Master’s Thesis of Landscape Architecture

Merging multiple sensing platforms and deep learning empowers individual tree mapping and tree species detection at the city scale

다중 센싱 플랫폼과 딥러닝을 활용한 도시 규모의 수목 맵핑 및 수종 탐지

February 2023

Graduate School of
Seoul National University

Department of Landscape Architecture and Rural Systems Engineering, Landscape Architecture Major

Kwon, Ryoungseob

Merging multiple sensing platforms and deep learning empowers individual tree mapping and tree species detection at the city scale

Under the Direction of Adviser, Prof. Ryu, Youngryel

Submitting a master's thesis of
Landscape Architecture

February 2023

Graduate School of
Seoul National University
Department of Landscape Architecture and Rural Systems Engineering, Landscape Architecture Major

Kwon, Ryoungseob

Confirming the master's thesis written by
Ryoungseob Kwon
February 2023

Chair	이동근	(Seal)
Vice Chair	류영렬	(Seal)
Examiner	강준석	(Seal)

Abstract

Precise estimation of the number of trees and individual tree location with species information all over the city forms solid foundation for enhancing ecosystem service. However, mapping individual trees at the city scale remains challenging due to heterogeneous patterns of urban tree distribution. Here, we present a novel framework for merging multiple sensing platforms with leveraging various deep neural networks to produce a fine-grained urban tree map. We performed mapping trees and detecting species by relying only on RGB images taken by multiple sensing platforms such as airborne, citizens and vehicles, which fueled six deep learning models. We divided the entire process into three steps, since each platform has its own strengths. First, we produced individual tree location maps by converting the central points of the bounding boxes into actual coordinates from airborne imagery. Since many trees were obscured by the shadows of the buildings, we applied Generative Adversarial Network (GAN) to delineate hidden trees from the airborne images. Second, we selected tree bark photos collected by citizen for species mapping in urban parks and forests. Species information of all tree bark photos were automatically classified after 'non-tree' parts of images were segmented. Third, we classified species of roadside trees by using a camera mounted on a car to augment our species mapping framework with street-level tree data. We estimated the distance from a car to street trees from the number of lanes detected from the images. Finally, we assessed our results by comparing it with Light Detection and Ranging (LiDAR), GPS and field data. We estimated over 1.2 million trees existed in the city of 121.04 km^2 and generated more accurate individual tree positions, outperforming the conventional field survey methods. Among them, we detected the species of more than 63,000 trees. The most frequently detected species was *Prunus yedoensis* (21.43 %) followed by *Ginkgo biloba* (19.44 %), *Zelkova serrata* (18.68 %), *Pinus densiflora* (7.55 %) and *Metasequoia glyptostroboides* (5.97 %). Comprehensive experimental results demonstrate that tree bark photos and street-level imagery taken by citizens and vehicles are conducive to delivering accurate and quantitative information on the distribution of urban tree species.

Keyword: urban trees, tree mapping, tree species detection, city-scale, multiple sensing platforms, deep learning

Student Number: 2021-25367

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Urban trees, which play pivotal roles in increasing urban biodiversity (Rhodes et al., 2011; Roy et al., 2012; Savard et al., 2000), storing carbon (Bae & Ryu, 2015; Burns et al., 1997; Chen et al., 2020; Edmondson et al., 2012), moderating local climate (Gaffin et al., 2012; Li et al., 2019; Schwaab et al., 2021) and improving citizens' mental well-being (Bratman et al., 2019; Kardan et al., 2017; Moreira et al., 2021), are a critical asset to cities. Hence, there are crucial needs to estimate the number and locations of urban trees (Escobedo et al., 2011; Song, 2005). To overcome periodic labor-intensive and time-consuming field surveys (Alonzo et al., 2014), recent studies increasingly applied airborne- or satellite-derived datasets as they provide abundant information on a wide range of environments (Fassnacht et al., 2016; Jensen et al., 2012). One of the most largely used techniques is Airborne Laser Scanning (ALS) because of its effectiveness at extracting structural traits of individual trees (Kaartinen et al., 2012). However, preprocessing point cloud data or merging other data was prerequisite for vegetation segmentation due to an influence of urban infrastructures (Alonzo et al., 2014; Budei et al., 2018; Dalponte et al., 2014; Liu et al., 2017). Deep learning and high spatial-resolution imagery for tree mapping has also been explored in various ecosystems (Aval et al., 2018; Brandt et al., 2020; Schiefer et al., 2020; Sun et al., 2022; Tong et al., 2021; Yao et al., 2021; C. Zhang et al., 2020; Zheng et al., 2020). Yet, none of them covered hidden trees under the shadow of buildings, which led to potential discrepancies in the city. Indeed, there is an urgent need to develop an innovative alternative for tree mapping at a city-scale.

Moreover, species information is a key element included in urban tree inventories. In most cases, multi- or hyperspectral images have been regarded as a data sources for efficient tree species mapping since physiological traits can be extracted (Ferreira et al., 2019; Hemmerling et al., 2021; Key et al., 2001; Pu & Landry, 2012; Tigges et al., 2013). These promising approaches, however, have critical shortcomings. For instance, spectral signals are easily influenced by artificial structures (Alonzo et al., 2014) and similarity among species hampers precise classification in heterogeneously planted areas (Pu & Landry, 2012). Thus, most studies predicted the approximate distribution of each species in the narrow area of the forests, which implies cities were still challenging sites (La Rosa et al., 2021; Mäyrä et al., 2021; B. Zhang et al., 2020). Another rapidly growing platform is street-level imagery, which enables fine-grained vegetation analysis along urban street networks (Li et al., 2017, 2018; Li et al., 2015; Richards & Edwards, 2017; Seiferling et al., 2017; Xia et al., 2021). Yet, vehicle-borne data was insufficient to locate target trees since street images provide few cues for measuring the exact locations of trees (Choi et al., 2022; Laumer et al., 2020; Lumnitz et al., 2021). In this respect, species detection at a ground-level while leveraging a conventional remotely-sensed data can be a solution for producing city-scale tree maps which provide accurate geolocations and species information of individual trees.

Here, we present a novel method merging multiple sensing platforms (i.e., airborne, citizen and vehicle) for producing city-scale tree maps. We automated the whole process of handling big data by using deep learning techniques, which has powerful feature learning capacity (Krizhevsky et al., 2012). Our framework can be divided into 3 parts; individual tree mapping from the air, tree species detection through ground-level data and evaluation of the results. First, we localized urban trees whose crown width is greater than 2 m that was clearly detected from airborne imagery. As many trees were obscured by the shadows of buildings, we applied Generative Adversarial Network (GAN) to delineate hidden trees. From these images, we estimated the geolocations of individual trees by simply extracting center points of deep learning results, which were commonly called 'bounding boxes'. Second, geotagged species information collected by citizens and a vehicle was mapped onto a previously generated tree maps. In detail, we classified species from tree bark photos which were transmitted by citizens. We also classified species of street trees from street−level imagery. Citizens and vehicles are promising sensors for mapping urban tree species as they could provide abundant data covering urban parks (Dickinson et al., 2012; Theobald et al., 2015) and streets (Branson et al., 2018; Wegner et al., 2016) respectively. Finally, we extensively assessed our results by comparing with other data sources. Owing to inaccuracy of Global Positioning System (GPS) under dense canopies, individual tree maps were evaluated by GPS− and ALS−based tree location measurements in streets and forests respectively. We validated tree species detection results through field surveys. We highlight that exploiting deep learning models and merging data from multiple platforms can facilitate the process of generating detailed urban tree maps.

Chapter 2. Methodology

2.1. Data collection

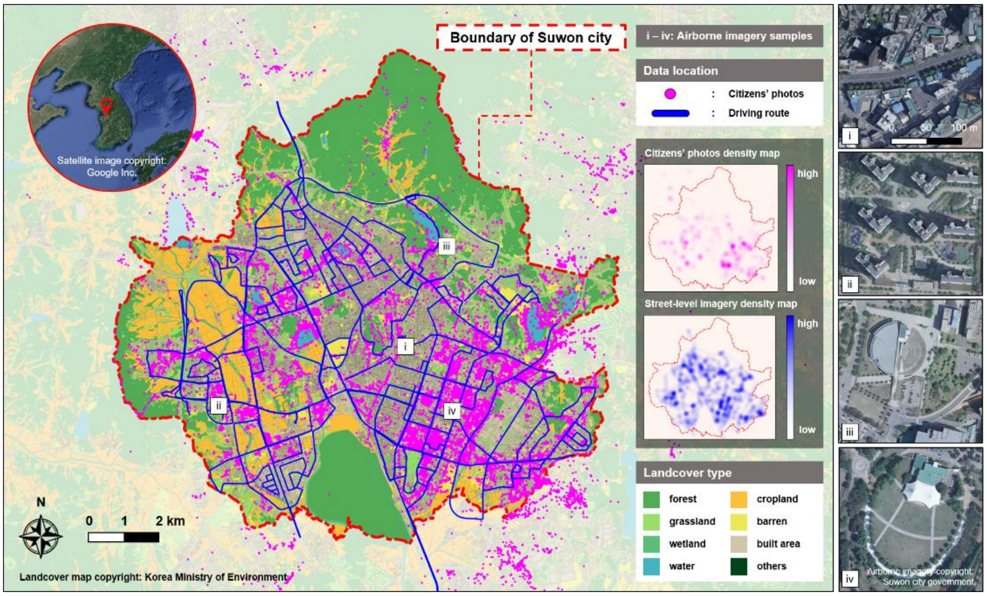


Fig. 1. Location of Suwon city, Republic of Korea. 1.18 million of citizens live in an area of 121.04 km^2 . Land cover types were presented based on the data provided by the Korea Ministry of Environment. Most of the study site consists of built areas, with forests in the north and croplands in the west. The airborne image examples, recorded GPS locations of citizens' photos and driving routes for collecting street-level images are also presented.

ID	Scientific name	Citizen sensed	Vehicle sensed
1	*Acer buergerianum* (Ab)	O	O
2	*Acer palmatum* (Ap)	O	O
3	*Aesculus turbinata* (At)	O	O
4	*Betula platyphylla* (Bp)	O	
5	*Cercidiphyllum japonicum* (Cj)	O	
6	*Chionanthus retusus* (Cr)		O
7	*Cornus officinalis* (Co)	O	
8	*Ginkgo biloba* (Gb)	O	O
9	*Lagerstroemia indica* (Li)	O	
10	*Liriodendron tulipifera* (Lt)	O	
11	*Metasequoia glyptostroboides* (Mg)	O	O
12	*Pinus densiflora* (Pd)	O	O
13	*Pinus koraiensis* (Pk)	O	
14	*Pinus strobus* (Ps)	O	
15	*Platanus occidentalis* (Po)	O	O
16	*Prunus yedoensis* (Py)	O	O
17	*Pseudocydonia sinensis* (Pi)	O	
18	*Quercus acutissima* (Qa)	O	
19	*Quercus palustris* (Qp)	O	O
20	*Styphnolobium japonicum* (Sj)	O	O
21	*Zelkova serrata* (Zs)	O	O
22	*OTHERS* (Ot)	O	O
Total		21	13

Table. 1. Target tree species. For tree species detection, dominant species in urban parks and streets were determined respectively based on field survey. Accordingly, a total of 20 tree species were selected for parks and urban forests, and a total of 12 tree species were selected for streets.

Our target site of this study is Suwon city, Republic of Korea. About 1.18 million of citizens live in an area of 121.04 km² . Landcover types were presented based on the data provided by the Korea Ministry of Environment in Fig. 1. Most of the study site consists of built areas, with forests in the north and croplands in the west. The airborne image examples, recorded GPS locations of citizens' photos and driving routes for collecting street-level images are also presented. For tree species detection, dominant species in urban parks and streets were determined respectively based on field survey. Accordingly, a total of 20 tree species were selected for parks and urban forests, and a total of 12 tree species were selected for streets.

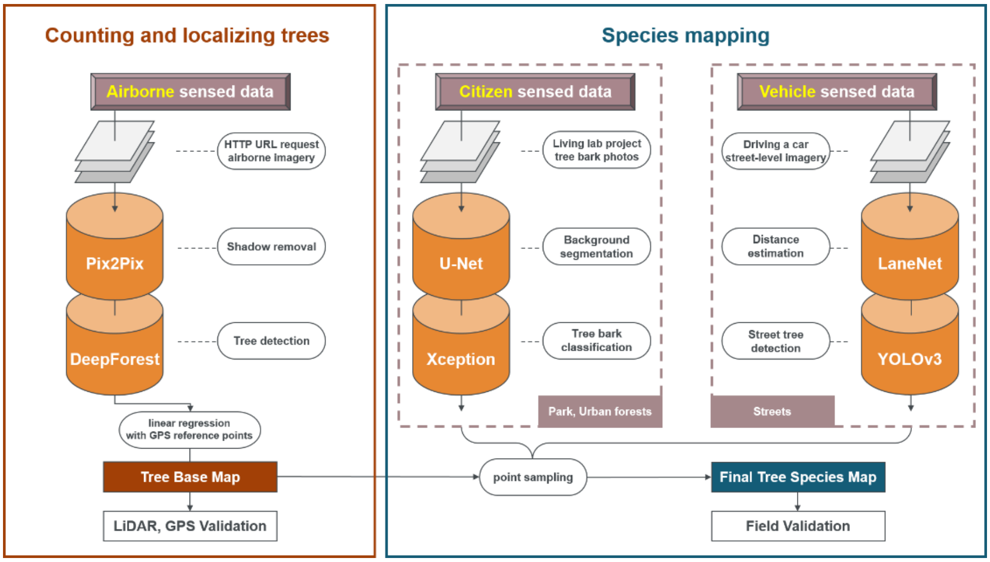


Fig. 2. Overall flowchart of methodology

For tree counting and mapping, airborne images taken in May 2020 under clear sky conditions were automatically downloaded through an HTTP URL request from the website of Suwon city[①]. These orthorectified images were acquired in three channels (i.e., Red, Green, Blue), with a spatial resolution of approximately 0.25 m. The website also provides the airborne imagery which was taken over the past several years. Therefore, it was possible to clearly identify the tree canopies hidden in the shadows of buildings at a certain year by observing the imagery of other years.

