Establishing a citywide street tree inventory with street view images and computer vision techniques

Dongwei Liu [a], Yuxiao Jiang [a,b], Ruoyu Wang [c], Yi Lu [a,*]

[a] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong
[b] School of Architecture, Tianjin University, Tianjin, China
[c] UKCRC Centre of Excellence for Public Health/Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

ARTICLE INFO

ABSTRACT

Trees in urban areas have diverse ecological, social, and health benefits. The establishment of up-to-date and accurate street-tree inventories that list the species and locations of individual street trees is critical to urban tree management and tree-planting campaigns. However, street-tree inventories are incomplete or lacking altogether in most cities. This is partly because conventional field assessment is laborious or expensive. In this study, we developed and validated a novel and cost-effective method to establish a city-wide tree inventory based on computer vision and freely available street view images (SVIs). Tree information such as species, height, crown diameter, and geographical coordinates at the individual tree level can be assessed. Based on an object detection model, we adopted a species-based loss function to address the challenges of long-tailed class distribution of species, which is caused by imbalance among sample size of different tree species and can lead to poor performance of the model. Compared with other research in urban tree species recognition, the modified model shows a higher accuracy. In order to calculate quantitative features of street trees, we employed a deep learning algorithm, which is pretrained on stereo dataset and validated on Google Street View images, to estimate the depth of each pixel in SVIs. Furthermore, as a demonstration, we established the citywide tree inventory and conducted tree diversity analysis for Jinan, China. Compared with new developed area, the old town has more street trees and more diverse tree species which can improve biodiversity and walkability. We also found that plane trees, which can cause allergic reactions, are dominant in northern new developed urban area.

1. Introduction

1.1. Significance of street trees and street tree inventory

Urban street trees are recognized as indispensable elements in cities (Benninger, 2002; Ellis, 2002; Schuyler, 1986) and benefit people and our environment in multiple aspects (Johnson, 2006). Planting and maintaining street trees brings not only direct benefits such as comfort (Lu, Sarkar, & Xiao, 2018; Y. Yang et al., 2019), shade (Chen et al., 2019), aesthetics (Pauleit, 2003), and better air quality (Grundström & Pleijel, 2014; McDonald et al., 2007) but also indirect benefits, e.g., social and cultural values (Wolf et al., 2020) and mental and physical health (Nehme, Oluyomi, Calise, & Kohl III, 2016; Pataki et al., 2021; Ulmer et al., 2016; Wolf et al., 2020). However, rapid global urbanization and increasing urban populations have become a threat to urban ecology and are leading to a decline in urban street trees (McKinney,

2002; Savard, Clergeau, & Mennechez, 2000).

The benefits of urban trees vary across various sizes, species, locations and other individual parameters of the trees (Escobedo, Kroeger, & Wagner, 2011; Escobedo & Nowak, 2009). For example, different tree species have different capabilities to mitigate air pollution and the greenhouse effect (Gillner, Vogt, Tharang, Dettmann, & Roloff, 2015). Different species of trees are also associated with different species of birds, small mammals and pests, which are important to biodiversity and the maintenance of urban ecological systems (Raupp, Cumming, & Raupp, 2006; Watson & Adams, 2010). Therefore, it is necessary to quantify individual parameters of the trees rather than just focusing on their numbers.

An accurate and timely updated tree inventory containing individual tree-level parameters is also critical for sustainable urban development and planning for three reasons (D. Li, Ke, Gong, & Li, 2015; Pu & Landry, 2012). First, it provides indispensable information for urban greenery

---

management and planning activities, such as risk management, landscape planning, precious and ancient tree protection, and pest and disease prevention (Song, He, & Zhang, 2019; Xiao, Ustin, & McPherson, 2004). Second, it can be a valuable data source for researchers from multiple subjects, such as biology, environmental science, urban studies and even economics (Nielsen, Östberg, & Delshammar, 2014). Among these subjects, individual and collective tree information such as species, sizes and amounts can be used to evaluate ecosystem service, biomass and carbon stock, climate change, and their impacts on human behaviors (McPherson et al., 1997; D. J. Nowak, Crane, & Stevens, 2006; Sjöman, Östberg, & Bühler, 2012). Third, a transparent and publicly accessible tree inventory can promote public interest and engagement in tree planting and preservation (Alpan & Sekeroglu, 2020). Existing studies highlighted the importance of bringing citizen science into the process of tree inventory, which means that the public can participate in the process of creating and maintaining tree inventory by providing timely information on individual trees (Roman et al., 2017).

1.2. Methods used to collect data for tree assessment

There are four common approaches to conduct urban street trees assessment, including questionnaires, field audits, using cameras and advanced sensors, and citizen science. Different approaches have different levels of precision and cost, time, and labor consumption (Fassnacht et al., 2016). Questionnaires usually collect information through local institutions or local residents, which is cost-efficient but less accurate (Ewald, 2001; Schaminée, Hennekens, Chytry, & Rodwell, 2009). Field audits, which are usually conducted by botanists or trained staff (Martin, 2011) and directly examine each part of a tree, such as its leaves, flowers, trunk, and fruit (Wäldchen & Mäder, 2018), are laborious and time-consuming but are more accurate, robust and complete.

Benefiting from the advanced algorithms and computational power, the dataset collected via cameras and sensors such as LiDAR are now popular in automatic tree assessment (M. Li & Yao, 2020; Lu, 2019; Nielsen et al., 2014) and has shown potential in tree species identification (Song et al., 2019; Sothe et al., 2019; Wäldchen & Mäder, 2018). Due to its high accuracy and feasibility, LiDAR (Bauwens, Bartholomeus, Calders, & Lejeune, 2016, 2016; Sankey, Donager, McVay, & Sankey, 2017), hyperspectral camera (Anderson et al., 2008; Sankey et al., 2017; Sothe et al., 2019) and other advanced sensors (Abdollahnejad & Panagiotidis, 2020; Amiri, Heurich, Krzystek, & Skidmore, 2018; Leckie et al., 2003) have been applied in forestry inventory studies widely.

However, considering about cost-effectiveness, normal RGB camera (Branson et al., 2018; Culman, Delalieux, & Van Tricht, 2020) is still the most affordable sensors for collecting data for large-scale tree assessment in urban environment. These cameras can be satellite-based, aerial-based or terrestrial-based. Satellite data are suitable for large-scale tree assessment (Ryherd & Woodcock, 1990) and identify trees based on the spectral radiance and surface reflectance from different species of trees. However, this approach can only provide images from one angle and the resolution is relatively low. Aerial-based sensors, especially via unmanned aerial vehicles (UAV), have unique advantages over satellite-based sensors because of high resolution, cost-efficiency and flexibleness (Colomina & Molina, 2014; Dandois, Olano, & Ellis, 2015; Jaakkola et al., 2010; Sankey et al., 2017; Sothe et al., 2019; Torresan et al., 2017). However, it is extremely time, labor-consuming to conduct city-wide assessment. Terrestrial-based data (Bauwens et al., 2016; Liang et al., 2016; Y. Lin & Herold, 2016) also attracted attention for their high-resolution capability, feasibility, and an approximate human perspective. Terrestrial vehicles can also access urban areas where UAVs and other aircraft are restricted. Researchers can obtain terrestrial tree data by using sensors on ground vehicles, such as terrestrial laser scanning (TLS) (Liang et al., 2016). They can also obtain freely available street view images (SVIs) from online mapping servers such as Google and Baidu (Branson et al., 2018). As a publicly accessible resource, the acquisition of SVIs does not require any priority

investments in equipment or labor cost. SVIs therefore have enormous potential for use in assessing street trees. Indeed, one study reported 93% agreement between trees identified in field audits and those identified in SVI audits (Berland & Lange, 2017).

Utilizing crowd-sourced data and public participation, citizen science is also increasingly used for urban street tree data collection especially in cities or communities lacking of resources and labors to conduct inventories (Crown, Greer, Gift, & Watt, 2018; Hauer et al., 2018). This approach is widely used in tree assessment including species recognition (Bancks, North, & Johnson, 2018; Bloniarz, 1996; Roman et al., 2017) and quantitative feature measurement (Bancks et al., 2018). Compared with expert audit, citizen science is low-cost and flexible, which can also be combined with other data sources such as SVIs (Berland, Roman, & Vogt, 2019). Some recent research has confirmed the accuracy and feasibility of citizen science in collecting urban tree information (Bancks et al., 2018; Bloniarz, 1996; Cozad, McPherson, & Harding, 2006; Roman et al., 2017).

1.3. Machine learning application on tree classification

There has been increasing interest in using machine learning techniques to classify tree species. Combining machine-learning techniques with various data sources creates opportunities for the large-scale automated classification of tree species (Colomina & Molina, 2014; Nielsen et al., 2014). The early researchers employed traditional machine learning methods, including k-nearest neighbors ($K-$NN) (Prasad, Peddoju, & Ghosh, 2013), random forest (RF) (Caglayan, Guclu, & Can, 2013) and support vector machine (SVM) (X.-M. Ren, Wang, & Zhao, 2012), to classify trees based on single or multiple pre-specified features which extracted manually.

In order to fully automatically classify tree species, advanced machine learning techniques such as deep learning-based computer vision techniques are introduced (LeCun, Bengio, & Hinton, 2015). These approaches can automatically extract features from imagery and classify them based on the feature map. Recent advancements in deep learning algorithm and computational power made it feasible to assess urban street trees with both RGB images and imagery collected with advanced sensors such as LiDAR. For instance, convolutional neural network (CNN), has been widely applied to detect and classify trees on satellite-based and aerial-based imagery (Ardila, Bijker, Tolpekin, & Stein, 2012; Balková, Bajer, Patočka, & Mikita, 2020; Pu, Landry, & Yu, 2018; Sankey et al., 2017; Schiefer et al., 2020). Also, two studies employed this approach on SVIs and other terrestrial-based images to assess urban vegetation, which focused on plants in home gardens and farmland, and achieved a 0.553 mean average precision (mAP) for seven tree species (Ringland, Bohm, & Baek, 2019; Ringland, Bohm, Baek, & Eichhorn, 2021).

Challenges such as backlighting, blocking issues, and varied viewpoint and object scales (Fig. 1) can lead to a high intra-class variance, while the similar shapes and colors can lead to a low inter-class variance. This can pose threat on the efficiency of the models. Therefore, more advanced and specified algorithms with sufficient and high-quality labeled dataset are necessary for training a fine-grained object detection model.

1.4. Fine-grained object detection models

Most object detection models focus on detecting objects that differ significantly with others, such as detecting vehicles, pedestrian and trees on the street. Fine-grained object detection tasks, which require localizing and classifying similar objects into more specific classes, can be challenging.

