DOI:10.1145/3762636

For years, conducting an urban tree inventory was a luxury only large, highly resourced cities could afford. A new AI-based approach not only brings this capability within reach to smaller cities but also requires significantly less time.

BY ADNAN FIROZE, AKSHAJ UPPALA, LINDSAY DARLING, RAYMOND A. YEH, BEDRICH BENES, BRADY HARDIMAN, SONGLIN FEI, AND DANIEL ALIAGA

Where Are the City Trees?

Monitoring Urban Trees across the U.S. Using Generative AI

TREES PLAY A critical role in urban environments today. Growth in U.S. cities[9] has occurred alongside a rise in extreme weather conditions,[8] driven by critical factors such as subpar air quality, excessive heat islands, and unmoderated rainwater overflow—all of which can be partially mitigated by urban trees.[31] Trees can also improve urban residents' physical and mental health and provide habitat for wildlife.[8]

Historically, only resource-rich U.S. cities have collected data about where their public trees are, usually through labor-intensive manual surveys or via coarse canopy-cover estimation. However, a significant portion of city trees are on private property, making them difficult to quantify with surveys, yet they contribute uniquely to species diversity and ecosystem service distribution.[11] Further, canopy-cover estimation cannot provide information about tree density, locations of trees across different land types, or changes in tree counts. Cities are under continual change, and the mean mortality rate of urban trees is twice that of rural trees.[30] Thus, frequent updating of tree analytics is critical for sustainable, habitable cities.

Method. Recent advances in computing—in particular, generative artificial intelligence (AI)—have enabled our multidisciplinary team, spanning computer science, engineering, and forestry, to develop a first-of-its-kind computational method that can individually locate and maintain an inventory of trees in at least 330 U.S. cities (Figure 1). Using satellite data, this approach can complete the inventory process in less than a day of automated computing. Individual trees are challenging to discern in satellite

» key insights

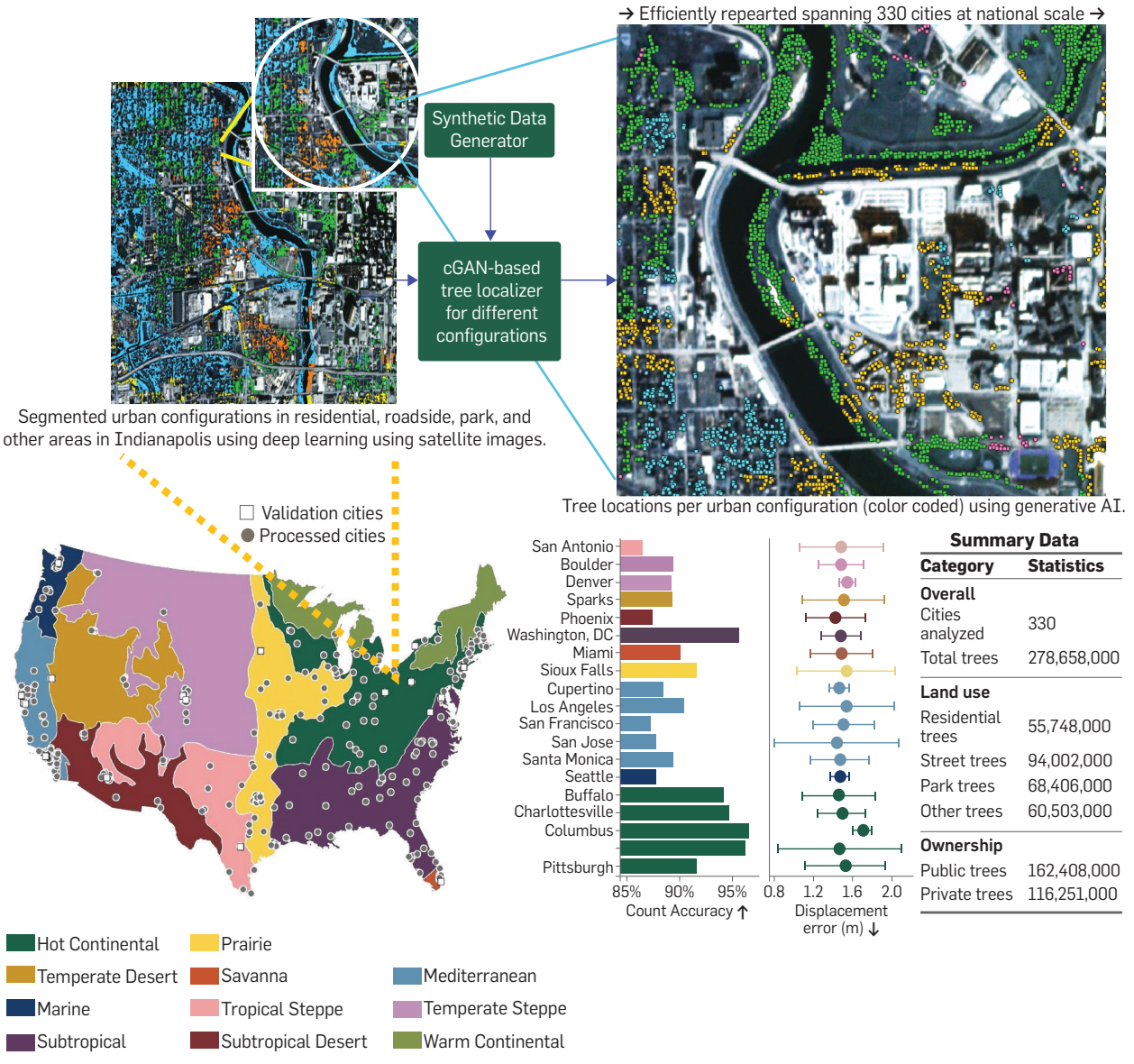
- **Urban trees are critically important to mitigate poor air quality, excessive heat islands, and unmoderated rainwater overflow, which collectively can lead to undesired extreme weather conditions.**
- **The use of generative AI enables a novel computational approach to localize individual trees in all cities, despite their mutual occlusion and overlap.**
- **A new monitoring approach facilitates updating a national-scale database of 278 million urban trees spanning 330 U.S. cities in less than a day of computing.**
- **Providing individual cities with the ability to frequently update their tree analytics is critical for planning sustainable, habitable cities.**

ILLUSTRATION BY JACEYTEC

images due to occlusion and resolution limitations, which in turn limits traditional segmentation-based approaches. Our approach leverages several key insights to enable a scalable generative AI solution. First, a frequent capture rate of satellite imagery (e.g., daily, monthly, etc.) provides spatiotemporal vegetation footprints, yielding richer information than single images. Our method includes a deep spatiotemporal vegetation cover classification using satellite images that classifies a city into tree, grass, and background, followed by a cluster-creation process and then individual tree localization using a set of *conditional generative adversarial networks (cGANs)*. Further, our method can be applied to current or archived satellite imagery, allowing for change detection and historical analysis.

