TITLE: Continental-scale computer vision models reveal generalizable patterns and pitfalls for urban tree inventories with street-view images

RUNNING TITLE: Generalizable computer vision models for urban tree inventories

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

Word count: 7926

ABSTRACT

Accurate, up-to-date catalogs of urban tree populations are crucial for quantifying ecosystem services and enhancing the quality of life in cities. However, mapping tree species cost-effectively remains challenging. In response, remote sensing researchers are developing general-purpose tools to survey 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. Our computer vision models identified most ground-truthed street trees (67.1%). Performance varied across the 23 cities (67.4% ± 9.3%) and 100 genera (50.9% ± 23.0%) and improved denser street-view coverage, simpler 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. Using widely available street-level imagery is a generalizable and promising avenue for mapping tree 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 and quality of life in cities (Edgar et al., 2021; Nowak and Dwyer, 2007; Roy et al., 2012). These forests provide necessary ecosystem services, including climate regulation, stormwater management, and pollution sequestration, valued at over $18 billion annually in the United States (Nowak and Dwyer, 2007; Nowak and Greenfield, 2018). Understanding tree distribution, species composition, and health is essential for quantifying and managing these benefits (Ma et al., 2021; McPherson et al., 2016). Traditionally, field surveys are used to document forest conditions, but collecting data on individual trees is labor-intensive and quickly becomes outdated due to urban development and pest or disease impacts (Nielsen et al., 2014; Pretzsch et al., 2017; Raum et al., 2023; Tubby and Webber, 2010). As such, there is a growing need for efficient, scalable approaches to monitor urban forests.

Tree inventories remain 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 and forecasting pest or pathogen risk (Hargrave et al., 2022; Jones et al., 2021; Jones et al., 2022; Ma et al., 2021; Miller et al., 2015). However, few communities maintain updated inventories (Hargrave et al., 2022). National programs such as the USDA Urban Forest Inventory and Analysis provide systematic plot-based estimates of urban forest characteristics (Edgar et al., 2021), but not a full census. Consequently, many cities rely on opportunistic surveys and community science observations to supplement records and track forest conditions (Bloniarz, 1996; Crown et al., 2018; Roman et al., 2017). Researchers are

increasingly turning to remote sensing and machine learning for more comprehensive and timely assessments of urban trees.

Advances in remote sensing have improved our ability to map and monitor urban forests (Fassnacht et al., 2016; Ward and Johnson, 2007). Tree species with distinct morphological, phenological, and spectral traits can be identified from aerial and satellite imagery (Fassnacht et al., 2016; Nagendra, 2001). Numerous studies have combined high-resolution images with light detection and ranging (LiDAR) to inventory urban tree species (Hartling et al., 2019; Pu et al., 2018; Pu and Landry, 2020; Zhang and Hu, 2012). Yet, extending these approaches beyond individual cities remains limited by the cost of fine-resolution data and inconsistent image quality, acquisition timing, and coverage (Velasquez-Camacho et al., 2021). As a result, these methods are typically applied at local scales (e.g., individual parks, neighborhoods, or cities) and rarely extended across multiple urban areas for broader comparison (Beery et al., 2022; Li et al., 2019; Velasquez-Camacho et al., 2021).

Street-level imagery provides a complementary perspective by offering ground-level views of urban trees (Berland and Lange, 2017; Berland et al., 2019; Roberts et al., 2019; Seiferling et al., 2017). Google Street View (GSV) captures high-resolution panoramas every 10-20 meters along roads worldwide (Anguelov et al., 2010; Biljecki and Ito, 2021). Similar platforms (e.g., Mapillary; https://www.mapillary.com/), provide analogous views with crowdsourced imagery. These images capture canopy and tree characteristics, including leaf arrangement, branching, and bark texture, that are typically obscured in aerial imagery. Yet, most studies examine individual neighborhoods or cities (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), and it remains unclear how well street-level surveys can generalize across broader regions.

Machine learning models are increasingly used to detect and classify plant species from remotely-sensed imagery, 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 region but often perform less reliably when applied to new areas (Beery et al., 2022; Marconi et al., 2022). Model generalizability is constrained by environmental and seasonal variation, and by factors related to image and site characteristics (Capecchi et al., 2023; Choi et al., 2022; Kim and Jang, 2023; Liu et al., 2023; Ventura et al., 2024; Velasquez-Camacho et al., 2024). As a result, developing models that generalize reliably across regions remains a 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 GSV and iNaturalist images. We trained tree detection and classification models 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 variation in accuracy, and (3) determine how ecological conditions and street-view image coverage influence model errors and generalization across cities. We used model predictions to generate 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 U.S. and Canadian cities. These cities have existing tree inventories, GSV imagery, and span diverse geographic, ecological, and socioeconomic contexts. Cities with smaller populations in northern, continental climates (e.g., Sioux Falls, South Dakota, and Kitchener, Ontario) have lower tree diversity, 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 higher tree diversity 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 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 genera to detect, classify, and geolocate across 23 cities (Table S1). We worked at the genus level to avoid distinguishing between numerous varieties, hybrids, and subspecies commonly found in urban areas. Tree species and genus diversity follows a long-tailed distribution, where a few taxa are abundant and most are infrequent or rare (Love et al., 2022; Ma et al., 2020; McCoy et al., 2022). Our selected genera account for over 95% of all genera reported in the 23 inventories. These include taxa that are widely distributed (e.g., *Acer* [maple], *Fraxinus* [ash]), 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 AutoArborist tree genera images

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

2.2.3 iNaturalist tree genera images

We augmented AutoArborist with iNaturalist images, a platform hosting millions of user-annotated plant images. We downloaded all available 'research-grade' images of our 100 genera captured in North America from their 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 community-sourced with images taken from a wide range of viewpoints. Many are close-range (i.e., within one meter) photos of leaves, twigs, flowers, and fruits. While these features are important for taxonomic identification, they do not match the street-level perspective of GSV, where trunk and canopy are visible. Prior work shows that mixing close-ups with broader views (e.g., UAV or satellite) introduces visual inconsistencies

that confuse computer vision models and degrade performance (Beery et al., 2022; Soltani et al., 2022; Soltani et al., 2024). Accordingly, we prioritized iNaturalist images with perspectives more consistent with GSV.