Region based Convolutional Neural Network (R-CNN) models (including R-CNN (Girshick, Donahue, Darrell, & Malik, 2015), Fast R-CNN (Girshick, 2015) Faster R-CNN (S. Ren, He, Girshick, & Sun, 2015)) and You Only Look Once (YOLO) (Redmon, Divvala, Girshick, &

Fig. 1. Examples of object detection challenges in SVIs: (a) backlighting, (b) obstructions by other objects, (c) varied object scales.

Farhadi, 2016) are two common series of models for fine-grained object detection. R-CNN divide object detection task into localization step and classification step. Using a region proposal algorithm, the models extract the most possible object locations and pass them to the classifier. Based on R-CNN, Fast R-CNN adds region of interest (ROI) pooling layer before the fully connected layer to reduce the processing time. Faster R-CNN employs a Region Proposal Network (RPN) to generate ROI, which achieves a 250 times improvement over R-CNN and 10 times improvement over Fast R-CNN.

Different from R-CNN series, YOLO (Redmon et al., 2016) divides image into G*G grids (the number of columns and rows are same) and detects objects in all the grids in one step. This straightforward pipeline improves the training efficiency significantly compared with R-CNN series and makes it possible for real time detection. This series is evolving and has released seven versions (from YOLOv2 to YOLOv7) until now.

1.5. Quantitative measurement by computer vision techniques

Estimating the object size is one fundamental task in computer vision. There are mainly three ways to measure object size from 2D pictures. One is geometric approaches (Benosman, Manière, & Devars, 1996; Mancini et al., 2013; Mur-Artal, Montiel, & Tardos, 2015; Zou & Li, 2010) which extract information of an object from multiple images containing the same object based on geometric constraints. Although these methods are accurate and efficient, they require one object appear in a series of images which is unfeasible when images are not abundant.

Another approach is to use advanced sensors such as LiDAR, which obtains the depth from object to the sensor and calculate other quantitative features based on this depth. These sensors are widely used in construction, robotic, unmanned driving and forestry industry. Google also has used LiDAR to provide depth images for their SVIs. However, these sensors are expensive and data processing is time-consuming, so are difficult to be applied in low-cost industries.

Recently, with the development of computer vision technology and the prevalence of consumer-grade monocular cameras, computer scientists have focused on estimate distance on single view images with deep learning methods, which is inspired by animals who can estimate distance based on their prior experience. Various algorithms with promising results were developed (Aleotti, Tosi, Poggi, & Mattoccia, 2018; Chakravarty, Narayanan, & Roussel, 2019; Garg, Bg, Carneiro, & Reid, 2016; Wang, Pizer, & Frahm, 2019).

1.6. Critical research gaps

To date, the importance of tree inventory has been underestimated by the public and some authorities, and only a few cities worldwide have completed citywide street tree inventories since large-scale tree assessments are expensive and laborious. The applications of computer vision techniques and SVIs can automatically assess trees in large scale with low cost. Some studies have already applied SVIs to obtain quantitative

features and species of street trees through deep learning techniques (Branson et al., 2018; Choi et al., 2022; Kang, Lee, & Zou, 2021; Laumer et al., 2020; M. Li & Yao, 2020; W. Wang et al., 2018). However, all studies using Google Street View (GSV) where researchers can access depth information from corresponding depth imagery. For countries and regions without GSV, such as mainland China, street view images from other platforms such as Baidu and Tencent do not have publicly access to such depth imagery. In addition, several challenges of fine-grained object detection in SVIs can undermine the efficiency of their classifiers, such as high intra-variance and low inter-variance of street tree species and long-tailed distribution problems.

In this study, we propose a fine-grained object detection model to assess citywide street trees from freely available SVIs without depth imagery. This method can assess not only quantitative parameters such as geographic coordinates, canopy and trunk diameters and height but also qualitative attributes such as species for each street tree in a city.

To estimate depth from single view SVI, we employed Monodepth2 (Godard, Mac Aodha, Firman, & Brostow, 2019), a novel deep learning model which can estimate 3D information based on 2D pictures. This model is pretrained in KITTI stereo dataset (Geiger, Lenz, Stiller, & Urtasun, 2013) which contains stereo imagery in the scenes of urban streets. The depth assessing method was validated with depth map in Google SVIs. To deal with high intra-variance and low inter-variance, we presented a new muti-category dataset based on SVIs which aimed at urban street tree detection and fine-grained classification. Considering the long-tailed distribution challenge, we also modified the loss function of YOLOv5 to improve the performance.

Using our proposed method, street tree inventory was established for the urban area of Jinan, China. Based on street tree inventory, some spatial analyses were conducted. We found that new development area has lower tree density and species richness compared with old town. A large part of new developed area is dominated by oriental plane trees which can pose allergy threat on citizens' health and the walkability of streets. Government is suggested to added more tree species to new developed area and replaced plane tree with other species.

2. Methodology

2.1. Study area

Jinan, a medium-sized provincial capital city, is representative of many urban areas in northern China. Jinan had a population of 9.20 million in 2020, and 3.76 million residents were living in inner city areas, including five districts: Lixia, Shizhong, Huaiyin, Tianqiao and part of Licheng (China National Bureau of Statistics, 2020). In this study, the study area is limited to these five urban districts, with a total area of 258.27 km^2 (Fig. 2) (China National Bureau of Statistics, 2020). Jinan has a continental monsoon climate with broad-leaved deciduous urban trees. The most common street trees in Jinan city are oriental plane, poplar, willow, locust tree, cypress, boxwood, wax tree and pine, which account for >90% of the whole street tree inventory (Jinan Forestry
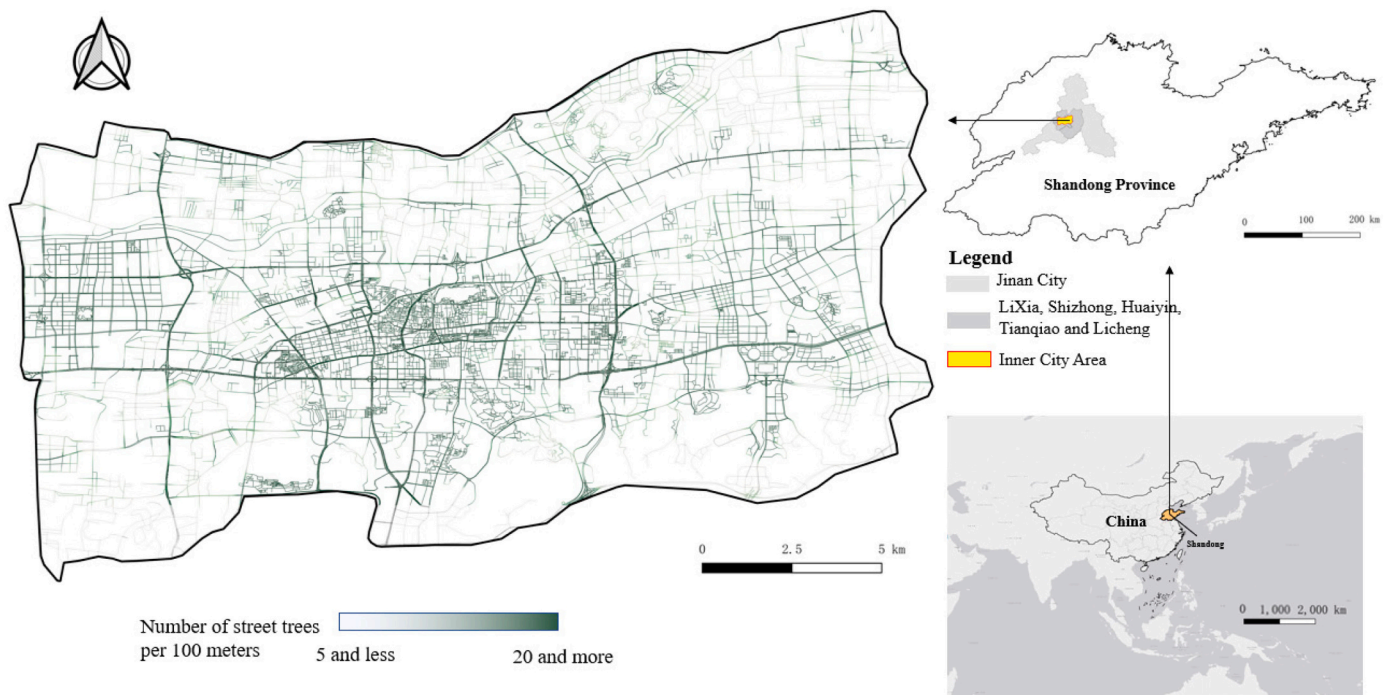
Fig. 2. The inner city of Jinan, China was selected as the study area. All street segments with SVIs were selected. The darker color means where more street trees were detected.
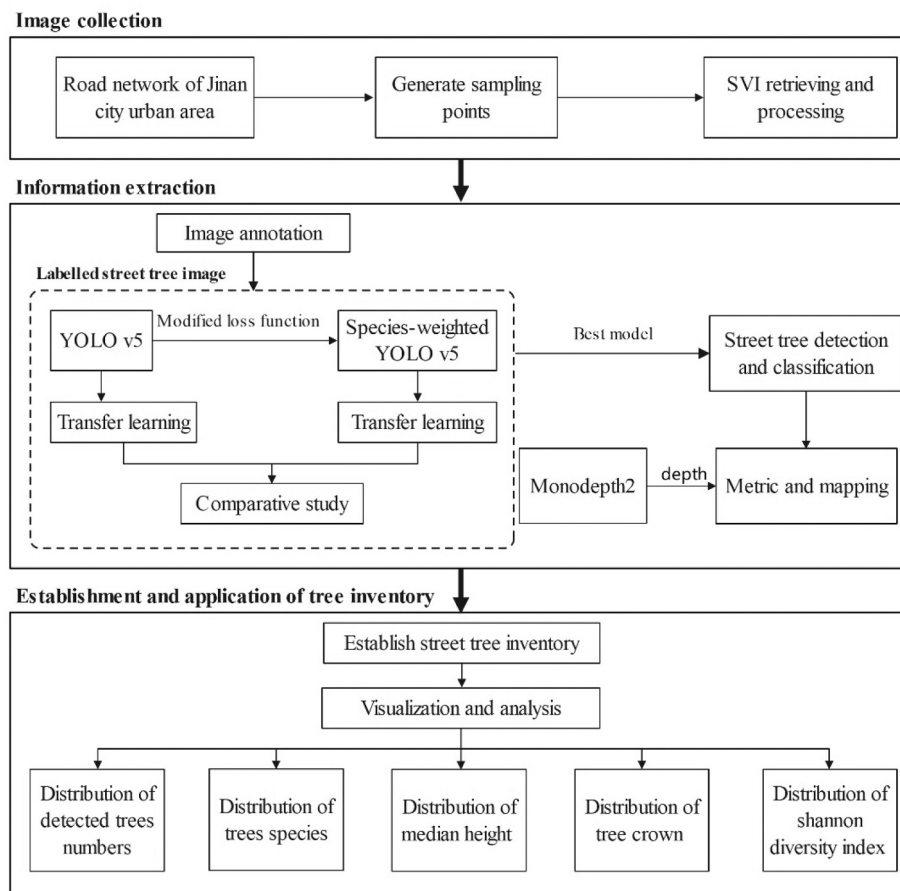

Fig. 3. The process flow of establishing a citywide street tree inventory from SVIs, which involves three steps: 1) collecting SVIs, 2) extracting tree information, and 3) establishing tree inventory.