For tree species detection in the parks and urban forests, we carried out a living lab project with the help of a CADA application[②] which collects tree bark photos with a GPS coordinates taken by citizens. Tree bark was chosen as the part of the tree that citizens can easily photograph from the closest distance. We obtained a total of 45,300 location data for tree bark images in our study area from May 2021 to March 2022, and our own 3,034 images were added as they were not used for training the deep learning models. Besides, we promoted the project with reward money of 50 (KRW) for each image.

For tree species detection in the streets, we drove over 333.75 km of the roads with a RGB camera (FLIR Blackfly) facing forward and a GPS device (APX15 for evaluation board and Trimble AV18 for antenna) mounted on the top of the car. For stable data collection, we tried to drive the vehicle at about 40 km/h and the camera was set to take a photo at a rate of 10 Hz. We obtained 705,897 frames of street view images in our study area from August 2021 to October 2021, and among them, 617,616 frames that contained GPS information correctly were selected for analysis.

① https://www.suwon.go.kr:38083/citizenIntranetMain.do
② © 2021. Paprika Inc., Seoul, Republic of Korea.

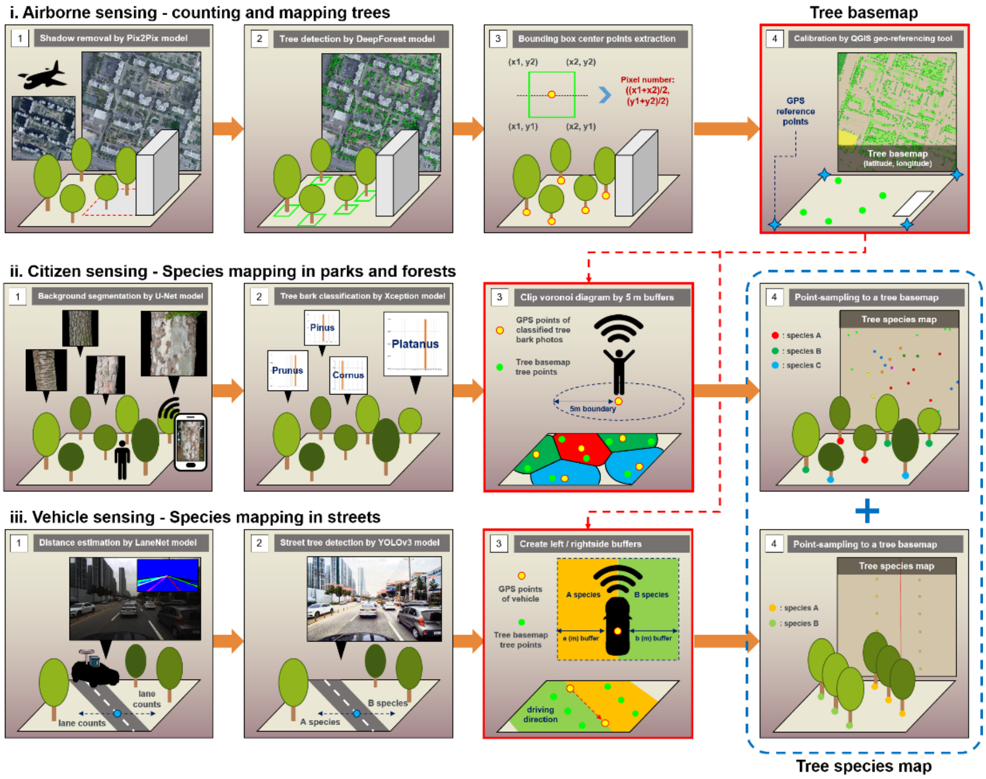


Fig. 3. Detailed diagrams of each part of methods

2.2. Deep learning overall

We did not change the architectures of the original networks. Whether to include the transfer learning process was determined by a subjective judgment (i.e., we started training the DeepForest model from pre−trained weights, as there was a model trained with other airborne images). By monitoring the change in loss values and figuring out the epoch where the training loss decreases but the validation loss increases, we determine the number of training epochs before the models begin to overfit. The implementation of 6 deep learning models adopts Python 3, TensorFlow (Abadi et al., 2016) and PyTorch (Paszke et al., 2019) according to frameworks implemented in original studies.

Objective	Airborne sensing		Citizen sensing		Vehicle sensing	
	Shadow removal	Counting and localizing trees	Image background segmentation	Tree bark classification	Distance estimation	Street tree detection
Network	Pix2Pix	DeepForest	U-net	Xception	LaneNet	YOLOv3
Backbone network	generator: U-Net discriminator: PatchGAN	object detector: RetinaNet classification: Resnet-50	U-Net	Inception	E-Net	DarkNet-53
Loss function	Binary Cross-Entropy loss	Focal loss	Binary Cross-Entropy loss	Categorical Cross-Entropy loss	Focal loss	bounding box: Generalized IoU classification, objectness: BCE loss
Transfer learning	X	O (NEON)	X	O (ImageNet)	O (TuSimple)	O (ImageNet)
Train/Val/Test	10,000 / 2000 / visual inspection (images)	3680 / 1229 / visual inspection (bounding boxes)	1200 / 400 / 400 (images)	6877 / 2290 / 1900 (images)	3780 / 1260 / visual inspection (images)	7133 / 2377 / 390 (images)
Input image size (w×h×c)	512×256×3	2560×2560×3	512×512×1	512×512×3	1280×720×3 (raw)	1280×800×3 (processed)
Batch size	4	1	2	8	64	16
Augmentation	Random [flip, shift, rotate, shear, brightness, contrast]	Random flip	Random flip	Random [flip, shift, scale, rotate, brightness, contrast, dropout, CHALE, blur]	Random [flip, brightness, contrast, saturation, hue]	Mosaic augmentation, Random [flip, translate, scale, hue]
Maximum learning rate	0.0002	0.001	0.001	0.0001	0.0001	0.001
Optimizer	Adam	SGD	Adam	Adam	SGD	Adam

Epochs	150	50	50	300 (Early stopped at 31)	300 (Early stopped at 58)	500
Final results	generator loss = 0.054, real image discriminator loss = 0.010, fake image discriminator loss = 0.002	train regression loss = 0.027, train classification loss = 0.086	train loss = 0.036	train loss = 0.008	train loss = 0.112	train bounding box loss = 0.034, train classification loss = 0.013, train objectness loss = 0.032
		validation regression loss = 0.227, validation classification loss = 0.302	validation loss = 0.092	validation loss = 0.003	validation loss = 0.214	validation bounding box loss = 0.031, validation classification loss = 0.011, validation objectness loss = 0.026
Framework	PyTorch	TensorFlow	PyTorch	TensorFlow	PyTorch	PyTorch
GPU	Single-GPU / NVIDIA GeForce RTX 3070 / CUDA Version: 11.3			Multi-GPU / NVIDIA A100 PCIE 40GB / CUDA Version: 11.4		

Table. 2. Training settings and performance evaluation of 6 deep neural networks.

2.3. Tree counting and mapping

To remove the shadows cast by the buildings in airborne imagery, we applied a conditional adversarial network designed for an image-to-image translation task (Isola et al., 2017). It was because the website did not provide original values of red, green and blue channels of images, which led to the unavailability of conventional image processing approaches. This generative model, called Pix2Pix, consists of a discriminator and a generator which learn to compete with each other. We prepared a set of train datasets with manually removed shadows and colored trees by referring to airborne imagery from past years and different seasons provided by the website. In these contexts, we expected the trained model to not only remove the shadows but also generate hidden trees with vivid colors.

To carry out tree counting and localizing automatically from the images with the shadow removed, we revised the DeepForest python package prebuilt for RGB-based tree crown delineation (Weinstein et al., 2020). This deep learning model performs tree detection thereby drawing bounding boxes on the detected objects. We fine-tuned the model with our hand labeled trees and conducted tree detection to extract the center pixels of the bounding boxes. To convert them into actual coordinates, we collected 120 GPS reference points around Suwon city and computed the linear regression between ground truth and pixel locations.

2.4. Tree species detection

ID	Scientific name	number of photos
1	*Acer buergerianum*	580
2	*Acer palmatum*	644
3	*Aesculus turbinata*	644
4	*Betula platyphylla*	662
5	*Cercidiphyllum japonicum*	360
7	*Cornus officinalis*	444
8	*Ginkgo biloba*	632
9	*Lagerstroemia indica*	384
10	*Liriodendron tulipifera*	632
11	*Metasequoia glyptostroboides*	640
12	*Pinus densiflora*	640
13	*Pinus koraiensis*	704
14	*Pinus strobus*	452
15	*Platanus occidentalis*	674
16	*Prunus yedoensis*	636
17	*Pseudocydonia sinensis*	594
18	*Quercus acutissima*	324
19	*Quercus palustris*	351
20	*Styphnolobium japonicum*	602
21	*Zelkova serrata*	638
Total		**11,237**

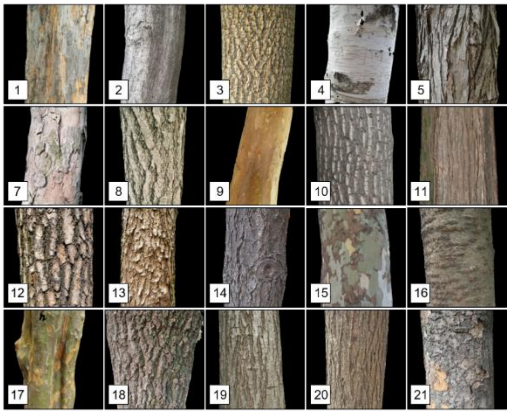
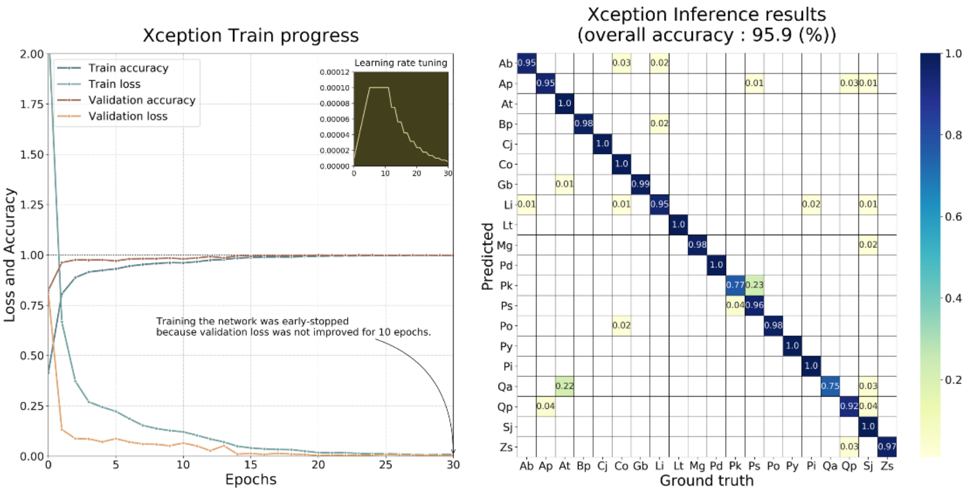


Table. 3. Target species and the number of collected photos to train the tree bark classification model and sample images



Ab : *Acer buergerianum* **Ap** : *Acer palmatum* **At** : *Aesculus turbinata* **Bp** : *Betula platyphylla*
Cj : *Cercidiphyllum japonicum* **Co** : *Cornus officinalis* **Gb** : *Ginkgo biloba* **Li** : *Lagerstroemia indica*
Lt : *Liriodendron tulipifera* **Mg** : *Metasequoia glyptostroboides* **Pd** : *Pinus densiflora* **Pk** : *Pinus koraiensis*
Ps : *Pinus strobus* **Po** : *Platanus occidentalis* **Py** : *Prunus yedoensis* **Pi** : *Pseudocydonia sinensis*
Qa : *Quercus acutissima* **Qp** : *Quercus palustris* **Sj** : *Styphnolobium japonicum* **Zs** : *Zelkova serrata*

Fig. 4. Recorded accuracy metrics for training the Xception network. Training process was early-stopped at 31st epoch as validation loss was not improved for 10 epochs. Learning rate was tuned while training the model. Confusion matrix of target species for tree bark classification. Performance of the trained network was evaluated through the proportion of correctly classified photos. Empty cells represent there is no prediction by the network.

For tree species detection from citizens' photos, we selected solely relying on bark due to its several advantages (Carpentier et al., 2018; Remeš & Haindl, 2019). First, tree bark images could be sustainably collected regardless of seasonal changes. Second, to take photography of bark, citizens have to get closer to trees thereby the recorded locations of photos can be used for useful references. Third, texture and color of bark are clearly different among tree species. However, several studies pointed out that intra-class differences and inter-class similarities between tree bark could be an obstruction to recognize species (Misra et al., 2020; Zhao et al., 2020). Therefore, we prepared various ages of tree bark datasets completely separated from citizens' photos.

To classify tree species from bark imagery, we applied Xception network (Chollet, 2017) after segmented background images using U-Net (Ronneberger et al., 2015). Segmentation of image background was necessary, since we took multiple bark pictures from a single tree which may lead the image classification model to learn the similar context of surroundings, not only bark texture itself. Annotations for training the U-Net model include two classes: tree or not tree. After that, these segmented images served as inputs of the Xception model for identifying tree species. The Xception model was chosen for its unique architecture which successfully reduced the number of parameters and the amount of computation so that overfitting and vanishing gradient problems could be improved. We monitored the training process of the Xception model with the Categorical Cross-Entropy loss function (Fig. 4). Training of the Xception network was early-stopped at 31st epoch as validation loss was not improved for 10 epochs. Categorical Cross Entropy is one of the most commonly used loss function for multi-class classification tasks, which is defined as:

$$\text{Categorical Cross Entropy} = -\sum_{i=1}^{n} y_i \log p_i$$

where n is the number of target classes, p_i is the probability of class i and y_i is 1 if the inference is correct, otherwise 0. To evaluate the performance of our classification model, we created a test dataset which contains 1,900 tree bark photos of 20 classes with the background removed (Fig. 4).