Tree inventories: State of the art. Existing tree-inventory-monitoring strategies are expensive; as a result, under-resourced communities rarely have access to this information and therefore are less able to manage their urban trees.[4] In addition, due to historic, systematic bias and other socioeconomic factors, low-income residents tend to have less access to urban trees and the benefits they provide.[34] In recognition of the importance of trees, there are increasing efforts to expand urban tree canopy globally while making its distribution more equitable, including a $1.5 billion investment in urban and community forestry in the U.S.[35] Researchers also assert that achieving an equitable presence of trees in a city requires that all communities can afford to obtain tree inventories, and across all land types. There are a few products that use aerial imagery to extract tree-cover data under for-profit business models for select cities, but they

Figure 1. AI-based framework for automated satellite-based urban tree inventory and localization.



are unaffordable for many cities and do not operate at a national scale.[5,22] Affordable tree monitoring—namely, a computational method requiring only satellite imagery that can be purchased quickly for any city and at relatively low cost—is pressingly needed.

Generative AI for Tree Localization

For our method, we compute tree locations within all mainland U.S. cities with a population over 100,000, as reported in the 2020 census.[32] For each of these 330 cities, we acquire a monthly series of satellite snapshots spanning 12 months (using Planet Labs satellite imagery, which has global and daily capture since 2010[27]). This image set enables capturing spatial and temporal characteristics that are not possible using approaches based on single-snapshot segmentation and/or the traditional Normalized Difference Vegetation Index (NDVI) (Figure 2). In the first phase, we train a deep segmentation network based on the U-NET architecture[29] and use clustering to obtain a classification of four urban configuration types. In the second phase, we use a set of trained cGANs for each of the four configuration types. Unlike wild forests, urban forests are subject to city management rules. Thus, the cGAN training process obtains realistic synthetic data from a generator, which leverages a tiered approach of producing example data from urban procedural rules for different urban configurations. The generator is then trained and tuned using batch processing spread across 100 GPUs during a month of training. The full pipeline is illustrated in Figure 3.

Segmentation. We use a customized U-Net-based architecture[29] and clustering to spatially temporally segment satellite image tiles—each 256

Figure 2. Regional highlights using our urban tree monitoring model. Regional statistics and trends with highlighted decomposition of tree location types (residential, roadside, park, others, public, and private) and notable findings. The map shows all 330 of our processed cities with several call-out cities. We report "trees per acre" because cities are of different sizes and their administrative boundaries are irregular.

Western Region: Seattle			North Central Region: Indianapolis			Northeastern Region: Baltimore			Southern Region: Jacksonville (Highest public tree density)			Southern Region: Montgomery (Highest private tree density)		
Category	Trees (n)	Trees per acre	Category	Trees (n)	Trees per acre	Category	Trees (n)	Trees per acre	Category	Trees (n)	Trees per acre	Category	Trees (n)	Trees per acre
Trees by land use			**Trees by land use**			**Trees by land use**			**Trees by land use**			**Trees by land use**		
Residential	179K	44	Residential	1,179K	23	Residential	160K	10	Residential	65K	31	Residential	65K	31
Park	136K	88	Park	992K	34	Park	22K	22	Park	5,346K	88	Park	5,346K	88
Roadside	108K	34	Roadside	339K	34	Roadside	144K	21	Roadside	2,852K	45	Roadside	2,852K	45
Others	16K	87	Others	106K	28	Others	92K	20	Others	456K	72	Others	456K	72
Trees by ownership			**Trees by ownership**			**Trees by ownership**			**Trees by ownership**			**Trees by ownership**		
Public	244K	61	Public	1,331K	34	Public	166K	21	Public	8,198K	66	Public	4,099K	66
Private	195K	66	Private	1,285K	26	Private	251K	15	Private	521K	51	Private	260K	51
Total trees	439K	63	Total trees	2,616K	30	Total trees	417K	18	Total trees	8,719K	59	Total trees	8,719K	59

