

The Auto Arborist Dataset: A Large-Scale Benchmark for Multiview Urban Forest Monitoring Under Domain Shift

Sara Beery*[†], Guanhang Wu[†], Trevor Edwards[†], Filip Pavetic[†], Bo Majewski[†],
Shreyasee Mukherjee[†], Stanley Chan[†], John Morgan[†], Vivek Rathod[†], Jonathan Huang[†]
California Institute of Technology* Google[†]

Abstract

Generalization to novel domains is a fundamental challenge for computer vision. Near-perfect accuracy on benchmarks is common, but these models do not work as expected when deployed outside of the training distribution. To build computer vision systems that truly solve real-world problems at global scale, we need benchmarks that fully capture real-world complexity, including geographic domain shift, long-tailed distributions, and data noise.

We propose urban forest monitoring as an ideal testbed for studying and improving upon these computer vision challenges, while working towards filling a crucial environmental and societal need. Urban forests provide significant benefits to urban societies. However, planning and maintaining these forests is expensive. One particularly costly aspect of urban forest management is monitoring the existing trees in a city: e.g., tracking tree locations, species, and health. Monitoring efforts are currently based on tree censuses built by human experts, costing cities millions of dollars per census and thus collected infrequently.

Previous investigations into automating urban forest monitoring focused on small datasets from single cities, covering only common categories. To address these shortcomings, we introduce a new large-scale dataset that joins public tree censuses from 23 cities with a large collection of street level and aerial imagery. Our Auto Arborist dataset contains over 2.5M trees and 344 genera and is >2 orders of magnitude larger than the closest dataset in the literature. We introduce baseline results on our dataset across modalities as well as metrics for the detailed analysis of generalization with respect to geographic distribution shifts, vital for such a system to be deployed at-scale.

1. Introduction

Urban forests provide critical benefits to the over 4B people living in urban areas worldwide [107]. They filter air and water, capture stormwater runoff, sequester atmospheric carbon dioxide, limit erosion and drought, and save energy in a variety of ways (e.g., by providing

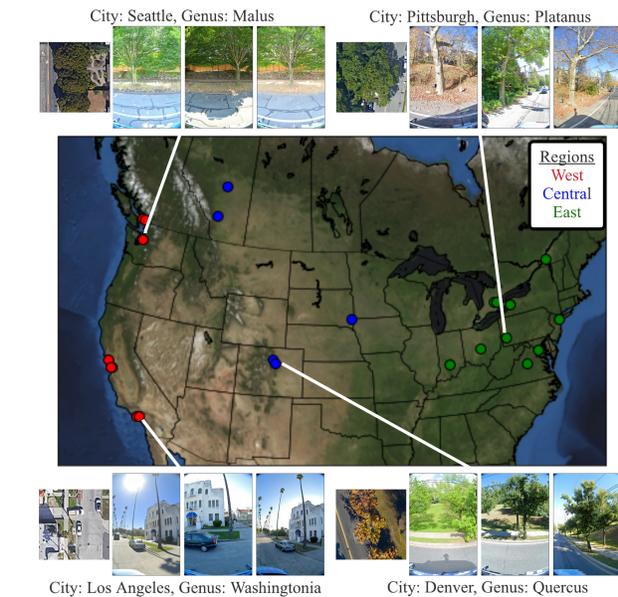


Figure 1. The Auto Arborist dataset covers 23 cities across North America, and contains paired aerial and multiview ground-level imagery for 2.6M trees across 344 unique genera.

shade and thus reducing cooling costs and urban heat islands [100, 120, 138, 141]). In the US alone, urban forests cover 127M acres and produce ecosystem services valued at >\$18B [106]. These forests make up the foundations of our urban ecosystems, and provide habitat for diverse urban wildlife and insect populations [44]. *Urban forest monitoring*, measuring the size, health and species distribution over time, allows us to (1) quantify ecosystem services including air quality improvement [20, 47], carbon sequestration [92, 106, 121], and benefits to public health [28, 47, 125, 125], (2) track damage from extreme weather events [8, 67, 99], and (3) target planting to improve robustness to climate change, disease and infestation [24, 64, 114, 115]. Further, lack of access to urban greenery is a key aspect of urban social inequality [56, 93, 104], including socioeconomic inequality [52, 73, 86] and racial inequality [21, 129]. Urban forest monitoring enables the quantification of this inequality and

Dataset	Ground-level images	Aerial Images	Sites	Classes	Forest Type
Registree [27]	46,321	28,678	1	40	Urban
Pasadena Re-ID [102]	6,141 (panoramas)	0	1	1	Urban
NEON Tree Evaluation [132]	0	25,949 (boxes)	22	1	Wilderness
IDTreeS Classification 2017 [98]	0	613 (boxes)	1	9	Wilderness
IDTreeS Classification 2020 [56]	0	452 (boxes)	2	34	Wilderness
Auto Arborist (Ours)	6,479,077	2,637,208	23	344	Urban

Table 1. Comparison of our dataset to previous tree detection and identification datasets. Note that for Pasadena Re-ID, there is only one class (“tree”) but the task is re-identification as opposed to categorization.

the pursuit of its improvement [22, 118].

To enable automated urban forest monitoring, we focus on the fundamental task of building a *tree census* (sometimes called a *tree inventory*). Due to their significant cost (a recent tree census in Los Angeles cost \$2M and took 18 months [2]), tree censuses are typically conducted only by cities with the means and will to invest in these undertakings, and even then they are conducted rarely (e.g., once every 5-10 years). We seek to dramatically lower the cost of a tree census by using computer vision to help find, label, and monitor individual trees using a combination of street level and aerial imagery. An automated system could democratize access to urban forest monitoring, providing this valuable information to under-resourced cities that are already disproportionately affected by climate change [109].

While there have been prior works on urban tree species recognition from aerial [4, 5, 63, 78, 124, 137, 139, 140] or street level [95] imagery (or both, in a limited number of cases [27, 130]), a major limitation has been a lack of large-scale labeled datasets. To our knowledge, all prior works have focused on single or limited numbers of cities, and have included only the most common categories. We introduce the *Auto Arborist dataset*, a multiview urban tree classification dataset which, at 2.6 million trees is two orders of magnitude larger than those in prior work [27, 130] and contains 344 genera (and many more species). To build Auto Arborist, we draw on public tree censuses from 23 cities in the US and Canada and merge these public records with street level and overhead RGB imagery. As the first urban forest dataset to cover multiple cities, Auto Arborist allows for previously-impossible detailed analysis of generalization with respect to geographic distribution shifts, vital to building systems that scale. We propose a set of metrics to evaluate performance with respect to these geographic distribution shifts and show the strengths and weaknesses of typical deep learning models when applied to the Auto Arborist dataset.

Going beyond its immediate application to sustainability and conservation, Auto Arborist can serve as an important challenge benchmark for computer vision. There has been increasing interest in domain generalization, which is ubiquitous in real-world applications [16, 51, 82, 97]. For example, prior works have observed that a model for self-driving cars that can drive safely in San Diego may not work equally well in Seattle [50, 68, 69]. In terms of num-

ber of domains, objects, classes, and images, Auto Arborist presents a scale not seen in previous real-world domain generalization benchmarks [13–17, 82]; it focuses on detailed cross domain analysis, and incorporates multiple views and modalities.

To summarize, our main contributions are as follows:

- We develop a pipeline for combining public tree census data with street level and aerial imagery.
- We introduce the Auto Arborist dataset built using this pipeline — the largest dataset of its kind covering >2.6M trees, >9.1M images and 344 categories and first one of its kind to cover multiple cities (23 cities).
- We show that for Auto Arborist, geographic domain shift and the category imbalance are major factors in performance of typical deep learning models.
- We show that diverse geographic coverage is important for generalization to a novel city, and that both multiple views and multiple data modalities are beneficial.

2. Related work

Tree detection, localization, and taxonomic identification have been studied in multispectral aerial imagery [49, 78, 140], ground-level imagery [95, 101], and LIDAR [46, 70], with some recent approaches combining data across modalities [6, 27]. Costly high-resolution data collected from low-flying aircraft has been shown to improve performance [18, 110], but this data is not available for much of the world. Though past studies have considered tree detection and categorization, many of these have been limited by perspective (aerial vs ground-level views), geospatial coverage, and taxonomic coverage. Our work seeks to expand upon all three, providing a testbed for urban forest monitoring that is broad in scope and relies on datatypes which are available across cities (aerial and street level RGB imagery) to enable the development of general models and methods which can be used off-the-shelf in novel cities.