Bureau, 2007) Therefore, these eight species are chosen as the tested tree species in our study.

2.2. Overall workflow

In this study, we developed and validated an improved method to automatically establish a city-wide tree inventory using SVIs and computer vision (Fig. 3). The workflow includes three steps: 1) collecting SVIs, 2) extracting tree information, and 3) establishing a tree inventory (Fig. 3).

2.2.1. Collecting SVIs

Baidu Street View and Tencent Street View are the two most popular SVI providers in China. Tencent SVIs provide only one image for each shotting point without a timeline, so we cannot ensure that all SVIs collected for tree inventory are in the same season. Therefore, this study chose Baidu Street View as the data source, which provides SVIs at multiple times and different seasons for each shotting point. All SVIs used for the tree inventory are shotted between March and August. The SVIs of the whole inner city were retrieved from the Baidu Street View API (Baidu, 2021). First, we constructed the street network of the study area with data provided by the local government in 2021 (Jinan City Planning Bureau, 2021).

Then, SVI sampling points were generated based on the threshold distance (T) along each street in our study area. The direction of sampling points depends on the topology and width of each street to avoid excessive or incomplete coverage along the street (Fig. 4). For each Baidu SVI, the heading parameters represent the horizontal directions of the camera, and 0°, 90°, 180°, and 270° represent the front, right, back and left sides of the shooting vehicle, respectively. Some roads have two lines of SVIs sampling points and others have only one line. We classified all streets into three types depending on the distribution of sampling points and the existence of middle green belt (Fig. 4). Type A streets have only one line of sampling points, so we retrieved both 90° and 270° directions for each sampling point to detect trees on both sides of the road. Type B streets have two lines of sampling points without greenbelts, so we retrieved only the 90° direction for each sampling point on both lanes. Type C streets have two lines of sampling points with

greenbelt in the middle, so we retrieved both 90° and 270° directions for one line and only 90° direction for another line to avoid repeated counting of street trees in the greenbelt.

We calculated the threshold distance (T) between street segments according to the distance (D) from the camera and curb, as shown in Fig. 4. As the range of view is 45° for every SVI, then T can be calculated as:

$$T = 2*D \times tan45° \tag{1}$$

where D is the distance from the camera to the curb estimated by a deep learning model in Section 2.4. Then, the closest segment distance was selected based on this threshold distance.

For the Baidu Street View, a total of 24 image tiles (512*512 pixels for each tile) for each sampling point were directly retrieved from the Baidu API, and the 24 image tiles were stitched together to create images with different horizontal directions, such as 90° and 270°. Each image is constructed by three tiles (1536 pixels) in the vertical direction and two tiles (1024 pixels) in the horizontal direction (Fig. 5).

Finally, a total of 185,830 SVIs were retrieved for this study. Among them, 2700 images (1350 images are street trees under built environment background and 1350 images are street trees under natural environment background (Fig. 6)) were labeled and randomly split into training, validation, and testing sets with proportions of 70%, 20% and 10%, respectively. The selected dataset covers a wide range of aspects, including context (urban, suburban and country), vegetation density (sparse and crowded), weather and lighting conditions.

2.2.2. Ground truth labeling

Because we focused on street trees, other plants (grass or bush) or nonstreet trees should be excluded from the tree assessment. To achieve this, SVIs containing only nonstreet trees and other plants were included in the training set and marked as unannotated images. To improve the robustness of the training model, street trees that are obscured objects (e.g., fences, vehicles, traffic signs) were also included in the dataset. The crown is the most distinguishing part of a tree, but computers tend to regard several trees as one tree if we label only the crown. Hence, both the crown and trunk of a tree were annotated with the same bounding
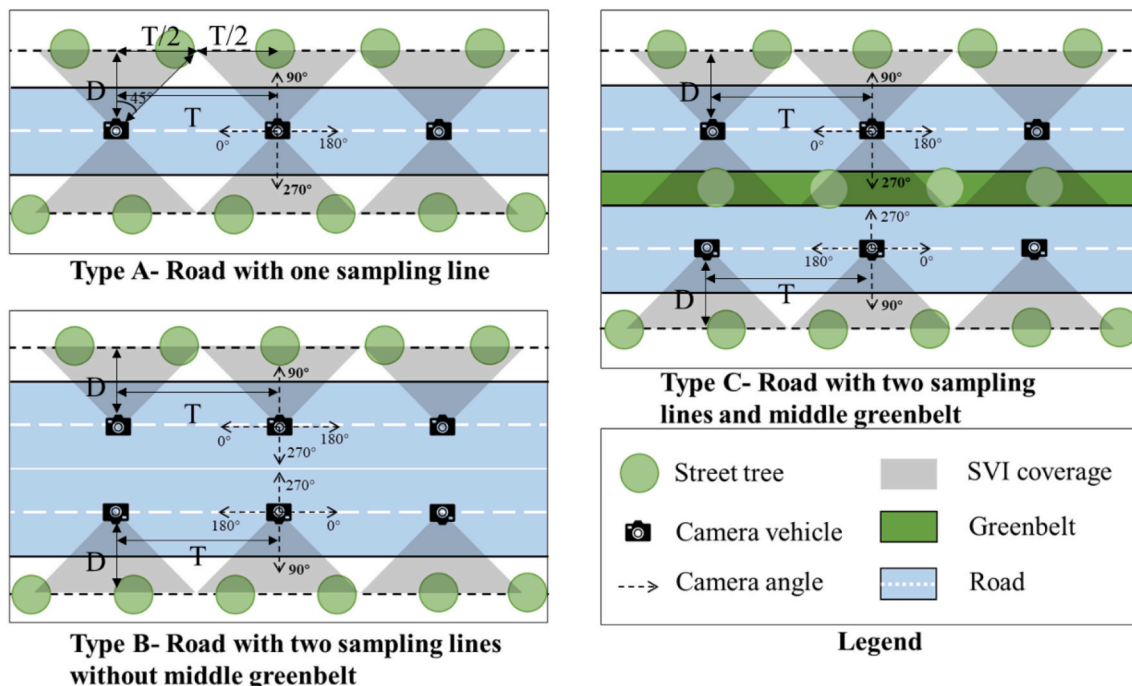

Fig. 4. Select sampling points and directions based on the type of roads and the distance between trees and cameras.

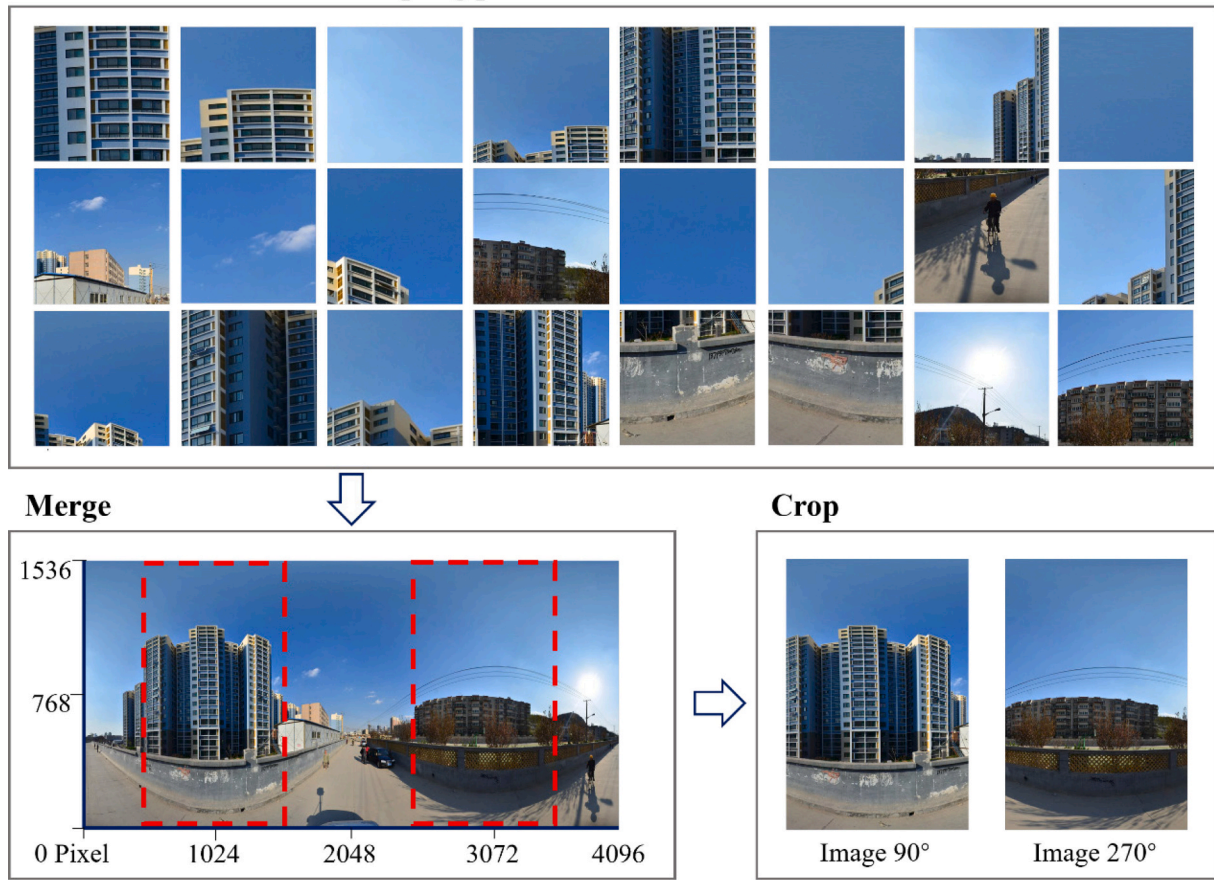## Retrieve tiles from each sampling point

Merge

Crop

Fig. 5. Retrieving and transforming Baidu SVIs.

box. An opensource labeling tool (Roboflow, 2022) was chosen as the labeling platform. Two research assistants in Jinan with adequate training on local trees labeled the nine classes (eight species and "other trees") of street trees in 2573 SVIs at all. Each of them labeled each SVI once and validated it in the field. Then, the mismatched SVIs are selected for further validation. The labeled SVIs were then divided into training, validating and testing datasets at a ratio of 7:2:1. Fig. 6 shows some labeled training images with different colors representing different species.

