $p < 0.001$) and genus richness (log-transformed; 0.93, 95% CI: 0.90–0.97, $p < 0.001$) reduced tree-matching success (Figure 5B–5C). Together, these results demonstrate that matching success is maximized when an inventory tree is well-represented in

the training dataset, particularly within similar geographic contexts, when street-level image coverage is dense, and when the local ecological context (tree composition and arrangement) is simpler.

3.3.5 Genus- and city-level deviations

Genus-level best linear unbiased predictions showed consistent over- and under-performance after adjusting for the fixed effects (Figure S15A). Here, "over-" and "under-performance" refer to group levels whose best linear unbiased predictions' estimates had a 95% confidence interval that did not overlap the overall mean. Among 100 genera examined, 29 genera over-performed and 21 under-performed. The strongest positive deviations were observed for *Ailanthus* (tree-of-heaven; best linear unbiased prediction ± 95% CI: 1.99 ± 0.39), *Melaleuca* (paperbark; 1.84 ± 0.37), *Abies* (fir; 1.56 ± 0.54), *Pseudotsuga* (douglas fir; 1.35 ± 0.38), and *Sequoia* (coast redwood; 1.22 ± 0.58). The strongest negative deviations occurred for *Ostrya* (hop hornbeam; –1.75 ± 0.76), *Cornus* (dogwood; –1.63 ± 0.26), *Chamaecyparis* (false cypress; –1.52 ± 0.67), *Syringa* (lilac; –1.44 ± 0.22), and *Citrus* (citrus; –1.33 ± 0.72).

City-level best linear unbiased predictions (Figure S15B) were more constrained, consistent with their minor contribution to explained variance (3.5%; Table 2). Only three cities over-performed (Vancouver = 1.20 ± 0.21; Charlottesville = 0.85 ± 0.54; Buffalo = 0.56 ± 0.28), while five underperformed (Kitchener = –0.89 ± 0.36; Los Angeles = –0.46 ± 0.16, Washington, DC = –0.35 ± 0.18, San Francisco = –0.26 ± 0.18, Edmonton = –0.22 ± 0.19). These results indicate that genus identity remains the primary source of residual variation in tree-matching performance, whereas deviations at the city level were few and of modest magnitude.

4. Discussion

Urban tree inventories are crucial for managing ecosystem services and responding to disturbances such as pest or disease outbreaks. While existing remote sensing models can create accurate inventories for single, well-characterized sites, their ability to generalize under varying imaging conditions, species compositions, and climates remains uncertain. Here, we trained computer vision models on Google Street View and iNaturalist images to detect, classify, and map 100 tree genera in 23 North American cities. Our analyses show that object detection and depth estimation models consistently identified and geolocated trees within urban landscapes. However, genus classification was more variable. We found that classification rates varied due to shifts in genus diversity and geography, and improved when images captured leaf-on conditions. Grad-CAM visualizations confirmed that classification relied on biologically meaningful and distinctive features, such as canopy structure, bark texture, and the presence of fruits. When we tested our approach in geographically withheld areas, our predicted trees matched 67.1% of AutoArborist inventory trees. Matching success was primarily driven by differences among genera, training data availability, and local context (image coverage and tree density), whereas city identity was less important. Taken together, our findings demonstrate that street-level imagery, combined with computer vision, offers a practical and scalable complement to traditional field-based inventories.

4.1. Opportunities and challenges for scaling urban tree inventories

Street-level imagery is emerging as an alternative modality to satellite and aerial imagery for surveying urban forests. While previous studies focused on single cities or neighborhoods (Branson et al., 2018; Capecchi et al., 2023; Choi et al., 2022; Liu et al., 2023; Velasquez-

Camacho et al., 2024; Wegner et al., 2016), our work extended these approaches to 23 North American cities and enabled a broader test of model generalizability. Notably, we found that our computer vision models achieved high precision and recall (>0.7) for eight of the most common genera in the United States (i.e., *Acer, Fraxinus, Quercus, Gleditsia, Ulmus, Tilia, Pyrus,* and *Platanus)* (Table S3). These genera represent ~ 45% of all public street trees in the United States (Ma et al. 2020), suggesting immediate operational value for many municipalities. For example, the city of Sioux Falls, South Dakota, had high classification performance because ~82% of public street trees were represented by these common genera (Figure 3).

We observed lower classification performance in large, biodiverse cities (e.g., Los Angeles, CA; New York City, NY; Denver, CO). These declines are likely due to visual similarity among co-occurring genera (i.e., low inter-class variability) and the presence of similar-looking species or cultivars (i.e., high intra-class variability). These findings are also consistent with previous field-based and remote sensing studies, which show that tree species classification is challenging in biodiverse, urban settings (Berland et al., 2019; Berland and Lange, 2017; Marconi et al., 2022; Weinstein et al., 2023). Although we classified at the genus level, future approaches could predict at multiple taxonomic levels (i.e., classifying trees to family, genus, and species) and leverage phylogenetic relatedness to potentially improve predictions (Gillespie et al., 2024).

4.2 Interpreting model predictions with trait-based evidence

Classification was shaped by both geography and seasonality. Regional differences in model performance likely reflect varying species compositions, management practices, and contextual scene cues (e.g., building styles, street layouts) (Beery et al., 2022). Seasonal

differences in GSV image timing also moderated classification rates. For example, for *Acer*, leaf-on months (July–August) reduced both false positives and false negatives relative to imagery captured in winter months independent of the monthly training set size (Table S5, Figure 4).

Grad-CAM visualizations provided additional insight to interpret genus classification. For some taxa, fruits and flowers were particularly informative features, whereas in other cases, subtler traits such as canopy architecture and bark texture were used to separate visually similar genera. These findings align with prior work highlighting the value of tree bark for species identification with CNNs (Kim et al., 2022) and extend them by suggesting that the relative importance of features shifts seasonally. For example, bark can provide consistent information year-round, while flowering or fruiting periods offer brief but highly informative classification windows. In addition to seasonal changes, feature salience may also vary with tree age. Although our dataset lacked information on age, the Grad-CAMs highlighted specific bark patterns and trunk characteristics (e.g., buttressing) that are pronounced in mature trees but less apparent or underdeveloped in young saplings. This suggests that for younger trees, foliage traits may be more informative for predictions. Such results point to clear opportunities for optimizing image acquisition by aligning data collection with periods or contexts when locally informative traits are most visible.

4.3 Factors shaping generalizability across cities

The representation of training data, street-level image coverage, and ecological complexity also drives the generalizability of our computer vision models. Genus-level training image availability was the strongest positive predictor of matching success. We observed that a probability of more than 50% in match success required a minimum of ~3000 genus-level

training images. City-specific genus-level training data further enhances matching success. For example, increasing the proportion of city-specific training data from 0.01 to 0.25 boosts matching ability from 61% to 74% (Figure 5B–C).

Tree abundance and stand structure also impacted generalizability. Dense stands and nearby neighboring trees reduced match success, likely due to spacing and occlusion by neighboring trees. Several studies have qualitatively noted similar patterns (Branson et al., 2018; Choi et al., 2022; Lumnitz et al., 2021); however, to our knowledge, this study is the first to evaluate the effect of stand structure on generalizability quantitatively. We find the probability of match success drops below 50% when 30 or more trees are found within a 20-meter radius of the inventory tree. Additionally, a minimum of ~1 meter distance from the nearest neighboring tree is required for the 50% probability of a successful match, while distances of 15 meters or more increase match success to about 75% (Figure 5B–C).