We applied a Contrastive Language-Image Pre-Training (CLIP) deep learning model to filter iNaturalist images (Radford et al., 2021). This model was trained on millions image-text pairs and ranks images based on alignment to a text prompt (Radford et al., 2021). We queried "a photo of a large mature tree in an urban landscape" to retain images consistent with the GSV perspective (Figure S1). We capped the number of retained images per genus to limit class imbalance. For genera with more than 10,000 images (e.g., *Acer* [maple]; 303,000 images), we kept the top 1,000 images based on prompt similarity. For genera with fewer than 10,000 images, we kept the top 10% of images. This ensured iNaturalist served as a supplement rather than a replacement for AutoArborist imagery. The curated subset totaled 59,132 images, which we incorporated into the training dataset for the tree classification model (see Section 2.3.2; Figure S1; Table S1).

2.3 Tree detection, classification, and geolocation models

2.3.1 Tree detection

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

We trained the object detection model on bounding box annotated tree images from AutoArborist (Beery et al., 2022). Because our goal was to build a single-class (tree) object detector that generalizes across all cities, we randomly sampled images across all genera to capture variation in tree appearance and image conditions. We sampled 750,000 images for training, 75,000 images for validation to fine-tune model parameters during training, and 75,000 images for testing. To improve generalization, we applied image augmentations (random horizontal flips, rotations, color shifts) during training to vary the appearance, scale, and illumination.

We trained 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 predicted and annotated boxes. Detections with an IoU of at least 0.5 were counted as true positives; otherwise they were classified as false positives. We selected this threshold to account for annotation inaccuracies in the AutoArborist dataset. For example, defining bounding boxes for complex objects such as trees is subjective (Everingham et al., 2010).

We calculated precision and recall to evaluate model performance. Precision quantifies resistance to false positives and is calculated as the number of true positives divided by the total predicted detections. Recall measures the ability to correctly detect trees, and is calculated as true positives divided the total number of annotated trees. Further details on the model's architecture, training, and evaluation are provided in Supplementary Methods 1.1.

2.3.2 Tree classification

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

We trained and evaluated the image classification model using the 100 genera sourced from AutoArborist and iNaturalist (range: 793 images for *Larix* [larch] to 45,743 for *Acer* [maple]; Table S1). To assess generalization within and across cities, we applied a two-stage splitting strategy. For each city, we split the dataset into two spatially distinct subsets (Figure S2). Approximately 70% of images (N = 789,820) were allocated to model development, and 30% (N = 356,341) were withheld as a geographically independent test set for final evaluation. This spatial partitioning prevented geographic overlap between training and testing images.

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 served as a holdout to assess cross-city classification performance (see Section *2.4.1*).

We trained the classification model for up to 100 epochs and terminated training when validation loss did not decrease after 10 consecutive epochs to limit overfitting. To improve generalization to unseen images, we applied data augmentations (random horizontal flips, rotations, and color shifts) during training. Model performance was assessed after training by computing recall (true positive rate), precision (proportion of predicted positives that are correct), and F1 score (a performance metric that accounts for both model recall and precision)

for each genus on the 30% geographically stratified test split. Additional details on the model architecture, training, and evaluation are provided in Supplementary Methods 1.2.

2.3.3. Tree classification model interpretation

Image classification performance generally improves with additional training data (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 performance (the F1 score calculated on the 30% geographically stratified test split).

To understand which visual features underlie correct classifications, we applied gradient-weighted class activation maps (Grad-CAMs) (Selvaraju et al., 2017). Class activation mapping highlights image regions most influential to a model's prediction (Zhou et al. 2016). Previous work using Grad-CAM indicates that bark and trunk features are important for species classification (Kim et al., 2022). We calculated Grad-CAMs from the penultimate convolutional layer in the image classification model for a random sample of 100 images per genus drawn from the AutoArborist 30% geographic testing split. We then visually inspected the Grad-CAMs to confirm that the model emphasized biologically meaningful features, such as leaves, branches, fruit, and crown structure, during inference.

2.3.4. Tree geolocation

We geolocated each detected and classified tree in the GSV image using the image's latitude and longitude, the angle of the tree, and the estimated camera-to-tree distance. We estimated distances with ZoeDepth, a monocular depth estimation model (Bhat et al. 2023).

Monocular depth estimation infers object distances from single-color images (Godard et al., 2018) and have been used to approximate distances between GSV images and urban objects, including trees (Krylov et al., 2018; Liu et al., 2023; Lumnitz et al., 2021). We employed the pre-trained ZoeD-NK model, which was optimized for outdoor street-level imagery (Bhat et al., 2023). For each tree, the camera-to-tree distance was calculated as the median predicted depth within its bounding box. To mitigate distortions from the GSV panoramic projection, we multiplied each depth estimate by three, following Lumnitz et al. (2021).

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

2.4.1 Effect of diversity

We examined how inter-city variation in urban tree diversity influences image classification rates. Within each city, genus diversity follows a long-tailed distribution in which a few genera are common and many are rare (Love et al. 2022; Ma et al. 2020). This class imbalance biases model predictions, typically yielding higher accuracy for common genera than for rare ones (Beery et al., 2022; Van Horn and Perona, 2017).

To characterize urban tree diversity, we calculated three metrics for each city from the full AutoArborist Inventory (Table 1): richness (number of genera), evenness (0-1; relative abundances of genera), and the Shannon diversity index (integrates richness and evenness; Hill, 1973). High richness with low evenness indicates many genera with a few dominant taxa, whereas high richness with high evenness reflects a balanced genus composition. Shannon values typically range from 1.5 to 4.0, with higher values indicating greater richness and a more even genus distribution (Love et al., 2022). All metrics were computed in R using the vegan package (version 2.6-4; Oksanen et al., 2013).

To evaluate inter-city model performance, we used the internal 10% test set (N = 74,631 images) defined during the initial model development split (*Section 2.3.2*). This test set provides representative samples of each city with respect to richness, evenness and diversity. 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. Weighting accounts for class imbalance in each city by giving greater weight to common genera while retaining contributions to rare ones.

To test whether classification performance was associated with urban tree diversity, we fitted 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

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

We grouped cities into four climate regions (Figure S6) using a principle component analysis (PCA) using 19 bioclimatic variables from WorldClim v2.1 at 2.5' (~4.5 km) resolution (https://worldclim.org/) that capture temperature and precipitation patterns relevant to tree distributions and phenology (Boucher-Lalonde et al., 2012; Fick and Hijmans, 2017). Bioclimatic variables were sampled at each city's centroid using the raster R package (Hijmans

et al., 2015). The first two PCA axes explained 77.6% of inter-city climate variation. Cities were then assigned to one of four regions based on their scores along these axes (Table 1, Figure S6).