2.2.3. YOLOv5 network

Considering both efficiency and robustness, we selected the YOLOv5 model, a one-step real-time multi-object detection algorithm (Redmon et al., 2016) as the base model for tree identification. Compared with other popular object detection algorithms, YOLOv5 has a higher inference speed and higher accuracy in object detection (Redmon & Farhadi, 2017). YOLOv5 also performs better at the detection of small or varied-scale objects and overlapping objects, e.g., street trees. These advantages are essential for this study because of the varied scales of street trees and the complexity of the urban environment in SVIs.

Fig. 7 shows the architecture of YOLOv5. The YOLOvs5s network includes five parts, which are input, backbone, neck, head, and output, and each part consists of different network layers with specific functions. The first part is the backbone, which is used to extract features from input images. The second part is neck, which is used to mix and combine features from backbone logical groups and transmit them for detection. The final part is the head, which is used to detect the object from the feature maps and output a vector with the class probability and the position and size of the objects.

2.2.4. Class-balanced loss function

In SVIs, street trees exhibit a long-tailed class distribution, where a small portion of dominating tree species have a relatively larger number of samples, while many other species have a relatively smaller number of samples. As the training dataset is generated from SVIs, such species imbalance of the training dataset can pose several problems in model training. The lack of tail-species samples makes it difficult to learn intraclass variance and can cause overfitting problems. The dominance of the head class leads to potentially biased feature spaces and decision boundaries so that the classifier tends to identify trees as head-class species instead of tail-class species.

One potential solution to this challenge is to replenish tail-species with images from extra sources. However, trees in non-street-view images (non-SVIs) have different distributions and backgrounds from those in SVIs, and the training model works best when trained on images from a similar distribution as the test images.

Instead, we modified the classification loss function of the YOLOv5 model to mitigate the long-tailed species distribution problem. The loss function of YOLOv5 is composed of three parts: bounding box regression loss, the confidence of object presence loss and the classification loss. The original classification loss function of YOLOv5 is the Cross-Entropy (CE) loss function, which can be written as:

$$CE_{softmax}(z, y) = -log\left(\frac{exp(z_y)}{\sum_{j=1}^{C} exp(z_j)}\right) \quad (2)$$

where z_y is the predicted probability for species y, and C is the total number of species.

Fig. 6. Samples of labeled images with different colors represent different species. a, b & c are street trees in built environment and d, e & f are street trees in natural environment.
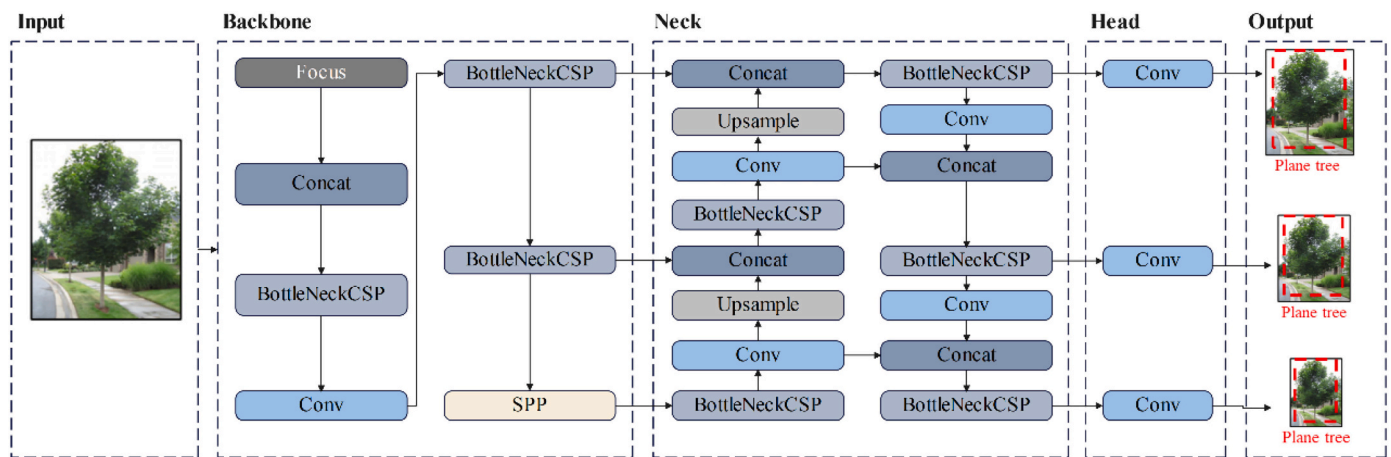

Fig. 7. Overview of the architecture of YOLOv5.

The species-weighted loss function introduces a weighting method inversely proportional to the effective number of samples of each species (Cui, Jia, Lin, Song, & Belongie, 2019); then, the model increases the probability of small classes, decreases that of large classes and closes the predicted results to the truth (Fig. 8). Suppose species y has N_y training samples and the total number of training samples is N. Then, the species-balanced (SB) cross-entropy loss can be written as:

$$SB_{softmax}(z, y) = -\frac{\left(1 - \frac{(N-1)}{N}\right)}{\left(1 - \frac{N_y - 1}{N_y}\right)} log \left(\frac{exp(z_y)}{\sum_{j=1}^{C} exp(z_j)} \right) \tag{3}$$

Head Class: Locust Trees   **Tail Class: Pine Trees**
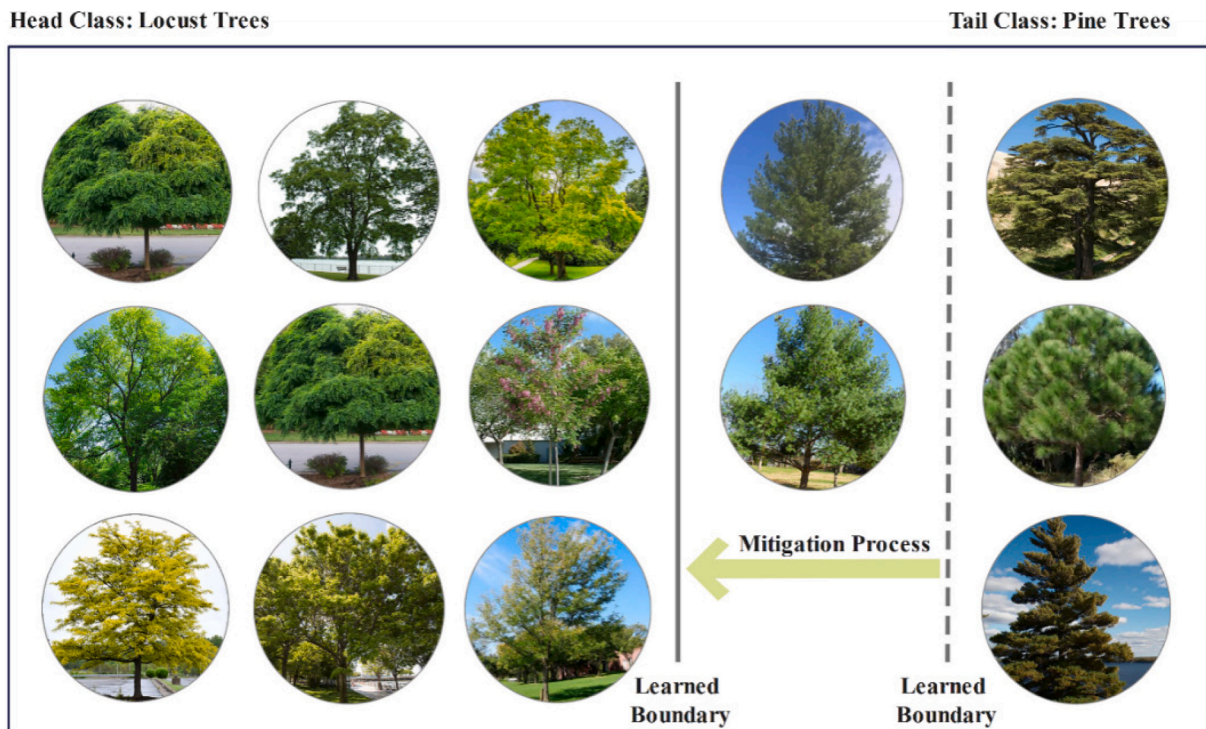

Fig. 8. By modifying the loss function, the learned boundary of the trained model is closed to the true boundary.

2.2.5. Metrics and mapping of street trees

The SVIs annotated by the object detection model are further processed to estimate each tree's location, height, and canopy diameter (Fig. 9). The distance between each pixel on the SVI and the camera is essential to the estimation of tree metrics. Hence, we employed Monodepth2 (Godard et al., 2019), a deep learning method that can generate scene depth maps based on 2D images, to estimate the distance (D) between each street tree trunk and the camera from the SVI. The model
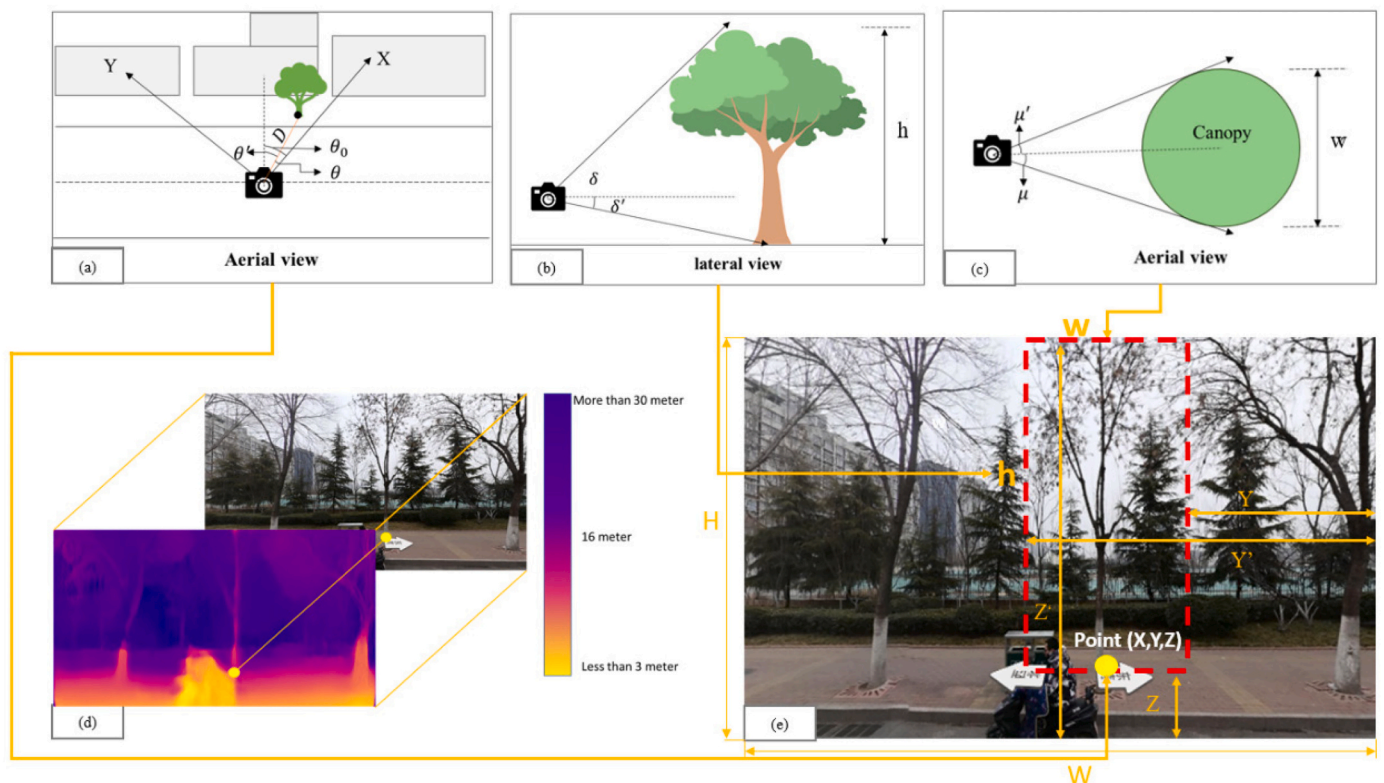