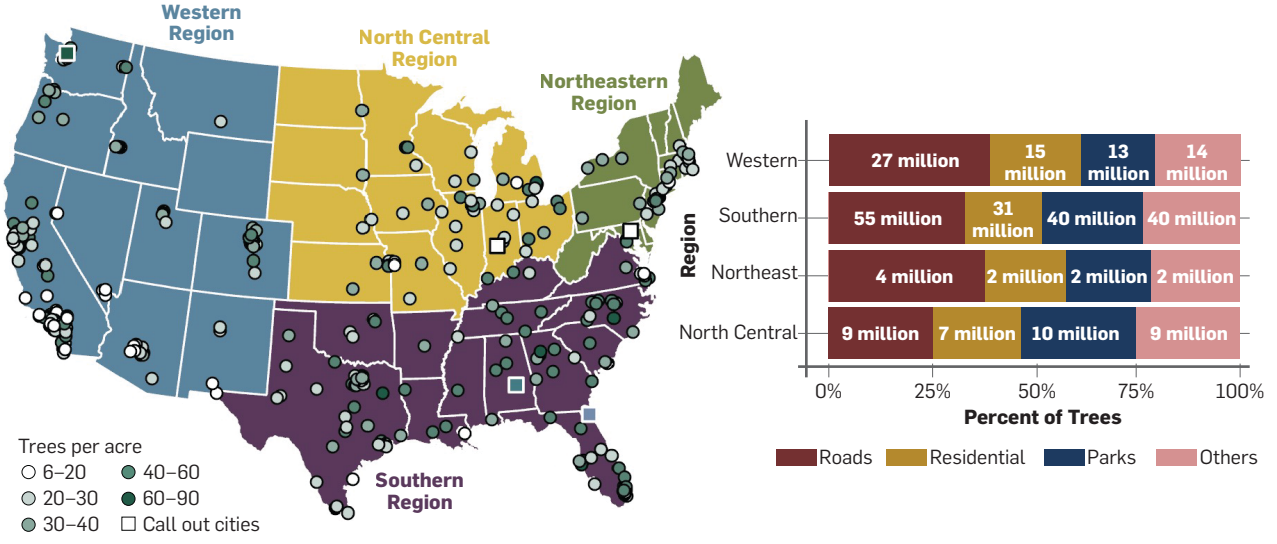
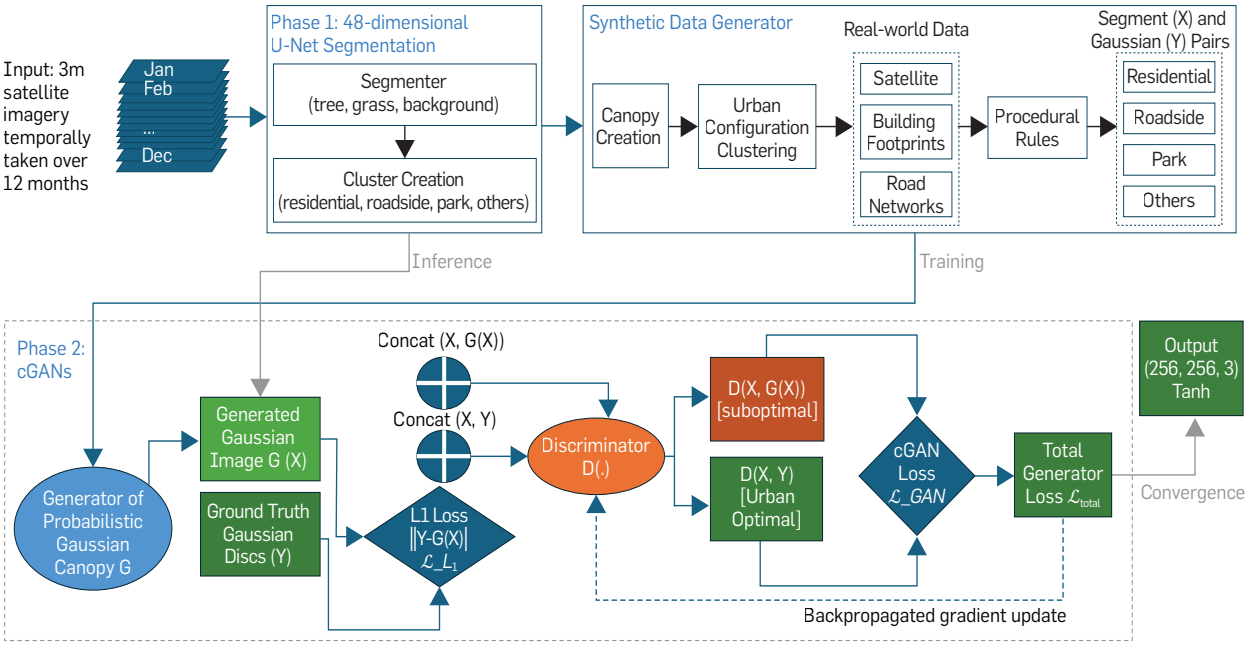


Figure 3. Pipeline. Phase 1 includes a 48-channel U-Net based segmentation engine and the synthetic data generator. Phase 2 consists of a set of customized and trained cGANs.



x 256 pixels, with 3m per pixel resolution—into regions of one of three classes: *trees*, *grass*, and *background*. We further classify the tree class into four configuration types: *residential*, *roadside*, *park*, and *others* (Figure 1, top left). Our experiments show that using 12 months of satellite image data significantly outperforms using a single snapshot across all cities (Figure S2 in supplement). This temporal fusion captures seasonal and structural changes, resulting in better generalization and accuracy in vegetation segmentation. We convert the input tiles of training data to 48 channels (four channels per monthly snapshot, resulting in 48D). The output is a segmentation of the same spatial size but with three channels corresponding to each of the region classes (see Figure S3 in the supplement for a pictorial example).

Table 1. Optimized parameters for universal (*U*) and cluster-specific rules (*V*). Four universal rules that are trivial and extracted from ANSI A300 standards parameterized below. Next, for each of the four configuration types (residential, roadside, park, and others), the optimized parameter values are presented. In the table, V_{res}, V_{str}, V_{prk}, and V_{oth} are procedural rules for the aforementioned four configuration types, respectively. Here d_{bldg} refers to average distance from buildings, d_{street} refers to average distance from streets/roads, w_{corner} refers to the ratio of trees along the corner of a vegetation segment, and w_{inside} refers to the ratio of tree on the inner area of a vegetation segment.

Universal rules (*U*)

U_1	**No overlap between trees**
U_2	**Min tree to tree distance is half of field of neighborhood (FON) [4] = 4m**
U_3	**Min tree to building distance (ANSI A300) [6]- heuristically extracted = 3.7m**
U_4	**Min tree to street distance (ANSI A300) [6]- heuristically extracted = 2.5m**

Cluster specific rules (*V*)

Parameter	V_{res}	V_{str}	V_{oth}	V_{prk}
d_{bldg}	2m	2m	3m	4m
d_{street}	1m	1m	1m	1m
w_{corner}	0.5	0.9	0.6	0.7
w_{inside}	0.5	0.1	0.4	0.3

After segmentation, we automatically label each tree class blob as one of the four configuration types. The types were determined by clustering a set of features describing the blobs, buildings, and roads within the training data. Using k-means clustering and the elbow method, we determined that four configuration types best represent the spatial variability within the tree class blobs. The four types—residential, roadside, park, and others—were named based on heuristics and visual inspection.

Conditional GANs. Our generative model uses a conditional generative adversarial network (cGAN) to map the tree class blobs within the tiles to individual tree locations. To facilitate cGAN training, we implement a *synthetic data generator* to produce training data for each of the four configuration types (Figure 3, top right). To generate realistic training data, we use a hierarchy of rules observed in cities (e.g., universal and configuration-type specific rules) (see Table

Table 2. Validation cities[1] with validation metrics.