The number and proximity of street-level images also influenced the generalizability of our approach. Denser image coverage improved match success, while greater distances of the inventory tree from the image reduced match success. This pattern is consistent with prior studies that show that multiple nearby viewpoints enhance the detection and localization of street trees (Lumnitz et al., 2021; Wegner et al., 2016) and that closer proximity of images reduces the likelihood of occlusion or missed detections (Laumer et al., 2020; Velásquez-Camacho et al., 2024). Unlike previous studies, our research is the first to quantify this relationship explicitly. We found that the probability of successful matching increases from 50% to 63% when the number of available images within a 20-meter radius of the inventory tree rises from one to three. Additionally, we found that the probability of a successful match drops below 50% when the closest street view image is more than 18 meters away (Figure 5B–C).

Taken together, these results highlight the interacting roles of training data representation, ecological structure, and street-level image availability in shaping model generalizability. From an operational standpoint, computer vision-based inventories will perform best in areas that are well imaged, when images are captured during leaf-on conditions that provide more discriminative features, and when stand structure is relatively simple (i.e., trees are not tightly packed or heavily occluded). Strategically expanding training images, particularly for underrepresented genera with examples drawn from the target geography, can further improve the reliability of these approaches.

4.4. Model limitations and implications for urban forestry

Despite the overall robustness of the computer vision models, several limitations warrant consideration. First, our evaluation relied on existing tree inventories, which may be incomplete, inconsistent across cities, or outdated. In this case, apparent model errors may sometimes reflect gaps or inaccuracies in the reference data rather than genuine misclassifications. Second, our matching rates should also be viewed as conservative because predictions were only compared to public inventories. The computer vision models were unable to distinguish between public and private street trees, so the true positives for private trees cannot be validated (Lumnitz et al., 2021). Third, our use of a 20-meter matching radius introduces some uncertainty by conflating nearby public and private trees of the same genus. Although a stricter 15-meter radius reduced overall agreement to 60.0%, the broader statistical patterns driving generalizability remained unchanged (Figure S12). Last, information about the health of individual trees is infrequently recorded in tree inventories. Without ground-truth measurements of these characteristics, our

models could not reliably distinguish between healthy, stressed, and diseased trees, thereby limiting their use for forest health monitoring (Kagan et al., 2021).

To reduce errors in remotely-sensed tree inventories, incorporating multiple image sources and geographic context into computer vision models could be particularly beneficial. For example, remotely-sensed images from the National Agricultural Imagery Program (NAIP) are freely available, orthorectified, and captured at high resolution (60 cm – 1-meter pixels) across the U.S. during the growing season. Prior studies have also demonstrated the effectiveness of aerial imagery in detecting and geolocating urban trees (Branson et al., 2018; Erker et al., 2019; Ventura et al., 2024). Several studies have found that combining image sources can improve classification rates by providing complementary perspectives on tree crown structure, foliage condition, and surrounding spatial context (Hartling et al., 2019; Wang et al., 2025). Conversely, others indicate a marginal benefit of incorporating high-resolution satellite imagery in classification performance (Beery et al., 2022). Besides image sources, adding geographic context (e.g., latitude and longitude) during model training could also improve the fine-grained classification of tree species. Given that tree species distributions are governed by regional climate and local soil conditions (Boucher-Lalonde et al., 2012), using geographic location as a prior can constrain model predictions to species that are ecologically plausible for a given area (Aodha et al., 2019; Cole et al., 2023).

Understanding where and when machine learning models generalize to novel domains is a fundamental challenge across various disciplines (Zhou et al., 2023; Beery et al., 2018). The AutoArborist dataset (Beery et al., 2022) was designed to assess the generalization of computer vision models in long-tailed, real-world classification problems, where few tree genera are abundant while most are uncommon or rare. This imbalance captures the natural heterogeneity of

urban forests and provides a realistic benchmark for evaluating model performance across varying ecological and imaging conditions. Our findings support the use of computer vision models and street-level imagery in urban forestry applications. Future work could extend these approaches to peri-urban and rural areas to estimate plant populations and rapidly document species' distributions (e.g., for mapping invasive plant species; Kotowska et al., 2021). In regions where street-level imagery is available sporadically, more sophisticated models (e.g., those that utilize both street-level and aerial imagery) can compensate for data gaps to enhance detection and mapping.

5. Conclusions

Our study demonstrates that computer vision models trained on publicly available street-level imagery can accurately detect, classify, and geolocate urban street trees across diverse cities in North America. While classification rates declined in areas with high genus richness or suboptimal image timing, our results reveal key sources of model error, such as inter- and intra-class variability, mirroring the same challenges faced by community observers and trained arborists. By quantifying where and when model performance deteriorates, we establish a framework for systematically improving future urban tree mapping efforts. Developing scalable and generalizable models for tree inventories will be essential for monitoring and safeguarding ecosystem services amid growing environmental pressures.

CRediT authorship contribution statement

Thomas Lake: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Project administration, Validation, Visualization, Writing – original draft,

Writing – review & editing. **Brittany Laginhas**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ross Meentemeyer**: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. **Chris Jones**: Conceptualization, Data curation, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Brennen Farrell**: Data curation, Formal analysis, Investigation.

Acknowledgements

This work was made possible by a data sharing agreement (https://google.github.io/auto-arborist). We are grateful to the community scientists who collected images used in the study. We acknowledge the computing resources provided by North Carolina State University's High Performance Computing Services Core Facility (RRID: SCR_022168).

Funding statement

This work is supported by the Agriculture and Food Research Initiative (AFRI), Data Science for Food and Agriculture Systems (DSFAS), project award no. 2022-67021-36465, from the U.S. Department of Agriculture's (USDA) National Institute of Food and Agriculture (NIFA). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

Data accessibility

Code and models are available on Github (*https://github.com/ncsu-landscape-dynamics/gsv_host_detector)* and will be made publicly available upon acceptance.

Tables

Table 1. Characteristics of the 23 cities analyzed in this study. The location of each city is given as latitude and longitude coordinates, along with its assigned geographic region (determined via principal component analysis of bioclimatic variables). Tree diversity is represented by genus richness (the number of tree genera reported in a city's tree inventory) and the Shannon Diversity Index. Additional city-level attributes include human population size (based on the 2020 U.S. Census or the 2021 Canadian Census), mean annual temperature (in °C, BIO1), and total annual precipitation (in mm, BIO12). Bioclimatic variables were obtained from WorldClim version 2.1.