We computed F1 scores for each genus by month and region combination using the 30% geographically independent test set (Figure S2). Because many genera lacked imagery across all month–region combinations, we restricted the analysis to genera represented in at least 50 month-region combinations (N = 17). This criterion increased temporal and spatial coverage for trend detection and reduced model failures due to missing data.

For the remaining 17 genera, we assessed the effects of month, region, and their interaction on classification performance using a two-way ANCOVA (Type II SS). The number of training images for each genus–month–region combination was included as a covariate to evaluate the influence of training data availability. Model assumptions were checked with standard diagnostics before inference. Pairwise comparisons of estimated marginal means for month and region used Tukey's adjustment via the 'emmeans' R package (Lenth, 2025).

2.5 Generalizability of computer vision-based tree inventories

To evaluate generalizability, we applied our computer vision models across cities (Figure 1) and compared predicted trees to AutoArborist inventory records. 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 analysis to inventory trees located within the 30% geographic test split of each city (Figure S2). To capture local variability, each city was tiled into 100-m grid cells, and

we retained only grids containing at least one tree. We then stratified grids by tree density using the 0.25, 0.5, 0.75 quantiles and drew a 5% random sample without replacement within each stratum. This design yielded 103,428 trees across 6,306 grids in 23 cities (Table S2).

For each sampled grid, we downloaded all available GSV images intersecting the grid and a 100-m outward buffer to reduce edge effects from missing imagery (Supplementary Methods 1.3). We then applied the computer vision models to these GSV images to detect, classify, and geolocate trees.

2.5.2 Matching 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 in meters were computed from WGS84 geographic coordinates for every inventory-prediction pair. We then iteratively selected the closest available pair, recorded the match, and removed both trees from future consideration, yielding a one-to-one assignment. Inventory trees remained unmatched if no congeneric predictions fell within their buffer or if all such predictions were already matched. We tested buffer radii from 5 to 30 meters in 5 meter increments and adopted 20 meters to balance matching performance with a local spatial scope (Figure S12).

2.5.3 Effects of ecological, image coverage, and prior-knowledge on match success

We modeled the probability that an inventory tree matched a congeneric prediction using binomial GLMM. The full model was:

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

where $\square_{\square\square\square\square}$ is the probability that an inventory tree $\square$ in a grid $\square$ of a city $\square$ belonging to a genus $\square$, was correctly matched (1 = match; 0 = no match). $\square_{0}$ is the intercept, and $\square_{\square,\square\square\square\square}$ are the 11 predictors. Random intercepts were specified as $\square_{\square} \sim \square\left(0, \square^{2}_{\square\square\square\square}\right)$ for city, $\square_{\square(\square)} \sim \square\left(0, \square^{2}_{\square\square\square\square}\right)$ sampling grids nested within city, and $\square_{\square} \sim \square\left(0, \square^{2}_{\square\square\square\square\square}\right)$ for genus.

We evaluated 11 fixed effects hypothesized to influence model generalizability: Ecological conditions (5): tree abundance, congeneric tree abundance, genus richness, nearest-neighbor (NN) tree distance (m), and temperature seasonality. These capture local tree density and spatial structure, which may influence matching outcomes by varying occlusion (tree abundance, NN tree distance), misclassification risk (genus richness), and match likelihood (congeneric abundance). Ecological predictors were computed within a 20 m buffer centered on each inventory tree, except seasonality, which 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).

Street-level image coverage (4): number of images, nearest image distance (m), modal capture year, and modal capture season (i.e., growing or dormant season; Supplementary Methods 1.4). These quantify the street-level information environment which may influence matching outcomes by affecting viewpoint availability and perspective diversity (number of images, nearest image distance), temporal recency and resolution (capture year), and visibility of diagnostic traits (capture season). Image coverage predictors were computed within a 20 m buffer centered on each inventory tree.

Prior knowledge bias (2): the number of AutoArborist training images for the tree's genus, and the proportion originating from the focal city (the city-level count divided by the genus-level count). These predictors quantify the model's genus- and location-specific familiarity, which impacts tree classification, and thus matching outcomes.

We included random intercepts for city and genus to account for data hierarchy, and a nested grid-within-city intercept to account for spatial autocorrelation. All numeric predictors were centered and scaled, and skewed predictors (all except temperature seasonality and capture year/season) were log-transformed to improve linearity on the logit scale.

We fit all 2,046 additive 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). Models were ranked by the Bayesian Information Criterion (BIC) as a conservative choice for large samples (Schwarz, 1978). Diagnostics for the top-ranked (minimum BIC) model included residual plots, uniformity, dispersion, and outlier checks with 'DHARMa' (Hartig, 2024), variance inflation factors with 'performance' (Lüdecke et al., 2021), and spatial autocorrelation tests with 'spdep' packages (Bivand, 2022) (Supplementary Methods 1.5)

We assessed consistency in variable inclusion, effect sizes, and directions for the full model and all submodels with $\Delta BIC < 10$. For the top model, we report marginal R^2 (variance explained by fixed effects) and the conditional R^2 (variance explained by fixed and random effects) using 'performance' package (Lüdecke et al., 2021). We decomposed random-effect variances to estimate the percentage of total variance attributable to city, grid, and genus. We visualized fixed-effect influences with partial effect plots, and deviations of city and genus levels from overall means were examined using best linear unbiased predictions (BLUPS).

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

The Grad-CAMs provided insight into different visual features underlying correct genus classifications (Figure 2). For example, correctly classified *Citrus* trees emphasized the characteristic orange fruit (Figure 2C). For genera where fruiting or flowering were absent, structural features, such as canopy shape and branching patterns, were important for classification (e.g., for deciduous genera like *Acer* and *Betula*, and for palms like *Phoenix*) (Figure 2). Likewise, bark color and texture informed correct classifications. The reddish-brown, furrowed bark in conifers (*Pinus*, *Thuja*), and white-brown papery bark in *Betula* (birch) guided

correct classifications (Figure 2). Across genera, Grad-CAMs revealed that classification relied on biologically meaningful and visually distinctive features.

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

3.2.1 Urban tree diversity affects performance among cities

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

3.2.2 Geography and capture month affect performance

Genus classification performance 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).

