

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

Supplementary Material

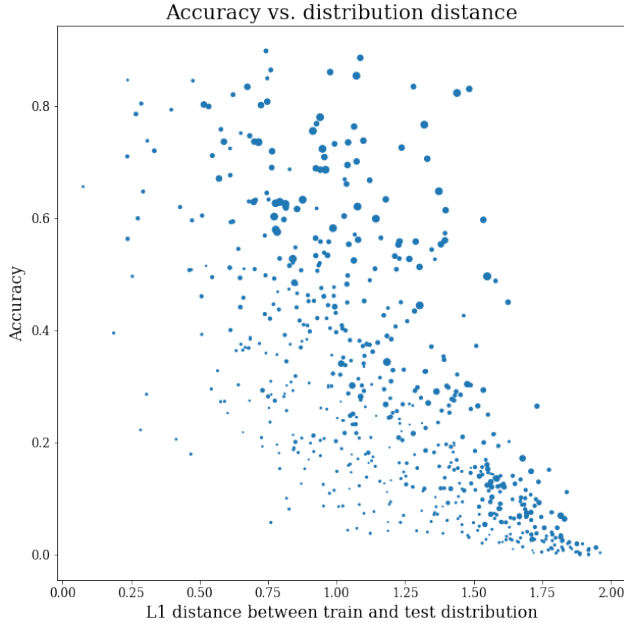


Figure 1. There is a strong correlation between distribution similarity and performance, notably models can achieve the same accuracy with significantly less training data if their distributions are similar.

A. Model training details

Each baseline model was trained with a batch size of 128 on 32 TPU cores on full-resolution images (input size 512x512 for aerial, 1152x768 for street level). Train time data augmentation included random horizontal flipping and minimal random cropping (at least 80% of the image maintained after the crop) for both data types. For simplicity, we used a linear warmup for one epoch, then trained for an additional 4 epochs at a 0.01 learning rate, then a final epoch at a 0.001 learning rate. We anticipate that training for additional epochs would improve all models, but that the relative performance would be maintained.

B. Further analysis

Accuracy vs. distribution distance. In Supplementary Figure 1 we show the pairwise accuracy vs distribution distance for all train/test pairs, with markersize denoting the number of training examples. We note the correlation between accuracy and distribution distance, as well as note

Dirichlet constant	AR	FAR	CAR	RAR
baseline	41.66	58.97	39.77	0.0
1	9.22	1.40	13.28	6.39
10	25.71	12.46	33.1	17.88
100	39.46	36.52	45.11	6.85
1K	43.13	49.49	46.09	0.58
10K	42.83	54.34	43.94	0.0
100K	41.98	60.48	40.04	0.02

Table 1. Logit-adjusted loss ablation study, training on Region W with Santa Monica holdout. Metrics are averaged across the test sets of all 8 cities in the region (see Table 2 in the main paper).

that the same accuracy level can be achieved between train and test pairs with less training data if the distributions are more similar (the markers tend to get larger as you move from left to right on any horizontal line).

Handling the data imbalance. We consider the logit-adjustment method for training on imbalanced data proposed in [1]. Eqn. 1 defines the logit adjusted softmax cross-entropy loss we used in our experiments:

$$l(y, f(x)) = -\log \frac{e^{f_y(x) + \log \pi_y}}{\sum_{y' \in [L]} e^{f_{y'}(x) + \log \pi_{y'}}}, \quad (1)$$

where $[L]$ is the set of all genera. We construct our logit adjustment term π based on the per-genus counts in the training set, and expand the logit-adjusted loss to include a Dirichlet smoothing term to regularize the highly long-tailed nature of our data.

$$\pi_y = (\text{count}(y) + c)^{-1} \text{ for } y \in [L] \quad (2)$$

where c is our Dirichlet smoothing constant. We experiment on Region W and find that while $c = 1000$ gives us a 2% boost in city-averaged AR, this is not a clear win, as it improves city-averaged AR-C (by 6.32%) but decreases city-averaged AR-F (by 9.48%). See Table 1 for ablation results on Region W, different parameters perform better for different rarity subsets and different cities (see Fig. 2 for additional detail).

C. Data Sources and Licenses

The public tree censuses for the 23 cities in our dataset are linked in Supplementary Table 2, along with the licensing information for each. We visualize the data per-city in