City	Location	Geographic Region	Genus Richness	Shannon Diversity Index	Human Population size	Mean Annual Temp (°C)	Annual Precipitation (mm)
Bloomington, IN	39.165, -86.536	East	47	2.99	79,077	11.6	90.57
Boulder, CO	40.025, -105.277	Central	49	2.95	108,123	9.7	79.86
Buffalo, NY	42.885, -78.874	East	57	2.74	277,551	8.4	91.19
Calgary, AB	51.044, -114.073	Central	28	2.4	1,481,806	3.4	90.23
Cambridge, ON	43.405, -80.329	East	51	2.66	138,479	7.0	94.46
Charlottesville, VA	38.032, -78.482	East	48	3.07	46,469	13.5	81.32
Columbus, OH	39.963, -83.002	East	64	3.26	906,418	10.9	90.9
Cupertino, CA	37.317, -122.046	Southwest	71	3.18	60,111	14.8	37.94
Denver, CO	39.737, -104.991	Central	64	3.4	717,606	10.2	83.5
Edmonton, AB	53.536, -113.495	Central	28	2.42	1,418,118	2.8	108.01
Kitchener, ON	43.447, -80.481	East	26	2.11	256,885	6.6	94.42
Los Angeles, CA	34.052, -118.253	Southwest	82	4.16	3,895,848	18.3	33.99
Montreal, QC	45.512, -73.586	East	56	3.14	1,762,949	6.5	108.04
New York, NY	40.720, -74.001	East	61	3.70	8,740,292	12.0	87.57
Pittsburgh, PA	40.440, -79.995	East	64	2.92	302,779	10.6	87.58
San Francisco, CA	37.768, -122.41	Southwest	83	3.99	870,518	14.1	22.48
San Jose, CA	37.331, -121.88	Southwest	90	4.02	1,009,319	15.4	40.72
Santa Monica, CA	34.023, -118.48	Southwest	60	3.54	92,913	17.2	26.28
Seattle, WA	47.604, -122.333	Northwest	76	3.70	740,565	10.9	48.39
Sioux Falls, SD	43.542, -96.731	Central	35	1.78	193,437	7.4	113.5
Surrey, BC	49.187, -122.846	Northwest	59	3.24	568,322	9.5	50.36
Vancouver, BC	49.279, -123.123	Northwest	66	3.35	662,248	9.9	52.91
Washington, DC	38.898, -77.038	East	56	3.38	689,545	13.2	85.44

Table 2. Model summary of the top generalized linear mixed model (GLMM; by minimum BIC) assessing local ecological and image characteristic predictors of successful prediction of inventory trees from out-of-sample street-level images. Street-level image distance, tree abundance, and genus richness had significant negative effects, whereas congeneric abundance, nearest-neighbor (NN) distance, training image ratio, and the number of training images per genus showed positive effects.

Fixed Effects					
Parameter	OR	SE	95% CI	Wald z	Sig.
Intercept	2.70	0.48	[1.91, 3.82]	5.59	***
ℓ(SL Img Distance)	0.67	0.01	[0.65, 0.68]	-37.97	***
ℓ(Tree Abudance)	0.74	0.02	[0.71, 0.77]	-13.10	***
ℓ(Genus Richness)	0.93	0.02	[0.90, 0.97]	-3.67	***
ℓ(Congeneric Tree Abundance)	1.42	0.02	[1.37, 1.46]	20.45	***
ℓ(NN Tree Dist)	1.44	0.02	[1.41, 1.48]	28.75	***
ℓ(N SL Img)	1.55	0.02	[1.51, 1.58]	39.01	***
Prop. AA Genus-Lvl Training Img (focal city)	1.60	0.03	[1.53, 1.67]	22.11	***
ℓ(AA Genus-Level Training Img)	2.01	0.16	[1.71, 2.35]	8.63	***

Random (Intercept) Effects				
Grouping Factor	N	Var.	SD	% Total Var.
Grid (City)	5864	0.62	0.78	12.7
Genus	100	0.78	0.89	10.0
City	23	0.46	0.46	3.5
Residual	-	-	-	73.8

Model Fit	
Metric	Value
N	90689
Degrees of Freedom	12
Marginal R^2	0.205
Conditional R^2	0.467

Notes: OR = odds ratio; SE = standard error; CI = confidence interval; Sig. = Significant; NS = Not significant; ℓ = log(x + 1); AA = AutoArborist; SL = Street-Level

*Significance codes: NS p>0.05; p<0.05; ** p<0.01; *** p<0.001*

Figure Legends

Figure 1. Workflow for computer vision-based tree inventories. A) Data collection: Images for 100 tree genera were sourced from the AutoArborist and iNaturalist datasets. AutoArborist images are obtained from Google Street View (GSV) and capture tree genera from 23 cities across the United States and Canada. iNaturalist images are sourced from research-grade observations in iNaturalist.org. B) Detection, classification, and geolocation: We developed and implemented an object detection (YOLOv5x) and image classification (EfficientNetV2S) model to identify tree genera from GSV images. Once identified and classified, we used a depth estimation model (ZoeDepth) to assign geographic coordinates to predicted trees. We applied the workflow to independent portions of each city to assess the generalizability of computer vision models.

Figure 2. Representative gradient-weighted class activation maps (Grad-CAMs) for six tree genera classified by the convolutional neural network (CNN). Each row shows three AutoArborist images (left) and three corresponding Grad-CAM heatmaps (right) for one genus: (A) *Acer* (Maple), (B) *Betula* (Birch), (C) *Citrus* (Citrus), (D) *Phoenix* (Date Palm), (E) *Pinus* (Pine), and (F) *Thuja* (Arborvitae). Grad-CAM visualizations highlight the image regions most influential for model predictions, with darker red colors indicating higher importance and blue indicating low importance. The CNN leveraged biologically meaningful and visually distinctive features, including reproductive traits (e.g., see white arrow: fruit on *Citrus*, panel C), structural characteristics (e.g., canopy shape and branching in *Acer*, *Betula*, and *Phoenix*, panels A–D), and bark texture (e.g., reddish-brown furrowed bark in *Pinus* and *Thuja*, and papery white bark in *Betula*, panels B, E, F). These results show that the model captured diagnostic morphological

features consistent with expert knowledge, supporting the interpretability of genus-level classifications.

Figure 3. Relationship between the Shannon diversity index and image classification model performance across 23 cities. We calculated the Shannon diversity index for each city using records of tree genera from the AutoArborist tree inventory dataset. We applied the image classification model to the 10% randomly withheld testing dataset for each city and calculated the weighted average F1 score across all genera as a metric of classification performance. A scatterplot illustrates the negative relationship between Shannon diversity and model performance, where cities with a greater diversity of tree genera have lower classification performance. The teal line indicates the predicted values, and the shaded area represents the 95% prediction interval.

Figure 4. Classification performance for maples (*Acer*) varies by month and geographic region. Boxplots show the distribution of F1 scores for *Acer* classification across months of image capture (left) and geographic regions (right). Regions were defined based on principal component analysis of bioclimatic variables, grouping cities by climate similarity. Each point represents the F1 score for testing images from a specific city-month combination within the 30% geographic testing split. Significant differences between months and regions were assessed using ANOVA and Tukey's post-hoc tests, with letters above boxplots indicating statistically distinct groups.

Figure 5**.** Model selection and fixed-effect summaries from generalized linear mixed models (GLMMs). Only the full model and submodels with a ΔBIC < 10 are shown. (A) Relative BIC weights (bars, left axis) and ΔBIC values (line, right axis) for each GLMM. (B) Heatmap of fixed-effect coefficients (odds ratio) with striped tiles denoting predictors excluded from a given model. Numeric predictors were log-transformed as [ℓ - log(x+1)]. (C) Partial effect plots showing predicted probabilities of tree matching across each predictor's observed range, with other predictors held at their mean. Solid lines and shaded ribbons represent model predictions and 95% confidence intervals, respectively. Predictor values were back-transformed to the original scale. Predictions were computed from the inverse link function with bias correction to account for Jensen's inequality. Abbreviations: SL = street-level; AA = AutoArborist.