Tree detection and localization from aerial data. There have been successful, broad-scale studies of tree density, canopy cover, and individual tree delineation from aerial data [9, 25, 38, 40, 46, 62, 66, 77, 94, 108, 110, 112, 122, 132], including tree crown detection across sites from the US National Ecological Observatory Network (NEON) [132, 133], tree canopy mapping in urban forests in cities across the US

Region W (West)			Region C (Central)			Region E (East)		
City	Trees	Genera	City	Trees	Genera	City	Trees	Genera
San Francisco, CA	154,698	195	Boulder City, CO	29,489	65	Washington, DC	152,983	71
San Jose, CA	225,655	201	Denver, CO	175,438	97	Charlottesville, VA	1571	56
Cupertino, CA	15,300	104	Calgary, AB	64,576	35	Pittsburgh, PA	23,382	79
Vancouver, BC	121,249	93	Sioux Falls, SD	13,277	37	Montreal, QC	208,097	61
Seattle, WA	150,983	142	Edmonton, AB	76,164	32	New York, NY	560,069	68
Surrey, BC	62,251	72				Buffalo, NY	29,527	74
Santa Monica, CA	25,381	126				Kitchener, ON	21,265	26
Los Angeles, CA	391,788	202				Cambridge, ON	14,757	57
						Columbus, OH	114,536	81
						Bloomington, IN	4,772	53
Total West	1,147,305	328	Total Central	358,944	104	Total East	1,130,959	102

Table 2. Cities by region. The holdout city for each region is in bold.

[93], and counting individual trees in Sub-Saharan Africa [25]. These methods rely on a diverse set of aerial data modalities, from low-resolution RGB or hyperspectral satellite data to high-resolution RGB, hyperspectral, and LIDAR data collected from low-flying aircraft and UAVs [18]. However, there are still open challenges in maintaining performance of methods in novel regions [98, 131, 134], and methods must be well-validated and possibly adapted for any novel region before use. Tree crown delineation in dense forests remains a challenge, leading to several studies (e.g., in sub-Saharan Africa [25]) focusing on low-tree-density regions or trees outside forests [43, 117]. Further, there is only so much that can be understood from an aerial view alone. A large amount of the woody vegetation in a forest is hidden under the tree canopy. Understory trees have been mapped with very-high-resolution UAV-collected data [32, 60, 61, 87], but this data is rarely available. Our combined approach allows us to use available ground level imagery to see under the canopy.

Tree taxonomic identification from aerial data. Automated tree identification in aerial data from satellite or low-flying aircraft, including RGB, hyperspectral, LIDAR, or some combination thereof, is well-studied in the remote sensing community. [49] is a thorough review of species classification from remote sensing data which notes the lack of studies considering large spatial extents. Many studies focus on predicting species occurrence, presence/absence, or abundance for a limited set of species [3, 23, 29–31, 33, 34, 41]. Detecting and categorizing individual trees presents further complexity [4, 26, 35, 36, 41, 42, 48, 48, 53, 55, 59, 65, 70–72, 74, 75, 80, 81, 83, 111, 113, 116, 128], and recently deep learning approaches have been shown to outperform more traditional methods on this task [39, 57, 63, 90, 114, 127, 142]. Generalization to novel regions is a known challenge with many of the proposed methods [110]. The IDTrees challenges [56, 98] were the first to propose a public benchmark for cross-site individual tree categorization, but provided limited labeled data (<1000 labeled trees from ≤ 34 tree categories at 3 NEON wilderness forest plots). Further challenge arises when predicting species in

an urban environment, where human intervention leads to a much higher diversity of tree species, with a much longer tail, than is seen in the wild [135]. For this reason, many studies of urban tree categorization focus only on common species [4, 5, 10, 11, 63, 78, 124, 137, 139, 140].

Tree detection and localization in the urban forest from ground-level data. Ground-level data, (e.g., from Google Street View [7], Mapillary [96], and iNaturalist [1]), have been identified as an important source of information for urban monitoring applications [19]. Automated measures of urban “greenness” and tree cover mapping from ground-level data have been proposed, with implications in social justice and public health [45, 88, 89, 119, 123]. Datasets such as Mapillary Vistas [105] and Cityscapes [37] facilitate semantic segmentation of urban categories, including vegetation, but do not provide instance-level information or fine-grained taxonomic labels. Similarly, most current computer vision studies of the urban forest focus on species-agnostic individual tree detection [76, 126] and localization [84, 95, 101–103] across multiple ground-level views of the same tree.

Tree taxonomic identification from combined aerial and ground-level data. Previous large-scale datasets that combine aerial and ground-level data, such as CVUSA [136], were designed for alternative tasks such as image geolocalization. Several methods exist for combining aerial + ground-level data, with tree identification as a key application [85, 117]. Here, ground-level data can include RGB imagery, LIDAR, and even physical measurements such as tree diameter or hyperspectral signature [54]. [130] and [27] proposed a system for identifying street trees using paired aerial and ground-level RGB imagery for urban forests and released a dataset of paired imagery for Pasadena. [6] proposed a class-agnostic tree detection method from aerial imagery and ground-level LIDAR. Recently, [79] used GNNs to map individual trees across aerial and ground-level community science imagery in forests. All of these prior works trained on a single city and could benefit from a much larger dataset such as ours.

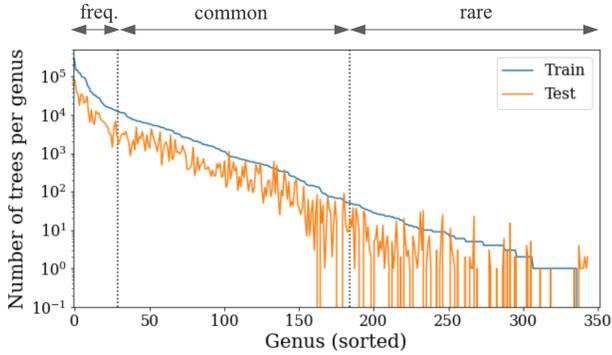


Figure 2. Distribution of genera in train and test, with frequent, common, and rare classes delineated.

3. The Auto Arborist Dataset

We have generated the largest, and most geographically diverse, computer-vision-ready multi-view dataset of urban trees to date. The *Auto Arborist* dataset contains 2,637,208 trees across 23 cities. Each tree is represented by a 512×512 pixel aerial image where each pixel is $5\text{cm} \times 5\text{cm}$, as well as up to three 768×1152 pixel street level images [7] of the tree (for a total of 9,116,285 images in the dataset)¹. To avoid taxonomic complexity arising from hybrid and subspecies when developing methods, we have chosen to focus on genus prediction (instead of species-level prediction) as our primary task and have confirmed with ecologists and city planners that a genus-level map would be highly useful as a first step. Our dataset includes 344 unique genera, with a real-world long-tailed class imbalance and unique class distribution for each city on the dataset (Figure 2).

3.1. Dataset curation

To curate Auto Arborist, we started from existing tree censuses which are provided by many cities online. For each tree census considered, we verified that the data contained GPS locations and genus/species labels, and was available for public use. This resulted in data from 23 cities which we then parsed into a common format, fixing common data entry errors (such as flipped latitude/longitude) and mapping groundtruth genus names (and their common misspellings) to a universal label map consisting of 344 categories. We also removed records with invalid genus names, such as “unidentified”. Aggregated into a single dataset, this process yielded localized records for $\sim 5\text{M}$ trees.

Figure 1 shows a map of the 23 selected cities as well as example imagery from the dataset. We partition the cities into three separate regions for evaluation purposes (discussed further in Section 4). Table 2 summarizes the contribution from each of the cities to the Auto Arborist

¹We are publishing all tree records (after curation/merging c.f. 3.1) and a subset of the imagery (verified to obtain consistent results to the full dataset) with personally identifiable information removed. For more information please visit <https://google.github.io/auto-arborist>

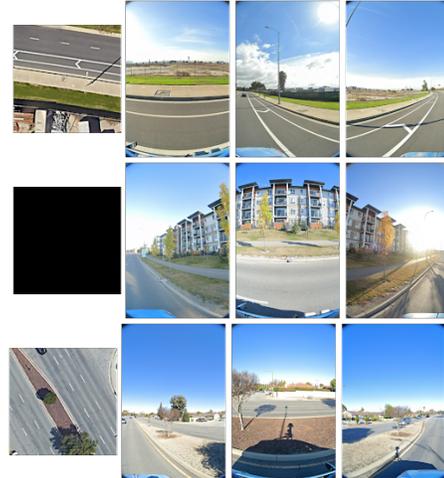


Figure 3. Noise in the Auto Arborist dataset includes trees that have died since the tree census was taken (top), aerial data quality, including failures causing black squares (middle), and temporal variation in deciduous trees (bottom — aerial image has leaves, but street level images are bare), which affects northern cities more than southern ones.

dataset organized by these regions. For this “v1” version of Auto Arborist, we restrict our focus to the US and Canada, with a single genus prediction task. There is room for Auto Arborist to grow in tasks and geographic area: many public tree censuses contain additional metadata (e.g. tree age, health, and trunk diameter), and there are many more cities we might include both in the US and Canada, and globally. We place our dataset in context with previously published tree classification datasets in Table 1, and emphasize the significantly enhanced scope in number of images across modalities, number of regions, and number of categories.