To estimate the tree species distribution, all collected citizens' data were mapped based on the deep learning inference results and recorded coordinates. To minimize the uncertainties of the quality of citizens' data, we filtered the photos over several steps. For instance, human resources manually filtered out photos that were difficult to be used such as 'taken from far away' or 'shaky'. We also implemented the Xception network to automatically exclude the images as 'others' class if they achieved lower than 0.4 of confidence score. Moreover, as the recorded location of citizen's data did not represent the actual tree points, we used a voronoi diagram (Aurenhammer, 1991) and point-sampling tool built in Geographic Information System (GIS) software to transmit tree species information to nearby tree points. Assuming that the average distance error of smartphone GPS is around 5 m, we fixed the radius of the voronoi diagram as 5 m and used euclidean distance.

ID	Scientific name	number of bounding boxes
1	*Acer buergerianum*	1,856
2	*Acer palmatum*	2,005
3	*Aesculus turbinata*	817
6	*Chionanthus retusus*	2,367
8	*Ginkgo biloba*	14,372
11	*Metasequoia glyptostroboides*	11,410
12	*Pinus densiflora*	2,882
15	*Platanus occidentalis*	3,559
16	*Prunus yedoensis*	7,510
19	*Quercus palustris*	1,148
20	*Styphnolobium japonicum*	1,291
21	*Zelkova serrata*	9,615
22	*OTHERS*	1,251
Total		**60,083**

Table. 4. Target species and the number of bounding boxes drawn to train the street tree detection model and sample images

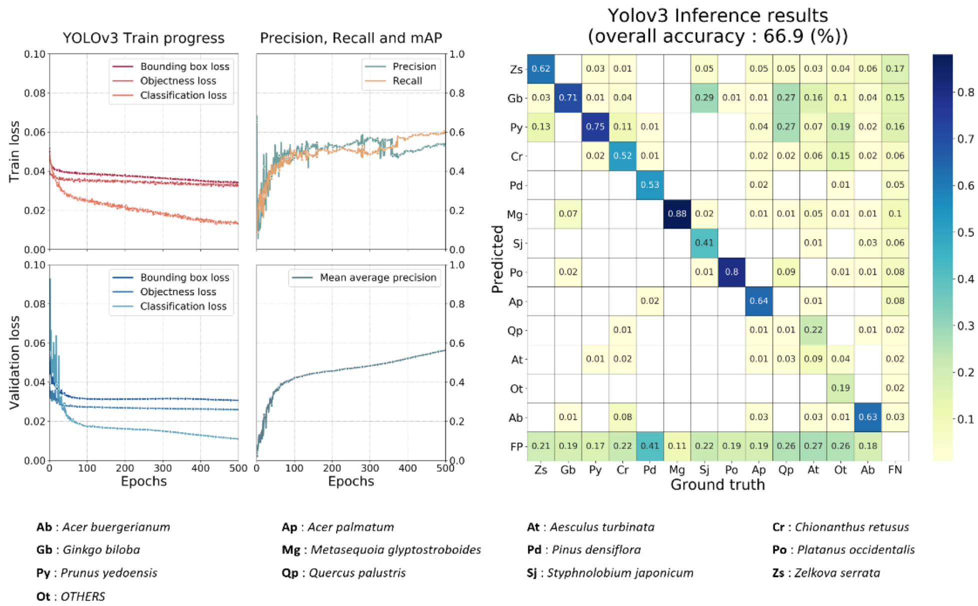


Fig. 5. Recorded accuracy metrics for training the YOLOv3 network. Each insets of the figure shows the decrease of the loss and the increase of the mean average precision (mAP), precision and recall. b, Confusion matrix of target species for street tree detection. Performance of the trained network was evaluated through the proportion of correctly detected street trees. Empty cells represent there is no prediction by the network.

For tree species detection from street-level imagery, we applied the YOLOv3 network (Redmon & Farhadi, 2018) to extract features on the pre-processed photos. Annotating the species for raw images of low brightness and contrast is problematic. Therefore, it is essential to pre-process the raw data by adopting image processing techniques, such as histogram equalization or contrast-limited adaptive histogram equalization. In contrast to tree bark photos, preparing an equal number of training data for each species poses a significant challenge since a few species comprise most of the street trees in the study site. After that, these annotated images served as train datasets of the YOLOv3 model for detecting street tree species.

We evaluated the performance of the YOLOv3 model by monitoring mAP, precision and recall, which are commonly used evaluation metrics for object detection frameworks (Table. 2). We measured mAP over multiple average precisions which is defined as the area under the precision−recall curve. Precision and recall are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}\ , \qquad Recall = \frac{TP}{TP + FN}$$

where TP, FP, and FN are the number of true positive, false positive, and false negative results, respectively. To evaluate the performance of our object detection model, we prepared a test dataset consisting of 390 street level images of 13 classes (Fig. 5).

We needed the distance references from the GPS points of the photos to the street trees which is similar to applying a voronoi diagram in a citizen sensing approach. Since the distance should be variable because the width of the road continuously changes, we used the LaneNet network (Neven et al., 2018) for counting the number of lanes from the images so that we could estimate the distance from a vehicle to street trees. We applied DBSCAN clustering (Ester et al., 1996) to LaneNet binary segmentation results due to its ability to discover clusters of arbitrary shape. After we counted the number of lanes from the images, we set the width between two lanes as 4 m and estimated the distances from a vehicle to trees on the left and right sides.

To estimate the tree species distribution in the streets, we used the deep learning inference results of all street−level imagery and converted them into text files that contain tree species and distance information on the left and right sides with coordinates recorded. However, there might exist errors of inference results such as misclassified tree species or miscounted number of lanes. We assumed that most street tree species were usually consistent in a single block. Therefore, we applied a majority voting approach to filter out the outliers and used the output as an input of a single−side buffer and point−sampling tool built in GIS software to transmit tree species information to nearby tree points.

2.5. Evaluation

For more reliable study results, point cloud data were scanned over the city by airborne laser scanning (ALS) and GPS coordinates were collected. The canopy height model derived by ALS and lidR R package (Roussel et al., 2020) served as the ground truth for the number and location of trees in urban parks and forests due to poor accuracy of GPS measurements under the dense canopy layer (Holopainen et al., 2013). Conversely, GPS data were used for the ground truth of the street trees, considering the feasibility and high accuracy of GPS devices under the sparse tree stands area. We selected 50 sites (30 x 30 m) in urban parks and forests and 50 sites (30 x 30 m) along streets, respectively, and compared tree counts results. Additionally, we validated produced tree geolocations by analyzing the closest distance between our results and reference tree positions which were based on ALS and GPS data (Branson et al., 2018). In Table. 5 and Fig. 18, we present the detailed information of LiDAR devices and image samples of processed point cloud data. We also conducted field survey to analyze the accuracy of the species map created by citizen- and vehicle-sensed datasets. We investigated about 10 % (1,572 trees) of trees mapped with citizens' data, and about 10 % (4,713 trees) of trees derived from street-level imagery. Additionally, we counted mis-classified species of trees within the buffer boundary as 'over-mapped' to analyze the feasibility of the point-sampling methodology using the buffer (see inset of Fig. 16).

Chapter 3. Results

3.1. Evaluation of deep learning performance

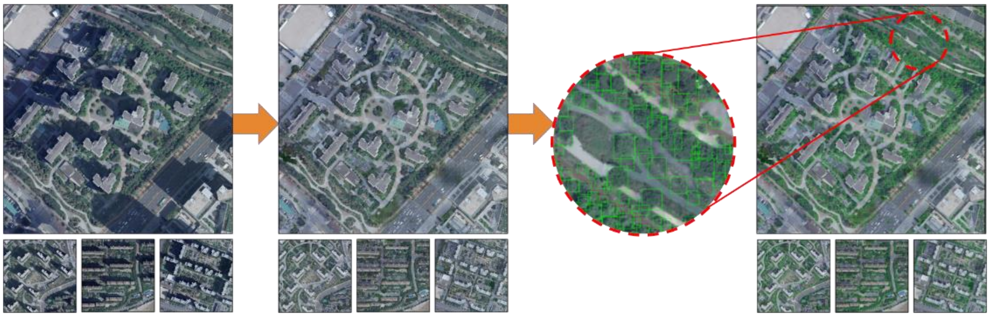


Fig. 6. Examples of deep learning results for shadow removal and tree detection (Green color of bounding boxes in the third image indicate tree detection from the airborne images)

Fig. 7. Examples of image background segmentation and tree bark classification results (Red horizontal lines in the plots indicate the threshold of confidence score to filter out low quality photos)

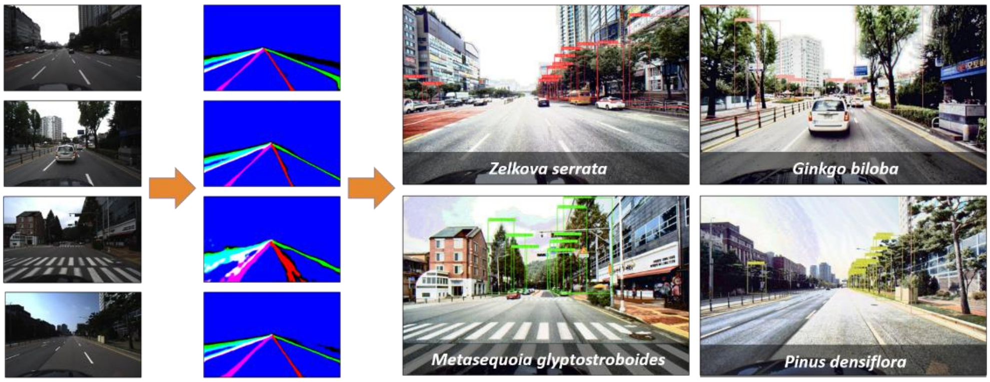


Fig. 8. Examples of deep learning results of lane segmentation and street tree detection (Various situations of roadside images were selected to show the performance of deep neural networks)

Automation of tree mapping and species detection at the city-scale were enabled, since all 6 deep neural networks showed considerable inference results after training process. For image shadow removal in complex urban situations, fake images synthesized by the Pix2Pix model showed hidden trees were generated realistic with colorization, although the values of the RGB channels were not provided. After fine-tuning with 50 epochs, the DeepForest model showed good performance for individual tree detection in the study site. Green color of bounding boxes in the third image indicate tree detection from the airborne images (Fig. 6). For tree bark classification, Fig. 7 presented that if the target tree was taken from far away, the performance of the U-Net model decreased. Then, the classification score became smaller accordingly and the photo would be filtered out automatically. Red horizontal lines in the plots indicate the threshold of confidence score to filter out low quality tree bark photos. Fig. 8 represented the consistent performance of the LaneNet model in various situations of street images. Our trained YOLOv3 network detected species with high accuracy for street trees with unique traits. Various situations of street images were selected to show the performance of deep neural networks.

3.2. Tree counting and mapping

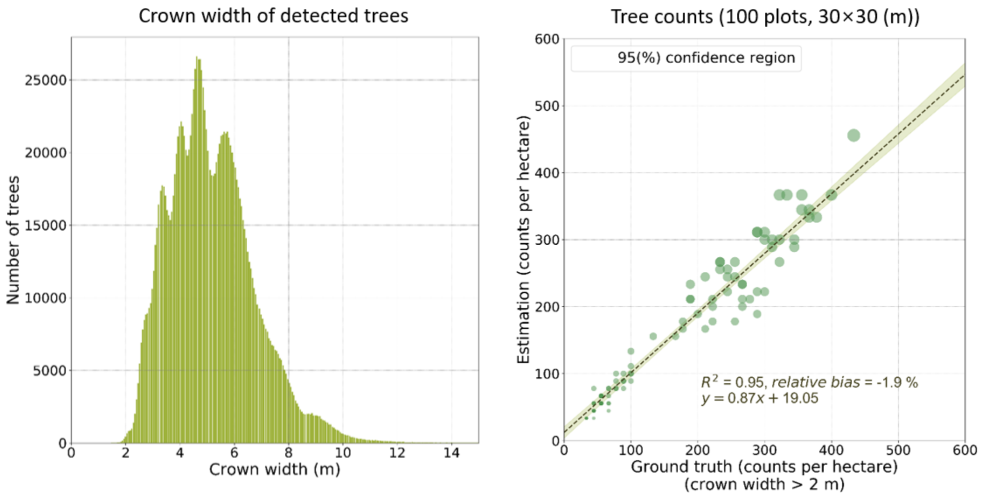


Fig. 9. The number of trees per crown width and the evaluation of tree counts. Crown width was defined as the mean value of the width and height of bounding box. Most detected trees showed around 4 to 6 m of crown width. Additionally, our estimation of the number of trees and the ground truth presented high correlation.

The final tree base map demonstrated that a total of about 1.29 million trees existed in the study area. In this study, we targeted trees whose crown width is greater than 2 m that was clearly detected from airborne imagery. Crown width of most detected trees were distributed around 5 m. Among them, the number of trees for each land cover type is the highest in the forest (57.4%) which was followed by grassland (21.0%), built area (14.6 %), cropland (4.0 %), barren (1.8 %). In Fig. 9, a comprehensive evaluation showed the comparison between our tree counting results and other data sources which were derived from Airborne Laser Scanning (ALS) and field survey with GPS devices (see Evaluation section in Methods). Notably, there appeared a strong linear relationship between our deep learning-based results and ground truth among 100 sites (30 x 30 m) in the study areas (R2 = 0.95, see inset in Fig. 9).