Figures

Figure 1.

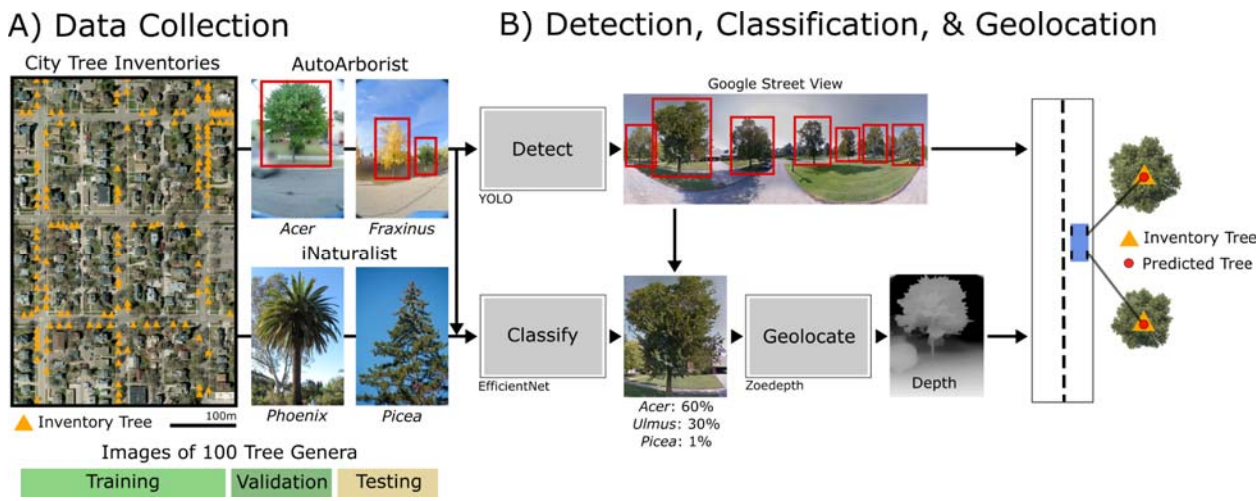


Figure 2.

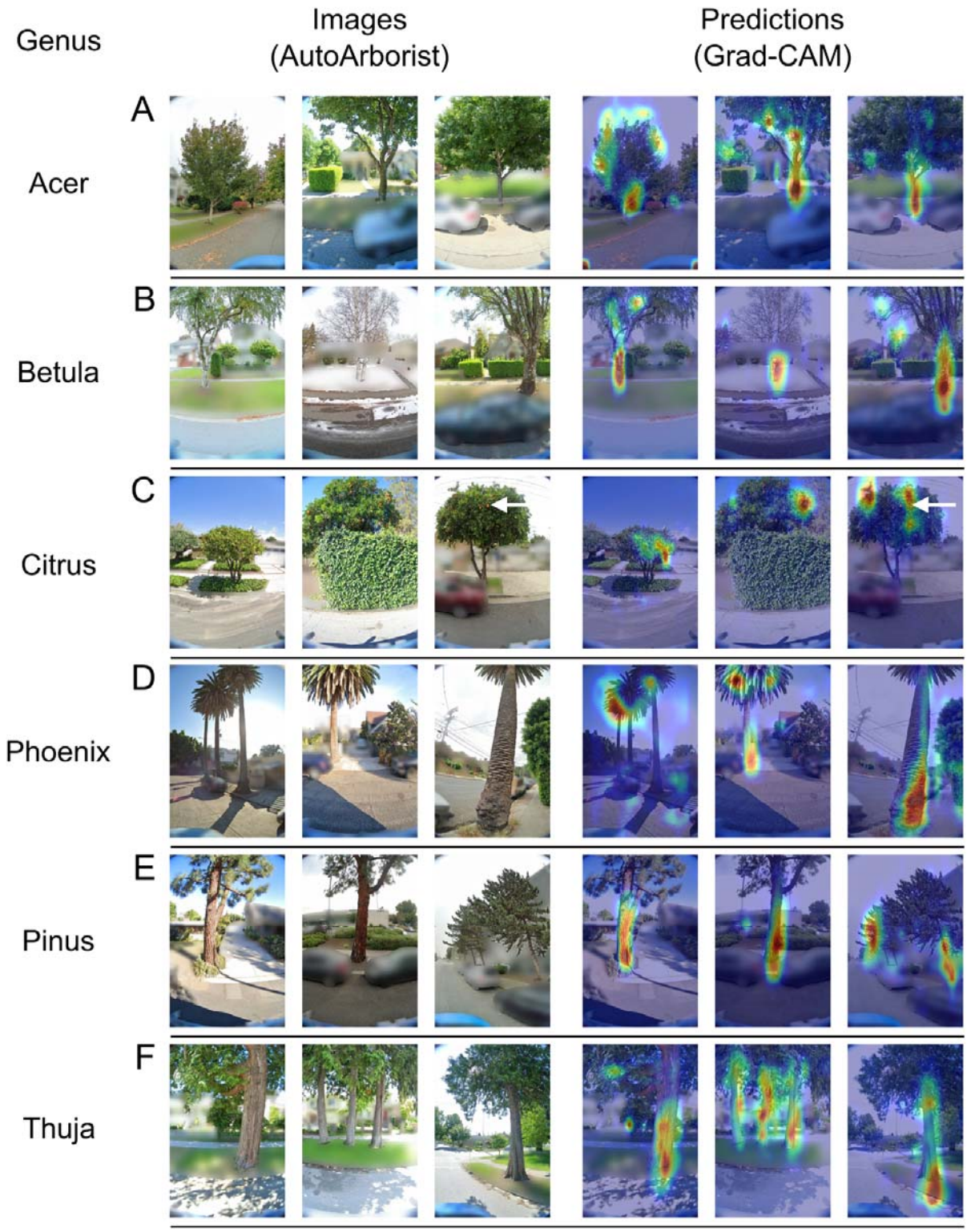

Genus
Images
(AutoArborist)
Predictions
(Grad-CAM)
A
Acer
B
Betula
C
Citrus
D
Phoenix
E
Pinus
F
Thuja

Figure 3.

