Neural network tree identification from street view images to estimate CO_2 sequestration

Sam Wilcock[1[0000-0002-8353-6219]], Ivan Smirnov[1[0009-0001-8650-9005]]

and Ornella Iuorio[1[0000-0003-0464-296X]].

[1] Dipartimento di Architettura, Ingegneria delle Costruzioni e Ambiente Costruito, Politecnico di Milano, Milan, Italy.
sam.wilcock@polimi.it

Abstract. This paper presents a novel framework for estimating urban tree dimensions and their associated CO_2 sequestration potential using neural network-based image segmentation applied to Google Street View imagery. Our approach combines a dual image acquisition strategy with geometric modelling based on the pinhole camera lens model to automatically extract key tree metrics such as trunk diameter and overall height. By leveraging road width data to approximate the distance to trees and applying trigonometric equations, pixel measurements are converted into real-world dimensions. To facilitate large-scale analysis, a custom neural network that builds on established models was developed — fine-tuning a Faster R-CNN/ResNet-50-FPN architecture with extensive data augmentation from both the Urban Street Tree Dataset and local imagery. Validation against ground truth measurements shows that while tree height estimates achieve high accuracy (within 3% error), diameter estimations exhibit greater variability, highlighting areas for future improvement. The derived tree dimensions are then used to compute CO_2 sequestration using standard forestry equations, demonstrating the method's potential to inform urban sustainability assessments. Overall, this integrated pipeline offers a scalable solution for urban tree inventory and environmental impact studies.

Keywords: Green Infrastructures, Nature-Based Solutions, Environmental Sustainability, Neural Networks, Machine Learning, Urban Tree Mapping, Tree Size Estimation, CO_2 Sequestration

1 Introduction

Green infrastructure (GI) is key in enhancing both cities' environmental performance and health. Multiple benefits of GI have been studied and recorded by a large set of researchers across the Architecture, Engineering, and Construction (AEC) community, as well as in medicine and psychology [1]. GI is indeed capable of mitigating urban heat island effect, and hence improving city's thermal comfort, it reduces particulate matter with the benefit of improving air quality, and specifically GI can sequestrate and store atmospheric carbon. GI includes woodlands, green corridors, street trees, green roofs and facades, as well as some blue infrastructure. Among them, trees are of particular interest as they can be spread around, utilizing also pocket areas in densely populated communities. Through evapotranspiration and shading, trees can contrast urban heat islands. For instance, a study done by Iuorio et al [2] through Brazil and north of England showed that trees are particularly influential in mitigating Urban Heat Island (UHI) effect. In tropical climates, the study showed that 21–25% of green areas in high-density areas can provide cooling effects on air temperatures up to 2.4 °C and promote thermal comfort up to 4 °C PET mitigation.

Trees are of particular benefit to urban zones due to their ability to capture Carbon Dioxide, and as part of anaerobic respiration breaking the CO_2 into constituent Oxygen (which is released back into the atmosphere) and Carbon (which becomes part of the tree mass). Known as CO_2 sequestration, this process allows the holding of global carbon stocks within tree inventories, providing benefits for global warming [3]. This is particularly prescient due to predictions that atmospheric CO_2 concentration will have doubled by the year 2100, and 50% of global CO_2 is formed through industry [4]. CO_2 sequestration is explicitly linked to tree species, biomass, and crown coverage – particularly in the case of biomass, it can then be related to tree geometry [5] .

The impact of tree covers has been quantified in terms of reduced death in densely populated cities. Like in the case of Tiwary et al. (2009) [6], who estimated that in London, a 10 × 10 km area with 25% tree cover can prevent two deaths and two hospitalizations annually by reducing PM10 concentrations. Having methodologies and tools to predict the impact of trees is important to influence green policies. To provide a methodology to facilitate large-scale, automated tree assessment for urban planning and environmental monitoring, this work proposes a systematic approach for extracting quantitative tree data from urban imagery. Aware that object recognition remains a complex domain with persistent challenges [7], this study employs Convolutional Neural Networks (CNNs). The tree recognition methodology was tested on a street in Milan, and obtained data about tree canopies and tree trunks were used to evaluate the overall carbon sequestration, as detailed in the following sections.