City	Count GT (| GT_i |)	Detected (| P_j |)	Correct (match)	Recall	Commission Rate	Omission Rate	Precision
Columbus	5482	5669	5414	0.99	0.04	0.01	0.96
NYC	1472	1517	1428	0.97	0.06	0.03	0.94
Washington, DC	3845	3680	3596	0.94	0.02	0.06	0.98
Charlottesville	2120	2232	2005	0.95	0.10	0.05	0.90
Buffalo	5125	4829	4663	0.91	0.03	0.09	0.97
Sioux Falls	2865	3104	2787	0.97	0.10	0.03	0.90
Pittsburgh	3070	3327	2991	0.97	0.10	0.03	0.90
Los Angeles	2721	2981	2682	0.99	0.10	0.01	0.90
Miami	1427	1493	1391	0.97	0.07	0.03	0.93
Santa Monica	1557	1721	1503	0.97	0.13	0.03	0.87
Boulder	1281	1316	1202	0.94	0.09	0.06	0.91
Sparks	1475	1512	1398	0.95	0.08	0.05	0.92
Denver	3677	3873	3531	0.96	0.09	0.04	0.91
Cupertino	1094	1042	995	0.91	0.05	0.09	0.95
San Jose	852	942	834	0.98	0.11	0.02	0.89
Phoenix	965	1008	910	0.94	0.10	0.06	0.90
San Francisco	8398	8768	8209	0.98	0.06	0.02	0.94
Seattle	816	932	797	0.98	0.14	0.02	0.86
San Antonio	1363	1406	1296	0.95	0.08	0.05	0.92

1). We extract the rule parameters from satellite images and from urban structural data, such as building footprints and road networks (e.g., OpenStreetMap[26]). Using this information, we generate 100,000 sample pairs per configuration type. Each sample pair contains a tile with tree blobs and a corresponding tile with Gaussian discs representing tree locations (with calibration shown in the surface plots in Figure S7 in the supplement). The generator captures both the local and global context of the tree blobs, obtaining their shape, size, and spatial arrangement in a compressed latent representation based on the inferred urban rules for the various configurations (see Figures S5 and S6 in the supplement for an illustration). For cGAN inference, the configuration type of each blob of an input tile is used to condition the generation of the most likely tree locations (Figure 3, bottom; also Figure S4).

Loss function and objective. The cGAN model is designed to translate tiles with tree blobs (X) into spatially continuous probability maps (Y), where each tree is represented as a Gaussian disc. These soft activations are post-processed via peak detection to produce precise tree-point locations. Our model is trained to produce structured spatial patterns that reflect the presence and relative positions of individual trees.

The loss function used to train the model is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{GAN}} + \lambda \cdot \mathcal{L}_{L1}$$

where

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{(X,Y)}[\log D(X,Y)] + \mathbb{E}_{(X,Y)}[\log(1 - D(X, G(X)))]$$

is the adversarial loss facilitating the generation of Gaussian discs representing trees, and

$$\mathcal{L}_{L1} = \mathbb{E}_{(X,Y)}[\|Y - G(X)\|]$$

is the pixel-wise loss that enforces spatial alignment and Gaussian smoothness.

Hyperparameter discussion. The cGAN is configured with hyperparameters for translating tree blob inputs into structured probabilistic outputs. An important hyperparameter is the L1 loss weight (λ), which ensures that the outputs remain quantitatively accurate and consistent with expected tree distributions. In our experiments, we set $\lambda = 100$ (where below 20 gave low precision and above 200 led to overfitting).[18] Tile size is another parameter (Figure S8 in supplement), which was found to work best at 256 x 256 pixels (or 80 x 80 meters). All models are optimized using the Adam optimizer, with a learning rate and β_1 set to $2\text{x}10^{-4}$ and 0.5, respectively. To improve generalization, dropout is applied in the decoding path of the generator. These hyperparameters collectively ensure that the model produces spatially aligned, point-representative outputs suitable for post-processing-based discretization of tree locations.

Results and Analysis

We validated our model by using existing, extremely time consuming ground-truth datasets for 19 U.S. cities[1] and partial ground-truth statistics for 79 U.S. cities (sources in Table 2 and supplementary Datasets S2 and S3), and by ensuring cities encompass all ecoregions in the mainland U.S. The complete dataset release of 330 cities is in Dataset S1 (see supplement). On average, our method's count accuracy is 92.5% (with high statistical significance of R^2=0.998) and the mean tree localization error is 1.49m with a standard deviation of only 0.42m (Figure 1). Using a metric proposed by researchers to detect individual trees in high-resolution aerial images,[37] our single-tree localization has a high accuracy rate, with a recall of 0.95, a precision of 0.92, a commission rate of 0.08, and an omission rate of 0.04 for the 19 validation cities.[1] A tabular and city-specific presentation is shown in Table 2.

Our analysis further showed the robustness of our approach across different spatial resolutions of satellite data. As we degraded the images to lower resolutions, we consistently see relatively accurate localization performance (Figure 4a). We moved from our native 3 meters per pixel (mpp) resolution (PlanetScope) to 5mpp (Rapideye), 10mpp (Sentinel-2), and 30mpp (Landsat). The difference between our native performance of normalized accuracy of 0.91 and the coarsest imagery (Landsat) was 0.16—a much smaller difference compared to competing models, even at native resolution (Figure 4b).

To understand computational

Figure 4. Robustness and comparison of accuracy and efficiency. (a) Box plots showing quantitative performance of our model using satellite imagery of different spatial resolutions. (b) Comparisons in terms of normalized localization accuracy and efficiency against 10 SOTA models, including multiple diffusion-based and transformer-based models (shown in black circles, whereas ours is shown in a green circle). The call-out boxes highlight statistics of our model and the next-best models in terms of efficiency and accuracy. Qualitative comparisons are also available in the supplementary material.