Extracting street level and aerial imagery. For each city, starting from the parsed tree census, we associate each tree census record to both street level and aerial images. For each tree in our dataset, we sample a $15\text{m} \times 15\text{m}$, 300×300 pixel RGB aerial image centered on the tree’s latitude and longitude. We consider all street level images taken within 2-10 meters of the record’s latitude and longitude, filtering out any images which do not meet all of the following criteria:

- Taken on or after Jan 1, 2018.
- Contains the base of the tree near the horizontal center of the image based on the projection of the tree’s latitude/longitude onto the image, based on estimated camera pose generated by the API.
- Contains a significant number of “tree” pixels based on a semantic segmentation model (to avoid cases when the tree has died or been removed, when possible) and does not contain any “person” or “bike rider” pixels based on a semantic segmentation model (to remove personally identifiable information).

After filtering, we have 2.6M tree records, each of which is

associated with one aerial image and 1-3 street level images, along with a date and GPS location.

3.2. Challenging aspects of the Auto Arborist data

By matching street level images from existing public records rather than collecting groundtruth labels from scratch, we have been able to achieve a scale much larger than any previous datasets. As we show, scale is important for generalizing to novel cities (which is the ultimate goal). But using public records to generate data across cities also introduces a number of challenges.

Sources of noise and ambiguity. First, we address several known sources of noise and ambiguity in our dataset. See Figure 3 for examples illustrating the following points.

- **Label noise:** There is a known discrepancy between label accuracy of volunteer citizen scientists vs. experts (e.g. with a PhD in Ecology) [12], and there is also no ecologically-agreed-upon definition of tree vs. bush. Cities differ in their labeling protocol.
- **Presence noise:** Tree records in censuses can often be outdated. Specifically, depending on the amount of time since the data was originally collected, there is increasing possibility that trees will have been removed or have died, and new trees planted.
- **Location noise:** Different cities use different data collection protocols and different sensors, leading to discrepancy in the accuracy of the position readings (e.g. by GPS). We estimate visually that they are usually accurate within ~ 3 meters.
- **Image quality:** Quality of aerial imagery varies for different cities. The primary tree in a street level image can sometimes be occluded — though we try to guard against this by removing images that are too far from the tree, sometimes vehicles block the tree from view. Qualitatively, access to multiple views frequently helps mitigate occlusion issues. Finally, deciduous trees vary in appearance across seasons, with leaves turning color and then dropping in the winter.
- **Unlabeled visible trees:** Trees on private property (e.g. yards) are not labeled in public censuses, but are visible in the background. While the tree of interest is often the most prominent, the presence of trees of other genera can create classification confusion.

Distribution shift and the long tail. One of our primary challenges is to be able to do well on novel cities that were not part of the training set, but in order for a model to do so, it will have to contend with distribution shift, where the training distribution of cities differs from the novel test distribution on some new city. We remark that there are two kinds of shift that we observe in our data — what we might call “label shift”, and “appearance shift”. Label shift refers to when the marginal distribution $P(y)$ of labels (genera) differs from city to city even if the appearance distribution

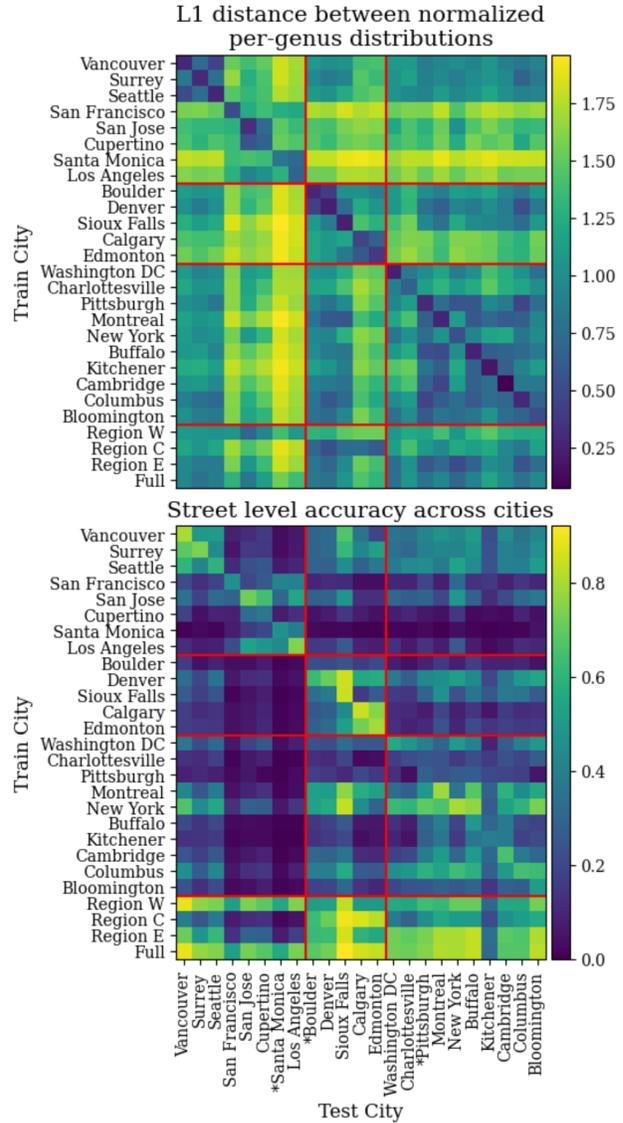


Figure 4. (Top) The distance between the distributions of training and test data for each training split and each test city (red lines represent regional boundaries). We use the L1 distance between the normalized per-class count vectors for each set as our measure of distributional distance. Because the class distribution is long-tailed and our test sets are split geographically within each city to prevent data poisoning, the train and test distribution are not identical within each city (the diagonal is non-uniform, and the matrix is not symmetric). (Bottom) Pairwise train/test accuracy from street level baselines.

of image x conditioned on a particular label $P(x|y)$ does not change (e.g., [91]). In our setting this simply can mean that species distributions vary geographically (e.g. we tend to see Palm trees in Southern California and less in Canada) but can also come from cities having different sizes (for example, Los Angeles is much larger than Santa Monica and thus contains many more species).

Figure 4 visualizes the distribution shift between every pair of cities (using L_1 distance between normalized genus distributions). In some cases we can see little overlap be-

tween genera from two cities, and for cities with similar location, i.e. Denver and Boulder, we tend to see high overlap in genus distribution. However even when two cities are very similar both in size and location, it is still generally the case that one city will contain a number of genera not found in the other due to the long tailed genus distribution. In the extreme setting of “train on one city, test on one city” we thus always have many test genera for which there are no training examples. And even in the regime of training on many cities and testing on a single holdout city we *still* typically have classes for which there are no training examples, implying value in expanding the dataset in future.

Beyond label shift, we also see “appearance shift” — the images of a particular genus can look different depending on the city. This is partly due to different backgrounds (which can in principle be handled by masking out the background pixels, but is out of scope for this work), but it can also be due to other external factors such as weather conditions (for example, we are likely to see more leafless trees from images in Edmonton than we are to see them in LA) or even “terroir” related factors like soil composition.

4. Evaluation Protocol

Since distribution shift is such a big factor in performance, we have chosen to set up our evaluation protocol to explicitly evaluate distribution shift based on 3 unique types of train/test splits, defined hierarchically:

1. **Per-city splits:** At the first level, we are interested in how well a city generalizes to itself. Here, each city has a defined training region and a defined test region, split geographically (usually based on latitude or longitude) to avoid overfitting on background characteristics. The test sets for each city are never used for training.
2. **Regional splits:** Next, we are interested in generalization within and across larger regions (e.g., how would we fare in cities on the East coast if we trained on West coast trees?) — for this level of evaluation, we split the cities into three regions, *Region W* (West), *Region C* (Central), and *Region E* (East) (Table 2). We build our regional training sets from the per-city training sets for that region. We hold out one city from each region (which we call “holdout cities”) to capture performance on an in-region novel city, and also show results on all out-of-region cities.
3. **Full dataset:** For the final and largest split, we combine training data across the three regions. We maintain the same holdout cities as the regional splits for training, and test on the test sets of all cities (including the holdouts).

Evaluation metrics. Due to the long-tailed distribution of the data across genera, a pure accuracy measure is insufficient to capture performance, as it is highly biased towards frequent species. Thus, we report accuracy alongside class-averaged recall (AR), calculated as average over all classes

of the proportion of correct predictions for the set of examples of that class (this is sometimes also called class-averaged accuracy). To capture performance in a more nuanced way, we also introduce an LVIS [58]-inspired breakdown of class-averaged recall for frequent ($n \geq 20,000$ examples), common ($100 \leq n < 20,000$ examples) and rare ($n < 100$ examples) subsets of our data. This results in 29 frequent, 150 common, and 165 rare genera, and we denote these metrics as FAR, CAR, and RAR respectively.