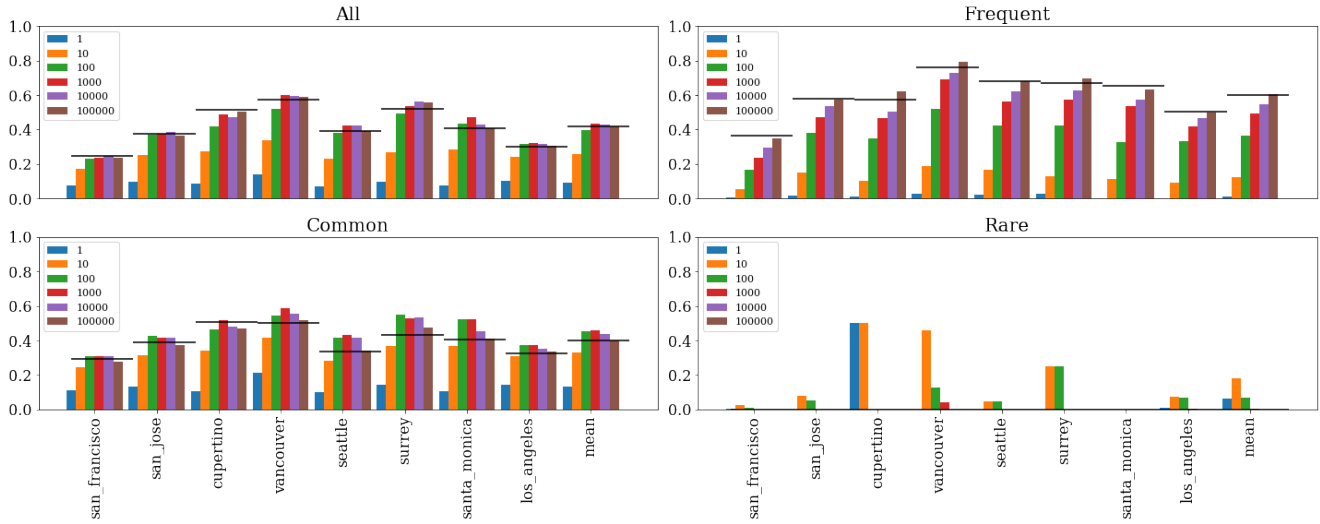


Figure 2. Loss balancing results per-city as well as averages across the Region W cities, for different values of our Dirichlet smoothing constant c . You can see the explicit tradeoff between frequent and common classes, with rare classes remaining a significant challenge. Baseline results are shown as a black line.

Figure 3, showing the coordinates of each tree as well as coloring the most common genera to show regional shifts in distribution.

References

- [1] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. *arXiv preprint arXiv:2007.07314*, 2020. 1

City	Data Source	License
Vancouver, BC	https://opendata.vancouver.ca/explore/dataset/street-trees/information/?disjunctive.species_name&disjunctive.common_name&disjunctive.height_range_id	https://opendata.vancouver.ca/pages/licence/
Surrey, BC	https://data.surrey.ca/dataset/park-specimen-trees	https://data.surrey.ca/pages/open-government-licence-surrey
Seattle, WA	https://www.seattle.gov/transportation/projects-and-programs/programs/trees-and-landscaping-program/seattle-tree-inventory-map	https://opendatacommons.org/licenses/pddl/summary/
San Francisco, CA	https://data.sfgov.org/City-Infrastructure/Street-Tree-List/tkzw-k3nq	https://opendatacommons.org/licenses/pddl/1-0/
San Jose, CA	https://gisdata-csj.opendata.arcgis.com/datasets/7db16e012fe8402db45074cd260c8f4e_510	https://gisdata-csj.opendata.arcgis.com/datasets/7db16e012fe8402db45074cd260c8f4e_510/explore
Cupertino, CA	https://gis-cupertino.opendata.arcgis.com/datasets/caa50a924b7d4b5ba8e8a4cbfd0d7f13	https://gis-cupertino.opendata.arcgis.com/datasets/Cupertino::trees-2/about
Santa Monica, CA	https://data.smgov.net/Public-Assets/Trees-Inventory/w8ue-6cnd	https://opendatacommons.org/licenses/by/1-0/
Los Angeles, CA	https://geohub.lacity.org/datasets/266c6255b1fc4ae8b8f100d8696e1fa4_0	Usage was approved by StreetsLA
Boulder City, CO	https://boulder.maps.arcgis.com/apps/opsdashboard/index.html#/328aac5a588840c99edee239672f7ca2	https://creativecommons.org/publicdomain/zero/1.0/
Denver, CO	https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-tree-inventory	https://creativecommons.org/licenses/by/3.0/
Sioux Falls, SD	https://opendata.arcgis.com/datasets/c880d62ae5fb4652b1f8e6cbca244107_10.csv	https://creativecommons.org/licenses/by/3.0/
Calgary, AB	https://data.smgov.net/Public-Assets/Trees-Inventory/w8ue-6cnd	https://data.calgary.ca/d/Open-Data-Terms/u45n-7awa
Edmonton, AB	https://data.edmonton.ca/Environmental-Services/Trees-Map/udbt-eiax	https://data.edmonton.ca/stories/s/City-of-Edmonton-Open-Data-Terms-of-Use/msh8-iff28/
Washington, DC	https://opendata.dc.gov/datasets/urban-forestry-street-trees/explore	https://creativecommons.org/licenses/by/4.0/
Charlottesville, VA	https://hub.arcgis.com/datasets/charlottesville::tree-inventory-point/explore?location=38.038850%2C-78.483500%2C13.89	https://creativecommons.org/licenses/by/4.0/
Pittsburgh, PA	https://data.wprdc.org/dataset/city-trees	https://creativecommons.org/licenses/by/4.0/
Montreal, QC	https://donnees.montreal.ca/ville-de-montreal/arbres	https://donnees.montreal.ca/licence-d-utilisation
New York, NY	https://data.cityofnewyork.us/Environment/2015-Street-Tree-Census-Tree-Data/pi5s-9p35	https://opendata.cityofnewyork.us/overview/
Buffalo, NY	https://data.buffalony.gov/Quality-of-Life/Tree-Inventory/n4ni-uuec/data	https://creativecommons.org/share-your-work/public-domain/cc0/
Kitchener, ON	https://data.waterloo.ca/datasets/KitchenerGIS::tree-inventory/about	https://data.waterloo.ca/datasets/KitchenerGIS::tree-inventory/about
Cambridge, ON	https://data.waterloo.ca/datasets/cityofcambridge::street-trees/explore	https://maps.cambridge.ca/images/opendata/Open%20data%20licence.pdf
Columbus, OH	https://opendata.columbus.gov/datasets/public-owned-trees/explore?location=39.974897%2C-82.996371%2C14.00&showTable=true	https://creativecommons.org/share-your-work/public-domain/cc0/
Bloomington, IN	https://data.bloomington.in.gov/dataset/public-tree-inventory	https://opendefinition.org/od/2.1/en/

Table 2. Data location and licensing information for the public tree censuses used to curate our dataset.