The main goal of this study is to develop a system that quickly and accurately extracts tree dimensions (like trunk diameter and overall height) using trigonometry and deep learning to then estimate how much CO_2 these trees store, offering a practical tool for urban planners and environmental managers. This work provides a scalable solution for urban tree inventory that can help inform better green infrastructure planning, support efforts to combat urban heat islands, and contribute to climate change mitigation strategies. Such a solution would allow estimating Carbon Sequestration in areas where municipal data are unavailable, enabling its application even in remote or underserved regions.

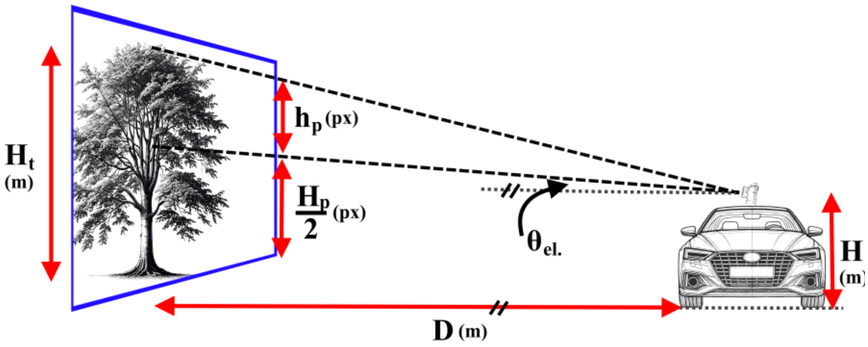

Fig. 1. Estimating tree height using trigonometry.

2 Methods

In this section, a methodology is presented to convert urban imagery into quantitative tree data. Geometric and trigonometric calculations are employed to transform pixel measurements into physical dimensions. Image acquisition via the Google Street View Application Programming Interface (API) is described with its corresponding configuration parameters. Automated segmentation of tree crowns and trunks is then detailed, providing the necessary inputs for carbon sequestration estimates.

2.1 Estimating tree dimensions

Dimensions of objects in images can be estimated using the following equation (1) derived using trigonometry:

$$W = 2D \tan\left(\frac{\theta_h}{2}\right) \times \frac{w_p}{W_p} \tag{1}$$

where parameters are described in Fig. 1. The distance to a tree can be estimated from Google Street View images through some basic assumptions:

- The images are taken from a camera mounted on the centre of a vehicle and the distance from the vehicle to the tree can be estimated from the width of the road.
- The tree is roughly in line with the camera, that is perpendicular to the side of the street view vehicle.
- The image is relatively undistorted, following a pinhole camera lens model.

Table 1. Parameters for estimating tree diameter

Symbol	Parameter	Known?
D	Distance to object (m)	Estimate from road width data
θ_h	Horizontal FOV angle of image (°)	Yes
w_p	Width of object in image (px)	Yes
W_p	Width of image (px)	Yes
W	Actual width of object (m)	No

In order to measure the height of a tree, multiple images can be used. First, the vertical field of view (FOV) must be known. Then, given an estimate for the distance D, if the number of pixels taken to view from the base to the top of the tree is known, then the angle from the camera to the top of the tree can be calculated from the distance to the tree and the number of pixels from the base to the top, allowing calculation of the height.