5. Experiments

We now demonstrate the benefits of having a multi-city, multiview dataset by training models on Auto Arborist. In this section we train separate aerial and street level baseline ResNet 101 models for each training split described in Section 4, including the training sets for each individual city, the regional splits, and the full dataset. Training details can be found in the supplementary.

Single city vs. regional vs. full dataset training. We begin by experimenting with single-view street level models (as the street level modality gives the most accurate results in isolation). In Figure 5, we compare performance on a city’s test set when training on that city’s training set (**city**), the aggregation of training sets from that city’s region (**region**), and all available training data (**full**). Unsurprisingly, we find that more data is better — we see an average improvement of 21.3% AR across cities when going from training on a single city to the full dataset. However we note that training on a region also gives strong performance gains over training on a city itself (average improvement of 18.3% AR), and for some test cities regional training can be on par with (or even slightly better than) training on the full set.

Cross-city generalization. Next we examine cross-city generalization, where we are interested in how effective it would be to train on a certain city A if we are interested in testing on B . For this analysis, we first perform all possible *cross training* combinations, training on every train split (including per-city, regional and full) and testing each model on the test set for each city. Results for these pairwise combinations are visualized in Figure 4 (bottom). Here we see regional “blocks” of strong generalization, reflecting that cities generalize well to cities in the same geographical area. For example, we tend to get good performance training on one of the Pacific Northwest cities (Seattle, Vancouver, Surrey) and testing on another.

We can also see that some cities tend to generalize quite well to other cities on average whereas some cities tend to generalize poorly to other cities. Figure 6 shows this effect in more detail — here we use a given city as a training set and report the spread of performance when applied to other cities’ test sets. In this plot, a larger gap between “self-test” (red stars) and the box implies less generalizability. Here, to remove confounding factors due to test genera not seen or

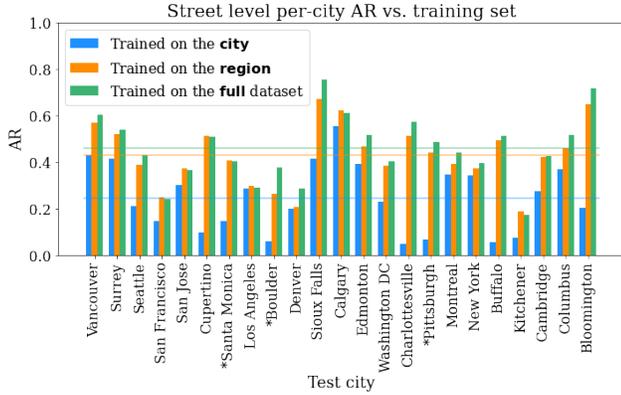


Figure 5. Performance growth when adding regional and continental diversity. For each city, we show test performance from a model trained on that city, trained on the respective region for that city, and trained on the full dataset. Note that performance improves on our holdout cities as well, despite the regional and full training sets not including data from those cities. Average performance from models trained on per-city, regional, and full are shown as horizontal lines.

rarely seen during training, we restrict computation of AR for train city A and test city B to only the “frequent” genera seen in the train split of A and the test split of B .

We observe that cities that are poor “training cities” (on the left side of Fig. 6) tend to be smaller and have poor performance overall, though this is not universally true (consider San Francisco). On the other hand, large cities (e.g. NYC) tend to generalize well on average. But we also see that there are no cities which generalize optimally to all others, and optimal generalization performance is only reached by training on the full dataset. Even restricting our attention to frequent, shared classes, we find that generalization ability continues to be highly correlated with label distribution similarity. In Figure 7, we compare AR across these shared, frequent genera with the L1 distribution distance for three cities and show they are negatively correlated — increased label distribution distance implies worse performance, even on frequently-seen classes shared between train and test cities.

Value-add of multiple views. Finally, in Table 3, we show that the multiview aspects of our dataset bring value. Overall, our street level models perform much better than the aerial models, generally with a difference of $>20\%$ AR and we see that using multiple views of a tree outperforms a single view. We have experimented with several techniques to combine information across street level views and aerial imagery, and find that while most of the predictive value comes from the street level imagery, there is benefit in incorporating aerial information. We combine the modalities via a simple method: average pooling the logits from multiple street level images and then combining with aerial logits via a Mixture of Experts (MoE) model:

$$f(x_{SL}, x_A) = x_{SL} \cdot \text{sigmoid}(w) + x_A \cdot (1 - \text{sigmoid}(w)). \quad (1)$$

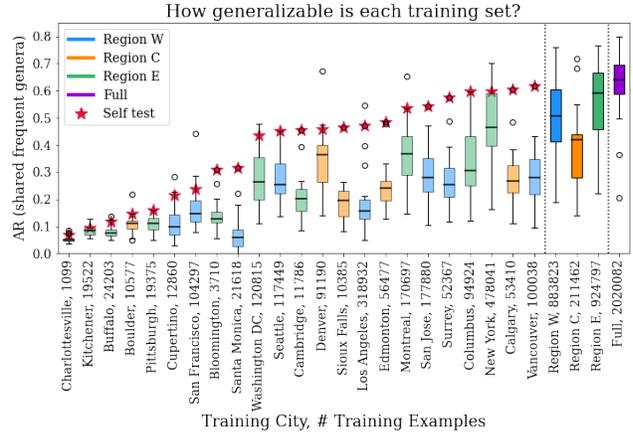


Figure 6. For each training set, we show the distribution of AR across the test cities, and highlight the “self test” case where a city is tested on its own test set. A larger gap between “self test” and the box implies less generalizability. Here, to remove confounding factors due to test genera not seen or rarely seen during training, we compute AR for train city A and test city B to be the average per-genus recall across the “frequent” genera seen in both train(A) and test(B).

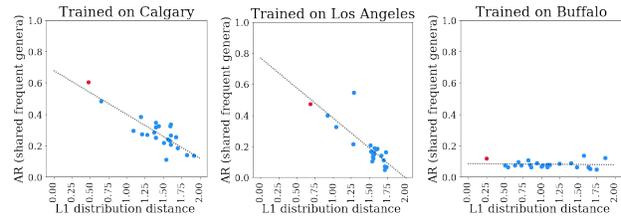


Figure 7. Digging further into the generalizability of a given training set, we visualize the generalization gap between AR on shared, frequent genera testing on that same city (“self-test”, in red) vs other cities, and plot against the L1 distance between the genus distribution of train vs. test, as seen in Fig. 4. We see that frequently these are anti-correlated, but training sets for some cities (like Buffalo) struggle to perform well across the board.

where x_{SL} and x_A are street level and aerial logits, $w \in \mathbb{R}^n$ are learned parameters, and n is the number of classes.

Combining modalities in this way yields an average $\sim 1\%$ boost for each regional model, compared to average-pooling logits across multiple street level views, and a 3-5% boost over predicting from a single street level image. For the full training set, we find that preserving the regional variations in learned MoE weights (w) is important — thus our best model (which achieves 49.96% AR) uses street level and aerial models trained on the full dataset but MoE weights specialized for the region to which a city belongs. We conjecture that this regional dependence is due primarily to regional variations in aerial image quality/availability. In Fig. 8 we visualize the per-genus weights learned by the MoE per region. Looking more closely at the MoE weights, we find that our models only assigns nonzero weights to aerial data for classes that have ≥ 400 training examples. Moreover, we see that we are able to rely on aerial images more in Region W compared to the other two regions.

We show results from our best model in Table 4, report-

Train Set	Aerial	1 SL	3 SL	A+SL
Region W	20.63	41.53	45.12	46.07
Region C	18.8	44.77	46.91	47.12
Region E	17.54	43.25	45.13	46.21
Full	18.7	46.13	49.0	49.23
Full w/ Regional MoE				49.96

Table 3. City-averaged percent AR for different regions and ensembling strategies. Street level imagery is much more informative than aerial, and combining multiple street level images gives a further boost. However, even though aerial performance on its own is quite low, we see benefit in adding the aerial imagery when making predictions. We find that while the features from the full model are more discriminative, we see best performance using full model features paired with region-specific Mixture of Experts to combine aerial and street level predictions.