Fig. 9. Calculating (a) geographic coordinates, (b) tree height, and (c) tree canopy diameter. Calculating (d) depth layers and (e) tree information from street view images.

is pretrained on KITTI stereo dataset (Geiger et al., 2013), which is the commonest benchmark and training dataset for depth estimation contains real street view images with corresponding LiDAR imagery. Then, we used the relative azimuth angle and distance between the tree and the camera location of the SVI in this tree to estimate a tree's location. The coordinates of the detected tree (x, y) will then be calculated using Eq. (4):

$$x = x' + D^* \cos\theta, y = y' + D^* \sin\theta \tag{4}$$

where (x', y') is the camera coordinate; D is the distance from the camera to the bottom of the tree; and θ is the yaw angle between the camera and the bottom of the tree, which equals $(\theta' + \theta_0)$ (where θ_0 is the default SVI tile yaw-angle and θ' is the yaw angle in depth image space).

We further estimated tree height (h) and canopy diameter (w) using Eqs. (5) and (6), respectively. The canopy diameter represents the width of crown from the perspective of camera.

$$h = D^* (tan\, \eth' - tan\, \eth) \tag{5}$$

where $\eth = 180°/H*(Z-H/2)$ and $\eth' = 180°/H*(Z'-H/2)$ (where H is the total number of pixels in the height of the image; Z is the number of pixels from the bottom of the tree to the bottom edge of the image; and Z' is the number of pixels from the top of the tree to the bottom edge of the image).

$$w = D^* (tan\, \mu' - tan\, \mu) \tag{6}$$

where $\mu = 360°/W*(Y-W/2)$, and $\mu' = 360°/W*(Y'-W/2)$ (where W is the total number of pixels in the weight of the image; Y is the number of pixels from the right end of the crown of the tree to the right edge of the image and Y' is the number of pixels from the left end of the crown of the tree to the right edge of the image).

As some SVIs overlap in streetscape coverage, some trees may be identified in two images. We use geographical coordinates to remove any duplicated trees (a tree within 0.5 m of another tree) in the inventory.

2.2.6. Model evaluation

2.2.6.1. Evaluation for tree recognition and classification model. To validate the accuracy and efficiency of the model and the whole process, we compared the performance of the detection model with the species-balanced loss function (Model A) and without the species-balanced loss function (Model B). We assessed two models on the testing dataset and calculated the precision, recall and average precision for comparison (Szegedy, Toshev, & Erhan, 2013).

Precision measures the ratio of relevant instances among the retrieved instances, while recall measures the ratio of relevant instances that were retrieved. In an object detection task, increasing precision can lead to a decrease in recall. Average precision (AP) measures the intersection area of prediction and ground truth boundaries divided by the

union area of prediction and ground truth boundaries, which can balance the precision-recall tradeoff and has been widely employed to assess the performance of object detection models. This study uses mean AP at IoU (intersection over union) threshold 0.5 (mAP0.5) as the major metric to evaluate the performance of the model.

The performance of our new model is promising (Table 1). In this study, to ensure the precision of information in tree inventory, more weight is assigned to precision than recall. Therefore, our model fitness is set as F = 0.8*precision+0.2*recall. We calculated precision, recall and average precision on the test SVI set for each species. The model with the species-balanced loss function (Model A) performed better (0.587 mAP0.5 and 0.715 highest AP0.5) than the model without the species-balanced loss function (Model B) (0.536 mAP0.5 and 0.710 highest AP0.5).

In general, species with larger sample sizes are easier to be identified. Locust trees, plane trees and willows have the best results. Locust trees and plane trees have 335 and 407 samples, respectively, and willows have a distinguishing appearance compared with other species. Although withered trees have 435 labeled samples, the test result is not as high as expected. The large variance in the appearance of withered trees may hinder the detection performance.

The performance of the model with the species-balanced loss function improved for nearly all species to varied extents. The most prominent improvements are on the recall of cypress and pine tree. These two species only have 93 and 96 labeled samples for training. This indicated that including a species-balanced loss function can effectively mitigate problems caused by long-tailed class distributions.

We compared the performance of our models with existing studies on street tree species classification with SVI as the only data source (Table. 2). Our model A exceeds both two previous studies in mAP0.5 (Branson et al., 2018; Choi et al., 2022).

2.2.6.2. Evaluation for quantitative assessment. The accuracy of depth estimation is critical because all other quantitative metrics are derived from it. To evaluate the performance of depth estimation, we use five evaluation indicators: RMSE, RMSE log, Absolute Relative difference, and Squared Relative difference and Accuracies (Eigen, Puhrsch, & Fergus, 2014).

Google API provides depth map corresponding to each SVI, which was obtained by LiDAR sensors and can be used as ground truth. We test

Table 2
Comparison of street tree species classification models' performance between our studies and existing studies using SVIs as the only data sources.

	mAP0.5	region
(Choi et al., 2022)	0.564	East Asia
(Branson et al., 2018)	0.581	North America
Model A	0.587	East Asia
Model B	0.536	East Asia

Table 1
Comparison of Model A (with the species-balanced loss function) and Model B (without the species-balanced loss function) for the training results.

Species	Labeled samples number	Model A			Model B		
		Precision	Recall	Average precision	Precision	Recall	Average precision
Locust tree	335	0.796	0.543	0.693	0.784	0.541	0.684
Willow	97	0.837	0.603	0.715	0.747	0.612	0.71
Poplar	150	0.631	0.513	0.528	0.52	0.464	0.471
Pine tree	96	0.697	0.515	0.528	0.536	0.326	0.331
Planes	407	0.602	0.719	0.623	0.594	0.722	0.611
Wax tree	98	0.583	0.509	0.521	0.581	0.459	0.503
Boxwood	525	0.639	0.582	0.594	0.634	0.583	0.593
Cypress	93	0.517	0.496	0.497	0.434	0.309	0.386
Mean (all)				0.587			0.536
Mean (built environment)				0.603			0.579
Mean (natural environment)				0.571			0.493

the accuracy of pretrained Monodepth2 on 741 Google Street View images. The evaluation outcome is compared with other state-of-the-art studies in depth estimation (Table 3). Overall, our model shows a satisfactory accuracy of 0.783, which is close to, if not exceeds other models.

3. Case study in Jinan city

3.1. Descriptive statistics

A total of 144,736 street trees were detected from 185,830 SVIs in Jinan city. The data were compared with official survey data conducted by the Jinan Forestry Bureau in 2007 (Jinan Forestry Bureau, 2007). Table. 4 shows that the proportions of eight species and "other trees" in this study are approximately consistent with those of the official field survey.

Quantitative results such as average height and average crown diameter are compared with data from a worldwide tree attribute database (World Agroforestry Centre (ICRAF), 2016). Most assessed results are consistent with the database except the average crown diameter of some species.

3.2. Street tree inventory and applications

All obtained information above was then used to establish the city-wide tree inventory and was visualized on a map based on their accurate coordinates. After the establishment of the tree inventory, we analyzed the street tree diversity of the study area.

The street tree inventory can be visualized in different ways. Fig. 10 shows the detected street trees around the Huancheng Park area. Each circle represents one tree, while the color of each circle represents the species of the tree, and the circle size represents the crown diameter of the tree. Street trees in this area have diverse species and crown diameters. Willows and boxwoods are two major species there.

Furthermore, the data can be aggregated and visualized in any spatial unit, e.g., 1 km *1 km grids (Fig. 11). The inner city of Jinan consists of the old town and the new development area surrounding the old city (Fig. 11). The old town features dense and narrow streets and low-rise buildings, while the new development area is dominated by high-rise buildings. Another significant boundary in this area is Jingshi Road. The north side of the road is a highly populated plain area, while the south side mountainous area has the most high-quality and low-density residences.

Fig. 11a reveals that locust trees, boxwood and plane trees are the three major species of street trees in the whole study area. We also observed spatial disparity in street tree provision within Jinan. The old town is dominated by boxwood, while the new development area is dominated by plane trees. The old town had the highest number of detected street trees and street tree species (Fig. 11b & e). Fig. 11d shows that the median tree height in the new development area is higher than that in the old town. To further understand the street tree diversity of Jinan city, this study introduces the Shannon diversity index as follows (Nolan & Callahan, 2006).

$$H' = - \sum_{i=1}^{s} p_i ln p_i \qquad (7)$$

The higher the value of the Shannon index is, the higher the diversity of species in the specific area. A higher Shannon index illustrates that the old town has more diverse tree species than the new development area (Fig. 11f).

4. Discussion

4.1. Model of fine-grained object detection on street trees

Previous studies have shown promising results when using computer vision technology to conduct image segmentation or object detection on SVIs. However, most studies focused on general element segmentation or coarse-grained object detection, such as for greenery, buildings and pedestrians. Fine-grained recognition tasks, such as tree species, remain challenging for two reasons. First, tree species have low interclass variance and high intraclass variance, especially in SVIs. Second, street tree species usually feature a long-tailed distribution, i.e., most street trees belong to a small number of dominating tree species, while a small proportion of trees belong to many relatively rare tree species. Such a skewed distribution leads to insufficient training data for these relatively rare tree species.

In this study, to cope with the long-tailed class distribution of the training dataset, we modified the classification loss function. We assigned the weights of each species inversely proportional to the its sample size to push the learned boundary more favorable to the tail classes. Such modification leads to significant improvement in identifying most species. Using a smaller training dataset, our modified model obtained similar, if not higher, performance compared with existing studies (Branson et al., 2018; Ringland et al., 2021). We concluded that with appropriate processing and parameters, our novel approach can achieve reasonable accuracy with low cost in establishing citywide tree inventories for many cities worldwide.