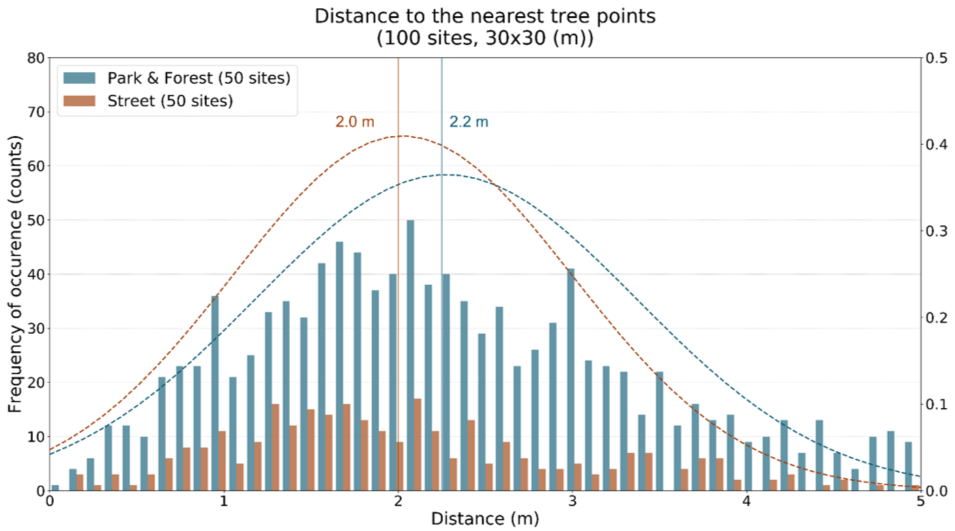


Fig. 10. Distance errors of the expected geolocations of urban trees. We selected 50 sites for urban parks and 50 sites for streets. The distributions of closest distance between our results and reference tree positions which were based on ALS and GPS data are presented.

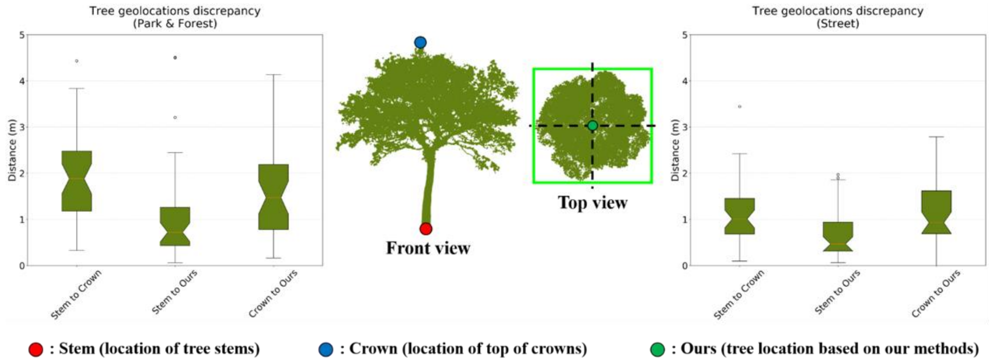


Fig. 11. Comparison of tree positions among our results, stem locations and top of crowns. Discrepancy of tree geolocations was calculated. Definition of tree location based on three methods is presented in the center part of the figure.

Our results provided reliable individual tree positions as the mean distance discrepancy between ground truth and predicted tree geo-locations were around 2.0 m. Assuming that the distance from predicted tree geo-locations to the closest ground truth larger than 5 m was measured from two different trees, we filtered out those occurrences. Fig. 10 indicated that the range of location error is slightly higher for trees in park and forest areas than the error for trees in streets. Additionally, assuming the lowest part of the tree stem as ground truth, we compared tree geolocation accuracy between ours and ALS-based results from point cloud data collected by Terrestrial Laser Scanning (TLS, Fig. 11). As a result, tree positions created by central points of bounding boxes were much closer to ground truth compared to the top of crowns. This indicates our methods performs better than ALS-based conventional approach which needs to be pre-processing steps and site-specific parameters to locate individual trees. Detailed information about the outcomes of individual tree mapping was described in Fig. 13. Also, visualization samples of TLS data and tree positions are presented in Fig. 12.

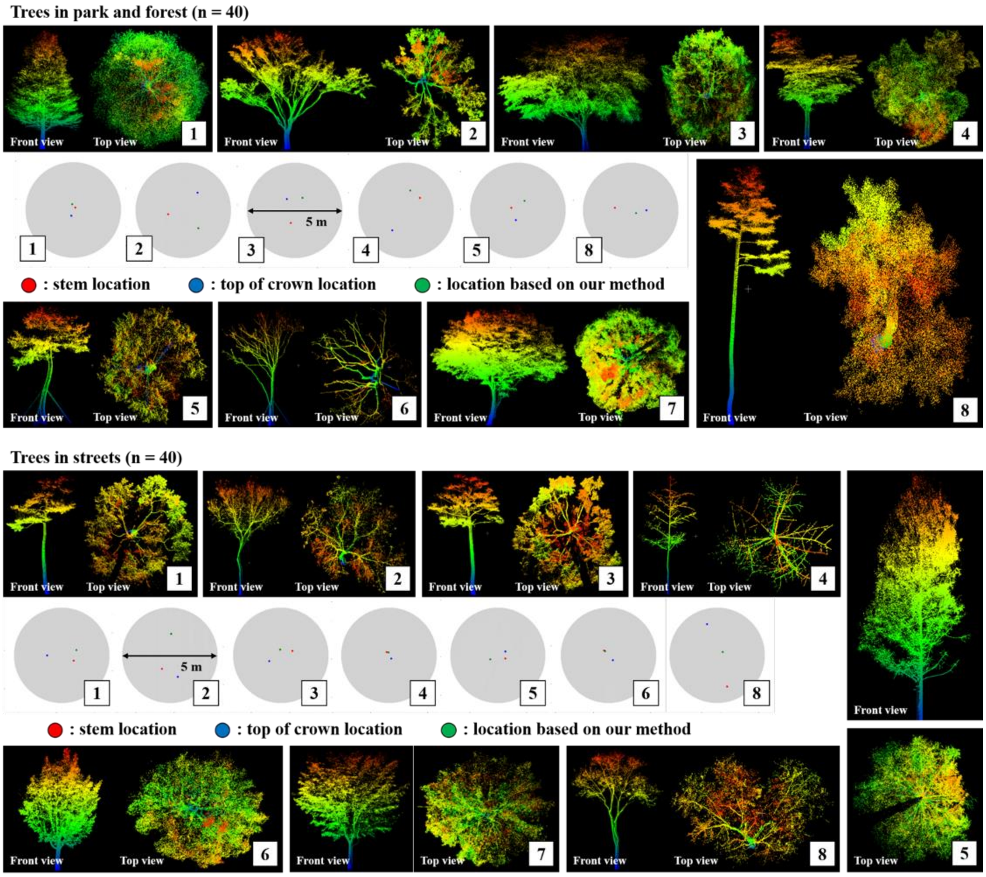


Fig. 12. Visualization samples of TLS point cloud data and tree positions based on each method (ours, stem and crown-based approaches).

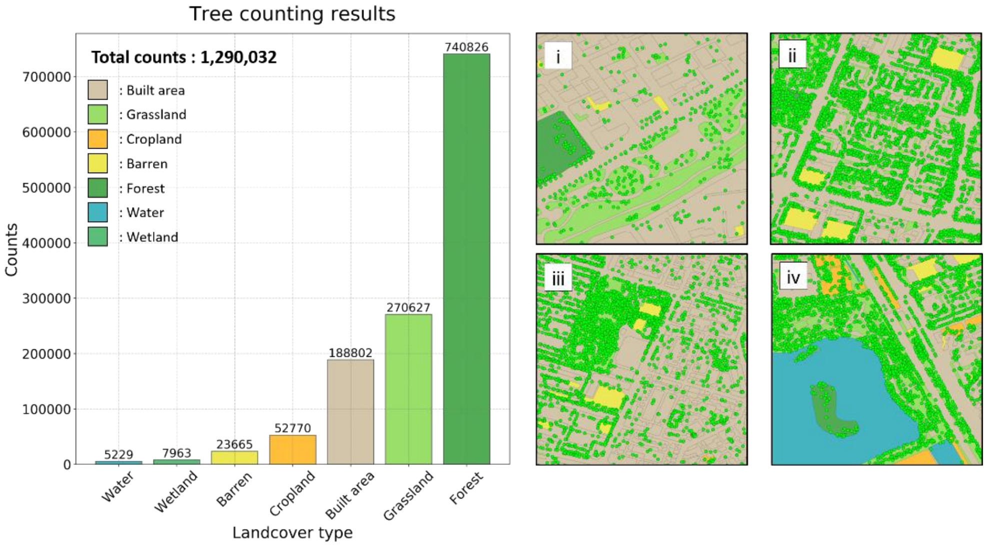


Fig. 13. Detailed information of tree counts results according to land cover types and image samples of tree positions (Each number in the samples represents locations of sites in Fig. 16)

3.3. Tree species detection with citizen and vehicle sensors

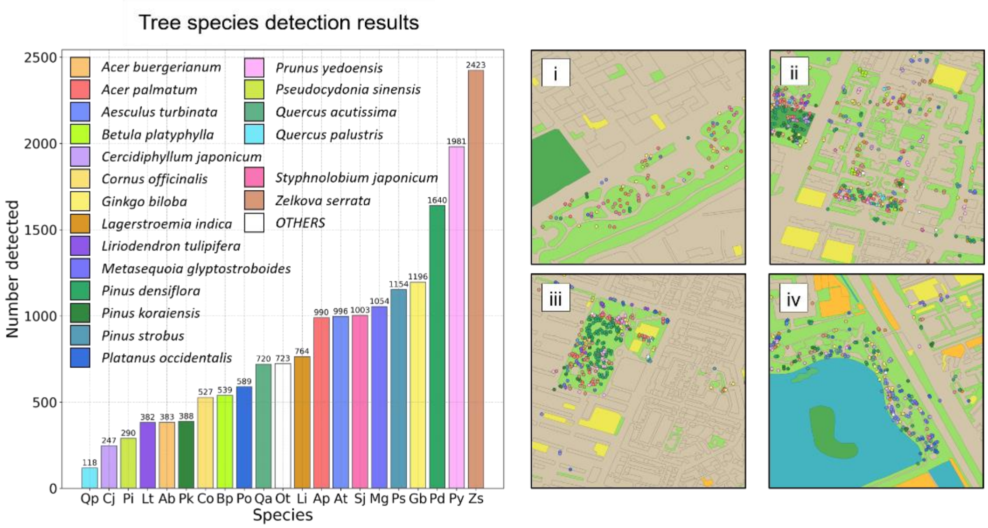


Fig. 14. Detailed information of the tree species map created by citizens' tree bark photos. Bar chart (left) indicates the number of trees by species which were detected by citizens. Representative tree species map results (right) visualize the tree species positions of the sample sites. Each number in the samples represents locations of sites in Fig. 16.

Tree bark photos, which were taken by citizens, provide unprecedented detailed information of tree species distribution in the urban areas. Fig. 4 presents our experimental evaluations of the trained network, which show the majority of classes are perfectly predicted. Although some errors arose due to misclassification of *Aesculus turbinata* as *Quercus acutissima* and *Pinus strobus* as *Pinus koraiensis*, we achieved the 95.9 % of overall accuracy for tree species detection from the bark images. After species mapping process was completed, the most frequently detected species from the species map was *Zelkova serrata* (13.38 %), followed by *Prunus yedoensis* (10.94 %), *Pinus densiflora* (9.06 %) and *Ginkgo biloba* (6.61 %). The least detected species was *Quercus palustris*, with 118 trees mapped. In particular, 3.99 % of trees were classified as 'OTHERS' because the classification score was lower than the confidence score threshold of 0.4. Our field survey indicated that 80.29% of trees were mapped to the correct species, and 15.27% of the trees were classified as wrong species. Additionally, Fig. 16 shows that only 2.54% of the trees were 'over-mapped', implying the feasibility of our approach, which simply transmits species classification results to nearby tree points. And about 4% of trees were found to be a minority species not included in deep learning training.

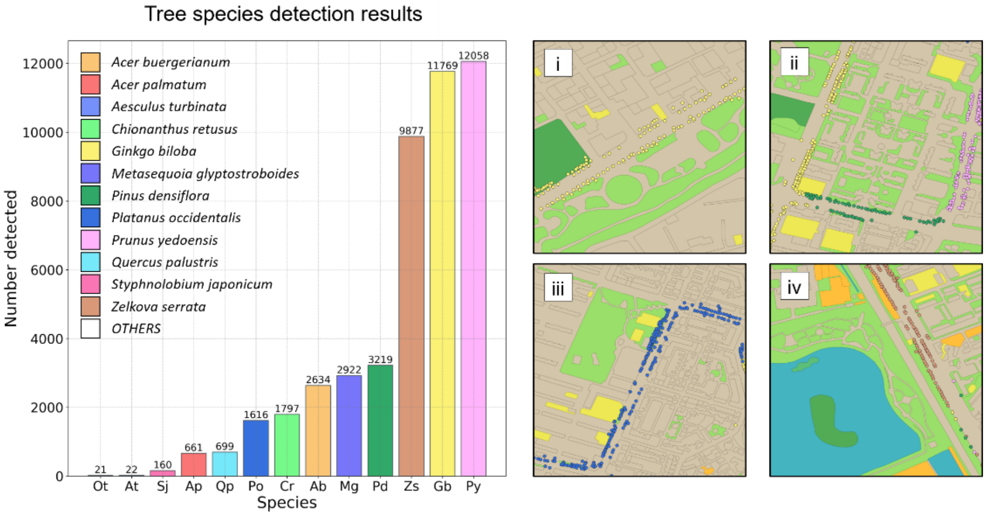


Fig. 15. Detailed information of the tree species map created by street-level imagery. Bar chart (left) indicates the number of trees by species which were detected by a vehicle. Representative tree species map results (right) visualize the tree species positions of the sample sites. Each number in the samples represents locations of sites in Fig. 16.

Our species mapping framework for roadside trees was robust, since a single species exists in a single block for most cases in urban areas. We achieved the 66.9 % of overall accuracy for street tree detection from street-level imagery. Fig. 5 implies the inaccuracy of the model was observed for *Quercus palustris* and *Aesculus turbinata* for which a few training data were collected. Only *Metasequoia glyptostroboides* and *Platanus occidentalis*, which have clearly distinct tree shapes, present high scores of inference results. Although we had an insufficient number of train datasets for several minor species, which led to misclassification results of the YOLOv3 model, a large proportion of them were converted to correct species through a majority voting process. As a result, most street trees were detected as *Prunus yedoensis* (25.41 %), *Ginkgo biloba* (24.8 %) or *Zelkova serrata* (20.81The evaluation of the vehicle-sensed species map shows 66.29 % of trees were mapped to the correct species and 17 % of trees were over-mapped (Fig. 16). The proportion of 'out-of-class' was lower (2.71 %) for street trees, implying that several major species make up the most roadside trees.