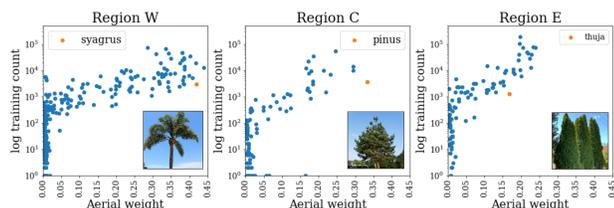


Figure 8. Each regional MoE learns to use aerial information only for genera with more than ~ 400 training examples. Notably, the distribution of the three is quite different, and there are certain genera that are more “aerially distinctive” (we have highlighted one for each region).

ing accuracy and AR for the full dataset, and broken down by frequent and common genera. Notably, many cities have $>80\%$ accuracy, and Vancouver and Sioux Falls see $>90\%$. There is still significant room to improve on AR across the board. Rare class performance was 0.0 for every city, unsurprising given most rare classes have <10 examples. This points to potential gain from low-shot and long-tail learning methods such as logit-adjustment, but we find that such methods struggle to perform well under such a high degree of imbalance (see Supplementary).

6. Limitations and future work

We have presented a baseline modeling approach meant to highlight the performance of a typical CNN and present simple methods for combining signals from multiple views — there is much room for improvement, particularly on rare classes. In future, to predict on cities with no past census, we would need to first localize and geocode the trees to be classified. We also hope to expand our dataset to include more cities, both in North America and worldwide, and include species level predictions and additional features such as tree size and health.

Auto Arborist represents an important first step towards global-scale urban forest monitoring. This has implications for environmental justice: given that marginalized communities have less access to urban greenery, systems trained on Auto Arborist could help equitize access to urban forests by empowering quantifiable analysis and targeted replanting. However we must be responsible with our technology

City	Acc	AR	FAR	CAR
Vancouver, BC	93.28	67.51	82.76	63.35
Surrey, BC	82.35	58.96	75.82	48.80
Seattle, WA	79.68	46.55	74.65	43.08
San Francisco, CA	58.71	26.39	37.87	31.37
San Jose, CA	77.71	40.07	63.35	41.13
Cupertino, CA	74.14	56.86	65.28	55.40
Santa Monica, CA	56.26	43.29	64.93	44.09
Los Angeles, CA	76.24	32.62	52.56	35.80
Boulder City, CO	73.23	42.23	58.61	32.88
Denver, CO	76.46	29.72	57.16	22.02
Sioux Falls, SD	93.78	76.76	81.52	62.50
Calgary, AB	88.81	62.18	70.92	52.32
Edmonton, AB	87.55	56.67	62.58	43.99
Washington, DC	77.44	44.49	67.31	30.05
Charlottesville, VA	73.52	57.77	73.38	42.16
Pittsburgh, PA	78.84	54.93	71.83	43.97
Montreal, QC	85.51	49.49	64.99	39.08
New York, NY	82.54	42.77	66.38	28.16
Buffalo, NY	86.03	54.01	71.92	43.41
Kitchener, ON	33.96	17.94	21.31	4.49
Cambridge, ON	72.16	47.69	65.84	34.38
Columbus, OH	69.28	55.71	68.29	47.32
Bloomington, IN	85.50	73.52	79.82	64.46

Table 4. Per-city performance (%) with our best model trained on the full dataset combining aerial and multiview street level modalities. AR is class-averaged recall for each city, averaged over the test classes for that city. FAR is “Frequent” AR, CAR is “Common” AR, which serve to further disentangle the commonality of a species in the training data with its per-city performance. Holdout cities in bold.

— to this end, we protect the privacy of residents of these urban and suburban areas by explicitly filtering out any imagery containing humans, and blur vehicle plates. Secondly we will need efficient human-in-the-loop validation protocols before such a system could be trusted, to ensure science policy is not based on poorly-generalized ML predictions.

7. Seeing the forest for the trees (Conclusions)

Climate change and loss of ecological diversity are among the most pressing issues of our time. Monitoring is a first crucial step to understanding and mitigating the effects of global warming on urban forests, but many cities cannot afford regular tree censuses. Towards the goal of broad, accessible, and affordable urban forest monitoring, we have introduced the Auto Arborist dataset. This dataset is the first of its kind to expand beyond a single city and common categories: Auto Arborist contains 2.6 million trees across 23 cities, covering 344 unique genera. This dataset will enable the computer vision community to tackle urban forest monitoring at scale, and our evaluation protocols help us measure performance without data poisoning, and to evaluate generalization to novel cities.

Acknowledgements We would like to thank R. Alcantara, T. Birch, V. Birodgar, T. Duerig, K. Gill, J. Hea, A. Kundu, P. Lall, M. Manolides, D. Morris, S. Marquardt, P. Perona, J. Quintero, D. Ross, T. Taylor, J. Thamkul, and X. Yin for their support. This work was supported in part by the Resnick Sustainability Institute and undertaken while S. Beery was a Student Researcher at Google.

References

- [1] inaturalist. <https://www.inaturalist.org>. Accessed: 2021-11-11. 3
- [2] La begins massive street tree census. 2
- [3] Samuel Adelabu, Onesimo Mutanga, Elhadi E Adam, and Moses Azong Cho. Exploiting machine learning algorithms for tree species classification in a semiarid woodland using rapideye image. *Journal of Applied Remote Sensing*, 7(1):073480, 2013. 3
- [4] Michael Alonzo, Bodo Bookhagen, and Dar A Roberts. Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, 148:70–83, 2014. 2, 3
- [5] Mike Alonzo, Keely Roth, and Dar Roberts. Identifying santa barbara’s urban tree species from aviris imagery using canonical discriminant analysis. *Remote Sensing Letters*, 4(5):513–521, 2013. 2, 3
- [6] Daniel Amigo, David Sánchez Pedroche, Jesús García, and José M Molina. Automatic individual tree detection from combination of aerial imagery, lidar and environment context. In *International Workshop on Soft Computing Models in Industrial and Environmental Applications*, pages 294–303. Springer, 2021. 2, 3
- [7] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. Google street view: Capturing the world at street level. *Computer*, 43(6):32–38, 2010. 3, 4
- [8] EC Anyanwu and I Kanu. The role of urban forest in the protection of human environmental health in geographically-prone unpredictable hostile weather conditions. *International Journal of Environmental Science & Technology*, 3(2):197–201, 2006. 1
- [9] Mélaïne Aubry-Kientz, Anthony Laybros, Ben Weinstein, James GC Ball, Toby Jackson, David Coomes, and Grégoire Vincent. Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3927–3936, 2021. 2
- [10] Bulut Aygüneş, Selim Aksoy, and Ramazan Gökberk Cinbiş. Weakly supervised deep convolutional networks for fine-grained object recognition in multispectral images. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 1478–1481. IEEE, 2019. 3
- [11] Bulut Aygunes, Ramazan Gokberk Cinbis, and Selim Aksoy. Weakly supervised instance attention for multisource fine-grained object recognition with an application to tree species classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176:262–274, 2021. 3
- [12] Nick Bancks, Eric A North, and Gary R Johnson. An analysis of agreement between volunteer-and researcher-collected urban tree inventory data. 2018. 5
- [13] Sara Beery, Arushi Agarwal, Elijah Cole, and Vighnesh Birodkar. The iWildCam 2021 competition dataset. *The Eighth Fine-Grained Visual Categorization Workshop at CVPR*, 2021. 2
- [14] Sara Beery, Elijah Cole, and Arvi Gjoka. The iWildCam 2020 competition dataset. *The Seventh Fine-Grained Visual Categorization Workshop at CVPR*, 2020. 2
- [15] Sara Beery, Dan Morris, and Pietro Perona. The iWildCam 2019 challenge dataset. *The Sixth Fine-Grained Visual Categorization Workshop at CVPR*, 2019. 2
- [16] Sara Beery, Grant Van Horn, Oisín MacAodha, and Pietro Perona. The iWildCam 2018 challenge dataset. *The Fifth Fine-Grained Visual Categorization Workshop at CVPR*, 2018. 2
- [17] Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *Proceedings of the European conference on computer vision (ECCV)*, pages 456–473, 2018. 2
- [18] Even Bergseng, Hans Ole Ørka, Erik Næsset, and Terje Gobakken. Assessing forest inventory information obtained from different inventory approaches and remote sensing data sources. *Annals of forest science*, 72(1):33–45, 2015. 2, 3
- [19] Filip Biljecki and Koichi Ito. Street view imagery in urban analytics and gis: A review. *Landscape and Urban Planning*, 215:104217, 2021. 3
- [20] J Blum. Contribution of ecosystem services to air quality and climate change mitigation policies: the case of urban forests in barcelona, spain. In *Urban Forests*, pages 21–54. Apple Academic Press, 2017. 1
- [21] Bob Bolin, Sara Grineski, and Timothy Collins. The geography of despair: Environmental racism and the making of south phoenix, arizona, usa. *Human Ecology Review*, pages 156–168, 2005. 1
- [22] Christopher G Boone, Geoffrey L Buckley, J Morgan Grove, and Chona Sister. Parks and people: An environmental justice inquiry in baltimore, maryland. *Annals of the Association of American Geographers*, 99(4):767–787, 2009. 2
- [23] M Boschetti, L Boschetti, S Oliveri, L Casati, and I Canova. Tree species mapping with airborne hyper-spectral mivis data: the ticino park study case. *International Journal of Remote Sensing*, 28(6):1251–1261, 2007. 3
- [24] Leslie Brandt, Abigail Derby Lewis, Robert Fahey, Lydia Scott, Lindsay Darling, and Chris Swanston. A framework for adapting urban forests to climate change. *Environmental Science & Policy*, 66:393–402, 2016. 1
- [25] Martin Brandt, Compton J Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, et al. An unexpectedly large count of trees in the west african sahara and sahel. *Nature*, 587(7832):78–82, 2020. 2, 3
- [26] Tomas Brandtberg. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5):325–340, 2007. 3
- [27] Steve Branson, Jan Dirk Wegner, David Hall, Nico Lang, Konrad Schindler, and Pietro Perona. From google maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135:13–30, Jan 2018. 2, 3