However in the test images, parked cars can obscure the base of the tree, and another option is to estimate the height of a car (taken to be 2.0m) with added camera (taken to be 0.4m), and measure the height in pixels to the top of the tree from half a frame above horizontal as shown in Fig. 1. Thus, the angle between the point on the tree horizontally in line with the camera and the top of the tree from the camera, θ_t, can be found. By considering that a full vertical span of the image spans an angle subtended by the vertical FOV angle θ_v, h_p is the height of the object above the horizontal centre of the image in pixels, H_{car} is the height of the camera on the car above the road, H_{ground} is the height of the tree base above ground (the height of the pavement) and H_p is the height of a single image in pixels. Then the tree height H_t can be estimated for camera pitch angle of θ_{el} as

$$H_t = \left(D \tan \theta_{el.} + H_{car} - H_{ground}\right) + \left(2\left(\frac{D}{\cos \theta_{el.}}\right) \tan\left(\frac{\theta_v}{2}\right) \cdot \frac{h_p}{H_p}\right) \qquad (2)$$

2.2 CO_2 Sequestration Calculation

CO_2 sequestration is the process of capturing and storing CO2 from the atmosphere, which trees perform to aid in photosynthesis. A commonly used formula to estimate carbon sequestration for a tree is

$$Carbon\ mass = \frac{\pi}{4} \times D^2 \times H \times \rho_{wood} \times F_{carbon} \times \mu_{DRY}\ (kg) \qquad (3)$$

where D is the trunk breast diameter in m (measured approx. 1.3m above ground), H is tree height in m, ρ_{wood} is the wood density measured in kg/m^3, F_{carbon} is the carbon fraction, representing the percentage of carbon in the wood when dry (standardly set as 50% based on IPCC guidelines [5]) and μ_{DRY} is the percentage mass of the tree when dried compared to when wet. Note that the equation is simply multiplying the mass of the tree by the carbon fraction. The equation applies specifically to the mass of the tree when wood is dry. Dry matter represents around 72.5% of the tree, hence μ_{DRY} can be taken to be 0.725 [8].

Since the carbon mass storage per tree can now be calculated, the associated CO_2 can also be calculated. Carbon is made up of one molecule of carbon which has atomic mass of $12.011amu$, and 2 molecules of oxygen with atomic mass of 15.999 each. Hence CO_2 can be calculated from the initial carbon storage equation as

$$CO_2\ storage = Carbon\ mass \times \frac{12.011 + 2 \times 15.999}{12.011}$$
$$= Carbon\ mass \times 3.66 \quad (kg) \tag{4}$$

The tree population of Via Ponzio is dominated by European nettle trees (*celtis australis*) which grow ~10-25m tall, and have the following properties (Table 2):

Table 2. Average properties of European nettle trees.

Wood density (kg/m^3) [9]	**Max. diameter (cm)** [10]	**Max. height (m)** [10]
550	60	25

2.3 Image collection

The Google Street View API provides a RESTful interface that enables the retrieval of street-level images by specifying key parameters such as:

- Location: A geographic coordinate (latitude, longitude)
- Size: The image resolution in [width, height] format
- FOV (Field-Of-View): The extent of the observable scene captured by the camera
- Heading: The compass direction in which the camera is pointed
- Pitch: The vertical angle relative to the horizon

These parameters offer the flexibility to target specific views along urban streets, ensuring that images capture the desired portions of the scene (the trees on both sides of the road).

A Python script was developed to automate the image retrieval process. The code constructs an API query with the appropriate parameters and saves the resulting image to disk. The acquisition strategy involves stepping through a street segment by incrementing geographic coordinates - approximating a fixed distance between image captures - and capturing images from both sides of the street by adjusting the heading parameter. Table 3 summarizes the key configuration parameters used in this process

By employing this dual-acquisition method of capturing both east and west-facing images, the pipeline ensures that all the greenery of the street is captured, thereby enhancing the data completeness for the later segmentation and analysis.

When a request is made, the API returns an image based on the coordinates closest to the specified point rather than the exact coordinates requested, potentially leading to discrepancies. To address this issue, Google Street View offers a secondary REST endpoint that provides metadata for each location, including the precise coordinates where the image was captured. Further details on this solution are provided in Section 3.2.