Image capture month significantly influenced classification for the most common genus, *Acer* ($p < 0.001$; Table S5, Figure 4). Images from summer months (e.g., July - August) produced fewer false positives and false negatives than winter images (e.g., January - February) (Figure 4). This effect was independent of the monthly sample size 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

Across cities and genera, the computer vision-based inventory matched 67.1% of the inventory trees to predicted congeneric trees. Matching varied more across genera (mean = 50.9%, SD = 23.0%; Table S6) than across cities (mean = 67.4%, SD = 9.4%; Figure S12).

3.3.2 Model Selection

We evaluated 2,046 candidate models to test how ecological, image coverage, and prior knowledge predictors affected generalizability. The best-supported model (BIC weight = 0.702) retained eight predictors from the full specification: four ecological (tree abundance, genus richness, congeneric abundance, nearest-neighbor tree distance), two image coverage (nearest image distance, number of images), and two prior-knowledge predictors (the number of AutoArborist training images for the tree's genus, and the proportion originating from the focal city).

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

Effect sizes for predictors shared across supported models (ΔBIC < 10) were highly consistent in 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. Given

the stability of effects and the strong support for the top model, all subsequent inferences are based on that model.

3.3.3 Model diagnostics and goodness of fit

Diagnostics for the top model indicated that assumptions were met: no overdispersion, non-uniform residuals, influential outliers, or spatial autocorrelation (Figures S13–S14). Fixed effects explained 20.5% of the variance in matching (marginal R^2), increasing to 46.7% when random effects were included (conditional R^2). Random effect variance was dominated by genus (Var = 0.78, SD = 0.89), with a smaller contribution from city (Var = 0.46, SD = 0.46). These results are consistent with variance partitioning estimates, which attribute 12.7% of the total variance to genus and 3.5% to city (Table 2).

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

All fixed effects retained in the top model significantly affected the probability of correctly matching inventory trees to predicted trees (Table 2). Prior-knowledge exerted the strongest effects. A multiplicative increase in the number of genus-level training images (log-transformed) more than doubled the odds of a match (OR = 2.01, 95% CI: 1.71–2.35, $p < 0.001$) and a higher proportion of training images originating from the focal city also increased success (OR = 1.60, 95% CI: 1.53–1.67, $p < 0.001$).

Local image coverage had comparably strong effects. More images (log-transformed) improved matching (OR = 1.55, 95% CI: 1.51-1.58, $p < 0.001$), while greater distance to the nearest image (log-transformed) reduced matching (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 abundance (log-transformed; OR = 1.42, 95% CI 1.37–1.46, $p < 0.001$) increased matching odds, whereas higher tree abundance (log-transformed; OR = 0.74, 95% CI: 0.71–0.77, $p < 0.001$) and greater genus richness (log-transformed; 0.93, 95% CI: 0.90–0.97, $p < 0.001$) decreased matching (Figure 5B–5C). Together, matching is maximized when the tree's genus is well represented in the training dataset, particularly with images from similar geographic contexts, when image coverage is dense, and when the local tree composition and spatial arrangement are simpler.

3.3.5 Genus- and city-level deviations

Genus-level best linear unbiased predictions (BLUPs) showed consistent overperformance and underperformance after accounting for fixed effects (Figure S15A). Among the 100 genera, 29 genera overperformed and 21 underperformed (defined as BLUP 95% CIs not overlapping the overall mean). Largest positive deviations were observed for *Ailanthus* (tree-of-heaven; BLUP ± 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 BLUPs (Figure S15B) were more constrained, consistent with their smaller share of explained variance (3.5%; Table 2). Three cities overperformed (Vancouver = 1.20 ± 0.21; Charlottesville = 0.85 ± 0.54; Buffalo = 0.56 ± 0.28), and 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). Overall, genus identity remains the primary source of residual variation in matching performance, whereas city-level deviations are fewer and modest in magnitude.

4. Discussion

Urban tree inventories underpin management of ecosystem services and responses to disturbances such as pest or disease. Although existing remote sensing methods accurately map trees within individual cities, their ability to generalize across variable imaging conditions, species compositions, and climates remains uncertain. We addressed this knowledge gap by training computer vision models on GSV and iNaturalist images to detect, classify, and map 100 tree genera across 23 North American cities. Object detection and depth estimation reliably identified and geolocated trees, but genus classification performance was more variable. Accuracy declined with shifts in genus diversity and geography but improved under leaf-on conditions. Grad-CAM indicated that classification leveraged biologically interpretable traits (e.g., canopy structure, bark texture, and fruit). In geographically independent evaluations, predictions matched 67.1% of AutoArborist inventory records. Variation in matching success was driven primarily by differences among genera, training data availability, and local ecological and image context rather than city identity. Collectively, these results show that street-level imagery coupled with computer vision offers a scalable complement to traditional inventories.

4.1. Opportunities and challenges for scaling urban tree inventories

Street-level imagery provides a promising alternative to aerial or satellite data for urban forest monitoring. While prior applications were city-specific (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 analysis across 23 cities enabled a broader assessment of model generalizability. The models achieved high precision and recall (>0.7) for eight common U.S. genera (*Acer, Fraxinus, Quercus, Gleditsia, Ulmus, Tilia, Pyrus,* and *Platanus)* (Table S3). Given these taxa represent ~45% of public street trees in the U.S., (Ma et al. 2020), this approach has operational value for many municipalities. For instance, our approach performed strongly in Sioux Falls, SD, because ~82% of trees belong to these genera (Figure 3).

Performance declined in large, biodiverse cities such as Los Angeles, New York, and Denver, likely reflecting visual similarity among co-occurring genera (low inter-class variability) and numerous species and cultivars (high intra-class variability). This pattern mirrors findings from field and remote sensing studies showing reduced classification performance in taxonomically diverse areas (Berland et al., 2019; Marconi et al., 2022; Weinstein et al., 2023). Although we classified at the genus level, future work could predict across multiple taxonomic levels and leverage phylogenetic relationships to improve accuracy (Gillespie et al., 2024).

4.2 Interpreting model predictions with trait-based evidence

Model performance varied by geography and seasonality, likely reflecting regional differences in tree composition, management practices, and contextual cues such as street layouts (Beery et al., 2022). Image timing also strongly influenced classification. For *Acer*, leaf-on months (July–August) reduced both false positives and negatives compared with winter imagery, independent of training sample size (Table S5, Figure 4).