- [28] Gregory N Bratman, Christopher B Anderson, Marc G Berman, Bobby Cochran, Sjerp De Vries, Jon Flanders, Carl Folke, Howard Frumkin, James J Gross, Terry Hartig, et al. Nature and mental health: An ecosystem service perspective. *Science advances*, 5(7):eaax0903, 2019. 1
- [29] DJ Brus, GM Hengeveld, DJJ Walvoort, PW Goedhart, AH Heidema, GJ Nabuurs, and K Gunia. Statistical mapping of tree species over europe. *European Journal of Forest Research*, 131(1):145–157, 2012. 3
- [30] Alexandre Carleer and Eléonore Wolff. Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering & Remote Sensing*, 70(1):135–140, 2004. 3
- [31] Dominic Chambers, Catherine Périé, Nicolas Casajus, and Sylvie de Blois. Challenges in modelling the abundance of 105 tree species in eastern north america using climate, edaphic, and topographic variables. *Forest Ecology and Management*, 291:20–29, 2013. 3
- [32] Francesco Chianucci, Andrea Cutini, Piermaria Corona, and Nicola Puletti. Estimation of leaf area index in understory deciduous trees using digital photography. *Agricultural and forest meteorology*, 198:259–264, 2014. 3
- [33] Moses Azong Cho, Pravesh Debba, Renaud Mathieu, Laven Naidoo, JAN Van Aardt, and Gregory P Asner. Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4133–4142, 2010. 3
- [34] Moses Azong Cho, Renaud Mathieu, Gregory P Asner, Laven Naidoo, JAN Van Aardt, Abel Ramoelo, Pravesh Debba, Konrad Wessels, Russell Main, Izak PJ Smit, et al. Mapping tree species composition in south african savannas using an integrated airborne spectral and lidar system. *Remote Sensing of Environment*, 125:214–226, 2012. 3
- [35] Matthew L Clark and Dar A Roberts. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensing*, 4(6):1820–1855, 2012. 3
- [36] Matthew L Clark, Dar A Roberts, and David B Clark. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote sensing of environment*, 96(3-4):375–398, 2005. 3
- [37] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 3
- [38] Thomas W Crowther, Henry B Glick, Kristofer R Covey, Charlie Bettigole, Daniel S Maynard, Stephen M Thomas, Jeffrey R Smith, Gregor Hintler, Marlyse C Duguid, Giuseppe Amatulli, et al. Mapping tree density at a global scale. *Nature*, 525(7568):201–205, 2015. 2
- [39] María Culman, Andrés C Rodríguez, Jan Dirk Wegner, Stephanie Delalieux, and Ben Somers. Deep learning for sub-pixel palm tree classification using spaceborne sentinel-2 imagery. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXIII*, volume 11856, pages 45–50. SPIE, 2021. 3
- [40] Darius S Culvenor. A spatial clustering approach to automated tree crown delineation. In *Proceedings of the forum on automated interpretation of high spatial resolution digital imagery for forestry, 1999*, pages 67–88. Canadian Forest Service, 1999. 2
- [41] Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle. Tree species classification in the southern alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and lidar data. *Remote sensing of environment*, 123:258–270, 2012. 3
- [42] Michele Dalponte, Hans Ole Ørka, Liviu Theodor Ene, Terje Gobakken, and Erik Næsset. Tree crown delineation and tree species classification in boreal forests using hyperspectral and als data. *Remote sensing of environment*, 140:306–317, 2014. 3
- [43] Hubert de Foresta, Eduardo Somarriba, August Temu, Désirée Boulanger, Hélène Feuilly, Michelle Gauthier, and D Taylor. Towards the assessment of trees outside forests: a thematic report prepared in the framework of the global forest resources assessment 2010. *World*, 7, 2020. 3
- [44] Steve H Dreistadt, Donald L Dahlsten, and Gordon W Frankie. Urban forests and insect ecology. *BioScience*, 40(3):192–198, 1990. 1
- [45] Fábio Duarte and Carlo Ratti. What big data tell us about trees and the sky in the cities. In *Humanizing Digital Reality*, pages 59–62. Springer, 2018. 3
- [46] Laura Duncanson and Ralph Dubayah. Monitoring individual tree-based change with airborne lidar. *Ecology and evolution*, 8(10):5079–5089, 2018. 2
- [47] Theodore S Eisenman, Galina Churkina, Sunit P Jariwala, Prashant Kumar, Gina S Lovasi, Diane E Pataki, Kate R Weinberger, and Thomas H Whitlow. Urban trees, air quality, and asthma: An interdisciplinary review. *Landscape and urban planning*, 187:47–59, 2019. 1
- [48] Robin Engler, Lars T Waser, Niklaus E Zimmermann, Marcus Schaub, Savvas Berdos, Christian Ginzler, and Achilleas Psomas. Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. *Forest Ecology and Management*, 310:64–73, 2013. 3
- [49] Fabian Ewald Fassnacht, Hooman Latifi, Krzysztof Stereńczak, Aneta Modzelewska, Michael Lefsky, Lars T Waser, Christoph Straub, and Aniruddha Ghosh. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186:64–87, 2016. 2, 3
- [50] Angelos Filos, Panagiotis Tigkas, Rowan McAllister, Nicholas Rhinehart, Sergey Levine, and Yarin Gal. Can autonomous vehicles identify, recover from, and adapt to distribution shifts? In *International Conference on Machine Learning*, pages 3145–3153. PMLR, 2020. 2
- [51] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020. 2