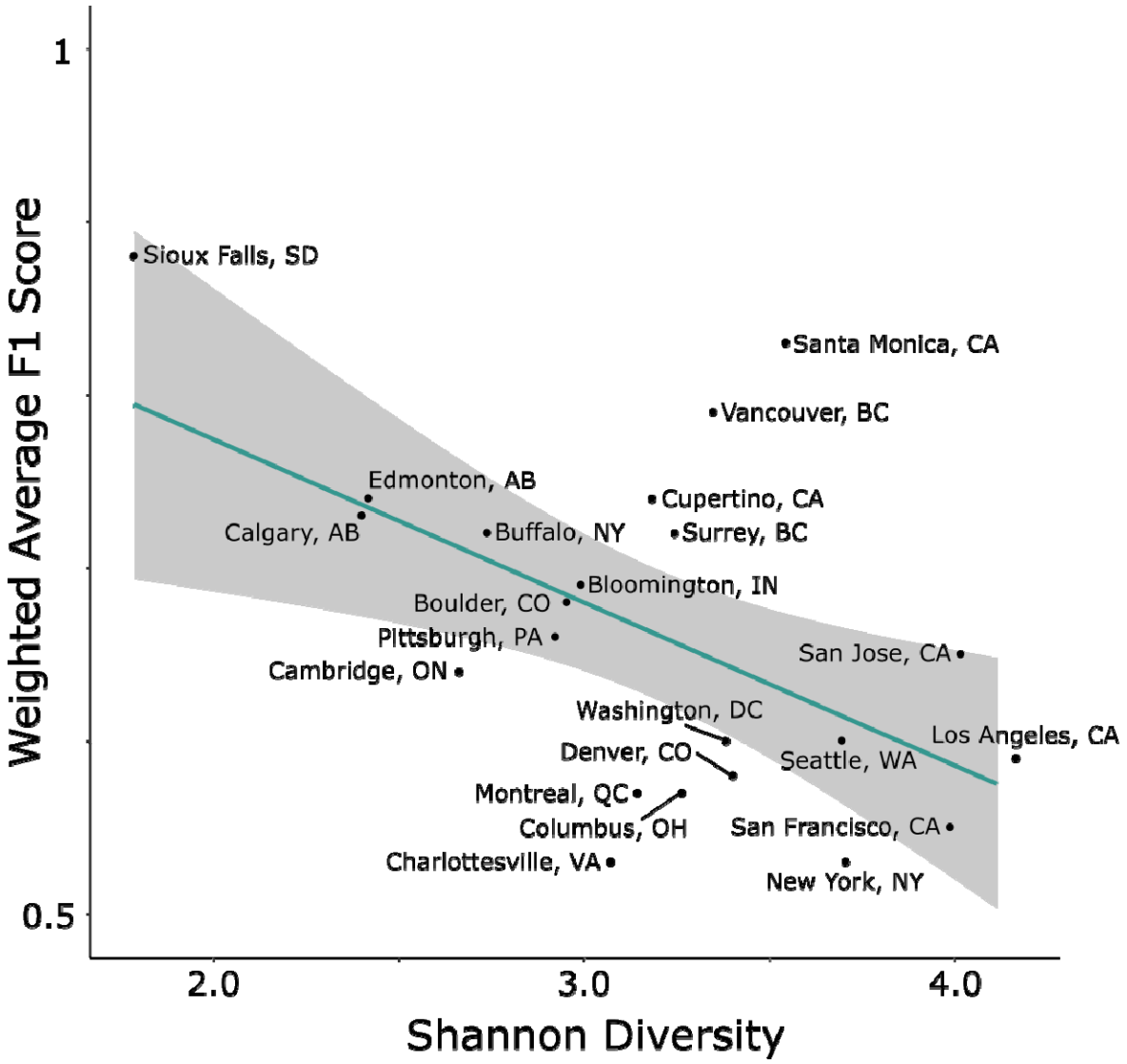

Weighted Average F1 Score
1
0.5
2.0
3.0
4.0
Shannon Diversity
Sioux Falls, SD
Santa Monica, CA
Vancouver, BC
Edmonton, AB
Cupertino, CA
Calgary, AB
Buffalo, NY
Surrey, BC
Bloomington, IN
Boulder, CO
Pittsburgh, PA
San Jose, CA
Cambridge, ON
Washington, DC
Los Angeles, CA
Denver, CO
Seattle, WA
Montreal, QC
Columbus, OH
San Francisco, CA
Charlottesville, VA
New York, NY

Figure 4.

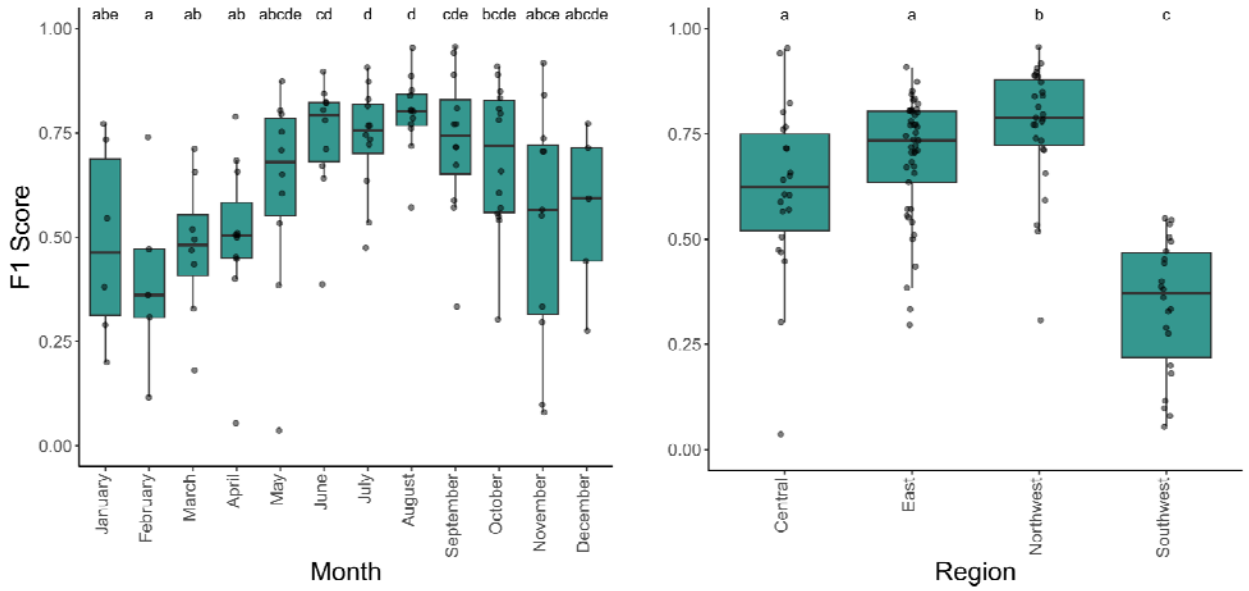
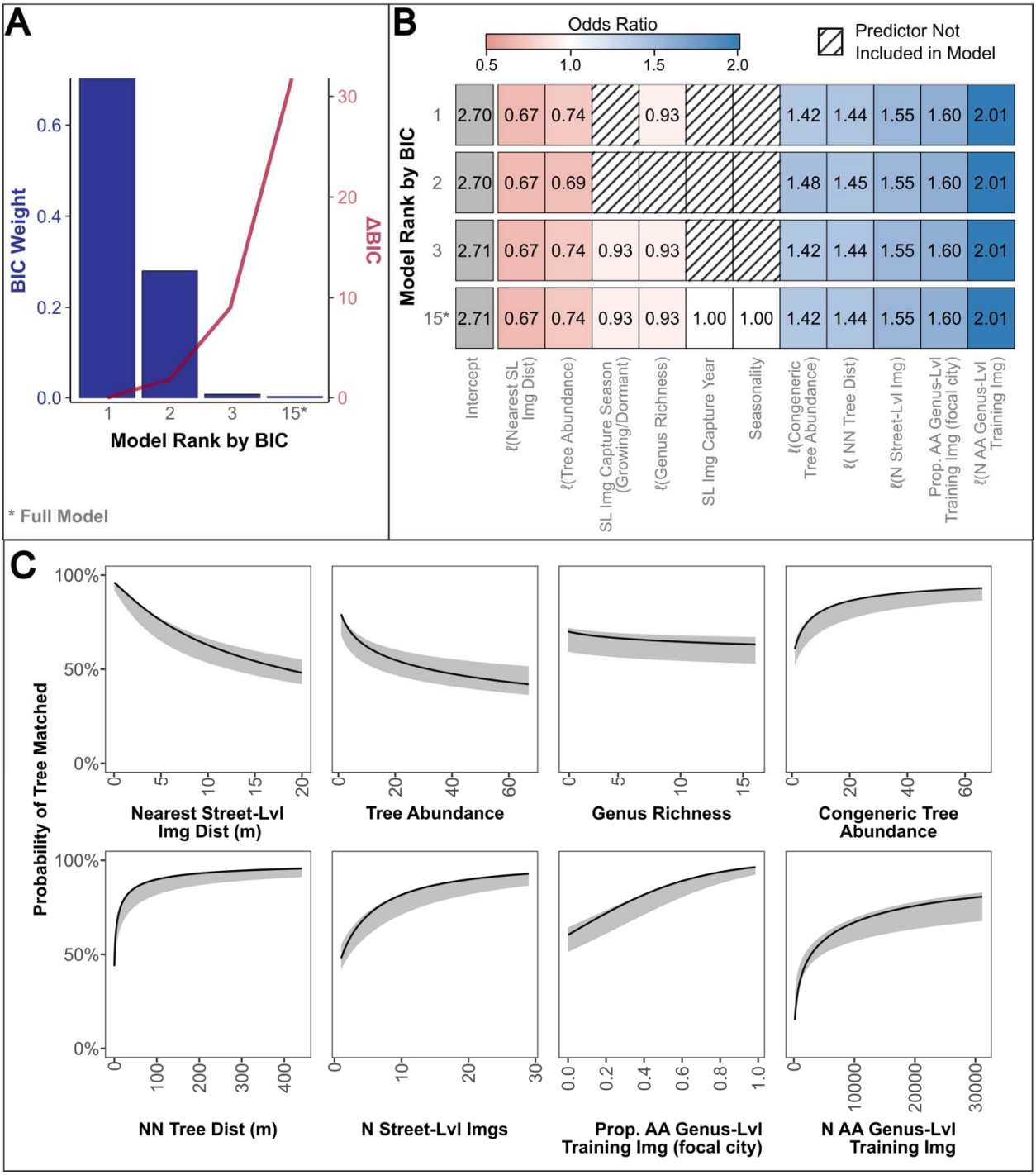