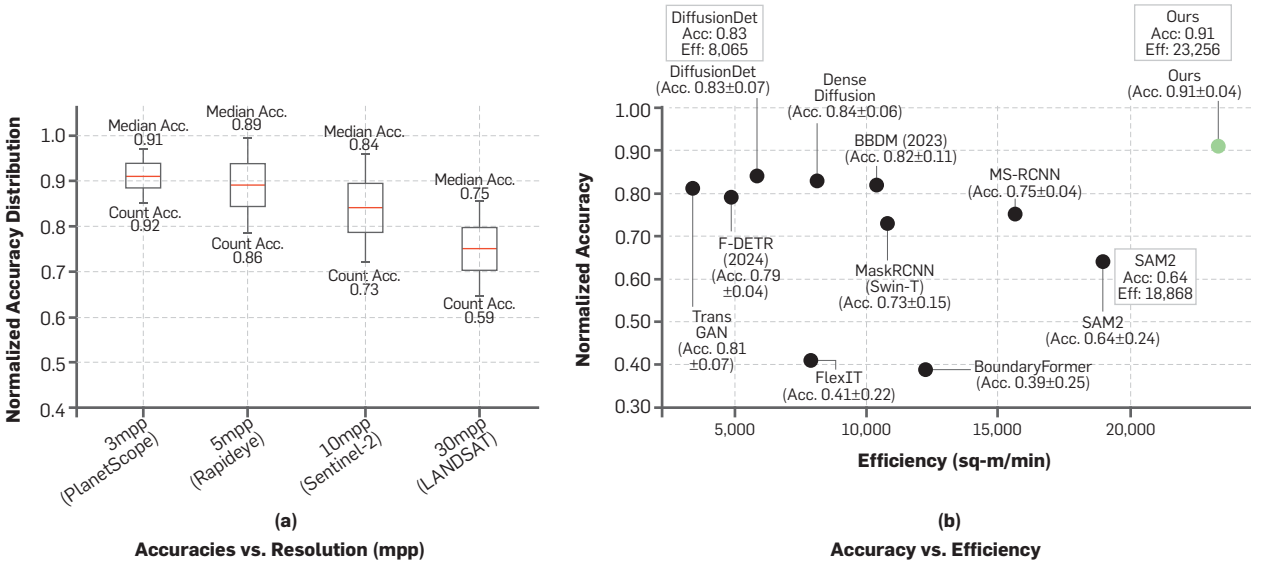
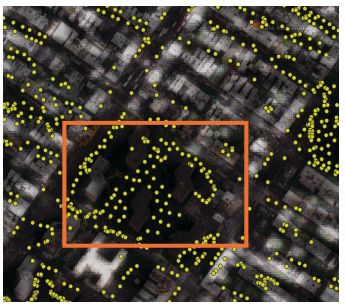
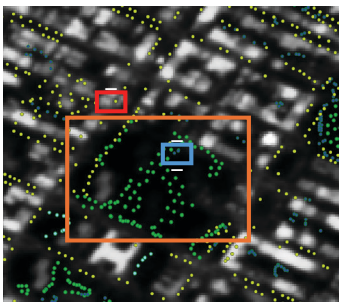


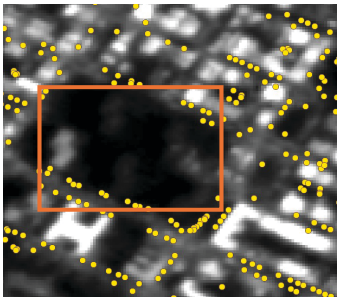
Figure 5. Qualitative comparisons with diffusion- and transformer-based approaches. Comparative snapshot of 6th Ave. Community Garden/Park in Midtown Manhattan, NY. Red box = case where a tree was localized over a building (less than 0.3%), blue box = park tree identified as residential, orange box = subjective superior performance of our model in the presence of occlusion and unresolvable areas where trees are significantly missing in transformer- and diffusion-based models. Note: Only our model offers such configuration-segmented localization. More examples can be found in the supplementary material (Figure S9).



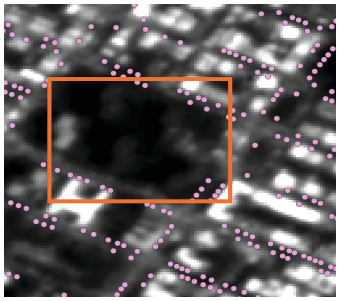
Ground Truth [Google UAV + Street View]



Ours
(green dots = park, yellow dots = roadside, blue dots = residential, teal dots = others)



Diffusion based — DenseDiffusion (ICCV, 2023)

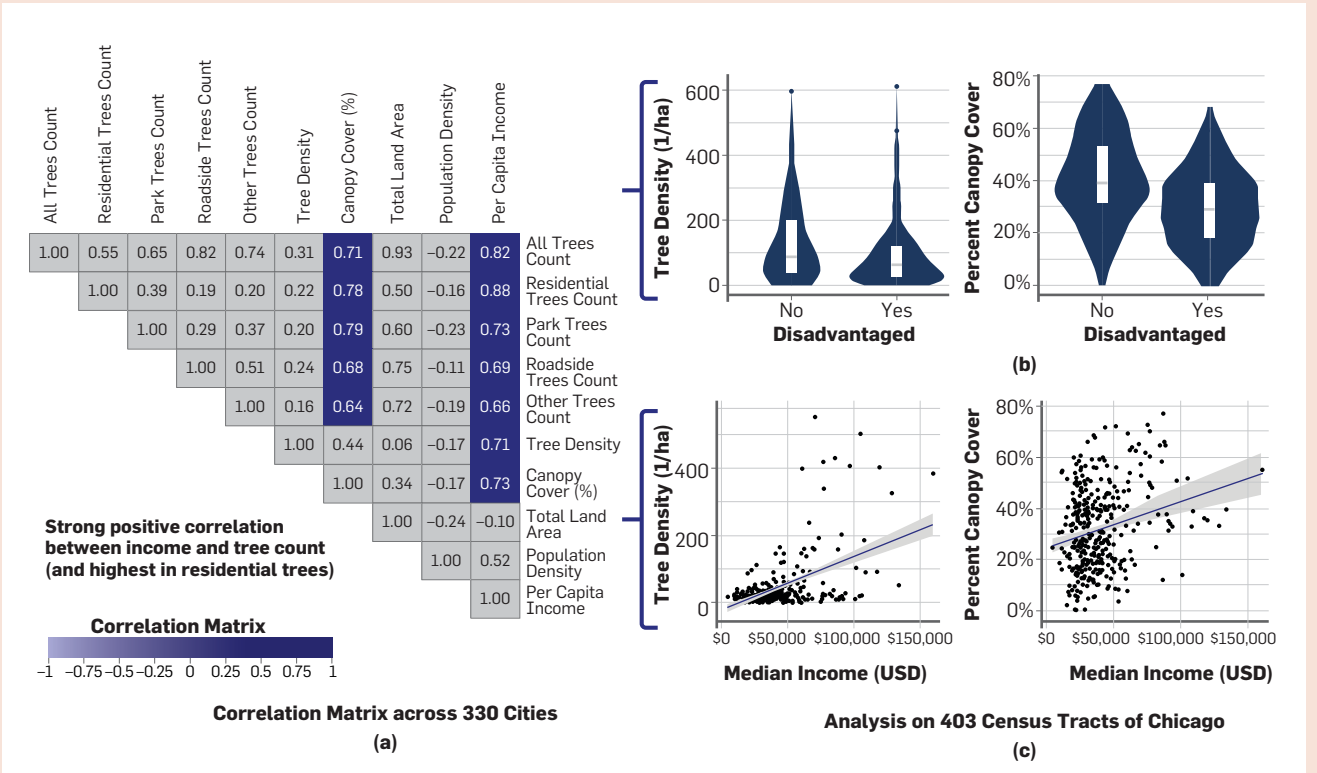


Transformer-based F-DETR (NeurIPS, 2024)