- [52] Ed Gerrish and Shannon Lea Watkins. The relationship between urban forests and income: A meta-analysis. *Landscape and Urban Planning*, 170:293–308, 2018. 1
- [53] Azadeh Ghiyamat, Helmi Zuhaidi M Shafri, Ghafour Amouzad Mahdiraji, Abdul Rashid M Shariff, and Shattri Mansor. Hyperspectral discrimination of tree species with different classifications using single-and multiple-endmember. *International journal of applied earth observation and geoinformation*, 23:177–191, 2013. 3
- [54] Peng Gong, Ruiliang Pu, and Bin Yu. Conifer species recognition: An exploratory analysis of in situ hyperspectral data. *Remote sensing of Environment*, 62(2):189–200, 1997. 3
- [55] François A Gougeon. Comparison of possible multispectral classification schemes for tree crowns individually delineated on high spatial resolution multis images. *Canadian journal of remote sensing*, 21(1):1–9, 1995. 3
- [56] Sarah J Graves, Sergio Marconi, Dylan Stewart, Ira Harmon, Ben G Weinstein, Yuzi Kanazawa, Victoria M Scholl, Maxwell B Joseph, Joseph McClinchy, Like Browne, et al. Data science competition for cross-site delineation and classification of individual trees from airborne remote sensing data. *bioRxiv*, 2021. 1, 2, 3
- [57] Haiyan Guan, Yongtao Yu, Zheng Ji, Jonathan Li, and Qi Zhang. Deep learning-based tree classification using mobile lidar data. *Remote Sensing Letters*, 6(11):864–873, 2015. 3
- [58] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5356–5364, 2019. 6
- [59] Arto Haara and Mika Haarala. Tree species classification using semi-automatic delineation of trees on aerial images. *Scandinavian Journal of Forest Research*, 17(6):556–565, 2002. 3
- [60] Hamid Hamraz, Marco A Contreras, and Jun Zhang. Forest understory trees can be segmented accurately within sufficiently dense airborne laser scanning point clouds. *Scientific reports*, 7(1):1–9, 2017. 3
- [61] Hamid Hamraz, Nathan B Jacobs, Marco A Contreras, and Chase H Clark. Deep learning for conifer/deciduous classification of airborne lidar 3d point clouds representing individual trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:219–230, 2019. 3
- [62] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, Stephen V Stehman, Scott J Goetz, Thomas R Loveland, et al. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013. 2
- [63] Sean Hartling, Vasit Sagan, Paheding Sidike, Maitiniyazi Maimaitijiang, and Joshua Carron. Urban tree species classification using a worldview-2/3 and lidar data fusion approach and deep learning. *Sensors*, 19(6):1284, 2019. 2, 3
- [64] Richard J Hauer, Ian S Hanou, and David Sivyer. Planning for active management of future invasive pests affecting urban forests: the ecological and economic effects of varying dutch elm disease management practices for street trees in milwaukee, wi usa. *Urban Ecosystems*, 23(5):1005–1022, 2020. 1
- [65] Ville Heikkinen, Timo Tokola, Jussi Parkkinen, Ilkka Korpela, and Timo Jaaskelainen. Simulated multispectral imagery for tree species classification using support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3):1355–1364, 2009. 3
- [66] Johannes Heinzel and Barbara Koch. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. *International Journal of Applied Earth Observation and Geoinformation*, 18:101–110, 2012. 2
- [67] Daniel Henstra. Toward the climate-resilient city: extreme weather and urban climate adaptation policies in two canadian provinces. *Journal of Comparative Policy Analysis: Research and Practice*, 14(2):175–194, 2012. 1
- [68] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. CyCADA: Cycle-consistent adversarial domain adaptation. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1989–1998. PMLR, 10–15 Jul 2018. 2
- [69] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv preprint arXiv:1612.02649*, 2016. 2
- [70] Markus Hollaus, Werner Mücke, Bernhard Höfle, Wouter Dorigo, Norbert Pfeifer, Wolfgang Wagner, Christoph Bauerhansl, and Bruno Regner. Tree species classification based on full-waveform airborne laser scanning data. *Proceedings of SILVILASER*, pages 54–62, 2009. 2, 3
- [71] Johan Holmgren and Åsa Persson. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, 90(4):415–423, 2004. 3
- [72] Johan Holmgren, Å Persson, and U Söderman. Species identification of individual trees by combining high resolution lidar data with multi-spectral images. *International Journal of Remote Sensing*, 29(5):1537–1552, 2008. 3
- [73] Diane Hope, Corinna Gries, Weixing Zhu, William F Fagan, Charles L Redman, Nancy B Grimm, Amy L Nelson, Chris Martin, and Ann Kinzig. Socioeconomics drive urban plant diversity. *Proceedings of the national academy of sciences*, 100(15):8788–8792, 2003. 1
- [74] A Hovi, L Korhonen, J Vauhkonen, and I Korpela. Lidar waveform features for tree species classification and their sensitivity to tree-and acquisition related parameters. *Remote sensing of environment*, 173:224–237, 2016. 3
- [75] Markus Immitzer, Clement Atzberger, and Tatjana Koukal. Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. *Remote sensing*, 4(9):2661–2693, 2012. 3
- [76] Kenta Itakura and Fumiki Hosoi. Automatic tree detection from three-dimensional images reconstructed from 360 spherical camera using yolo v2. *Remote Sensing*, 12(6):988, 2020. 3

- [77] Marek K Jakubowski, Wenkai Li, Qinghua Guo, and Maggi Kelly. Delineating individual trees from lidar data: A comparison of vector-and raster-based segmentation approaches. *Remote Sensing*, 5(9):4163–4186, 2013. 2
- [78] Ryan R Jensen, Perry J Hardin, and Andrew J Hardin. Classification of urban tree species using hyperspectral imagery. *Geocarto International*, 27(5):443–458, 2012. 2, 3
- [79] Björn Lütjens Dava Newman Tom Crowther Ce Zhang Kenza Amara, David Dao. One forest: Towards a global species dataset by fusing remote sensing and citizen science data with graph neural networks, 2020. 3
- [80] Thomas Key, Timothy A Warner, James B McGraw, and Mary Ann Fajvan. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. *Remote Sensing of Environment*, 75(1):100–112, 2001. 3
- [81] Sooyoung Kim. *Individual tree species identification using LIDAR-derived crown structures and intensity data*. University of Washington, 2008. 3
- [82] Pang Wei Koh, Shiori Sagawa, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanus Phillips, Irena Gao, Tony Lee, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*, pages 5637–5664. PMLR, 2021. 2
- [83] Ilkka Korpela, Lauri Mehtätalo, Lauri Markelin, Anne Seppänen, Annika Kangas, et al. Tree species identification in aerial image data using directional reflectance signatures. *Silva Fenn*, 48(3):1–20, 2014. 3
- [84] Daniel Laumer, Nico Lang, Natalie van Doorn, Oisín Mac Aodha, Pietro Perona, and Jan Dirk Wegner. Geocoding of trees from street addresses and street-level images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:125–136, 2020. 3
- [85] Sébastien Lefèvre, Devis Tuia, Jan Dirk Wegner, Timothée Produit, and Ahmed Samy Nassar. Toward seamless multi-view scene analysis from satellite to street level. *Proceedings of the IEEE*, 105(10):1884–1899, 2017. 3
- [86] Misha Leong, Robert R Dunn, and Michelle D Trautwein. Biodiversity and socioeconomics in the city: a review of the luxury effect. *Biology Letters*, 14(5):20180082, 2018. 1
- [87] Linyuan Li, Jun Chen, Xihan Mu, Weihua Li, Guangjian Yan, Donghui Xie, and Wuming Zhang. Quantifying understory and overstory vegetation cover using uav-based rgb imagery in forest plantation. *Remote Sensing*, 12(2):298, 2020. 3
- [88] Xiaojiang Li. Examining the spatial distribution and temporal change of the green view index in new york city using google street view images and deep learning. *Environment and Planning B: Urban Analytics and City Science*, 48(7):2039–2054, 2021. 3
- [89] Xiaojiang Li, Chuanrong Zhang, Weidong Li, Robert Ricard, Qingyan Meng, and Weixing Zhang. Assessing street-level urban greenery using google street view and a modified green view index. *Urban Forestry & Urban Greening*, 14(3):675–685, 2015. 3
- [90] Wenzhi Liao, Fricke Van Coillie, Lianru Gao, Liwei Li, Bing Zhang, and Jocelyn Chanussot. Deep learning for fusion of apex hyperspectral and full-waveform lidar remote sensing data for tree species mapping. *IEEE Access*, 6:68716–68729, 2018. 3
- [91] Zachary Lipton, Yu-Xiang Wang, and Alexander Smola. Detecting and correcting for label shift with black box predictors. In *International conference on machine learning*, pages 3122–3130. PMLR, 2018. 5
- [92] Changfu Liu and Xiaoma Li. Carbon storage and sequestration by urban forests in shenyang, china. *Urban Forestry & Urban Greening*, 11(2):121–128, 2012. 1
- [93] Dexter H Locke, Billy Hall, J Morgan Grove, Steward TA Pickett, Laura A Ogden, Carissa Aoki, Christopher G Boone, and Jarlath PM O’Neil-Dunne. Residential housing segregation and urban tree canopy in 37 us cities. *npj Urban Sustainability*, 1(1):1–9, 2021. 1, 3
- [94] Dexter H Locke, Shawn M Landry, J Morgan Grove, and Rinku Roy Chowdhury. What’s scale got to do with it? models for urban tree canopy. *Journal of Urban Ecology*, 2(1), 2016. 2
- [95] Stefanie Lumnitz, Tahia Devisscher, Jerome R Mayaud, Valentina Radic, Nicholas C Coops, and Verena C Griess. Mapping trees along urban street networks with deep learning and street-level imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:144–157, 2021. 2, 3
- [96] Dawei Ma, Hongchao Fan, Wenwen Li, and Xuan Ding. The state of mapillary: An exploratory analysis. *ISPRS International Journal of Geo-Information*, 9(1):10, 2020. 3
- [97] Oisín Mac Aodha, Elijah Cole, and Pietro Perona. Presence-only geographical priors for fine-grained image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9596–9606, 2019. 2
- [98] Sergio Marconi, Sarah J Graves, Dihong Gong, Morteza Shahriari Nia, Marion Le Bras, Bonnie J Dorr, Peter Fontana, Justin Gearhart, Craig Greenberg, Dave J Harris, et al. A data science challenge for converting airborne remote sensing data into ecological information. *PeerJ*, 6:e5843, 2019. 2, 3
- [99] Kendra Marshman. The eye of the storm: extreme weather events and sustainable urban forest management. *Dalhousie Journal of Interdisciplinary Management*, 14, 2018. 1
- [100] Robert I McDonald, Timm Kroeger, Ping Zhang, and Perrine Hamel. The value of us urban tree cover for reducing heat-related health impacts and electricity consumption. *Ecosystems*, 23(1):137–150, 2020. 1
- [101] Ahmed Samy Nassar, Stefano D’aronco, Sébastien Lefèvre, and Jan D Wegner. Geograph: Graph-based multi-view object detection with geometric cues end-to-end. In *European Conference on Computer Vision*, pages 488–504. Springer, 2020. 2, 3
- [102] Ahmed Samy Nassar, Sébastien Lefèvre, and Jan Dirk Wegner. Simultaneous multi-view instance detection with learned geometric soft-constraints. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6559–6568, 2019. 2, 3