Grad-CAM visualizations clarified how models distinguished genera. Some taxa were identified by fruits or flowers, while others relied on subtler traits like canopy architecture and

bark texture. These findings align with prior work showing that CNNs can reliably identify species using tree bark imagery (Kim et al., 2022) and suggest that the visual cues driving classification shift seasonally. Bark provides consistent information year-round, whereas flowering and fruiting periods offer brief but highly informative windows for genus identification. Feature salience may also vary with tree age. Grad-CAMs highlighted age-related cues such as trunk buttressing and coarse bark patterning that are pronounced in mature trees but less apparent in young saplings. For younger trees, foliage traits may therefore provide more informative signals for classification. These results point to opportunities for optimizing image acquisition by aligning data collection with periods and contexts when locally diagnostic traits are most visible.

4.3 Factors shaping generalizability across cities

Generalizability was driven by training data volume, local ecological structure, and street-level image coverage. Genus-level image availability was the strongest predictor of matching success: > 3,000 training images per genus were typically required to exceed a 50% match probability. Incorporating city-specific training images further improved performance: matching rose from 61% to 75% as the share of focal-city training images increased from 1% to 25% (Figure 5B–C). Denser stands reduced match success, consistent with occlusion effects by neighboring trees: match probability fell below 50% when > 30 trees occurred within 20 meters of the inventory tree and a minimum spacing of ~1 m was needed to achieve a 50% match probability, increasing to 75% at 15 m (Figure 5B–C). Image density and camera-to-tree distance also mattered: match probability rose from 50% to 63% as available viewpoints increased from one to three, but dropped below 50% when the nearest image distance exceeded 18 m

These quantitative relationships extend qualitative observations in previous studies (e.g., Lumnitz et al., 2021; Wegner et al., 2016). Operationally, computer vision inventories will perform best in well-imaged areas, with leaf-on imagery that shows discriminative features, and in structurally simple stands (i.e., limited crowding and occlusion). Targeted expansion of training images, particularly for underrepresented genera with examples from the target geography, can further improve reliability of these approaches.

4.4. Model limitations and implications for urban forestry

Despite strong performance, several limitations remain. First, public tree inventories may be incomplete or outdated; some apparent model errors may reflect inaccuracies in the reference data. Second, our matching rates are conservative because predictions were compared only to public inventories, and private trees could not be assessed. Third, our 20-m matching radius introduces uncertainty by potentially conflating nearby public and private trees. A stricter 15-meter radius reduced overall agreement to 60.0%, but the broader statistical patterns remained unchanged (Figure S12). Last, most inventories lacked tree-health attributes, limiting our ability to distinguish between healthy, stressed, and diseased trees (but see Kagan et al., 2021).

Integrating multiple image sources and spatial context could reduce errors. National Agricultural Imagery Program imagery (NAIP) provides high-resolution, leaf-on coverage suitable for urban tree mapping (Erker et al., 2019; Rice et al., 2025; Ventura et al., 2024). Fusing NAIP with street-level imagery can improve classification rates by providing complementary canopy and contextual information (Hartling et al., 2019; Wang et al., 2025). Conversely, others report marginal gains from incorporating high-resolution imagery (Beery et al., 2022). Adding geographic context during training could also improve fine-grained

classification. Given that species distributions are governed by environmental gradients (Boucher-Lalonde et al., 2012), incorporating location can constrain predictions to ecologically plausible taxa (Aodha et al., 2019; Cole et al., 2023).

Understanding where and when models generalize remains a core challenge, particularly under class imbalance (Zhou et al., 2023; Beery et al., 2018). The AutoArborist dataset (Beery et al., 2022) exemplifies this challenge by reflecting the long-tailed distribution of urban forests, where few genera dominate and most are rare. Our results demonstrate computer vision applied to street-level imagery has scalable applications for urban forestry management despite this imbalance. Extending these methods to peri-urban and rural landscapes could support biodiversity assessments and invasive species monitoring (Kotowska et al., 2021). Where street-level imagery is sparse, fusing street-level and aerial data offers a practical strategy to fill coverage gaps.

5. Conclusions

Computer vision models trained on publicly available street-level imagery can accurately detect, classify, and geolocate urban street trees across diverse North American cities. Where classification rates declined, namely in areas with high genus richness or off-season imagery, we identified failure patterns, such as inter-class confusion (between-genus look-alikes) and intra-class variability (visual variation within a genus) that mirror challenges faced by community observers and arborists. By quantifying where and when performance deteriorates and identifying practical levers for improvement (e.g., leaf-on acquisition, denser image coverage, targeted expansion of training images), we establish a framework for systematically improving future urban tree mapping. Developing scalable, 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 study cities. For each city, we report latitude and longitude coordinates; its assigned geographic region (determined via principal component analysis of WorldClim v2.1 bioclimatic variables); tree diversity (genus richness and the Shannon diversity index); human population (2020 U.S. Census or the 2021 Canadian Census); mean annual temperature (in °C; BIO1); and total annual precipitation (in mm; BIO12).

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. Top GLMM (selected by minimum BIC) assessing how ecological, street-level image coverage, and prior-knowledge influence the probability that computer-vision predictions match AutoArborist inventory records in a geographically independent test set. Numeric predictors were transformed as [ℓ - log(x+1)]. Predictors associated with lower odds: nearest image distance, tree abundance, and genus richness. Predictors associated with higher odds: congeneric abundance, nearest-neighbor (NN) tree distance (m), the number of training images from the tree's genus, and the proportion of training images.

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 Abundance)	1.42	0.02	[1.37, 1.46]	20.45	***
ℓ(NN 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

** p<0.05; ** p<0.01; *** p<0.001*

Figure Legends

Figure 1. Computer vision-based tree inventories. A) Data collection: Images for 100 tree genera were sourced from AutoArborist (Google Street View [GSV] images from 23 cities in United States and Canada) and research-grade observations on iNaturalist.org. B) Detection, classification, and geolocation: An object detection (YOLOv5x) model and image classification (EfficientNetV2S) model identify tree genera in GSV images. Next, a depth estimation model (ZoeDepth) assigns geographic coordinates to predicted trees. The workflow is evaluated on spatially independent holdout areas within each city to assess out-of-sample generalizability.