abe
a
ab
ab
abcde
cd
d
d
cde
bcde
abce
abcde
1.00
0.75
0.50
0.25
0.00
F1 Score
January
February
March
April
May
June
July
August
September
October
November
December
Month
a
a
b
c
Central
East
Northwest
Southwest
Region

Figure 5.


A
BIC Weight
ΔBIC
Model Rank by BIC
* Full Model
B
Odds Ratio
Predictor Not Included in Model
Model Rank by BIC
Intercept
ℓ(Nearest SL Img Dist)
ℓ(Tree Abundance)
SL Img Capture Season (Growing/Dormant)
ℓ(Genus Richness)
SL Img Capture Year
Seasonality
ℓ(Congeneric Tree Abundance)
ℓ(NN Tree Dist)
ℓ(N Street-Lvl Img)
Prop. AA Genus-Lvl Training Img (focal city)
ℓ(N AA Genus-Lvl Training Img)
C
Probability of Tree Matched
Nearest Street-Lvl Img Dist (m)
Tree Abundance
Genus Richness
Congeneric Tree Abundance
NN Tree Dist (m)
N Street-Lvl Imgs
Prop. AA Genus-Lvl Training Img (focal city)
N AA Genus-Lvl Training Img

References

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, 148, 70–83.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google Street View: Capturing the World at Street Level. *Computer*, 43(6), 32–38.

Aodha, O. M., Cole, E., & Perona, P. (2019). Presence-only geographical priors for fine-grained image classification. *IEEE International Conference on Computer Vision*, 9595–9605.

Arevalo-Ramirez, T., Alfaro, A., Figueroa, J., Ponce-Donoso, M., Saavedra, J. M., Recabarren, M., & Delpiano, J. (2024). Challenges for computer vision as a tool for screening urban trees through street-view images. *Urban Forestry & Urban Greening*, 95, 128316.

Augustinus, B. A., Abegg, M., Queloz, V., & Brockerhoff, E. G. (2024). Higher tree species richness and diversity in urban areas than in forests: Implications for host availability for invasive tree pests and pathogens. *Landscape and Urban Planning*, 250, 105144.

Bartoń, K. (2024). Multi-Model Inference [R package MuMIn version 1.48.4]. https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48.

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in terra incognita. *Proceedings of the European Conference on Computer Vision (ECCV)*, 456–473.

Beery, S., Wu, G., Edwards, T., Pavetic, F., Majewski, B., Mukherjee, S., Chan, S., Morgan, J., Rathod, V., & Huang, J. (2022). The Auto Arborist Dataset: A Large-Scale Benchmark

for Multiview Urban Forest Monitoring Under Domain Shift. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21294–21307.

Berland, A., & Lange, D. A. (2017). Google Street View shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening*, 21, 11–15.

Berland, A., Roman, L. A., & Vogt, J. (2019). Can field crews telecommute? Varied data quality from citizen science tree inventories conducted using street-level imagery. *Forests*, 10(4), 349.

Bhat, S. F., Birkl, R., Wofk, D., Wonka, P., & Müller, M. (2023). ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth. *In arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/2302.12288

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, 104217.

Bivand, R. (2022). R packages for analyzing spatial data: A comparative case study with areal data. *Geographical Analysis*, 54(3), 488–518.

Bloniarz, D. V. (1996). The use of volunteer initiatives in conducting urban forest resource inventories. Thesis. University of Massachusetts Amherst.

Boucher□Lalonde, V., Morin, A., & Currie, D. J. (2012). How are tree species distributed in climatic space? A simple and general pattern. *Global Ecology and Biogeography*, 21(12), 1157-1166.

Branson, S., Wegner, J. D., Hall, D., Lang, N., Schindler, K., & Perona, P. (2018). From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135, 13–30.

Capecchi, I., Borghini, T., & Bernetti, I. (2023). Automated urban tree survey using remote sensing data, Google street view images, and plant species recognition apps. *European Journal of Remote Sensing*, 56(1), 2162441.

Choi, K., Lim, W., Chang, B., Jeong, J., Kim, I., Park, C.-R., & Ko, D. W. (2022). An automatic approach for tree species detection and profile estimation of urban street trees using deep learning and Google street view images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190, 165–180.

Cole, E., Van Horn, G., Lange, C., Shepard, A., Leary, P. R., Perona, P., Loarie, S., & Aodha, O. M. (2023). Spatial Implicit Neural Representations for global-scale species mapping. *International Conference on Machine Learning*, 6320–6342.

Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2).

Crown, C. A., Greer, B. Z., Gift, D. M., & Watt, F. S. (2018). Every tree counts: Reflections on NYC's third volunteer street tree inventory. *Arboriculture & Urban Forestry*, 44(2), 49–58.

Edgar, C. B., Nowak, D. J., Majewsky, M. A., Lister, T. W., Westfall, J. A., & Sonti, N. F. (2021). Strategic National Urban Forest Inventory for the United States. *Journal of Forestry*, 119(1), 86–95.

Erker, T., Wang, L., Lorentz, L., Stoltman, A., & Townsend, P. A. (2019). A statewide urban tree canopy mapping method. *Remote Sensing of Environment*, 229, 148–158.