- [103] Ahmed Samy Nassar, Sébastien Lefèvre, and Jan Dirk Wegner. Multi-view instance matching with learned geometric soft-constraints. *ISPRS International Journal of Geo-Information*, 9(11):687, 2020. 3
- [104] Lorien Nesbitt, Michael J Meitner, Cynthia Girling, Stephen RJ Sheppard, and Yuhao Lu. Who has access to urban vegetation? a spatial analysis of distributional green equity in 10 us cities. *Landscape and Urban Planning*, 181:51–79, 2019. 1
- [105] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Buló, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017. 3
- [106] David J Nowak and Eric J Greenfield. Us urban forest statistics, values, and projections. *Journal of Forestry*, 116(2):164–177, 2018. 1
- [107] UN. Department of Economic and Social Affairs. 2018 revision of world urbanization prospects, 2018. 1
- [108] Jarlath O’Neil-Dunne, Sean MacFaden, and Anna Royar. A versatile, production-oriented approach to high-resolution tree-canopy mapping in urban and suburban landscapes using geobia and data fusion. *Remote sensing*, 6(12):12837–12865, 2014. 2
- [109] C Ordóñez and Peter N Duinker. Assessing the vulnerability of urban forests to climate change. *Environmental Reviews*, 22(3):311–321, 2014. 2
- [110] Alin-Ionuț Pleșoianu, Mihai-Sorin Stupariu, Ionuț Șandric, Ileana Pătru-Stupariu, and Lucian Drăguț. Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model. *Remote Sensing*, 12(15):2426, 2020. 2, 3
- [111] Eetu Puttonen, Paula Litkey, and Juha Hyypä. Individual tree species classification by illuminated—shaded area separation. *Remote Sensing*, 2(1):19–35, 2010. 3
- [112] Ilya Sherstyuk Sergio Parra Rebekah Loving, Arushi Agarwal. A network fusion model pipeline for multi-modal, deep learning for tree crown detection, 2020. 2
- [113] WG Rhode. Multispectral sensing of forest tree species. *Photogrammetric Engineering*, 38(12), 1972. 3
- [114] Anastasiia Safonova, Siham Tabik, Domingo Alcaraz-Segura, Alexey Rubtsov, Yuriy Maglinets, and Francisco Herrera. Detection of fir trees (*abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote sensing*, 11(6):643, 2019. 1, 3
- [115] Frank S Santamour Jr. Trees for urban planting: diversity uniformity, and common sense. *C. Elevitch, The Overstory Book: Cultivating connections with trees*, pages 396–399, 2004. 1
- [116] L Sayn-Wittgenstein. Recognition of tree species on aerial photographs. 1978. 3
- [117] Sebastian Schnell, Christoph Kleinn, and Göran Ståhl. Monitoring trees outside forests: a review. *Environmental monitoring and assessment*, 187(9):1–17, 2015. 3
- [118] Kirsten Schwarz, Michail Fragkias, Christopher G Boone, Weiqi Zhou, Melissa McHale, J Morgan Grove, Jarlath O’Neil-Dunne, Joseph P McFadden, Geoffrey L Buckley, Dan Childers, et al. Trees grow on money: urban tree canopy cover and environmental justice. *PloS one*, 10(4):e0122051, 2015. 2
- [119] Ian Seiferling, Nikhil Naik, Carlo Ratti, and Raphaël Proulx. Green streets- quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, 165:93–101, 2017. 3
- [120] Kurt Shickman and Global Cool Cities Alliance. Cool policies for cool cities: Best practices for mitigating urban heat islands in north american cities. 2014. 1
- [121] Nathan L Stephenson, AJ Das, R Condit, SE Russo, PJ Baker, Noelle G Beckman, DA Coomes, ER Lines, WK Morris, Nadja Rüger, et al. Rate of tree carbon accumulation increases continuously with tree size. *Nature*, 507(7490):90–93, 2014. 1
- [122] Dylan Stewart, Alina Zare, Sergio Marconi, Ben G Weinstein, Ethan P White, Sarah J Graves, Stephanie Ann Bohlman, and Aditya Singh. Randcrowns: A quantitative metric for imprecisely labeled tree crown delineation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021. 2
- [123] Esra Suel, John W Polak, James E Bennett, and Majid Ez-zati. Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific reports*, 9(1):1–10, 2019. 3
- [124] Gencer Sumbul, Ramazan Gokberk Cinbis, and Selim Aksoy. Fine-grained object recognition and zero-shot learning in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):770–779, 2017. 2, 3
- [125] Mark S Taylor, Benedict W Wheeler, Mathew P White, Theodoros Economou, and Nicholas J Osborne. Research note: Urban street tree density and antidepressant prescription rates—a cross-sectional study in london, uk. *Landscape and Urban Planning*, 136:174–179, 2015. 1
- [126] Andrew Thirlwell and Ognjen Arandjelović. Big data driven detection of trees in suburban scenes using visual spectrum eye level photography. *Sensors*, 20(11):3051, 2020. 3
- [127] Dong Tianyang, Zhang Jian, Gao Sibin, Shen Ying, and Fan Jing. Single-tree detection in high-resolution remote-sensing images based on a cascade neural network. *ISPRS International Journal of Geo-Information*, 7(9):367, 2018. 3
- [128] Lars T Waser, C Ginzler, M Kuechler, Emmanuel Baltasvias, and Lorenz Hurni. Semi-automatic classification of tree species in different forest ecosystems by spectral and geometric variables derived from airborne digital sensor (ads40) and rc30 data. *Remote Sensing of Environment*, 115(1):76–85, 2011. 3
- [129] Shannon Lea Watkins and Ed Gerrish. The relationship between urban forests and race: A meta-analysis. *Journal of environmental management*, 209:152–168, 2018. 1
- [130] Jan D. Wegner, Steve Branson, David Hall, Konrad Schindler, and Pietro Perona. Cataloging public objects using aerial and street-level images — urban trees. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6014–6023, 2016. 2, 3
- [131] Ben G Weinstein, Sarah J Graves, Sergio Marconi, Aditya Singh, Alina Zare, Dylan Stewart, Stephanie A Bohlman,

- and Ethan P White. A benchmark dataset for canopy crown detection and delineation in co-registered airborne rgb, lidar and hyperspectral imagery from the national ecological observation network. *PLoS computational biology*, 17(7):e1009180, 2021. 3
- [132] Ben G Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, and Ethan White. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11):1309, 2019. 2
- [133] Ben G Weinstein, Sergio Marconi, Stephanie A Bohlman, Alina Zare, Aditya Singh, Sarah J Graves, and Ethan P White. A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network. *Elife*, 10:e62922, 2021. 2
- [134] Ben G Weinstein, Sergio Marconi, Stephanie A Bohlman, Alina Zare, and Ethan P White. Cross-site learning in deep learning rgb tree crown detection. *Ecological Informatics*, 56:101061, 2020. 3
- [135] Roy WELCH. Spatial resolution requirements for urban studies. *International Journal of Remote Sensing*, 3(2):139–146, 1982. 3
- [136] Scott Workman, Richard Souvenir, and Nathan Jacobs. Wide-area image geolocalization with aerial reference imagery. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–9, 2015. Acceptance rate: 30.3%. 3
- [137] Q Xiao, SL Ustin, and EG McPherson. Using aviris data and multiple-masking techniques to map urban forest tree species. *International Journal of Remote Sensing*, 25(24):5637–5654, 2004. 2, 3
- [138] Qiuyan Yu, Wenjie Ji, Ruiliang Pu, Shawn Landry, Michael Acheampong, Jarlath O’Neil-Dunne, Zhibin Ren, and Shakhawat Hosen Tanim. A preliminary exploration of the cooling effect of tree shade in urban landscapes. *International Journal of Applied Earth Observation and Geoinformation*, 92:102161, 2020. 1
- [139] Caiyun Zhang and Fang Qiu. Mapping individual tree species in an urban forest using airborne lidar data and hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 78(10):1079–1087, 2012. 2, 3
- [140] Kongwen Zhang and Baoxin Hu. Individual urban tree species classification using very high spatial resolution airborne multi-spectral imagery using longitudinal profiles. *Remote Sensing*, 4(6):1741–1757, 2012. 2, 3
- [141] Carly D Ziter, Eric J Pedersen, Christopher J Kucharik, and Monica G Turner. Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proceedings of the National Academy of Sciences*, 116(15):7575–7580, 2019. 1
- [142] Xinhui Zou, Ming Cheng, Cheng Wang, Yan Xia, and Jonathan Li. Tree classification in complex forest point clouds based on deep learning. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2360–2364, 2017. 3