Chapter 4. Discussion

4.1. Multiple sensing platforms for urban areas

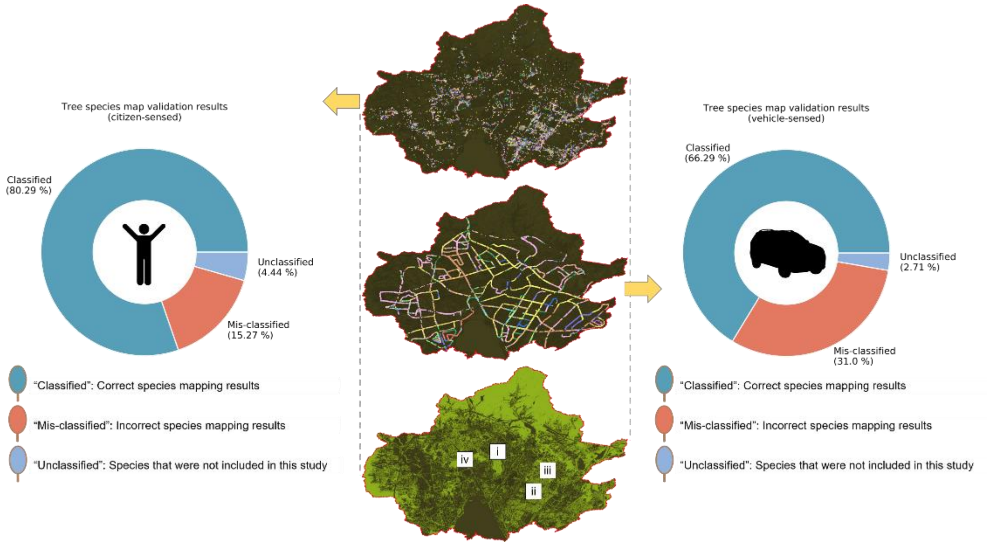
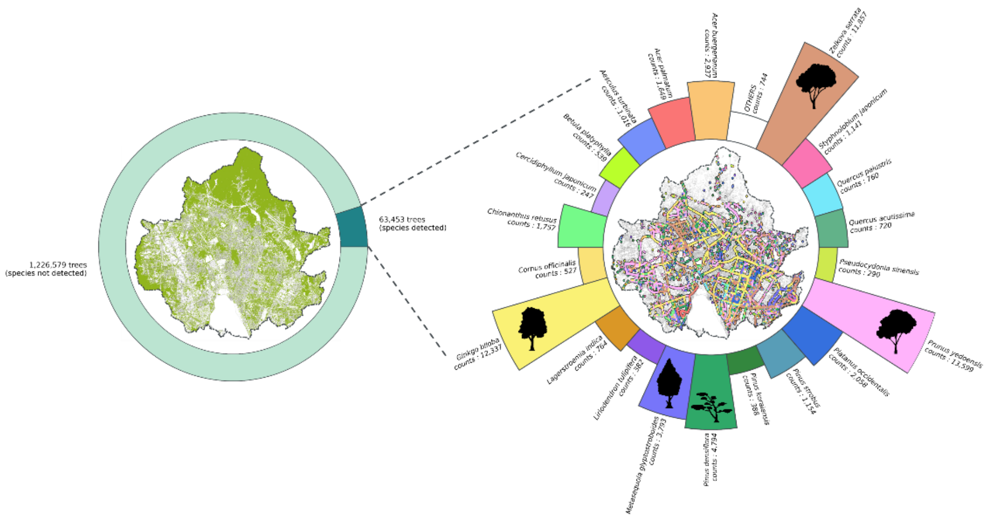


Fig. 16. Validation of species-mapping results derived from citizens' photos and vehicle-sensed imagery. The doughnut plots on the left and right indicate the proportion of "classified", "misclassified" and "unclassified" tree species. The definition of each term is also described in the insets of the figure. In the middle of the figure, the top map is a tree species map generated from citizens' datasets, and the middle one is generated from vehicle-sensed imagery.

In this study, we demonstrated that merging multiple sensing platforms was effective for heterogeneously planted urban areas, since remotely-sensed imagery and ground-level data complement each other in terms of spatial extents and resolution (Suel et al., 2021). The remote sensing approach still performs well for individual tree delineation at the city-scale. Our tree localization results are also promising, because the accuracy of tree positions outperforms some previous studies (Lumnitz et al., 2021), which only relied on a single sensing platform and were even conducted in small areas of sites. These locations can be used as the starting point for generating tree inventories (La Rosa et al., 2021; Laumer et al., 2020), since each position can include its metadata such as species, canopy height and diameter of breast height (Fassnacht et al., 2016). Particularly, tree species maps are essential for urban practitioners, since species diversity and spatial distribution are important indicators of the environment and citizens' well-being. We proved that ground-level data collected by citizens and vehicles could provide individual tree-level species information in urban areas. We conducted all these processes of individual tree counting and localizing with species detection at large spatial extents, and our multiple platforms-based approach was compared to other studies, as shown in Table. 7.



Species	*Prunus yedoensis*	*Ginkgo biloba*	*Zelkova serrata*	*Pinus densiflora*	*Metasequoia glyptostroboides*	Others
Proportion (%)	21.43	19.44	18.68	7.55	5.97	26.93

Fig. 17. Total number of trees whether their species were detected and circular bar chart representing the number of trees by species. Final species map is visualized inside the chart (right). Top 5 tree species which were frequently detected by citizen and vehicle are also presented.

We have shown that around 1.29 million trees exist in the study area (Fig. 13). Meanwhile, comparison between our results and other data sources implied there might be more trees under the dense canopies. Although our estimation of tree detection from the airborne imagery is highly correlated to GPS- and ALS-derived tree counts (Fig. 9), we expect that the number of tree tops estimated by ALS is underestimated in the sites with high planting density. This is because, when several trees are densely planted, they often share canopies with each other and underlying vegetation is prone to be undetectable. Considering that a large proportion of trees exist in forests (Fig. 13), the actual number of trees in the site will be more than expected. Additionally, since Pine trees are frequently planted in the site, the gap between our results and ground truth can be attributed to the insufficiency of the DeepForest model to identify these conifers. Although there were trees, we could not annotate ill-defined conifers from the image when labeling tree bounding boxes for fine-tuning the model (see Fig. 22). Therefore, even if the model has been trained well, it will inevitably be different from the real number of trees. Nevertheless, we considered the trees under the top canopies were out of extents of this study, since trees which had clearly visible canopies from the aerial images were pre-defined as target trees.

We stress that our localizing process for individual tree position provides an efficient baseline to elucidate the current status of urban trees at a city-scale. As mentioned above, the positioning database of these individual trees is the essential element of building an urban tree inventory, since we can add numerous information. We simply extracted the center points of the bounding boxes which were the inference results of the deep neural networks and the mean distance errors were around 2 m, which showed the remarkable performance of our method for localizing trees at the city-scale. As tree detection within a 4 m radius of ground truth was regarded as true positive (Branson et al., 2018), our deep learning based results are reliable. This approach is a more simplified version of the tree mapping framework compared to conventional ALS-based methods which need preprocessing to segment vegetation and selecting the site-specific parameters to locate tree tops. As shown in Fig. 18, it is noteworthy that the tree does not grow upright under the influence of sunlight or water, and the location of the stem and the top of the canopy are inevitably different. In addition, in urban areas with low planting density, such as parks, man-made structures such as light poles may be detected as trees in the ALS data. Thus, we concluded that ALS could not be a universal solution for individual tree localization in urban areas. It was one of the reasons why most of the previous studies prepared the ground truth of tree positions by visual interpretation of images (Brandt et al., 2020; Weinstein et al., 2020; Zhao et al., 2022; Zheng et al., 2020). In this study, we considered tree stem points and tree canopy tops as 'geolocation' of trees for sparsely and densely planted areas, respectively. We still need further exploration to define which can represent the location of trees for future studies. Comparison between ALS and TLS based tree point cloud data are visualized in Fig. 18.

	Airborne Laser Scanning	Terrestrial Laser Scanning	Global Positioning System
Device name	Leica TerrainMapper (Leica-Geosystems Inc., Heerbrugg, Switzerland)	Leica BLK360 (Leica-Geosystems Inc., Heerbrugg, Switzerland)	Trimble R4s GNSS Receiver (Trimble Inc., California, U.S.)
Operational settings	Flight altitude: 4,800 ft (reference plane elevation 20 m) Scanning angle: 28 ° Scanning rate per second: 1,563 kHz Mean point density: 20.4 pts/m2	Scanning distance: 0.6 – 60 m Scanning angle: horizontal 360 ° , vertical 300 ° Scanning rate per second: about 360,000 points Mean point density: 65 pts/m2	240 channels: GPS, GLONASS, SBAS, Galileo, Beidou Less than 1 m horizontal error in VRS, SBAS survey
Data acquisition period (year/month)	2020/09	2021/04, 2022/03	2021/11, 2022/02, 2022/03

Table. 5. Name and operational settings of each device

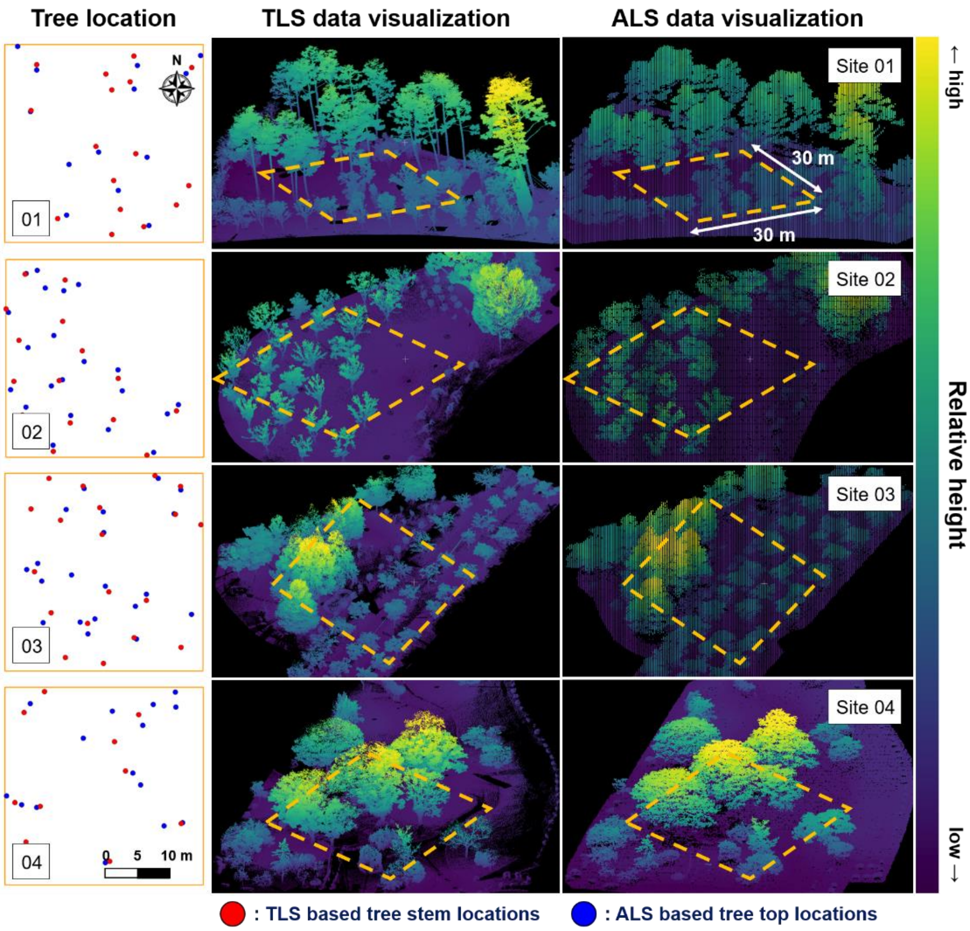


Fig. 18. Comparison between ALS and TLS based tree point cloud data

4.2. Potential of citizen and vehicle sensors

Our proposed new method presents the potential of citizen and vehicle sensors for producing fine−grained urban tree species maps. Most conventional studies have focused on forest sites (Fassnacht et al., 2016) since it has been difficult to map a large number of species in cities using multi− or hyperspectral imagery due to the influence of human−made ground, soil conditions, and the complexity of the underlying vegetation (Alonzo et al., 2014; Pu & Landry, 2012). In this respect, we highlight that citizens and vehicles are the most commonly found sensors in densely populated and road−meshed cities as they can cover different types of green spaces (i.e., urban parks and streets) although only 63,453 (4.91 %) trees were detected with species information (Fig. 17). This was expected results since most trees were located in mountain areas (Fig. 13) which was difficult to be accessed by citizens or vehicles. Note that, the precise geolocations of each tree species collected over a long period also enables time series analysis of urban vegetation. For example, phenological changes of plants were analyzed via citizen data over two centuries (Fuccillo Battle et al., 2022) and the variation of street trees was estimated via vehicle data during two periods (Branson et al., 2018). In addition, we expect that our method for generating the tree species map could perform as ground truth for training the model to classify the other areas, such as mountains. However, since the major tree species constituting the mountains were highly different from the urban areas, we considered this analysis to be outside the boundaries of this study.

We proved that mapping species information derived from citizens' tree bark photos is a promising framework for tree species classification in the city. Our classification model was trained quickly even with insufficient data, and showed high inference performance (Fig. 4). This is because image segmentation of U-Net is applied and only the textural features of tree bark images are used for training the Xception model. Fig. 7 shows that the Xception model is successfully classifying tree species, by filtering out the photos of low quality or other species which were not included in the training. Our assessment of citizen-sensed species maps implies that only 4.44 % of all trees are not filtered properly, although they are minority species. In addition, only 2.54 % of tree points were 'over-mapped' in the entire citizen-sensed species map, which shows that 5 m is suitable as the radius of the boundary from the geographical coordinates of acquired citizen data (Fig. 16). The high accuracy of the citizen-sensed species map (80.29 %) is attributed to the characteristics of urban landscape design in Korea, which is prone to colonize the same species in parks and forests. Therefore, although the classification result of the tree bark photo collected from one location was mapped to all trees within a radius of 5 m, there were few errors since trees of the same species existed nearby. Most of the trees evaluated as ‘wrong’ were not due to misclassification of the deep neural network but rather because of GPS errors in the citizens’ smartphones under the dense canopy. For instance, even if a tree bark photo of certain species was correctly classified, there were some cases where it was inaccurately mapped to other trees.