cost and efficiency across different models, we compared our model's efficiency against 10 state-of-the-art (SOTA) baselines that included both diffusion-based and transformer-based models. Using the same computational resources, we processed 23,256m^2 per minute, whereas the next-best model processed 18,868m^2 per minute and with approximately 30% less accuracy (see Figure 4b across the horizontal axis).

We then compared our model to SOTA baselines in terms of both normalized localization accuracy, count accuracy, and efficiency (see Figure 4b across the vertical axis). We selected the recent best performing transformer-, diffusion-, and convolution-based models, including F-DETR,[20] Swin-Transformer-based MaskRCNN,[21] SegmentAnything2,[28] DiffusionDet,[2] Mask Scoring RCNN,[10] BBDM,[19] BoundaryFormer,[16] DenseDiffusion,[13] FlexIT,[3] and TransGAN.[12] All models used four RTX 4090 GPUs for processing/prediction. Although there are limited failure cases (e.g., missing some trees that overlap with buildings), our model outperforms diffusion-based and transformer-based models (Figure 5). This analysis further highlights the

Figure 6. Disparity of tree presence in relationship to income level. We show the positive relationship between urban tree abundance and income level. (a) Correlation matrix among tree counts from our method, with population and economic data across 330 cities. (b) Violin plots showing current tree density and canopy cover disparity between disadvantaged and non-disadvantaged census tracts across 403 census tracts in Chicago. (c) Regression plot showing the same Chicago census tracts plotted to show tree density across varying income levels.



strength of our model when it comes to occlusion handling and challenging cases such as limited visibility.

Urban Tree Analytics

As a demonstration of the analytics enabled, and easily repeatable and updateable by our method, we provide the following example studies.

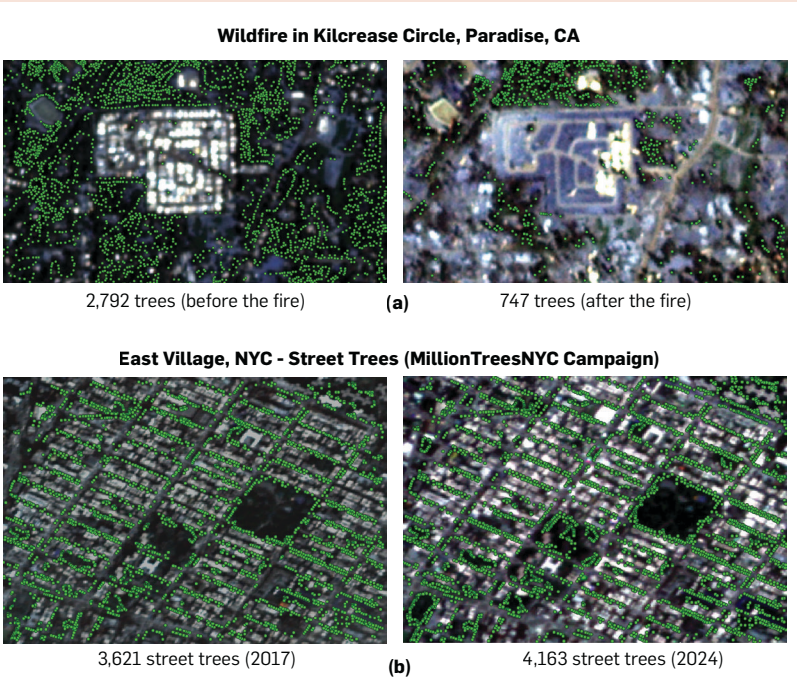
Nationwide urban tree perspective. Given the broad-scale yet still per-tree localization ability of our method, we provide for the first time nationwide urban tree statistics that can be updated regularly (Figure 2). We partition the 330 cities into four geographic regions: Northeast (NE), Northwest (NW), Southeast (SE), and Southwest (SW) using the geographic center of mainland U.S. as that set by the U.S. National Geodetic Survey (NGS) – 39°50'N 98°35'W. Figure 2 illustrates the counts in all regions.

Our method showed that the NE region had the highest mean density of 41 urban trees per acre, with a standard deviation of 11 trees per acre. Within this region, the highest density was 69 trees per acre in Ann Arbor, MI, and the lowest density was Newark, NJ, with 17 trees per acre. The second highest mean tree density region was the NW region, with a mean density of 37 trees per acre and a standard deviation of eight trees per acre. In this region, the maximum density was 63 trees per acre in Seattle, WA, whereas the minimum was 20 trees per acre in Provo, UT. The third-highest mean tree density region was the SE, with a mean density of 31 trees per acre and a standard deviation of 10 trees per acre. In this region, the city with highest tree density was Sandy Springs, GA, with 86 trees per acre. In contrast, the lowest density was seven trees per acre for Corpus Cristi, TX. Lastly, the region that exhibited the lowest mean tree density was the SW, with 28 trees per acre and a standard deviation of 11 trees per acre. In this region, the highest density was 53 trees per acre in Lakewood, CO, and the lowest was eight per acre in Inglewood, CA.

We further observed that currently the cities in the NE region have the lowest relative presence of residential trees (with a mean of 19%), whereas the NW has the largest residential relative presence of 25%. In another observation, NE cities were dominated by urban park trees, with a 25% relative presence in a region boasting more than 43 million trees. NW comes last in this category, with only 17% park trees.