Escobedo, F. J., Kroeger, T., & Wagner, J. E. (2011). Urban forests and pollution mitigation: analyzing ecosystem services and disservices. *Environmental Pollution*, 159(8–9), 2078–2087.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2), 303-338.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1□km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12), 4302-4315.

Galle, N. J., Halpern, D., Nitoslawski, S., Duarte, F., Ratti, C., & Pilla, F. (2021). Mapping the diversity of street tree inventories across eight cities internationally using open data. *Urban Forestry & Urban Greening*, 61(127099), 127099.

Gillespie, L. E., Ruffley, M., & Exposito-Alonso, M. (2024). Deep learning models map rapid plant species changes from citizen science and remote sensing data, *Proc. Natl. Acad. Sci. U.S.A.* 121 (37) e2318296121.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

Godard, C., Aodha, O. M., Firman, M., & Brostow, G. (2018). Digging Into Self-Supervised Monocular Depth Estimation. *In arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1806.01260

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). *MIT press Cambridge.*

Hargrave, J. R., Harper, R. W., Butler, B. J., & Mullins, J. T. (2022). Municipal Forest Program Management in the United States of America: A Systematic Review. *Forests, Trees and Livelihoods*, 14(1), 35.

Hartig, F. (2024). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. *CRAN: Contributed Packages*.

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., & Carron, J. (2019). Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors*, 19(6).

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M. M. A., Yang, Y., & Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. *In arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1712.00409

Hijmans, R. J., Bivand, R., Forner, K., Ooms, J., Pebesma, E., & Sumner, M. D. (2022). Package 'terra'. Maintainer: Vienna, Austria, 384.

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., Shortridge, A., & Others. (2015). Package 'raster.' R Package, 734, 473.

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432.

Jensen, R. R., Hardin, P. J., & Hardin, A. J. (2012). Classification of urban tree species using hyperspectral imagery. *Geocarto International*, 27(5), 443–458.

Jha, R. K., Nölke, N., Diwakara, B. N., Tewari, V. P., & Kleinn, C. (2019). Differences in tree species diversity along the rural-urban gradient in Bengaluru, India. *Urban Forestry & Urban Greening*, 46(126464), 126464.

Jocher, G., Ayush Chaurasia, Stoken, A., Borovec, J., NanoCode012, Yonghye Kwon, Kalen Michael, TaoXie, Jiacong Fang, Imyhxy, Lorna, 曾逸夫(Zeng Yifu), Wong, C., Abhiram V, Montes, D., Zhiqiang Wang, Fati, C., Jebastin Nadar, Laughing, … Jain, M. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. *Zenodo*. https://doi.org/10.5281/ZENODO.3908559

Jones, C. M., Jones, S., Petrasova, A., Petras, V., Gaydos, D., Skrip, M. M., Takeuchi, Y., Bigsby, K., & Meentemeyer, R. K. (2021). Iteratively forecasting biological invasions with PoPS and a little help from our friends. *Frontiers in Ecology and the Environment*, 19(7), 411–418.

Jones, C., Skrip, M. M., Seliger, B. J., Jones, S., Wakie, T., Takeuchi, Y., Petras, V., Petrasova, A., & Meentemeyer, R. K. (2022). Spotted lanternfly predicted to establish in California by 2033 without preventative management. *Communications Biology*, 5(1), 558.

Kagan, D., Fuhrmann Alpert, G., & Fire, M. (2021). Automatic large scale detection of red palm weevil infestation using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 182, 122–133.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 24-49.

Kim, T. K., Hong, J., Ryu, D., Kim, S., Byeon, S. Y., Huh, W., ... & Kim, H. S. (2022). Identifying and extracting bark key features of 42 tree species using convolutional neural networks and class activation mapping. *Scientific Reports*, 12(1), 4772.

Kim, J., & Jang, K. M. (2023). An examination of the spatial coverage and temporal variability of Google Street View (GSV) images in small- and medium-sized cities: A people-based approach. *Computers, Environment and Urban Systems*, 102(101956), 101956.

Kotowska, D., Pärt, T., & Żmihorski, M. (2021). Evaluating Google Street View for tracking invasive alien plants along roads. Ecological Indicators, 121, 107020.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Krylov, V. A., Kenny, E., & Dahyot, R. (2018). Automatic Discovery and Geotagging of Objects from Street View Imagery. *Remote Sensing*, 10(5), 661.

Lake, T. A., Briscoe Runquist, R. D., & Moeller, D. A. (2022). Deep learning detects invasive plant species across complex landscapes using Worldview□2 and Planetscope satellite imagery. *Remote Sensing in Ecology and Conservation*, 8(6), 875-889.

Laumer, D., Lang, N., van Doorn, N., Aodha, O. M., Perona, P., & Wegner, J. D. (2020). Geocoding of trees from street addresses and street-level images. *ISPRS Journal of Photogrammetry and Remote Sensing, 162*, 125–136.

Lenth, R. V. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means. https://rvlenth.github.io/emmeans/

Li, X., Chen, W. Y., Sanesi, G., & Lafortezza, R. (2019). Remote sensing in urban forestry: Recent applications and future directions. *Remote Sensing*, 11(10), 1144.

Liu, D., Jiang, Y., Wang, R., & Lu, Y. (2023). Establishing a citywide street tree inventory with street view images and computer vision techniques. *Computers, Environment and Urban Systems*, 100, 101924.

Love, N. L. R., Nguyen, V., Pawlak, C., Pineda, A., Reimer, J. L., Yost, J. M., Fricker, G. A., Ventura, J. D., Doremus, J. M., Crow, T., & Ritter, M. K. (2022). Diversity and structure in California's urban forest: What over six million data points tell us about one of the world's largest urban forests. *Urban Forestry & Urban Greening*, 74, 127679.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139.

Lumnitz, S., Devisscher, T., Mayaud, J. R., Radic, V., Coops, N. C., & Griess, V. C. (2021). Mapping trees along urban street networks with deep learning and street-level imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 144–157.

Ma, B., Hauer, R. J., Östberg, J., Koeser, A. K., Wei, H., & Xu, C. (2021). A global basis of urban tree inventories: What comes first the inventory or the program. *Urban Forestry & Urban Greening*, 60, 127087.

Ma, B., Hauer, R. J., Wei, H., Koeser, A. K., Peterson, W., Simons, K., Timilsina, N., Werner, L. P., & Xu, C. (2020). An assessment of street tree diversity: Findings and implications in the United States. *Urban Forestry & Urban Greening*, 56(126826), 126826.

Marconi, S., Weinstein, B. G., Zou, S., Bohlman, S. A., Zare, A., Singh, A., Stewart, D., Harmon, I., Steinkraus, A., & White, E. P. (2022). Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network. *Remote Sensing of Environment*, 282, 113264.

McCoy, D. E., Goulet-Scott, B., Meng, W., Atahan, B. F., Kiros, H., Nishino, M., & Kartesz, J. (2022). Species clustering, climate effects, and introduced species in 5 million city trees across 63 US cities. *ELife*, 11.

McPherson, E. G., van Doorn, N., & de Goede, J. (2016). Structure, function and value of street trees in California, USA. *Urban Forestry & Urban Greening*, 17, 104–115.