For the purposes of this study, 40 images were collected - 20 on each side of Via Ponzio street in Milan. Given that the street is oriented along a south-to-north axis, adjustments to this methodology will be necessary when it is applied to the entire city.

2.4 Tree segmentation within images using neural networks

Measuring the trees dimensions as pixels can be done manually with image manipulation software, but the process is time-consuming and doesn't scale for larger sets of data input. To automatically select trees from the images, a solution was searched for through image segmentation. Neural network-based segmentation solutions already exist to produce similar results. The Segment Anything (SAM) network can produce segmentation masks using text-based prompts (e.g. "tree") [11], however an additional network is required to perform the processing of text input, and the network lacks specificity to the problem. The PercepTree network is trained to work specifically on trees, however the training set of real images was taken from Canadian forests and only the trunks are labelled [12] which is useful for estimating trunk diameter, but not for height. Hence it was decided to train a new network specific to tree segmentation from an existing dataset.

A dataset of particular note is the Cityscapes dataset [13] which has 25,000 labelled images with various annotations including vehicles, people, and vegetation.

Table 3. Parameters for street view images acquisition.

Parameter	Description	Value
$Step\ distance$	Distance between successive images	$10m$
$Number\ of\ steps$	Total number of images per side	20
Image size	Resolution of acquired images	$400x640px$
$Heading$	Camera orientation for each street side	90°/270°
$Pitch$	Camera angle from horizontal	30°
FOV	Camera field of view	120°

Fig. 2. Image from the USTD dataset with tree mask annotations.

The vegetation labels however include bushes and verges etc. which are unsuitable for this work. Instead, a new neural network was trained to specifically segment trees. The urban street tree dataset (USTD) consists of 3,949 images (and 386 test images), and associated image masks marking the polygon outlines of the trees within the image, with data taken from photos of cityscapes in China [14] (see Fig. 2).
The original paper associated with the USTD demonstrates segmentation of trees within a test set using the fully convolutional network (FCN) [15]. It was decided to extend the original dataset using data augmentation. The 3,949 images in the training dataset were randomly augmented for a number of training epochs and used to fine tune the Faster R-CNN/ResNet-50-FPN-3x network within Detectron2 [16]. Augmentations included random modifications of image brightness, flipping images horizontally, image cropping and rotations within 30°, to improve the robustness of the network and increase the number of potential training images [17]. Training was performed using a T4 GPU within Google CoLab, using the original paper parameters (100 epochs and constant learning rate 0.001), with augmentation under the same parameters, with 500 epochs and an initial warming stage where the learning rate is increased for the first 20% of iterations to encourage wider initial exploration of the potential weights for optimisation, and in the same configuration but with a halved initial learning rate of 0.0005.

Automated dimension extraction The trained R-CNN networks were used to segment and label both tree crowns and trunks. Using maximum bounding boxes of the trees as shown in Fig. 3, the maximum height of the tree in pixel space can be found easily, either by manual measurement from images or automatically by taking the bounding box maximum y-value in pixels. Meanwhile, the pretrained PercepTree network which excels at identifying tree-trunks can be used similarly to estimate the tree diameter. Note that for simplicity, trunk diameter in pixel space was found