Urban tree socioeconomic impact. Our approach enables correlating current tree localization with the Climate and Economic Justice Screening Tool (CEJST)[33] based on the current U.S. census (Figure 6). In recognition of the pivotal role trees play in urban settings and as a response to the disparity in tree presence between wealthier urban areas and disadvantaged areas, city governments have undertaken substantial investments to expand their tree cover, particularly focusing on marginalized neighborhoods. Our method enables quantifying and updating the trends for the first time, to

Figure 7. Before/after major events monitoring. (a) The 2018 "Camp Fire" near Paradise, CA, is one of the deadliest fires on record. Our method was automatically applied to satellite imagery before and after the fire, resulting in a tree loss estimate for this relatively small region of 1.2 square kilometers. (b) The MillionTreesNYC initiative started in 2007, ended in 2017, and was recently reinstated until 2030. We show the past and current performance of urban planting projects, such as the NYC campaign.



the best of our knowledge. Currently, we found strong positive Pearson correlation coefficients (CC) between per capita income and the total number of trees (CC=+0.82), number of residential trees (CC = +0.88), street trees (CC = +0.69), park trees (CC = +0.73), and other trees (CC = +0.66). A similar trend was seen with a strong CC of +0.73 between per capita income and canopy cover. Figure 6 illustrates a detailed validation conducted in Chicago. It reveals that within Chicago's 403 census tracts, disadvantaged regions exhibit substantially lower median tree density: approximately 42% fewer trees per hectare. The same trend is found for canopy cover, showing a difference of 11% in Chicago. Besides census tracts labeled as disadvantaged, a similar pattern emerges when assessing the census tracts in relation to income levels in Chicago: Higher-income tracts consistently exhibit greater tree presence. These trends can be regularly updated by tree data extracted by our approach.

Trees after disasters and planting initiatives. As a demonstration, we performed fine-grained monitoring tasks in response to urban disasters and tree-planting initiatives (Figure 7). For example, the "Camp Fire" in 2018 was the deadliest and most destructive wildfire in California history. With our approach, we localized and counted the trees before and after the fire by processing the corresponding satellite imagery. Figure 7a illustrates the micro-scale situation of tree analytics from a temporal perspective for a part of a city, namely the Kilcrease Circle apartment complex area in Paradise, CA, in the wake of the Camp Fire. To generalize even further, we conducted a before/after analysis for the full city of Vacaville, CA, which went ablaze during the 2020 LNU Lightning Complex fires. For the full city, we observed a loss of 80,916 trees across its 7,975 acres. The fire decreased the canopy cover in the city from 41% in 2019 to 35% in 2021. These estimates match up with small, ground-based efforts but provide the first-ever tool to assess tree loss at such a scale, and with minimal effort.

As another example, one of the most prominent urban tree-planting drives was the MillionTreesNYC campaign.[25] It first ran from 2007 until 2017, and was then revived and reactivated with another million-tree target for 2030. These planting campaigns focus heavily on planting street trees, as they are public and accessible. Our method is the first one to extract fine-grained analytics for such initiatives. It was reported that New York City (NYC) had approximately 5.2 million trees in 2005.[25] The same report stated that it took more than 2,000 volunteers a total of 30,000 hours to count the street trees in NYC. Later, in 2017, the USDA released a plot-level count of NYC trees at approximately seven million, based on an aggregation from 296 land plots.[24] We processed more than 500km^2 of NYC and found approximately 6,795,102 trees, of which 688,088 were street trees. A repeat analysis for 2024 yielded 6,985,902 trees, of which 720,839 were street trees (a 4.75% increase in street trees compared to 2017, but also a decline of 5.3% in residential trees). Our density agreed with a recent work by Niese et al.[23] Therefore, the trees were not only counted but also categorized into four categories, of which street trees are particularly interesting due to the MillionTreesNYC campaign. Overall, our method showed (and localized) a net increase of 2.3% trees in all of NYC between 2017 and 2024. Figure 7b shows an illustration of the localized street trees in the East Village, NYC, in 2017 and 2024. Our automatic approach took less than five hours (and thus far less than the 30,000 hours of human effort reported) to accomplish the localization and categorization of urban trees for the full city, showcasing the efficient application of our novel approach in extracting fine-grained analytics from tree-planting campaigns.

Moving Forward

As urbanization continues, understanding the quantity and distribution of urban trees is urgently needed for sustainable, effective management and policymaking. Using our AI-based method, for the first time we can update urban tree inventories at will on both public and private lands at 1.5m positional accuracy and 92.5% count accuracy, spanning at least 330 U.S. cities with a population more than 100,000. Instead of relying solely on high spatial resolution, our approach

uses spatiotemporal data and generative AI modeling to overcome occlusion, abutting trees, leaf-on/leaf-off conditions, and fall color assumptions. The system enables updating trends in the distribution of trees and identifying differences in tree density and canopy cover across ecoregions, providing a compelling tool for anomaly detection, among many other applications. Our approach is automatic and repeatable, implying that the monitoring for any city can be computed in a few hours and quickly repeated at will. The same city can be processed in an average of three to five hours (subject to the area) and yield a mean positional accuracy of 1.5m per tree. Nationwide processing can be done in under a day by using multiple computers.

We acknowledge the existence of high-resolution point cloud data such as LiDAR, but it is either airborne or backpack based and covers limited spatial areas and at limited temporal frequency—which is why, for a national scale, our work scales well. However, for places of particular interest (where fine-grained data collection is possible), we see positive potential to fuse our methodology of overhead localization with stem-based (understory) localization through multimodal approaches.

It is also notable that our method can serve well for the rapid development of urban planning policies and can simulate impacts using "what if" situations. This makes a difference, especially in smaller cities with limited budgets for tree analytics and planning. Since recent approaches create simulation-ready 3D reconstruction datasets from single images,[17] our work can serve as one of the key sources for photorealistic simulation of urban twins. Moving forward, and with the help of an NSF cyberinfrastructure grant, we are incorporating our methods into a Web portal for timely national-scale tree inventory estimation and what-if planning tools.

Looking forward, we anticipate that timely tree inventory and localization data can also be used to improve weather/climate modeling, by building upon our prior works[6,7] and by integrating with, for example, WRF Urban,[15,36] air quality (e.g., Kim et al.,[14] Li et al.[18]), and particulate matter (PM) absorption and dispersion models. We envision the novel and scalable nature of this work will be adopted across the globe toward the broad goal of greener, more sustainable cities.

Acknowledgments

This work was made possible by NSF Grants 1835739, 2106717, 2417510, 2412928, and 2411273, USDA NIFA Grants 1032672 and 2023-68012-38992, and Purdue University Institute for Digital Forestry.