On the other hand, mapping street tree species with street−view imagery alone needs further improvements. Since street trees are pruned frequently, even different species often show similar tree shapes. As shown in Fig. 5, only *Metasequoia glyptostroboides* and *Platanus occidentalis* present high inference score, because of their unique tree shape and colorful bark, respectively. Biased number of training data due to only a few tree species constituting the majority of street trees also led to the inaccuracy of the YOLOv3 model. For example, it showed high accuracy for *Ginkgo biloba*, *Prunus yedoensis* and *Zelkova serrata*, which were labeled frequently. Nevertheless, since street trees of a certain species are often planted for each block, our approach was effective. Although we found our model missed some trees since there was a large proportion of False Negative (Fig. 5), the ‘correct’ ratio of species map validation results reached 66.29 % by simply filtering out the outliers within a single road. This approach also complements the uncertainty of the YOLOv3 model, because the created buffers covered tree positions that the model missed. In addition, a slight lane detection error did not significantly affect the mapping result, as shown in Fig. 16. If the expected number of lanes is greater than the real situation, there will be a larger boundary of the buffer. There is a possibility that species information will be transmitted to tree positions that are not street trees. But in most cases, the results were mapped to adjacent roadside trees only because street trees are usually planted in a row next to buildings.

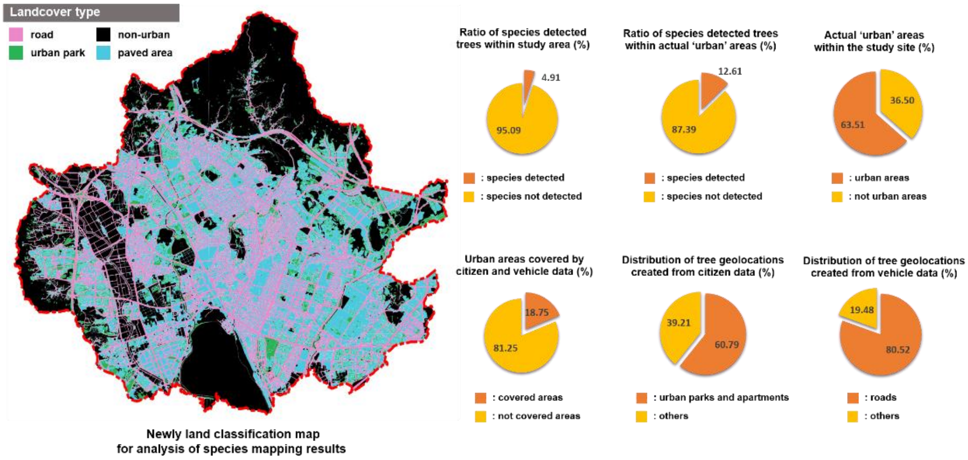
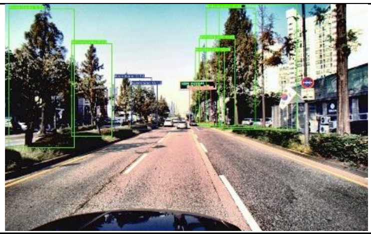


Fig. 19. Revised land classification map and the coverage of citizen and vehicle data

Additionally, we analyzed the coverage of citizens' and street-level data for tree species detection in urban areas. Species information which was inferred from citizens and vehicle data covered 11.9% of the total area of the study site. However, there were some landcover types which could not be regarded as urban areas such as mountainous sites, croplands, and water bodies within the target site. Therefore, we produced our new land classification map composed of 'urban park', 'paved area', 'road' and 'non-urban' by revising the landcover map and road data provided by the Korean government. Since Korean law stipulates that the minimum width of sidewalks is about 2 m in consideration of traffic safety, we defined the road data and a buffer of 2 m from road as 'road'. In addition, in the case of 'paved area', apartment complexes and residential areas are included except for urban parks. 'Non-urban' areas include mountainous sites, military bases, lakes and croplands which are difficult to access by citizens and vehicles. The newly produced land classification map showed that 'urban park' was 14.99 km^2 (12.4 %), 'paved area' was 35.27 km^2 (29.1 %), 'road' was 26.79 km^2 (22.1 %), and 'non-urban' was 44.27 km^2 (36.5 %). In other words, the ratio of the actual urban areas within the study site is about 63.51%. Using this, we estimated about 18.75% of the urban areas within the study site was covered by citizens and vehicles data. In detail, we proved that about 60.79% of trees of which species detected from citizens' data exist in urban parks and apartment complexes. Additionally, we found that 80.52% of the trees of which species detected from the street-level imagery were street trees located on the road. Notably, citizens and vehicles data detected species of only 7.51% and 2.63% of total trees in mountains and croplands, respectively, indicating that there are limitations for citizens and vehicles to cover 'non-urban' areas. Then, we analyzed how many trees got species information from citizens and vehicles. We found that 28.2% of trees out of about 149,000 trees existing on 'road' were detected from citizens and vehicles. Among them, 90.7% were mapped from vehicles, showing that most of the street trees were mapped from street-level imagery. Also, species information of 5.5% and 4.9% of trees were mapped in 'urban park' and 'paved area', respectively. Citizens' tree bark photos mapped tree species information to about 3.1% of all trees detected in 'urban park' and 'paved area'. Although the majority of citizen data was recorded in 'urban park' and 'paved area', only 3.1% of trees were mapped, implying that alternatives are needed to boost citizens to participate in living lab projects across a wide range of areas. As a result, excluding the 'non-urban' in the study site, about 12.61% of trees among all the trees existing in the urban areas got the species information from inference results of deep neural networks. Detailed information about the coverage of citizen and vehicle data is presented in Fig. 19.

ID	1	2	3	4	5	6	7
Citizens' photos							
Tree bark classification	*Acer palmatum* (correct)	*Ginkgo biloba* (correct)	*Metasequoia glyptostroboides* (correct)	*Aesculus turbinata* (correct)	*Zelkova serrata* (correct)	OTHERS (correct)	*Zelkova serrata* (correct)
Vehicle-sensed imagery							
Street tree detection	*Ginkgo biloba* (correct)		*Prunus yedoensis* (correct)		*Platanus occidentalis* (correct)		
Cause of species differences	GPS inaccuracy of smartphones	–	GPS inaccuracy of smartphones	Inaccuracy of LaneNet inference results	GPS inaccuracy of smartphones	–	

ID	8	9	10	11	12	13
Citizens' photos						
Tree bark classification	*Zelkova serrata* (correct)	*Prunus yedoensis* (correct)	*Pseudocydonia sinensis* (correct)	*Prunus yedoensis* (correct)	*Platanus occidentalis* (correct)	*Cornus officinalis* (**wrong**)
Vehicle–sensed imagery						
Street tree detection	*Metasequoia glyptostroboides* (correct)			*Ginkgo biloba* (correct)		
Cause of species differences	GPS inaccuracy of smartphones		Inaccuracy of LaneNet inference results	Xception error		–

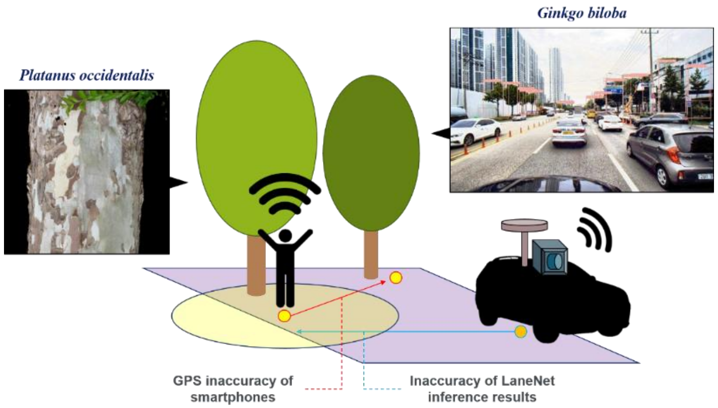
ID	14	15	16	Example image of species differences between citizen− and vehicle−based species maps
Citizens' photos				
Tree bark classification	*Zelkova serrata* (correct)	*Prunus yedoensis* (correct)	*Prunus yedoensis* (correct)	
Vehicle−sensed imagery				
Street tree detection		*Ginkgo biloba* (correct)		
Cause of species differences	Inaccuracy of LaneNet inference results	GPS inaccuracy of smartphones		

Table. 6. Field survey of tree species which were mapped from citizen and vehicle datasets.

We also tried to identify the potential and limitations of each data source by analyzing trees which have tree species information mapped from both citizen and vehicle data. As expected, most trees with overlapping tree species information from both data sources exist in the 'road' (86.5%). We selected some trees with different species information detected from citizen data and vehicle data and performed on-site survey. As a result, 30.9% of trees showed citizen data detected correctly, and 52.2% of trees showed vehicle data detected correctly. Notably, 16.9% of trees were found to be other species that were not covered in this study, which implied the need for improvement of deep learning-based methodologies that depend heavily on train data. Even if the tree species information based on vehicle data was more precise than citizen data, citizens could access the inside of a block that vehicles could not cover. Therefore, in this study, when different tree species were mapped from citizens and vehicles, we followed the results based on citizens' data. In general, the cause of the error in the tree species map based on citizen data was the inaccuracy of the smartphone GPS. While performing field survey, we found that most of the tree bark photos was precisely classified by the Xception network. However, most of tree geolocations which were mapped by both citizen and vehicle data showed these tree bark photos were taken from the trees inside the block, not the roadside trees. In other words, even if the tree bark classification was performed well, the tree species information was incorrectly mapped to the adjacent street tree due to the smartphone GPS error. Therefore, in order to increase the accuracy of this study, which used citizens' data to map urban tree species, the smartphone GPS performance needs to be improved. Therefore, we believe that if citizen data is accumulated over a long period of time and the performance of smartphones is improved, this study will emerge as an innovative alternative for managing trees on a city scale. Similar to Fig. 16, we found that one drawback of the street-level imagery-based methodology was a lane detection error on a curved road. Some of trees which got species information from both tree bark photos and street-level imagery showed that street tree detection results excessively transmitted to trees inside the blocks. Since the lane width or the shape of lanes are inconsistent in the city, the distance from a vehicle to street trees were overestimated when the vehicle turned right or left. In this case, the citizen data results were often correctly mapped. An example of a tree in which citizen data and vehicle data are mapped to different tree species can be seen in Table. 6.

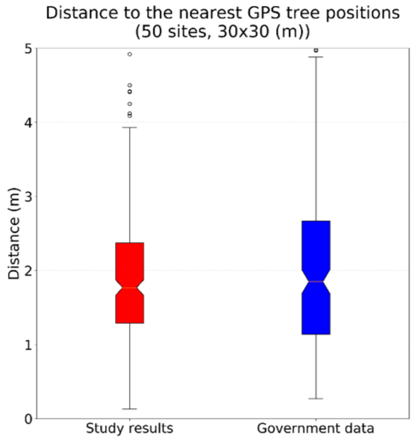


Fig. 20. Comparison of distance errors between study results and field data (only street trees) provided by Suwon city government

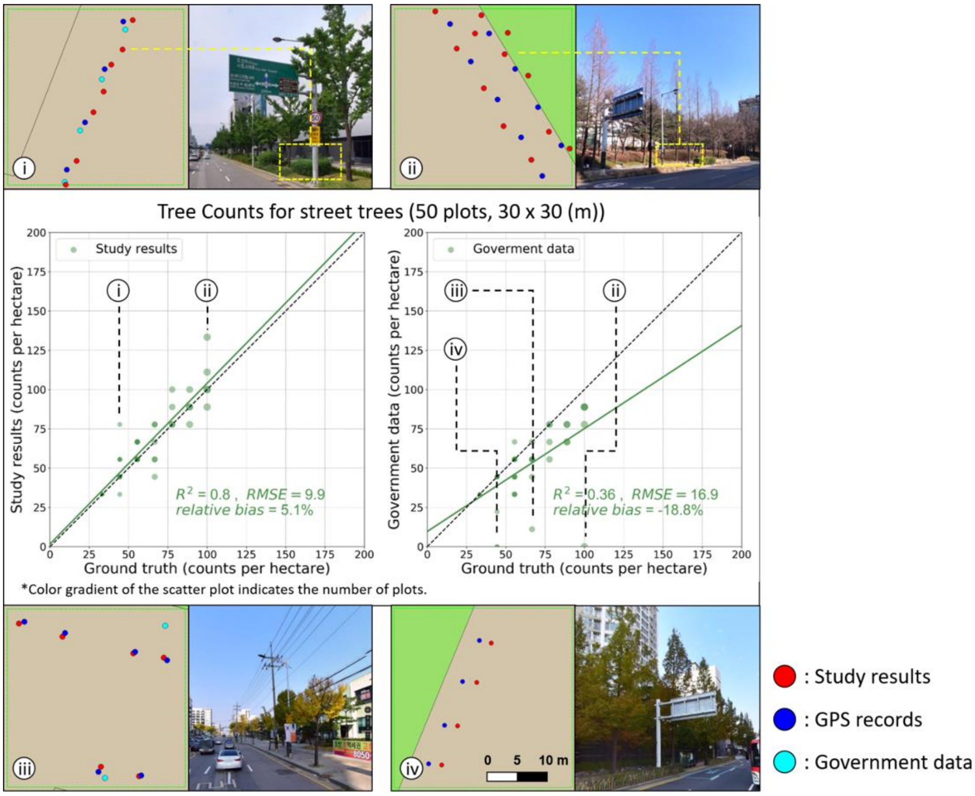


Fig. 21. Comparison of the number of street trees between study results and field data provided by Suwon city government. The sample plots i-iv, which show big difference with ground truth, are also presented with site photos provided by NaverMap (https://map.naver.com/v5/). The yellow boxes in the above figures indicate the cause of tree detection errors in this study (e.g., large areas of shrubs were often detected as trees).