4.2. Influences of the characteristics of SVIs

The characteristics of SVIs, such as weather conditions, collection seasons, and the surrounding environment, may influence the accuracy of street tree identification and assessment. For instance, the appearance of street trees, especially deciduous trees, differs significantly in different seasons in temperate climate zones, including northern China. Therefore, all retrieved SVIs, which are used for the establishment of tree inventory, are generated between March and August according to the timeline of Baidu SVIs. In further studies, we will train models to recognize street trees in different seasons.

This study also illustrated that the surrounding environments of SVIs can influence the accuracy of the model. The accuracy of detecting street trees in urban background outperforms that of detecting trees in natural environment background (such as suburban areas) or green spaces (such as natural parks). This result suggests that our method may work well in built environment and that other tools (such as the application of LiDAR) may be more suitable for tree detection in natural environment.

Table 3
Comparison of evaluation results with other state-of-the-art studies in monocular depth estimation with deep learning methods.

Studies	Absolute Relative difference	Squared Relative difference	RMSE	RMSE log	Accuracy ($\delta < 1.25$)
	Lower is better				Higher is better
(Zhou, Brown, Snavely, & Lowe, 2017)	0.208	1.768	6.865	0.283	0.678
(Casser, Pirk, Mahjourian, & Angelova, 2019)	0.109	0.825	4.750	0.187	0.874
(Wang, Wang, Liu, & Chen, 2019)	0.158	1.277	5.858	0.233	0.785
(Yin & Shi, 2018)	0.155	1.296	5.857	0.233	0.793
(Z. Yang, Wang, Xu, Zhao, & Nevatia, 2017)	0.182	1.481	6.501	0.267	0.725
This study	0.170	1.167	5.596	0.247	0.783

Table 4
Descriptive statistics of detected street trees compared with field audit data.

Species	Detected number (proportion)	Field study in 2007 number (proportion)	Height (m) Mean (SD)	Benchmark height (m)	Crown diameter (m) Mean (SD)	Benchmark diameter (m)
Locust tree	15,819 (19%)	21,904 (26%)	9.09 (7.78)	8–24	5.48 (5.34)	~ 8
Willow	5387 (7%)	1661 (2%)	5.71 (5.17)	8–15	4.83 (5.04)	7–12
Poplar	14,219 (17%)	23,563 (28%)	9.48 (7.78)	12–20	5.43 (4.12)	~ 6
Pine tree	1721 (2%)	NA	9.43 (5.72)	3–24	4.34 (5.00)	6–12
Plane tree	31,726 (39%)	29,082 (34%)	5.35 (4.43)	8–30	6.25 (5.48)	5–24
Wax tree	1526 (2%)	6570 (7%)	5.43 (5.72)	3–14	4.34 (5.01)	6–7.5
Boxwood	33,968	NA	0.76 (0.66)	0.5–3	2.38 (5.45)	NA
Cypress	7006 (9%)	NA	5.23 (5.51)	6–25	1.73 (1.32)	2–9
other	4272 (5%)	2742 (3%)	NA	NA	NA	NA
Sum (except boxwood)	81,676	85,522	4.39 (5.57)	NA	3.37 (4.32)	NA

Fig. 10. Distribution of detected street trees around the Huancheng Park area.

4.3. Indications for urban planning and management of street tree inventory

Based on this city-wide tree inventory in Jinan city, this study also conducts some spatial analysis that may provide planning implications for urban planners and city managers.

We conducted six city-wide analyses and data visualization based on tree inventory data. We analyzed the distribution of detected street trees based on number of trees, species and tree height.Fig. 11a reveals the dominant species of each grid. Boxwoods and locust trees are two major species in the old town and southern area, while plane trees occupy the northern part of the newly developed area. Studies (M. Nowak, Szymańska, & Grewling, 2012) found that plane tree pollen is responsible for a high rate of allergic reactions, which have a negative impact on walkability and pedestrian health. This indicated that government officials should reconsider the choices of street trees in newly developed areas. Fig. 11b shows the number of detected street trees in each grid. Some studies found that the quantity of street greenery has a positive association with residents' physical activity and active transportation, which benefits both citizens' health and the urban environment (Huang, Jiang, & Yuan, 2022; Jiang et al., 2021; Lu, 2019; Lu et al., 2018). Density of street trees in old town are higher than that in newly developed areas, which indicates that the government should put more effort into the green infrastructure of newly developed areas. Fig. 11c and Fig. 11d illustrate that street trees in old towns have a higher average height and longer average crown diameter. These results are logical, as street trees are planted later in newly developed areas than in old towns. Increasing tree crown diameter can significantly mitigate the urban heat island effect, so urban planners can choose tree species with larger crowns (Wang and Akbari, 2016). Fig. 11e and Fig. 11f show the number of species and diversity index of each grid. Both of these values are higher in old town areas than in newly developed areas. Studies have found that street tree species diversity has negative associations with crime (J. Lin, Wang, & Huang, 2021) and can benefit both pedestrians and city ecology. These spatial analyses and visualizations can provide intuitive implications that can be further analyzed by advanced
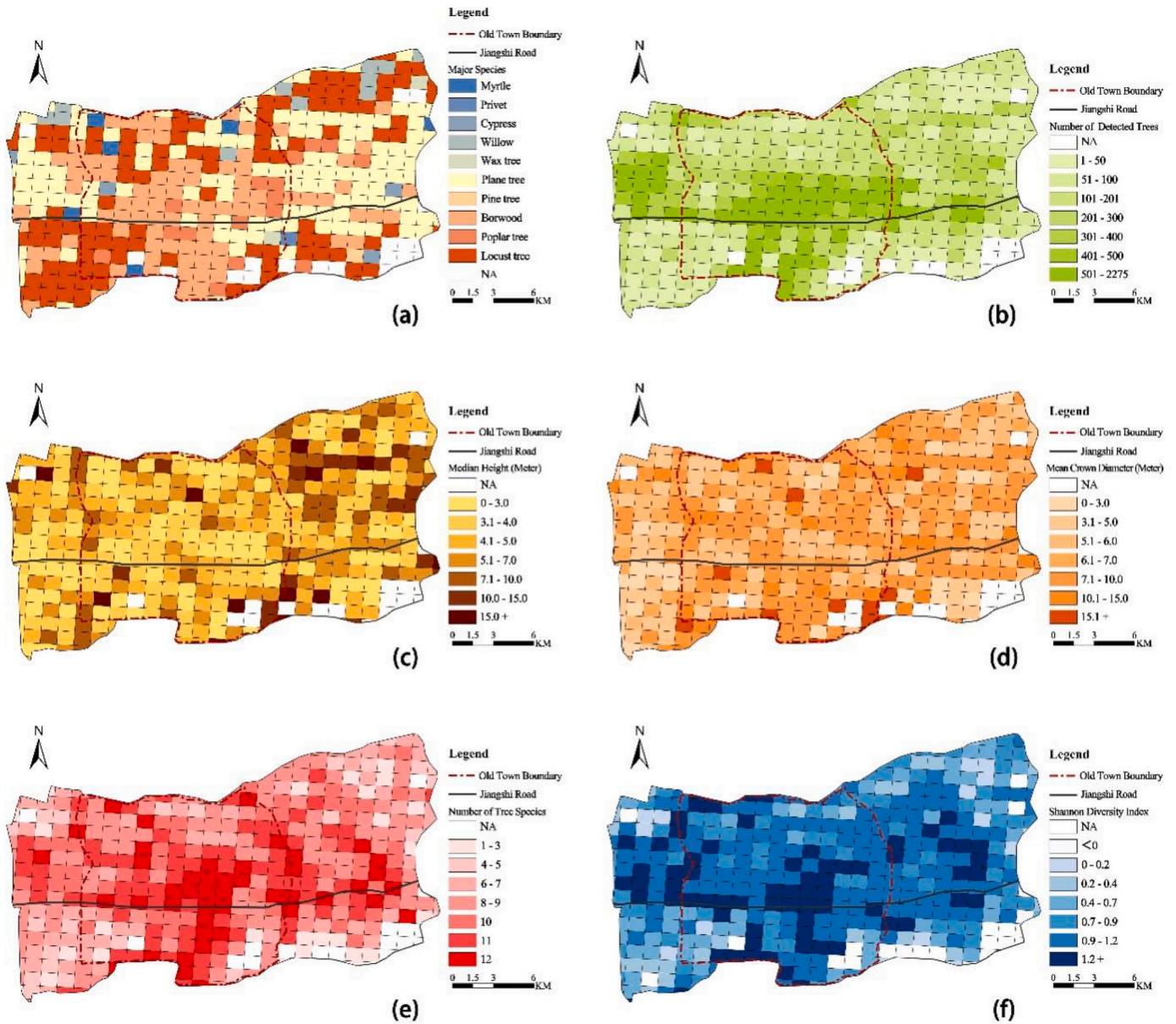
Fig. 11. Street tree inventory visualization of 1 km*1 km grids in the inner city of Jinan; (a) the most species of street trees; (b) the number of detected street trees; (c) the median height of detected street trees; (d) the mean crown diameter of detected street trees; (e) the species richness of street trees; (f) the Shannon diversity index of street tree species.

statistical techniques.

4.4. Limitations

Although we achieved encouraging results, both the model and the pipeline have potentials for improvement.

First, we used SVIs as the sole data source, while SVIs may be incomplete or inconsistent. For instance, Baidu SVIs cannot achieve full coverage of urban roads in some cities. Some studies tried to combine satellite-based and drone-based images. In addition, the SVIs of a city can be taken in multiple seasons (summer or winter), various weather conditions (sunny or rainy) and particular times of the day (daytime or evening), which can affect the appearance of street trees and model accuracy. We tentatively suggest that SVIs used for tree inventory should be taken in summer or spring when the characteristics of most tree species are most distinguishing. Our findings also suggest that Baidu should collect SVIs more frequently and provide more API parameters,

such as weather context and collection times.

Second, the distance between sampling points along each street was calculated based on the distance between the camera and the trees. However, the distances between trees on the middle greenbelt and the cameras are distinct from those on the sidewalk, so it is possible to repeat or miss counting these street trees. Hence, more specific adjustments are needed in future studies.