References

1. Beery, S. et al. The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift. In *Proceedings of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (2022), 21262–21275.
2. Chen, S., Sun, P., Song, Y., and Luo, P. DiffusionDet: Diffusion model for object detection. In *Proceedings of IEEE/CVF Int. Conf. Comput. Vis.* (2023), 3423–3433.
3. Couairon, R. et al. FlexIT: Towards flexible semantic image translation. In *Proceedings of IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2022), 14048–14058.
4. Drescher, M. Urban heating and canopy cover need to be considered as matters of environmental justice. *Proceedings of the National Academy of Sciences* 810, 116 (2019), 26153–26154.
5. Earth Define. US Tree Map (2023).
6. Firoze, A. et al. Tree instance segmentation with temporal contour graph. In *Proceedings of 2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2023), 2193–2202.
7. Firoze, A., Benes, B., and Aliaga, D. Urban tree generator: Spatio-temporal and generative deep learning for urban tree localization and modeling. *The Visual Computer* 38 (2022), 3327–3339.
8. Fischer, E., Sippel, S., and Knutti, R. Increasing probability of record-shattering climate extremes. *Nature Climate Change* 11 (2021), 689–695.
9. Gomes, C. et al. Computational sustainability: Computing for a better world and a sustainable future. *Commun. ACM* 62, 9 (2019), 56–65.
10. Huang, Z., Huang, L., Huang, Y., and Liu, W. Mask scoring R-CNN. In *Proceedings of IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2019), 6409–6418.
11. Hutt-Taylor, K. and Ziter, C. Private trees contribute uniquely to urban forest diversity, structure and service-based traits. *Urban Forestry & Urban Greening* 78 (2022).
12. Jiang, Y. et al. TransGAN: Two pure transformers can make one strong GAN, and that can scale up. arXiv preprint arXiv:2102.07074 (2021).
13. Kim, S., Kim, J., and Kwak, S. Dense text-to-image generation with attention modulation. In *Proceedings of IEEE/CVF Int. Conf. Comput. Vis.* (2023), 7700–7710.
14. Kim, Y. et al. MUNICH v2.0: A street-network model coupled with SSH-aerosol (v1.2) for multi-pollutant modelling (2022).
15. Kusaka, H. et al. The integrated WRF/urban modelling system: Development, evaluation, and applications to urban environmental problems. *Intern. J. of Climatology* 31, 1 (2011), 273–288.
16. Lazarow, J., Xu, W., and Tu, Z. Instance segmentation with mask-supervised polygonal boundary transformers. In *Proceedings of IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2022), 1292–1302.
17. Lee, J.J. et al. Tree-D Fusion: Simulation-ready tree dataset from single images with diffusion priors. In *ECCV 2024*. Springer-Verlag (2024), 439–460.
18. Li, Z., Wang, Y., and Liu, Q. Fabric defect data augmentation using conditional generative adversarial networks. *Soft Computing* 28 (2024), 6109–6122.
19. Li, B., Xue, K., Liu, B., and Lai, Y. BBDM: Image-to-image translation with Brownian bridge diffusion models. In *Proceedings of IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2023), 1952–1961.
20. Liu, Y., Wang, Z., and Wang, Y. F-DETR: Fusion detection transformer with heterogeneous scale multi-branch structure. *Applied Soft Computing* 147, (2024), 110990.
21. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030* (2021).
22. Nearmap. Aerial map coverage (2023).
23. Niese, T. et al. Procedural urban forestry. *ACM Trans Graph.* 41 (2022), 2.
24. Nowak, D. et al. The urban forest of New York City. U.S. Department of Agriculture, Forest Service, Northern Research Station (2018).
25. NYC Parks and Recreation Project. Street tree census (2017).
26. OpenStreetMap contributors and OpenStreetMap; https://www.openstreetmap.org
27. Planet. Planet explorer (2023).
28. Ravi, N. et al. SAM 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024).
29. Ronneberger, O. et al. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention* (2015), 234–241.
30. Smith, I. et al. Live fast, die young: Accelerated growth, mortality, and turnover in street trees. *PLoS ONE* 14 (2019).
31. Szewczyk, R. et al. Habitat monitoring with sensor networks. *Commun. ACM* 47, 6 (2004), 34–40.
32. U.S. Census Bureau. City and town population totals: 2020–2022. (2023).
33. Walker, K. and Herman, M. Tidycensus: Load US census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames (2023); https://walker-data.com/tidycensus/
34. Willis, K. and Petrokofsky, G. The natural capital of city trees. *Science* 356 (2017), 374–376.
35. Wolf, K. et al. Urban trees and human health: A scoping review. *Intern. J. of Environmental Research and Public Health* 17 (2020).
36. Wong, M. et al. Evaluation of uWRF performance and modeling guidance based on WUDAPT and NUDAPT UCP datasets for Hong Kong. *Urban Climate* 28 (2019), 100460.
37. Zheng, Y. and Wu, G. Single shot multibox detector for urban plantation single tree detection and location with high-resolution remote sensing imagery. *Frontiers in Environmental Science* 9 (2021).

more online

To view the supplement for this article, please visit https://dl.acm.org/doi/10.1145/3762636 and click on Supplemental Material.

Adnan Firoze is a Ph.D. student in the Department of Computer Science at Purdue University, West Lafayette, IN, USA.

Akshaj Uppala is an M.S. student in the Department of Computer Science at Purdue University, West Lafayette, IN, USA.

Lindsay Darling is a Ph.D. student in the Department of Forestry and Natural Resources at Purdue University, West Lafayette, IN, USA.

Raymond A. Yeh is an assistant professor in the Department of Computer Science at Purdue University, West Lafayette, IN, USA.

Bedrich Benes is a professor in the Department of Computer Science at Purdue University, West Lafayette, IN, USA.

Brady Hardiman is an associate professor in the Department of Department of Forestry and Natural Resources at Purdue University, West Lafayette, IN, USA.

Songlin Fei is a professor in the Department of Department of Forestry and Natural Resources at Purdue University, West Lafayette, IN, USA.

Daniel Aliaga (aliaga@purdue.edu) is an associate professor in the Department of Computer Science at Purdue University, West Lafayette, IN, USA.

Watch the authors discuss this work in the exclusive *Communications* video. https://cacm.acm.org/videos/city-trees