Nevertheless, we demonstrated strengths of our method compared to previous studies and field data provided by city government. In order to assess the novelty of our methodology, we compared several representative studies (Alonzo et al., 2014; Aval et al., 2018; Brandt et al., 2020; Laumer et al., 2020; Liu & An, 2019; Liu et al., 2017; Lumnitz et al., 2021; Martins et al., 2021; Pu & Landry, 2012; Sun et al., 2022; Tong et al., 2021; Yang et al., 2022) which were conducted for similar purposes. Most previous studies relied on a single platform, although the data types were different. Not only urban areas, we also covered research which were conducted in other land cover types of regions. As shown in Table.

7, a few studies performed counting, mapping individual trees and species detection in urban areas and the number of target species was small. One study achieved the highest accuracy targeting the largest number of tree species. Note that the price of airborne LiDAR and hyperspectral imagery is expensive compared to our data. Nevertheless, since the mountain areas were not included in our study and the locations of acquisition of citizen data were not evenly distributed, our results still need to be improved to cover the whole sites. In addition, we compared the results of this study with the street tree data provided by the Suwon city government. Note that although our study is not limited to street trees, the accuracy of our results outperformed the government data. Suwon city data in Figure. 21 recorded that there were no trees at all in some sample sites despite the existence of trees. We confirmed that the Suwon city data underestimated the actual number of trees. On the other hand, our results of the number of trees were similar to the ground truth. Additionally, the estimated locations of the detected trees were also more precise than those of city data. The average distance error between tree positions in GPS data which were collected by our field survey was concentrated around 1~2 m. However, in the distribution of errors, we confirmed that our results were more concentrated than the Suwon city data, which means that our methodology shows much more stable accuracy than labor-intensive data collection. We have demonstrated that our deep learning-based approach was superior to the conventional time-consuming and labor-intensive field survey which were conducted by the city government.

Name	Sensing platform	Data type	Study area (km²)	Landcover type	Counting trees	Individual tree mapping	Species detection	Number of target species	Accuracy		
									Tree counts	Positional error (m)	*Tree species detection (%)
Sun et al. (2022)	Airborne	RGB	7434.4	Urban	O	X	X	–	R2: 0.88	–	–
Aval et al. (2018)	Airborne, contextual information	Hyperspectral, DSM, GIS data	street-level-only	Urban	O	O	X	–	F-score: 0.91	–	–
Brandt et al. (2020)	Satellite	Multispectral	1.3 million	Dryland	O	O	X	–	R2: 0.97	–	–
Tong et al. (2021)	Unmanned Aerial Vehicle	RGB	no info	Cropland	O	O	X	–	RMSE: 3 to 6	–	–
Yang et al. (2022)	Airborne	RGB	3.41	Park	O	O	X	–	F-score: 0.90	–	–
Laumer et al. (2020)	Vehicle	RGB	street-level-only	Urban	O	O	X	–	–	–	–
Lumnitz et al. (2021)	Vehicle	RGB	street-level-only	Urban	O	O	X	–	Detect rate: > 0.70	4 to 6	–
Choi et al. (2022)	Vehicle	RGB	street-level-only	Urban	X	O	O	5	–	43.2	56.4
Pu and Landry (2012)	Satellite	Multispectral	30	Urban	X	X	O	7	–	–	63.0
Liu and An (2019)	Satellite	Hue, Saturation, Value	2.45	Urban	X	X	O	7	–	–	77.6

Martins et al. (2021)	Airborne	RGB	3.25	Urban	X	X	O	9	–	–	79.3 ± 8.6
Alonzo et al. (2014)	Airborne	LiDAR, hyperspectral	no info	Urban	X	O	O	29	–	–	83.4
Liu et al. (2017)	Airborne	LiDAR, hyperspectral	316	Urban	X	O	O	15	–	–	70.0 ± 3.1
This study	Airborne, Citizen, Vehicle	RGB	121.04	Urban	O	O	O	21	0.95	2.0 to 2.2	80.3 (park, forest), 66.3 (street)

Table. 7. Comparison between our study and others which were performed for similar purposes. Detailed methods of accuracy measurement are not the same.

4.3. Implications

We propose several suggestions for the improvement of future study. First, the citizens' data collected in this study is prone to be concentrated in a specific area (Fig. 1). They were mainly collected in parks that are visited by a lot of people or easily accessible apartments. We expected that this citizens' behavior implied the reward money for the living lab project was somewhat insufficient. Therefore, in the future study, we will induce active movement of citizens by providing higher rewards for areas that are difficult to access, or by setting a distance threshold of the data acquisition location rather than the time interval. Second, the performance of our camera mounted on a car was not enough to capture clear differences among street tree species. Although some species had distinctive traits, the features of most species were slightly different at the leaf-level rather than the canopy-level. We believed that these subtle differences might be omitted by resizing images to capture abstract features of objects on images while training CNNs.

Although deep learning is able to process city-scale data automatically, a simpler model and sufficient training data which contains a larger number of tree species are highly desirable for future study. Recent years have witnessed the development of ecological analysis led by incorporating deep learning into accessible sensors, thereby increasing available data sources (Tuia et al., 2022). In this study, we suggested solutions for each purpose (i.e. tree counting, localization, and species detection) step by step. Since deep learning has the problems of lacking theoretical explanation and so-called black box issues, the performance of multiple deep neural networks is hard to be identifiable (Shwartz-Ziv & Tishby, 2017). Therefore, improving the overall accuracy of this study was not easy because the uncertainty of each model stacked little by little. Designing a single network that performs multiple tasks is challenging due to the difficulty of preparing good quality data and huge computational costs. Deep neural networks are also heavily affected by the data itself. For instance, there were 4 % of remaining tree species not included in this study in urban parks (Fig. 16). In other words, we cannot detect every species all over the city since it is hard to include all species for training the models.

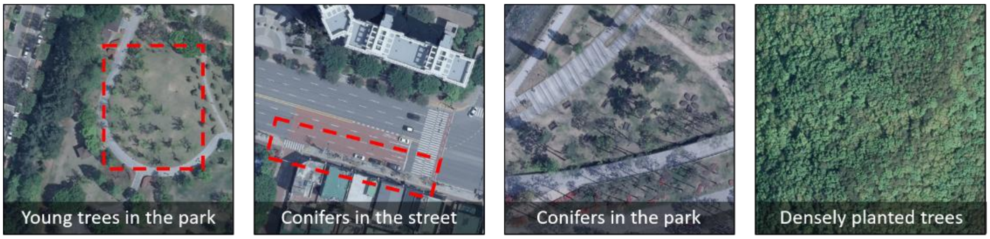
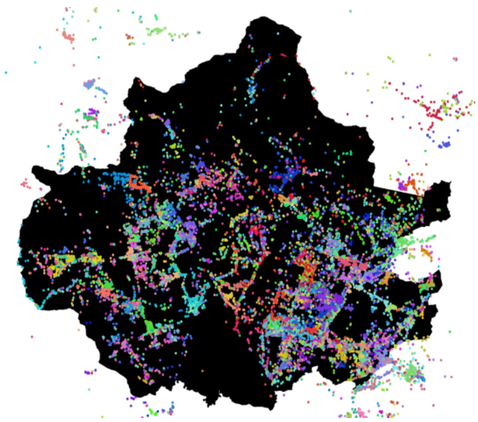


Fig. 22. Examples of airborne imagery which show various planting patterns in the study area



Ranking	Device ID	Participation count
1	lm4h**********************U2	1,779
2	Lxwc**********************R2	1,065
3	oP2s**********************02	1,009
4	IU5C**********************A3	917
5	s3BU**********************72	878
6	SkDw**********************h1	819
7	eMcX**********************62	781
8	FOry**********************I2	720
9	Gf2k**********************j2	671
10	TvsbQ**********************g1	655
	...	
889	Tpfx**********************12	1
Total number of photos		45,300

Fig. 23. Distributions of each smartphone device ID and the number of participation by citizens who participated in the living lab project

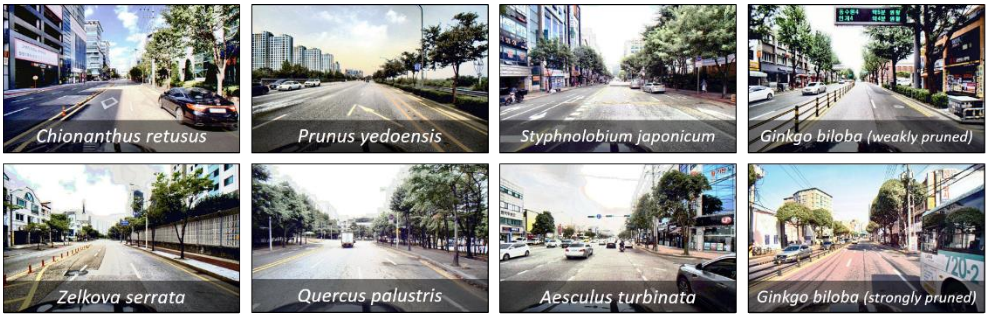


Fig. 24. Street-level imagery which show interspecies similarity and the intra-species variability

We analyzed detailed implications of airborne, citizen and vehicle sensing. Fig. 22 illustrates the heterogeneous planting patterns of urban trees. The first and second airborne photos show that young trees and conifer street trees which have ambiguous canopy boundaries are difficult to be distinguished clearly. Conifers in the park were one of the major causes of uncertainty in tree count and location results. In particular, shadows from conifers adversely affected the deep learning model's performance to detect trees. In mountain areas, since most trees have large canopy size, delineating the canopy boundaries is easier than in urban areas. However, aerial photos alone do not imply how many more trees there will be under the canopy. Fig. 23 shows the device information of citizens who participated in the living lab project. The color of each point represents each smartphone ID. That is, location records of the same color mean they were transmitted from the same device. We analyzed the total number of photos sent by actively participating citizens, and tabulated the results. Among 45,300 photos, 1,779 photos were sent by the most enthusiastically engaged citizens. We confirmed that the number of photos collected by the top 10 people reached 20.5 % of the total tree bark photos. This suggests that more effective public relations are needed for the collection of extensive citizen data. Street trees are strongly pruned in urban areas, since buildings are located next to trees in most cases. Fig. 24 present interspecies similarity of street trees. Although there were subtle differences in color, shape, and texture of the canopy among *Chionanthus retusus*, *Prunus yedoensis*, *Styphnolobium japonicum*, *Zelkova serrata* and *Quercus palustris*, they are pruned in such a similar shape that it is difficult to identify the species without an expertise. Even the same *Ginkgo biloba* looked more similar to another species depending on the strength of the pruning. *Aesculus turbinata* looked much similar to *Ginkgo biloba* which was strongly pruned in this figure.

Chapter 5. Conclusion

In this study, we developed a novel framework for mapping individual trees at the city−scale by merging multimodal data and deep learning. We demonstrated the applicability of various deep neural networks for producing urban tree maps; we also showed tree bark photos and street−level imagery for tree species mapping. First, our trained generative network successfully generated the trees, which were hidden in the shadows. Second, it was possible to create a tree map by extracting the central point of the bounding boxes, which were the results of the tree detection. Third, we demonstrated that tree bark photos with the background removed provided clear information for tree species detection, and the high accuracy of validation results implied the capability of citizens' data for species mapping. Finally, we confirmed that a street imagery−based approach should be improved for mapping tree species, since it was difficult to collect tree traits with a camera alone. Our results highlight that deep learning and data derived from multiple platforms could be used for generating a city−scale tree map. We believe our study provides a precise estimation of the number of trees and individual tree location with species information which is essential for urban tree management.

Bibliography

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M. (2016). {TensorFlow}: A System for {Large-Scale} Machine Learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16).
2. Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment, 148, 70-83.
3. Aurenhammer, F. (1991). Voronoi diagrams—a survey of a fundamental geometric data structure. ACM Computing Surveys (CSUR), 23(3), 345-405.
4. Aval, J., Demuynck, J., Zenou, E., Fabre, S., Sheeren, D., Fauvel, M., Adeline, K., & Briottet, X. (2018). Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. ISPRS Journal of Photogrammetry and Remote Sensing, 146, 197-210.
5. Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., & Diouf, A. A. (2020). An unexpectedly large count of trees in the West African Sahara and Sahel. Nature, 587(7832), 78-82.
6. Branson, S., Wegner, J. D., Hall, D., Lang, N., Schindler, K., & Perona, P. (2018). From Google Maps to a fine-grained catalog of street trees. ISPRS Journal of Photogrammetry and Remote Sensing, 135, 13-30.
7. Bratman, G. N., Anderson, C. B., Berman, M. G., Cochran, B., De Vries, S., Flanders, J., Folke, C., Frumkin, H., Gross, J. J., & Hartig, T. (2019). Nature and mental health: An ecosystem service perspective. Science advances, 5(7), eaax0903.
8. Budei, B. C., St-Onge, B., Hopkinson, C., & Audet, F.-A. (2018). Identifying the genus or species of individual trees using a three-wavelength airborne lidar system. Remote Sensing of Environment, 204, 632-647.

9. Burns, T. J., Davis, B., & Kick, E. L. (1997). Position in the world-system and national emissions of greenhouse gases. Journal of World-Systems Research, 3(3), 432-466.
10. Carpentier, M., Giguere, P., & Gaudreault, J. (2018). Tree species identification from bark images using convolutional neural networks. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
11. Chen, S., Chen, B., Feng, K., Liu, Z., Fromer, N., Tan, X., Alsaedi, A., Hayat, T., Weisz, H., & Schellnhuber, H. J. (2020). Physical and virtual carbon metabolism of global cities. Nature communications, 11(1), 1-11.
12. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition.
13. Dalponte, M., Bruzzone, L., & Gianelle, D. (2012). 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.
14. Dalponte, M., Ørka, H. O., Ene, L. T., Gobakken, T., & Næsset, E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sensing of Environment, 140, 306-317.
15. Edmondson, J. L., Davies, Z. G., McHugh, N., Gaston, K. J., & Leake, J. R. (2012). Organic carbon hidden in urban ecosystems. Scientific Reports, 2(1), 1-7.
16. Escobedo, F. J., Kroeger, T., & Wagner, J. E. (2011). Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. Environmental pollution, 159(8-9), 2078-2087.
17. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd.
18. Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment, 186, 64-87.