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-CAMs (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-CAMs highlight image regions most influential for the model's predictions (red = higher importance; blue = lower importance). The CNN leveraged biologically meaningful and visually distinctive features, including reproductive traits (white arrow indicating 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 patterns show that the model captured diagnostic morphological features consistent with expert knowledge, supporting the interpretability of genus-level classifications.

Figure 3. Relationship between the city-level Shannon diversity index and image classification performance across 23 cities. For each city, the Shannon diversity index was computed using from AutoArborist inventory records. We applied the image classification model to a 10% randomly withheld testing dataset for each city and evaluated performance by computing the weighted average F1 score across genera. The scatterplot illustrates the negative relationship between Shannon diversity and performance, where more diverse cities have lower weighted F1 scores. The teal line indicates the predicted values and shading denotes the 95% prediction interval.

Figure 4. Maple (*Acer*) classification performance by month and climate-defined geographic region. Boxplots show F1-score distributions across months of image capture (left) and regions (right). Regions were derived by 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. Differences in performance were assessed with a two-way ANCOVA (Month, Region; Type II), followed by Tukey-adjusted pairwise comparisons of estimated marginal means; letters above boxplots denote significantly different groups.

Figure 5**.** GLMMs evaluating how ecological context, street-level imagery coverage, and prior knowledge affect the probability that computer-vision predictions match AutoArborist inventory records from a geographically independent test set. 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 expressed as odds ratio; striped

tiles denoting predictors excluded from a given model. Numeric predictors were transformed as [ℓ - log(x+1)]. (C) Partial effects showing predicted probabilities of matching across each predictor's observed range, with all other predictors held at their mean. Solid lines and shaded ribbons denote model predictions and 95% confidence intervals, respectively. Predictor values are back-transformed to the original scale. Predictions are obtained from the inverse link function with bias correction to account for Jensen's inequality. Abbreviations: SL = street-level; AA = AutoArborist.

Figures

Figure 1.

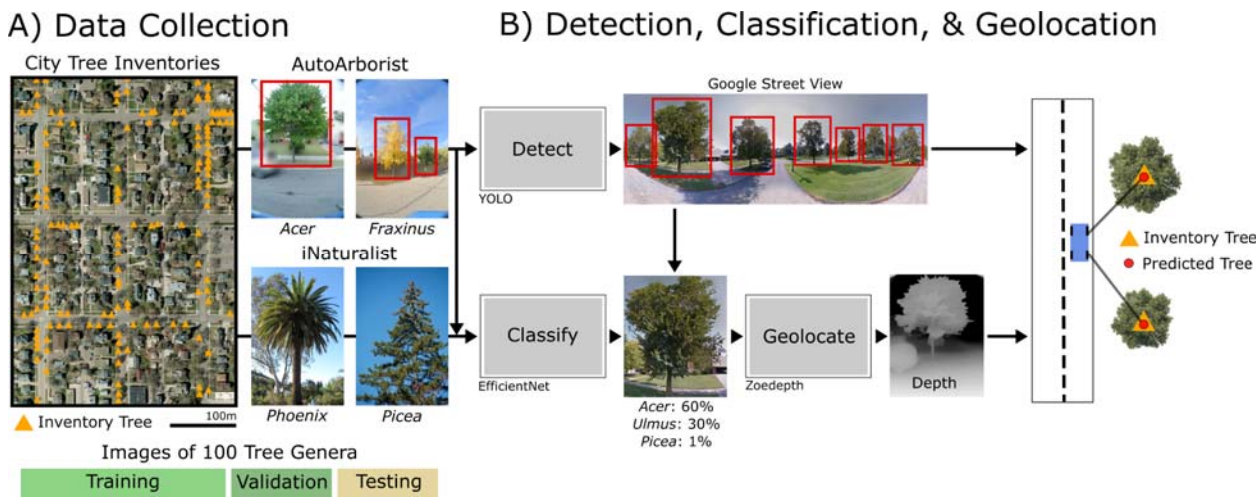


Figure 2.