Third, some tree information, including the height, crown diameter and coordination, was calculated based on the distance from the SVI camera to the street tree. Some studies used depth imagery of Google SVIs (Hebbalaguppe, Garg, Hassan, Ghosh, & Verma, 2017; Ning et al., 2022), which are data layer representing the distance from any pixel to a camera in a SVI (Fig. 12), to obtain the distance between objects and the camera. However, this method is impractical in mainland China because of the lack of coverage of Google Maps. Hence, we estimated this distance by a deep learning model, and such model should be further validated.
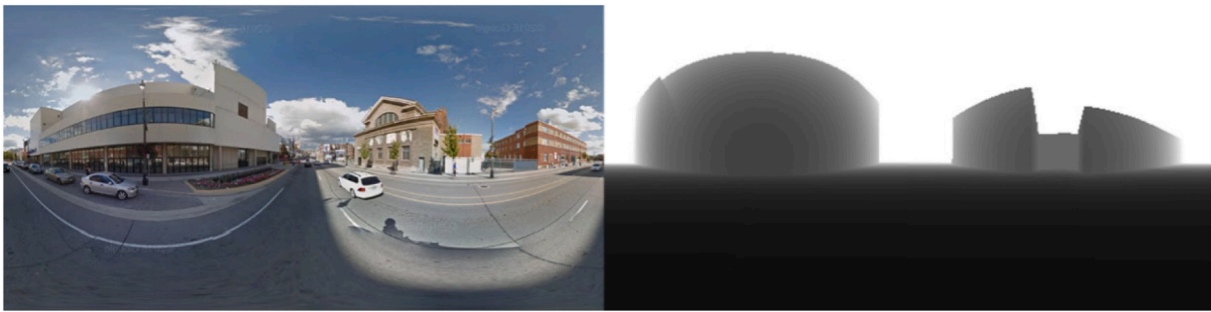
Fig. 12. An example of SVI (left) from Google Maps and its corresponding depthmap (right).

Fourth, although YOLOv5 is an efficient object detection model, some improvement in the model is feasible. For instance, although the species-based weighing function has increased precision and recalls for species with small sample sizes in this study, more advanced approaches, such as head-to-tail transfer learning, may better mitigate the issue of long-tail species distribution (Zhang, Kang, Hooi, Yan, & Feng, 2021).

5. Conclusion

This research proposes an improved method of automatedly detecting, classifying and measuring urban street trees using computer vision approaches and SVIs in large-scale practice. The results are encouraging. Compared with other methods, this approach is cost-effective because SVIs are freely available for most cities in the world. Compared with other similar studies, this research identifies more than eight tree species with SVIs and computer vision algorithms in such a large-scale practice and obtain a relatively high accuracy. We find that a species-based weighted loss function can improve the overall model performance by addressing the long-tail distribution of street tree species. This study is also the first one to adopt deep learning methods to estimate depth for tree inventory and validate its efficiency on SVIs.

Funding

Author contribution statement

Dongwei Liu took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript. Dongwei Liu and Yi Lu designed the pipeline critical methods and the computational framework, and analyzed the data. Dongwei Liu and Yuxiao Jiang collected the data and conduct the critical revision of the article.

References

Abdollahnejad, A., & Panagiotidis, D. (2020). Tree species classification and health status assessment for a mixed broadleaf-conifer forest with UAS multispectral imaging. *Remote Sensing, 12*(22), 3722. https://doi.org/10.3390/rs12223722

Aleotti, F., Tosi, F., Poggi, M., & Mattoccia, S. (2018). Generative adversarial networks for unsupervised monocular depth prediction. In *Proceedings of the European conference on computer vision (ECCV) workshops, 0–0*.

Alpan, K., & Sekeroglu, B. (2020). Tree inventory registration system. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 44,* 29–32.

Amiri, N., Heurich, M., Krzystek, P., & Skidmore, A. K. (2018). Feature relevance assessment of multispectral airborne lidar data for tree species classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42*(3).

Anderson, J., Plourde, L., Martin, M., Braswell, B., Smith, M., Dubayah, R., Hofton, M., & Blair, J. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment, 112*(4), 1856–1870. https://doi.org/10.1016/j.rse.2007.09.009

Ardila, J. P., Bijker, W., Tolpekin, V. A., & Stein, A. (2012). Context-sensitive extraction of tree crown objects in urban areas using VHR satellite images. *International Journal of Applied Earth Observation and Geoinformation, 15,* 57–69.

Baidu. (2021). Baidu street view API. https://lbsyun.baidu.com/.

Balková, M., Bajer, A., Patočka, Z., & Mikita, T. (2020). Visual exposure of rock outcrops in the context of a forest disease outbreak simulation based on a canopy height model and spectral information acquired by an unmanned aerial vehicle. *ISPRS International Journal of Geo-Information, 9*(5), 325.

Bancks, N., North, E. A., & Johnson, G. R. (2018). *An analysis of agreement between volunteer-and researcher-collected urban tree inventory data.*

Bauwens, S., Bartholomeus, H., Calders, K., & Lejeune, P. (2016). Forest inventory with terrestrial LiDAR: A comparison of static and hand-held mobile laser scanning. *Forests, 7*(6), 127.

Benninger, C. C. (2002). Principles of intelligent urbanism: The case of the new capital plan for Bhutan. *Ekistics,* 60–80.

Benosman, R., Manière, T., & Devars, J. (1996). Multidirectional stereovision sensor, calibration and scenes reconstruction. In *Proceedings of 13th international conference on pattern recognition, 1* (pp. 161–165).

Berland, A., & Lange, D. A. (2017). Google street view shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening, 21,* 11–15.

Berland, A., Roman, L. A., & Vogt, J. (2019). Can field crews telecommute? Varied data quality from citizen science tree inventories conducted using street-level imagery. *Forests, 10*(4), 349.

Bloniarz, D. V. (1996). *The use of volunteer initiatives in conducting urban forest resource inventories.* University of Massachusetts Amherst.

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.

Caglayan, A., Guclu, O., & Can, A. B. (2013). A plant recognition approach using shape and color features in leaf images. *International Conference on Image Analysis and Processing,* 161–170.

Casser, V., Pirk, S., Mahjourian, R., & Angelova, A. (2019). Depth prediction without the sensors: Leveraging structure for unsupervised learning from monocular videos. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), 8001–8008.

Chakravarty, P., Narayanan, P., & Roussel, T. (2019). GEN-SLAM: Generative modeling for monocular simultaneous localization and mapping. In *2019 International conference on robotics and automation (ICRA)* (pp. 147–153).

Chen, X., Zhao, P., Hu, Y., Ouyang, L., Zhu, L., & Ni, G. (2019). Canopy transpiration and its cooling effect of three urban tree species in a subtropical city-Guangzhou, China. *Urban Forestry & Urban Greening, 43,* Article 126368.

China National Bureau of Statistics. (2020). *Jinan City population.* China National Bureau of Statistics. http://www.stats.gov.cn/.

Choi, K., Lim, W., Chang, B., Jeong, J., Kim, I., Park, C.-R., & Ko, D. W. (2022). An automatic approach for tree species detection and profile estimation of urban street trees using deep learning and Google street view images. *ISPRS Journal of Photogrammetry and Remote Sensing, 190,* 165–180.

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 92,* 79–97.

Cozad, S. K., McPherson, E. G., & Harding, J. A. (2006). *STRATUM case study evaluation in Minneapolis.* Minnesota: University of California, Davis.

Crown, C. A., Greer, B. Z., Gift, D. M., & Watt, F. S. (2018). Every tree counts: Reflections on NYC's third volunteer street tree inventory. *Arboriculture & Urban Forestry, 44*(2).

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 9268–9277.

Culman, M., Delalieux, S., & Van Tricht, K. (2020). Individual palm tree detection using deep learning on RGB imagery to support tree inventory. *Remote Sensing, 12*(21), 3476.

Dandois, J. P., Olano, M., & Ellis, E. C. (2015). Optimal altitude, overlap, and weather conditions for computer vision UAV estimates of forest structure. *Remote Sensing, 7* (10), 13895–13920.

Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. *Advances in Neural Information Processing Systems, 27.*

Ellis, C. (2002). The new urbanism: Critiques and rebuttals. *Journal of Urban Design, 7*(3), 261–291.

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.

Escobedo, F. J., & Nowak, D. J. (2009). Spatial heterogeneity and air pollution removal by an urban forest. *Landscape and Urban Planning, 90*(3–4), 102–110.

Ewald, J. (2001). Der Beitrag pflanzensoziologischer Datenbanken zur vegetationsökologischen Forschung. *Ber. Reinhold-Tüxen-Ges, 13*, 53–69.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., … Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment, 186*, 64–87.

Garg, R., Bg, V. K., Carneiro, G., & Reid, I. (2016). Unsupervised cnn for single view depth estimation: Geometry to the rescue. *European Conference on Computer Vision*, 740–756.

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The Kitti dataset. *The International Journal of Robotics Research, 32*(11), 1231–1237.

Gillner, S., Vogt, J., Tharang, A., Dettmann, S., & Roloff, A. (2015). Role of street trees in mitigating effects of heat and drought at highly sealed urban sites. *Landscape and Urban Planning, 143*, 33–42.

Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(1), 142–158.

Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). Digging into self-supervised monocular depth estimation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3828–3838.

Grundström, M., & Pleijel, H. (2014). Limited effect of urban tree vegetation on NO2 and O3 concentrations near a traffic route. *Environmental Pollution, 189*, 73–76.

Hauer, R. J., Timilsina, N., Vogt, J., Fischer, B. C., Wirtz, Z., & Peterson, W. (2018). A volunteer and partnership baseline for municipal forestry activity in the United States. *Arboriculture & Urban Forestry, 44*(2).

Hebbalaguppe, R., Garg, G., Hassan, E., Ghosh, H., & Verma, A. (2017). Telecom inventory management via object recognition and localisation on Google street view images. In *2017 IEEE Winter conference on applications of computer vision (WACV)* (pp. 725–733).

Huang, D., Jiang, B., & Yuan, L. (2022). Analyzing the effects of nature exposure on perceived satisfaction with running routes: An activity path-based measure approach. *Urban Forestry & Urban Greening, 68*, Article 127480.

Jaakkola, A., Hyyppä, J., Kukko, A., Yu, X., Kaartinen, H., Lehtomäki, M., & Lin, Y. (2010). A low-cost multi-sensoral mobile mapping system and its feasibility for tree measurements. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(6), 514–522.

Jiang, Y., Chen, L., Grekousis, G., Xiao, Y., Ye, Y., & Lu, Y. (2021). Spatial disparity of individual and collective walking behaviors: A new theoretical framework. *Transportation Research Part D: Transport and Environment, 101*, Article 103096.

Jinan City Planning Bureau. (2021). *Road network of Jinan City*. Jinan City Planning Bureau. http://www.jinan.gov.cn/.

Jinan Forestry Bureau. (2007). *Jinan street tree population and distribution*. Jinan Forestry Bureau. http://www.jinan.gov.cn/col/col2151/index.html.

Johnson, S. (2006). *The ghost map: The story of London's most terrifying epidemic—And how it changed science, cities, and the modern world*. Penguin.

Kang, B., Lee, S., & Zou, S. (2021). Developing sidewalk inventory data using street view images. *Sensors, 21*(9), 3300.