19. Ferreira, M. P., Wagner, F. H., Aragão, L. E., Shimabukuro, Y. E., & de Souza Filho, C. R. (2019). Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 149, 119-131.
20. Fuccillo Battle, K., Duhon, A., Vispo, C. R., Crimmins, T. M., Rosenstiel, T. N., Armstrong-Davies, L. L., & de Rivera, C. E. (2022). Citizen science across two centuries reveals phenological change among plant species and functional groups in the Northeastern US. Journal of Ecology.
21. Gaffin, S. R., Rosenzweig, C., & Kong, A. Y. (2012). Adapting to climate change through urban green infrastructure. Nature Climate Change, 2(10), 704-704.
22. Hemmerling, J., Pflugmacher, D., & Hostert, P. (2021). Mapping temperate forest tree species using dense Sentinel-2 time series. Remote Sensing of Environment, 267, 112743.
23. Holopainen, M., Kankare, V., Vastaranta, M., Liang, X., Lin, Y., Vaaja, M., Yu, X., Hyyppä, J., Hyyppä, H., & Kaartinen, H. (2013). Tree mapping using airborne, terrestrial and mobile laser scanning–A case study in a heterogeneous urban forest. Urban Forestry & Urban Greening, 12(4), 546-553.
24. Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition.
25. Jensen, R. R., Hardin, P. J., & Hardin, A. J. (2012). Classification of urban tree species using hyperspectral imagery. Geocarto International, 27(5), 443-458.
26. Kälin, U., Lang, N., Hug, C., Gessler, A., & Wegner, J. D. (2019). Defoliation estimation of forest trees from ground-level images. Remote Sensing of Environment, 223, 143-153.
27. Kardan, O., Gozdyra, P., Misic, B., Moola, F., & Palmer, L. J. (2017). Neighborhood greenspace and health in a large urban center. In Urban Forests (pp. 77-108). Apple Academic Press.
28. Key, T., Warner, T. A., McGraw, J. B., & Fajvan, M. A. (2001). 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.
29. La Rosa, L. E. C., Sothe, C., Feitosa, R. Q., de Almeida, C. M., Schimalski, M. B., & Oliveira, D. A. B. (2021). Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing, 179, 35-49.
30. Laumer, D., Lang, N., van Doorn, N., Mac Aodha, O., Perona, P., & Wegner, J. D. (2020). Geocoding of trees from street addresses and street-level images. ISPRS Journal of Photogrammetry and Remote Sensing, 162, 125-136.
31. Leighton, G. R., Hugo, P. S., Roulin, A., & Amar, A. (2016). Just Google it: assessing the use of Google Images to describe geographical variation in visible traits of organisms. Methods in Ecology and Evolution, 7(9), 1060-1070.
32. Li, D., Liao, W., Rigden, A. J., Liu, X., Wang, D., Malyshev, S., & Shevliakova, E. (2019). Urban heat island: Aerodynamics or imperviousness? Science advances, 5(4), eaau4299.
33. Li, X., Ratti, C., & Seiferling, I. (2017). Mapping urban landscapes along streets using google street view. International cartographic conference.
34. Li, X., Ratti, C., & Seiferling, I. (2018). Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. Landscape and Urban Planning, 169, 81-91.
35. Liu, H., & An, H. (2019). Urban greening tree species classification based on HSV colour space of WorldView-2. Journal of the Indian Society of Remote Sensing, 47(11), 1959-1967.
36. Liu, L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. Remote Sensing of Environment, 200, 170-182.

37. Lumnitz, S., Devisscher, T., Mayaud, J. R., Radic, V., Coops, N. C., & Griess, V. C. (2021). Mapping trees along urban street networks with deep learning and street−level imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 175, 144−157.
38. Martins, G. B., La Rosa, L. E. C., Happ, P. N., Coelho Filho, L. C. T., Santos, C. J. F., Feitosa, R. Q., & Ferreira, M. P. (2021). Deep learning−based tree species mapping in a highly diverse tropical urban setting. Urban Forestry & Urban Greening, 64, 127241.
39. Misra, D., Crispim−Junior, C., & Tougne, L. (2020). Patch−based CNN evaluation for bark classification. European Conference on Computer Vision.
40. Moreira, T. C., Polize, J. L., Brito, M., da Silva Filho, D. F., Chiavegato Filho, A. D., Viana, M. C., Andrade, L. H., & Mauad, T. (2021). Assessing the impact of urban environment and green infrastructure on mental health: results from the São Paulo Megacity Mental Health Survey. Journal of exposure science & environmental epidemiology, 1−8.
41. Neven, D., De Brabandere, B., Georgoulis, S., Proesmans, M., & Van Gool, L. (2018). Towards end−to−end lane detection: an instance segmentation approach. 2018 IEEE intelligent vehicles symposium (IV).
42. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high−performance deep learning library. Advances in neural information processing systems, 32.
43. Pu, R., & Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView−2 imagery for mapping urban tree species. Remote Sensing of Environment, 124, 516−533.
44. Pu, R., & Landry, S. (2020). Mapping urban tree species by integrating multi−seasonal high resolution pléiades satellite imagery with airborne LiDAR data. Urban Forestry & Urban Greening, 53, 126675.
45. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
46. Remeš, V., & Haindl, M. (2019). Bark recognition using novel

rotationally invariant multispectral textural features. Pattern Recognition Letters, 125, 612–617.
47. Rhodes, J. R., Ng, C. F., de Villiers, D. L., Preece, H. J., McAlpine, C. A., & Possingham, H. P. (2011). Using integrated population modelling to quantify the implications of multiple threatening processes for a rapidly declining population. Biological Conservation, 144(3), 1081–1088.
48. Richards, D. R., & Edwards, P. J. (2017). Quantifying street tree regulating ecosystem services using Google Street View. Ecological Indicators, 77, 31–40.
49. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention.
50. Roussel, J.-R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R., Meador, A. S., Bourdon, J.-F., de Boissieu, F., & Achim, A. (2020). lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. Remote Sensing of Environment, 251, 112061.
51. Roy, S., Byrne, J., & Pickering, C. (2012). A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. Urban Forestry & Urban Greening, 11(4), 351–363.
52. Savard, J.-P. L., Clergeau, P., & Mennechez, G. (2000). Biodiversity concepts and urban ecosystems. Landscape and Urban Planning, 48(3-4), 131–142.
53. Schiller, C., Schmidtlein, S., Boonman, C., Moreno-Martínez, A., & Kattenborn, T. (2021). Deep learning and citizen science enable automated plant trait predictions from photographs. Scientific Reports, 11(1), 1–12.
54. Schwaab, J., Meier, R., Mussetti, G., Seneviratne, S., Bürgi, C., & Davin, E. L. (2021). The role of urban trees in reducing land surface temperatures in European cities. Nature communications, 12(1), 1–11.
55. Seiferling, I., Naik, N., Ratti, C., & Proulx, R. (2017). Green

streets– Quantifying and mapping urban trees with street-level imagery and computer vision. Landscape and Urban Planning, 165, 93-101.

56. Shwartz-Ziv, R., & Tishby, N. (2017). Opening the black box of deep neural networks via information. arXiv preprint arXiv:1703.00810.
57. Song, C. (2005). Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability? Remote Sensing of Environment, 95(2), 248-263.
58. Suel, E., Bhatt, S., Brauer, M., Flaxman, S., & Ezzati, M. (2021). Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. Remote Sensing of Environment, 257, 112339.
59. Sun, Y., Li, Z., He, H., Guo, L., Zhang, X., & Xin, Q. (2022). Counting trees in a subtropical mega city using the instance segmentation method. International Journal of Applied Earth Observation and Geoinformation, 106, 102662.
60. Suzuki-Ohno, Y., Westfechtel, T., Yokoyama, J., Ohno, K., Nakashizuka, T., Kawata, M., & Okatani, T. (2022). Deep learning increases the availability of organism photographs taken by citizens in citizen science programs. Scientific Reports, 12(1), 1-10.
61. Suzuki-Ohno, Y., Yokoyama, J., Nakashizuka, T., & Kawata, M. (2020). Estimating possible bumblebee range shifts in response to climate and land cover changes. Scientific Reports, 10(1), 1-12.
62. Tigges, J., Lakes, T., & Hostert, P. (2013). Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. Remote Sensing of Environment, 136, 66-75.
63. Tong, P., Han, P., Li, S., Li, N., Bu, S., Li, Q., & Li, K. (2021). Counting trees with point-wise supervised segmentation network. Engineering Applications of Artificial Intelligence, 100, 104172.
64. Toomey, A., Strehlau-Howay, L., Manzolillo, B., & Thomas, C.

(2020). The place-making potential of citizen science: Creating social-ecological connections in an urbanized world. Landscape and Urban Planning, 200, 103824.
65. Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., & Burghardt, T. (2022). Perspectives in machine learning for wildlife conservation. Nature communications, 13(1), 1-15.
66. van Ewijk, K. Y., Randin, C. F., Treitz, P. M., & Scott, N. A. (2014). Predicting fine-scale tree species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. Remote Sensing of Environment, 150, 120-131.
67. Wegner, J. D., Branson, S., Hall, D., Schindler, K., & Perona, P. (2016). Cataloging public objects using aerial and street-level images-urban trees. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
68. Weinstein, B. G., Marconi, S., Aubry-Kientz, M., Vincent, G., Senyondo, H., & White, E. P. (2020). DeepForest: A Python package for RGB deep learning tree crown delineation. Methods in Ecology and Evolution, 11(12), 1743-1751.
69. Wood, C., Sullivan, B., Iliff, M., Fink, D., & Kelling, S. (2011). eBird: engaging birders in science and conservation. PLoS biology, 9(12), e1001220.
70. Yang, M., Mou, Y., Liu, S., Meng, Y., Liu, Z., Li, P., Xiang, W., Zhou, X., & Peng, C. (2022). Detecting and mapping tree crowns based on convolutional neural network and Google Earth images. International Journal of Applied Earth Observation and Geoinformation, 108, 102764.
71. Yao, L., Liu, T., Qin, J., Lu, N., & Zhou, C. (2021). Tree counting with high spatial-resolution satellite imagery based on deep neural networks. Ecological Indicators, 125, 107591.
72. Zhang, C., Atkinson, P. M., George, C., Wen, Z., Diazgranados, M., & Gerard, F. (2020). Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 169, 280-291.

73. Zhao, H., Zhong, Y., Wang, X., Hu, X., Luo, C., Boitt, M., Piiroinen, R., Zhang, L., Heiskanen, J., & Pellikka, P. (2022). Mapping the distribution of invasive tree species using deep one-class classification in the tropical montane landscape of Kenya. ISPRS Journal of Photogrammetry and Remote Sensing, 187, 328-344.
74. Zhao, Y., Gao, X., Hu, J., Chen, Z., & Chen, Z. (2020). Tree species identification based on the fusion of bark and leaves. Mathematical Biosciences and Engineering, 17(4), 4018-4033.
75. Zheng, J., Fu, H., Li, W., Wu, W., Zhao, Y., Dong, R., & Yu, L. (2020). Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network. ISPRS Journal of Photogrammetry and Remote Sensing, 167, 154-177.

Abstract

도시 전역에 존재하는 모든 수목의 숫자와 개별 위치, 그리고 수종 분포를 정확하게 파악하는 것은 생태계 서비스를 향상시키기 위한 필수조건이다. 하지만, 도시에서는 수목의 분포가 매우 복잡하기 때문에 개별 수목을 맵핑하는 것은 어려웠다. 본 연구에서는, 여러가지 센싱 플랫폼을 융합함과 동시에 다양한 딥러닝 네트워크들을 활용하여 세밀한 도시 수목 지도를 제작하는 새로운 프레임워크를 제안한다. 우리는 오직 항공사진, 시민, 차량 등의 플랫폼으로부터 수집된 RGB 이미지만을 사용하였으며, 6가지 딥러닝 모델을 활용하여 수목을 맵핑하고 수종을 탐지하였다. 각각의 플랫폼은 저마다의 강점이 있기 때문에 전 과정을 세 가지 스텝으로 구분할 수 있다. 첫째, 우리는 항공사진 상에서 탐지된 수목의 딥러닝 바운딩 박스로부터 중심점을 추출하여 개별 수목의 위치 지도를 제작하였다. 많은 수목이 도시 내 고층 빌딩의 그림자에 의해 가려졌기 때문에, 우리는 생정적 적대적 신경망 (Generative Adversarial Network, GAN)을 통해 항공사진 상에 숨겨진 수목을 그려내고자 하였다. 둘째, 우리는 시민들이 수집한 수목의 수피 사진을 활용하여 도시 공원 및 도시 숲 일대에 수종 정보를 맵핑하였다. 수피 사진으로부터의 수종 정보는 딥러닝 네트워크에 의해 자동으로 분류되었으며, 이 과정에서 이미지 분할 모델 또한 적용되어 딥러닝 분류 모델이 오로지 수피 부분에만 집중할 수 있도록 하였다. 셋째, 우리는 차량에 탑재된 카메라를 활용하여 도로변 가로수의 수종을 탐지하였다. 이 과정에서 차량으로부터 가로수까지의 거리 정보가 필요하였는데, 우리는 이미지 상의 차선 개수로부터 거리를 추정하였다. 마지막으로, 본 연구 결과는 라이다(Light Detection and Ranging, LiDAR)와 GPS 장비, 그리고 현장 자료에 의해 평가되었다. 우리는 121.04 km² 면적의 대상지 내에 약 130만여 그루의 수목이 존재하는 것을 확인하였으며, 다양한 선행연구보다 높은 정확도의 개별 수목 위치 지도를 제작하였다. 탐지된 모든 수목 중 약 6만 3천여 그루의 수종 정보가 탐지되었으며, 이중 가장 빈번히 탐

지된 수목은 왕벚나무 (*Prunus yedoensis*, 21.43 %)였다. 은행나무 (*Ginkgo biloba*, 19.44 %), 느티나무 (*Zelkova serrata*, 18.68 %), 소나무 (*Pinus densiflora*, 7.55 %), 그리고 메타세쿼이어 (*Metasequoia glyptostroboides*, 5.97 %) 등이 그 뒤를 이었다. 포괄적인 검증이 수행되었고, 본 연구에서는 시민이 수집한 수피 사진과 차량으로부터 수집된 도로변 이미지는 도시 수종 분포에 대한 정확하고 정량적인 정보를 제공한다는 것을 검증하였다.