Miller, R. W., Hauer, R. J., & Werner, L. P. (2015). Urban Forestry: Planning and Managing Urban Greenspaces, Third Edition. *Waveland Press.*

Nagendra, H. (2001). Using remote sensing to assess biodiversity. International Journal of *Remote Sensing*, 22(12), 2377–2400.

Nielsen, A., Östberg, J., & Delshammar, T. (2014). Review of urban tree inventory methods used to collect data at single-tree level. *Arboriculture & Urban Forestry*, 40(2).

Nowak, D. J., Crane, D. E., Stevens, J. C., Hoehn, R. E., Walton, J. T., & Bond, J. (2008). A Ground-Based Method of Assessing Urban Forest Structure and Ecosystem Services. *Arboriculture & Urban Forestry*, 34(6), 347–358.

Nowak, D. J., & Dwyer, J. F. (2007). Understanding the benefits and costs of urban forest ecosystems. *In Urban and community forestry in the northeast* (pp. 25-46). Dordrecht: Springer Netherlands.

Nowak, D. J., & Greenfield, E. J. (2018). US urban forest statistics, values, and projections. *Journal of Forestry*, 116(2), 164–177.

Nowak, D. J., Hoehn, R. E., Bodine, A. R., Greenfield, E. J., & O'Neil-Dunne, J. (2016). Urban forest structure, ecosystem services and change in Syracuse, NY. *Urban Ecosystems*, 19(4), 1455–1477.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., & Oksanen M. (2013). Package 'vegan.' *Community Ecology Package*, Version, 2(9), 1–295.

Pretzsch, H., Biber, P., Uhl, E., Dahlhausen, J., Schütze, G., Perkins, D., Rötzer, T., Caldentey, J., Koike, T., van Con, T., Chavanne, A., Toit, B. du, Foster, K., & Lefer, B. (2017).

Climate change accelerates growth of urban trees in metropolises worldwide. *Scientific Reports*, 7(1), 15403.

Pu, R., & Landry, S. (2020). Mapping urban tree species by integrating multi-seasonal high resolution pléiades satellite imagery with airborne LiDAR data. *Urban Forestry & Urban Greening*, 53, 126675.

Pu, R., Landry, S., & Yu, Q. (2018). Assessing the potential of multi-seasonal high resolution Pléiades satellite imagery for mapping urban tree species. *International Journal of Applied Earth Observation and Geoinformation*, 71, 144–158.

R Core Team. (2025). R: A Language and Environment for Statistical Computing.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *International conference on machine learning* (pp. 8748-8763). PmLR.

Raum, S., Collins, C. M., Urquhart, J., Potter, C., Pauleit, S., & Egerer, M. (2023). Tree insect pests and pathogens: a global systematic review of their impacts in urban areas. *Urban Ecosystems*, 26(2), 587–604.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.

Roberts, J., Koeser, A., Abd-Elrahman, A., Wilkinson, B., Hansen, G., Landry, S., & Perez, A. (2019). Mobile terrestrial photogrammetry for street tree mapping and measurements. *Forests*, 10(8), 701.

Roman, L. A., McPherson, E. G., Scharenbroch, B. C., & Bartens, J. (2013). Identifying common practices and challenges for local urban tree monitoring programs across the United States. *Arboriculture & Urban Forestry*. 39 (6): 292-299, 39(6), 292–299.

Roman, Lara A., Pearsall, H., Eisenman, T. S., Conway, T. M., Fahey, R. T., Landry, S., Vogt, J., van Doorn, N. S., Grove, J. M., Locke, D. H., Bardekjian, A. C., Battles, J. J., Cadenasso, M. L., van den Bosch, C. C. K., Avolio, M., Berland, A., Jenerette, G. D., Mincey, S. K., Pataki, D. E., & Staudhammer, C. (2018). Human and biophysical legacies shape contemporary urban forests: A literature synthesis. *Urban Forestry & Urban Greening*, 31, 157–168.

Roman, Lara A., Scharenbroch, B. C., Östberg, J. P. A., Mueller, L. S., Henning, J. G., Koeser, A. K., Sanders, J. R., Betz, D. R., & Jordan, R. C. (2017). Data quality in citizen science urban tree inventories. *Urban Forestry & Urban Greening*, 22, 124–135.

Roy, S., Byrne, J., & Pickering, C. (2012). A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. *Urban Forestry & Urban Greening*, 11(4), 351–363.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Seiferling, I., Naik, N., Ratti, C., & Proulx, R. (2017). Green streets − Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, 165, 93–101.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Visual explanations from deep networks via gradient-based localization*. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.

Soltani, S., Feilhauer, H., Duker, R., & Kattenborn, T. (2022). Transfer learning from citizen science photographs enables plant species identification in UAV imagery. ISPRS Open Journal of Photogrammetry and Remote Sensing, 5, 100016.

Soltani, S., Ferlian, O., Eisenhauer, N., Feilhauer, H., & Kattenborn, T. (2024). From simple labels to semantic image segmentation: leveraging citizen science plant photographs for tree species mapping in drone imagery. Biogeosciences, 21(11), 2909-2935.

Tan, M., & Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. *International conference on machine learning* (pp. 10096-10106). PMLR.

Terven, J., & Cordova-Esparza, D. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *In arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/2304.00501

Tubby, K. V., & Webber, J. F. (2010). Pests and diseases threatening urban trees under a changing climate. *Forestry*, 83(4), 451–459.

Van Horn, G., & Perona, P. (2017). The Devil is in the Tails: Fine-grained Classification in the Wild. *In arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1709.01450

Velasquez-Camacho, L., Cardil, A., Mohan, M., Etxegarai, M., Anzaldi, G., & de-Miguel, S. (2021). Remotely sensed tree characterization in urban areas: A review. *Remote Sensing*, 13(23), 4889.

Velasquez-Camacho, L., Merontausta, E., Etxegarai, M., & de-Miguel, S. (2024). Assessing urban forest biodiversity through automatic taxonomic identification of street trees from

citizen science applications and remote-sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, 128, 103735.

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. (2024). Individual tree detection in large-scale urban environments using high-resolution multispectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 130(103848), 103848.

Wang, S., Liu, C., Wei, S., & Tang, H. (2025). Urban tree species classification using multisource satellite remote sensing data and street view imagery. *Geo-spatial Information Science*, 28(1), 164-184.

Ward, K. T., & Johnson, G. R. (2007). Geospatial methods provide timely and comprehensive urban forest information. *Urban Forestry & Urban Greening*, 6(1), 15–22.

Wegner, J. D., Branson, S., Hall, D., Schindler, K., & Perona, P. (2016). Cataloging public objects using aerial and street-level images — urban trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6014-6023).

Weinstein, B. G., Marconi, S., Graves, S. J., Zare, A., Singh, A., Bohlman, S. A., Magee, L., Johnson, D. J., Townsend, P. A., & White, E. P. (2023). Capturing long□tailed individual tree diversity using an airborne imaging and a multi□temporal hierarchical model. *Remote Sensing in Ecology and Conservation*, 9(5), 656–670.

Zhang, K., & Hu, B. (2012). Individual Urban Tree Species Classification Using Very High Spatial Resolution Airborne Multi-Spectral Imagery Using Longitudinal Profiles. *Remote Sensing*, 4(6), 1741–1757.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

Zhou, K., Liu, Z., Qiao, Y., Xiang, T., & Loy, C. C. (2023). Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4), 4396–4415.