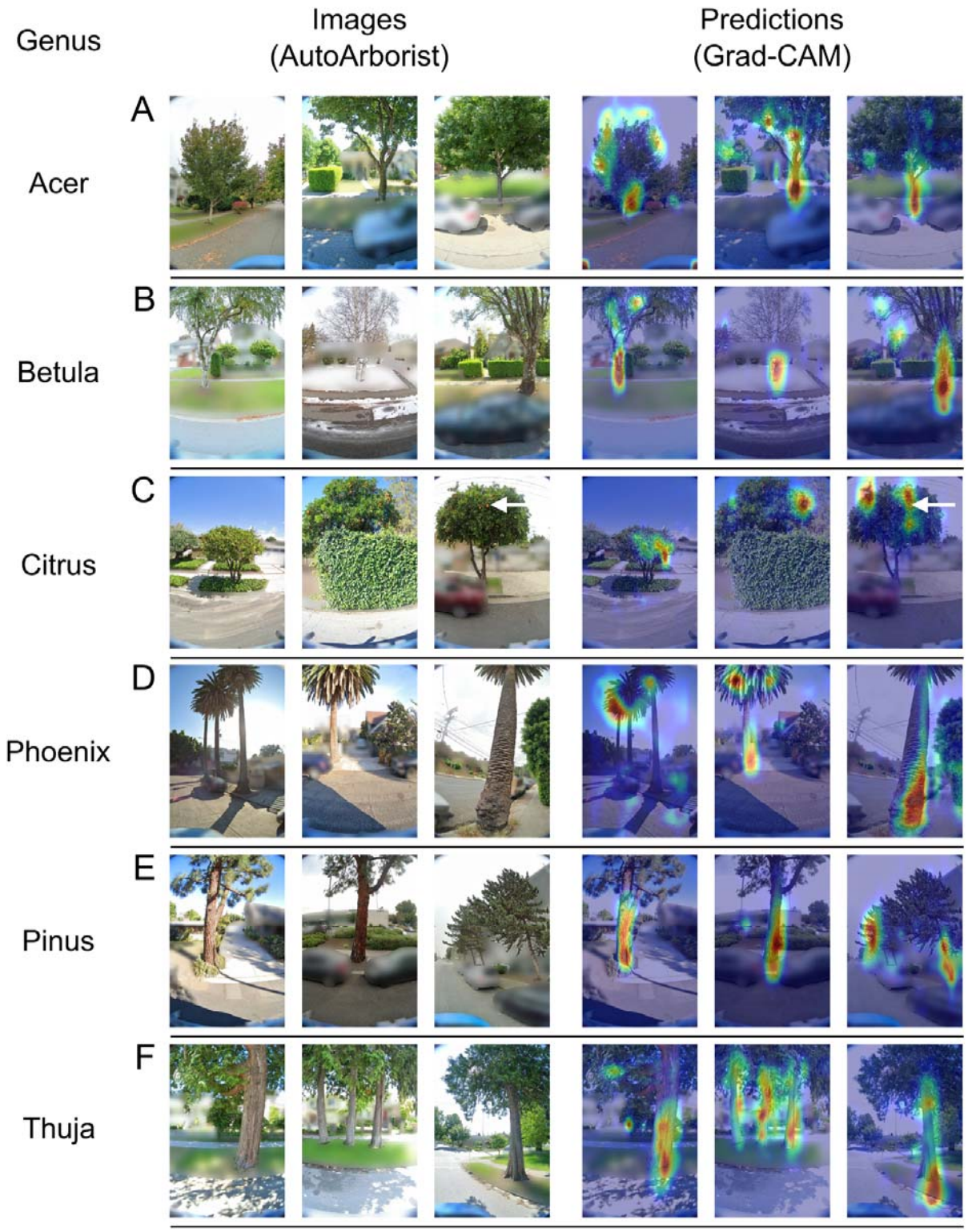

Genus
Images
(AutoArborist)
Predictions
(Grad-CAM)
A
Acer
B
Betula
C
Citrus
D
Phoenix
E
Pinus
F
Thuja

Figure 3.

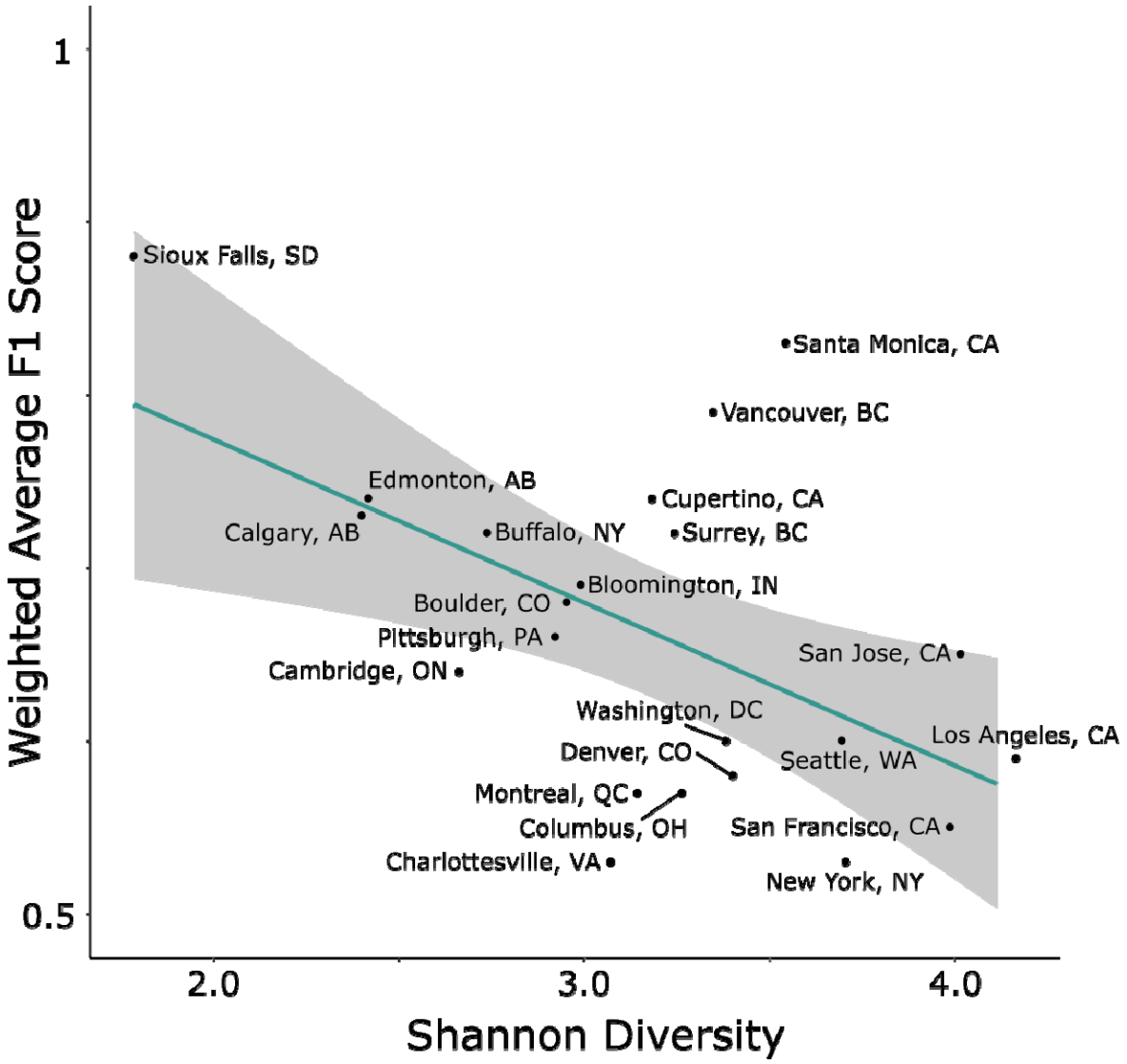

1
0.5
Weighted Average F1 Score
2.0
3.0
4.0
Shannon Diversity
Sioux Falls, SD
Santa Monica, CA
Vancouver, BC
Edmonton, AB
Cupertino, CA
Calgary, AB
Buffalo, NY
Surrey, BC
Bloomington, IN
Boulder, CO
Pittsburgh, PA
San Jose, CA
Cambridge, ON
Washington, DC
Los Angeles, CA
Denver, CO
Seattle, WA
Montreal,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

Figure 4.