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.

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., & Shreenan, R. (2003). Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing, 29*(5), 633–649.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Li, D., Ke, Y., Gong, H., & Li, X. (2015). Object-based urban tree species classification using bi-temporal WorldView-2 and WorldView-3 images. *Remote Sensing, 7*(12), 16917–16937.

Li, M., & Yao, W. (2020). 3D map system for tree monitoring in Hong Kong using Google street view imagery and deep learning. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 5*(3).

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., & Guan, F. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing, 115*, 63–77.

Lin, J., Wang, Q., & Huang, B. (2021). Street trees and crime: What characteristics of trees and streetscapes matter. *Urban Forestry & Urban Greening, 65*, Article 127366.

Lin, Y., & Herold, M. (2016). Tree species classification based on explicit tree structure feature parameters derived from static terrestrial laser scanning data. *Agricultural and Forest Meteorology, 216*, 105–114.

Lu, Y. (2019). Using Google street view to investigate the association between street greenery and physical activity. *Landscape and Urban Planning, 191*, Article 103435.

Lu, Y., Sarkar, C., & Xiao, Y. (2018). The effect of street-level greenery on walking behavior: Evidence from Hong Kong. *Social Science & Medicine, 208*, 41–49.

Mancini, F., Dubbini, M., Gattelli, M., Stecchi, F., Fabbri, S., & Gabbianelli, G. (2013). Using unmanned aerial vehicles (UAV) for high-resolution reconstruction of topography: The structure from motion approach on coastal environments. *Remote Sensing, 5*(12), 6880–6898.

Martin, N. A. (2011). *A 100% tree inventory using i-Tree Eco protocol: A case study at Auburn University, Alabama* [PhD Thesis].

McDonald, A. G., Bealey, W. J., Fowler, D., Dragosits, U., Skiba, U., Smith, R. I., … Nemitz, E. (2007). Quantifying the effect of urban tree planting on concentrations and depositions of PM10 in two UK conurbations. *Atmospheric Environment, 41*(38), 8455–8467.

McKinney, M. L. (2002). Urbanization, biodiversity, and conservation The impacts of urbanization on native species are poorly studied, but educating a highly urbanized human population about these impacts can greatly improve species conservation in all ecosystems. *BioScience, 52*(10), 883–890.

McPherson, E. G., Nowak, D., Heisler, G., Grimmond, S., Souch, C., Grant, R., & Rowntree, R. (1997). Quantifying urban forest structure, function, and value: The Chicago urban forest climate project. *Urban Ecosystem, 1*(1), 49–61.

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics, 31*(5), 1147–1163.

Nehme, E. K., Oluyomi, A. O., Calise, T. V., & Kohl, H. W., III (2016). Environmental correlates of recreational walking in the neighborhood. *American Journal of Health Promotion, 30*(3), 139–148.

Nielsen, A. B., Östberg, J., & Delshammar, T. (2014). Review of urban tree inventory methods used to collect data at single-tree level. *Arboriculture & Urban Forestry, 40*(2), 96–111.

Ning, H., Li, Z., Wang, C., Hodgson, M. E., Huang, X., & Li, X. (2022). Converting street view images to land cover maps for metric mapping: A case study on sidewalk network extraction for the wheelchair users. *Computers, Environment and Urban Systems, 95*, Article 101808.

Nolan, K. A., & Callahan, J. E. (2006). Beachcomber biology: The Shannon-Weiner species diversity index. *Proceedings of Workshop ABLE, 27*, 334–338.

Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening, 4*(3–4), 115–123.

Nowak, M., Szymańska, A., & Grewling, Ł. (2012). Allergic risk zones of plane tree pollen (Platanus sp.) in Poznan. *Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii, 29*(3), 156–160.

Pataki, D. E., Alberti, M., Cadenasso, M. L., Felson, A. J., McDonnell, M. J., Pincetl, S., … Whitlow, T. H. (2021). The benefits and limits of urban tree planting for environmental and human health. *Frontiers in Ecology and Evolution, 9*, 155.

Pauleit, S. (2003). Urban street tree plantings: Identifying the key requirements. *Proceedings of the Institution of Civil Engineers: Municipal Engineer, 156*(1), 43–50.

Prasad, S., Peddoju, S. K., & Ghosh, D. (2013). Mobile plant species classification: A low computational aproach. In *2013 IEEE second international conference on image information processing (ICIIP-2013)* (pp. 405–409).

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.

Pu, R., Landry, S., & Yu, Q. (2018). Assessing the potential of multi-seasonal high resolution Pléiades satellite imagery for mapping urban tree species. *International Journal of Applied Earth Observation and Geoinformation, 71*, 144–158.

Raupp, M. J., Cumming, A. B., & Raupp, E. C. (2006). *Street tree diversity in eastern North America and its potential for tree loss to exotic borers*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7263–7271.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 28*.

Ren, X.-M., Wang, X.-F., & Zhao, Y. (2012). An efficient multi-scale overlapped block LBP approach for leaf image recognition. *International Conference on Intelligent Computing*, 237–243.

Ringland, J., Bohm, M., & Baek, S.-R. (2019). Characterization of food cultivation along roadside transects with Google street view imagery and deep learning. *Computers and Electronics in Agriculture, 158*, 36–50.

Ringland, J., Bohm, M., Baek, S.-R., & Eichhorn, M. (2021). Automated survey of selected common plant species in Thai homegardens using Google street view imagery and a deep neural network. *Earth Science Informatics, 14*(1), 179–191.

Roboflow. (2022). Roboflow. http://app.roboflow.com.

Roman, L. A., Scharenbroch, B. C., Östberg, J. P. A., Mueller, L. S., Henning, J. G., Koeser, A. K., … Jordan, R. C. (2017). Data quality in citizen science urban tree inventories. *Urban Forestry & Urban Greening, 22*, 124–135. https://doi.org/10.1016/j.ufug.2017.02.001

Ryherd, S. L., & Woodcock, C. E. (1990). The use of texture in image segmentation for the definition of forest stand boundaries. In *Proceedings of the twenty-third international symposium on remote sensing of environment. 18–25 April 1990, Bangkok, Thailand. Volume II* (pp. 1209–1213).

Sankey, T., Donager, J., McVay, J., & Sankey, J. B. (2017). UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sensing of Environment, 195*, 30–43.

Savard, J.-P. L., Clergeau, P., & Mennechez, G. (2000). Biodiversity concepts and urban ecosystems. *Landscape and Urban Planning, 48*(3–4), 131–142.

Schaminée, J. H., Hennekens, S. M., Chytry, M., & Rodwell, J. S. (2009). *Vegetation-plot data and databases in Europe: An overview*.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., & Schmidtlein, S. (2020). Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 170*, 205–215.

Schuyler, D. (1986). *The new urban landscape. The redefinition of city form in nineteenth century America*.

Sjöman, H., Östberg, J., & Bühler, O. (2012). Diversity and distribution of the urban tree population in ten major Nordic cities. *Urban Forestry & Urban Greening, 11*(1), 31–39.

Song, Y., He, F., & Zhang, X. (2019). To identify tree species with highly similar leaves based on a novel attention mechanism for CNN. *IEEE Access, 7*, 163277–163286.

Sothe, C., Dalponte, M., de Almeida, C. M., Schimalski, M. B., Lima, C. L., Liesenberg, V., … Tommaselli, A. M. G. (2019). Tree species classification in a highly diverse subtropical forest integrating UAV-based photogrammetric point cloud and hyperspectral data. *Remote Sensing, 11*(11), 1338.

Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep neural networks for object detection. *Advances in Neural Information Processing Systems, 26*.

Torresan, C., Berton, A., Carotenuto, F., Di Gennaro, S. F., Gioli, B., Matese, A., … Wallace, L. (2017). Forestry applications of UAVs in Europe: A review. *International Journal of Remote Sensing, 38*(8–10), 2427–2447.

Ulmer, J. M., Wolf, K. L., Backman, D. R., Tretheway, R. L., Blain, C. J., O'Neil-Dunne, J. P., & Frank, L. D. (2016). Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription. *Health & Place, 42*, 54–62.

Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering, 25*(2), 507–543.

Wang, G., Wang, H., Liu, Y., & Chen, W. (2019). Unsupervised learning of monocular depth and ego-motion using multiple masks. In *2019 international conference on robotics and automation (ICRA)* (pp. 4724–4730).

Wang, R., Pizer, S. M., & Frahm, J.-M. (2019). Recurrent neural network for (un-)supervised learning of monocular video visual odometry and depth. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5555–5564.

Wang, W., Xiao, L., Zhang, J., Yang, Y., Tian, P., Wang, H., & He, X. (2018). Potential of internet street-view images for measuring tree sizes in roadside forests. *Urban Forestry & Urban Greening, 35*, 211–220.

Wang, Y., & Akbari, H. (2016). The effects of street tree planting on Urban Heat Island mitigation in Montreal. *Sustainable Cities and Society, 27*, 122–128.

Watson, D., & Adams, M. (2010). *Design for flooding: Architecture, landscape, and urban design for resilience to climate change*. John wiley & sons.

Wolf, K. L., Lam, S. T., McKeen, J. K., Richardson, G. R., van den Bosch, M., & Bardekjian, A. C. (2020). Urban trees and human health: A scoping review. *International Journal of Environmental Research and Public Health, 17*(12), 4371.

World Agroforestry Centre (ICRAF). (2016). *Tree functional attributes and ecological database*. World Agroforestry Centre (ICRAF). https://www.worldagroforestry.org/output/tree-functional-and-ecological-databases.

Xiao, Q., Ustin, S. L., & McPherson, E. G. (2004). Using AVIRIS data and multiple-masking techniques to map urban forest tree species. *International Journal of Remote Sensing, 25*(24), 5637–5654.

Yang, Y., He, D., Gou, Z., Wang, R., Liu, Y., & Lu, Y. (2019). Association between street greenery and walking behavior in older adults in Hong Kong. *Sustainable Cities and Society, 51*, Article 101747.

Yang, Z., Wang, P., Xu, W., Zhao, L., & Nevatia, R. (2017). *Unsupervised learning of geometry with edge-aware depth-normal consistency*. ArXiv Preprint ArXiv:1711.03665.

Yin, Z., & Shi, J. (2018). Geonet: Unsupervised learning of dense depth, optical flow and camera pose. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1983–1992.

Zhang, Y., Kang, B., Hooi, B., Yan, S., & Feng, J. (2021). *Deep long-tailed learning: A survey*. ArXiv Preprint ArXiv:2110.04596.

Zhou, T., Brown, M., Snavely, N., & Lowe, D. G. (2017). Unsupervised learning of depth and ego-motion from video. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1851–1858.

Zou, L., & Li, Y. (2010). A method of stereo vision matching based on OpenCV. In *2010 international conference on audio, language and image processing* (pp. 185–190).