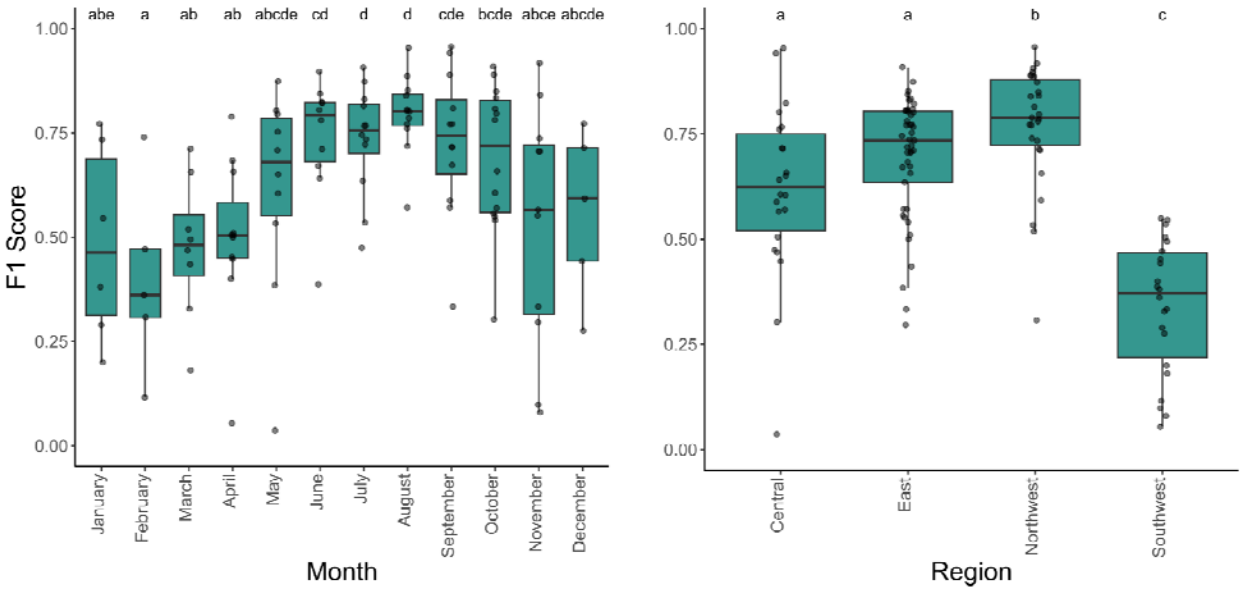
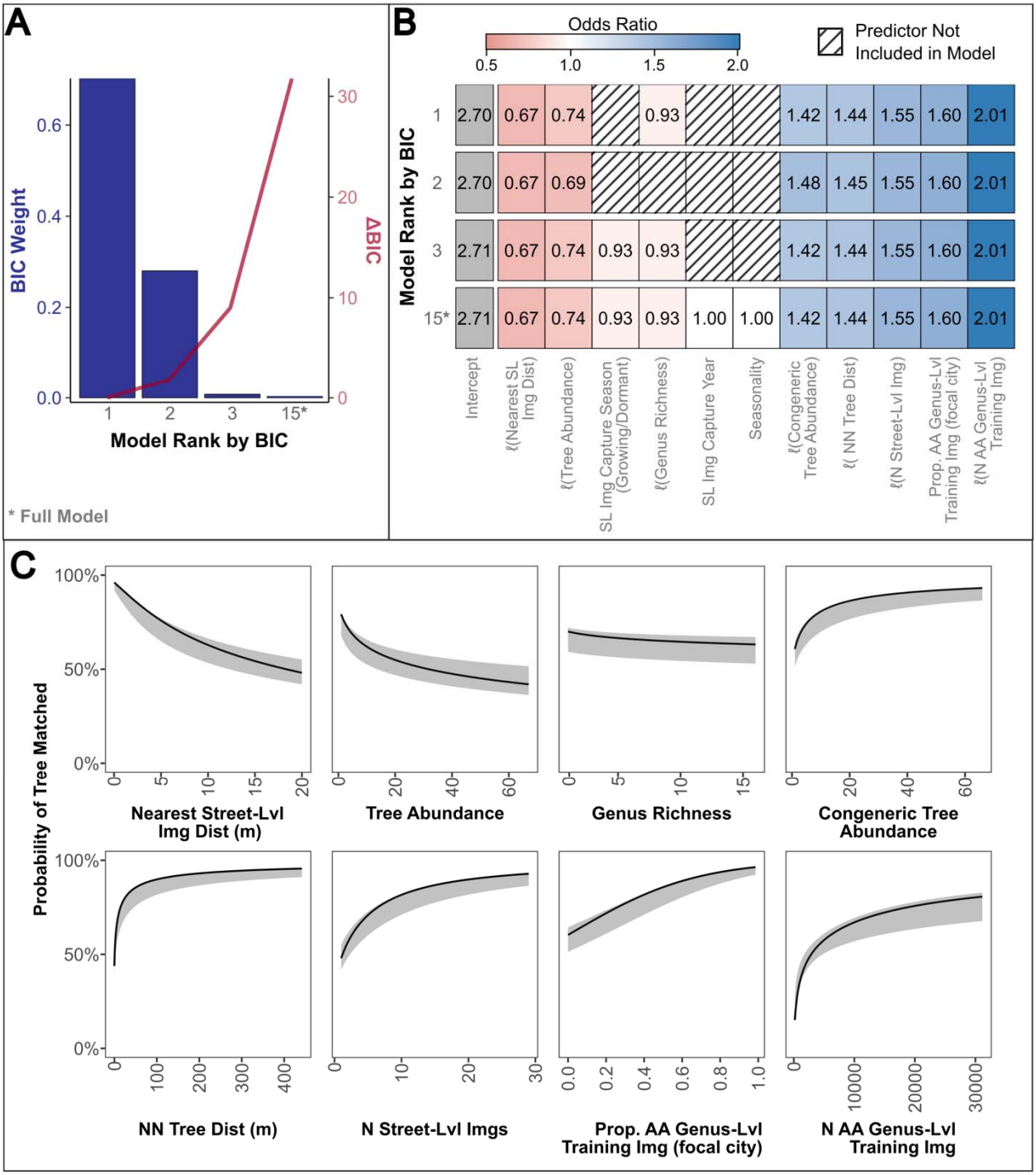

abe
a
ab
ab
abcde
cd
d
d
cde
bcde
abce
abcde
1.00
0.75
0.50
0.25
0.00
F1 Score
January
February
March
April
May
June
July
August
September
October
November
December
Month
a
a
b
c
Central
East
Northwest
Southwest
Region

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.

Rice, J. R., Fricker, G. A., & Ventura, J. (2025). An end-to-end deep learning solution for automated LiDAR tree detection in the urban environment. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 100092.

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.