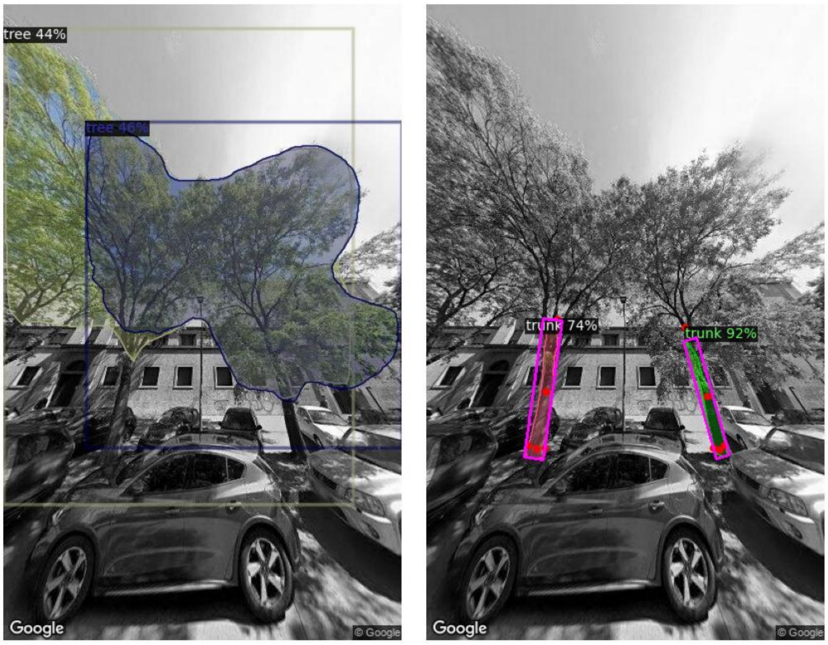

Fig. 3. Identification of tree crowns (left) and trunks (right) on Via Ponzio, using the trained R-CNN and pretrained PercepTree networks respectively. On the trunks, minimum bounding boxes used to measure trunk diameter are highlighted in magenta. Image capture: Apr 2023 © 2025 Google, Italy.

from the width of the trunk's minimum area bounding box, that is the smallest possible bounding box rotated to align with the trunk, and more accurate estimation based on finding the thinnest region (or the diameter at breast height – diameter at 1.3m above ground level) would be appropriate.

3 Results

In this section, results are summarized. The validity of tree dimension estimates is demonstrated through manual measurements and comparisons with survey data. Segmentation performance is evaluated using Intersection over Union (IoU) metrics, and CO_2 sequestration values are derived from the measured dimensions. Limitations and potential sources of error are also discussed.

3.1 Manual validation of tree dimension estimates

To test the validity of equation 1 and the assumptions made, an image of two trees was extracted from Google Street View featuring 2 trees outside of the Politecnico di Milano dABC building, with image dimensions of 640x480px and a reported FOV angle of 80° (Fig. 4). The tree widths were manually measured in pixels within the image.

Road map data was extracted from the OpenStreetMap API providing info such as road width, number of lanes and parking conditions (primarily, whether an additional vehicle width of road is used as parking). Road lane width in the surveyed area was listed as 2.95m, with parking lane width of 2m. The surveyed area has a parking lane on either side, with two lanes of active road in the centre. Assuming the camera vehicle is in the middle of the lane, and that trees are planted at least 50cm from the kerb edge, the distance to a tree is then estimated as 2.95×1.5+2+0.5=6.925m for far lane images, and 2.95×0.5+2+0.5=3.975m for near lane images.

From equation 1, for the left-hand tree the estimated diameter is then

$$Diameter = 2 \times 6.925 \times \tan\left(\frac{80}{2}\right) \times \frac{30}{640} = 0.55m \tag{5}$$

whilst similarly for the right-hand tree the diameter is estimated as 0.40m.

Actual measurements for tree diameter were then taken from the circumference of the trees, seen in Table 4. As can be seen, the diameter measurement exhibits a reasonably high level of error, and improvements should be considered for diameter estimation, particularly as the CO_2 sequestration calculations of equation 33 have a

Fig. 4. Image of Via Ponzio from Google Maps Street View API, using an 80° FOV. Pixel widths added manually. Image capture: Apr 2023 © 2025 Google, Italy.

squared relationship with the diameter, which will compound the effects of measurement error.

From Fig. 4 however, the tops of trees cannot be observed. Using a wider HFOV of 120° and a vertical image instead captures the full tree. The vertical field of view θ_v can be calculated from the horizontal field of view θ_h using the aspect ratio $(AR = width/height)$ as

$$\theta_v = 2\tan^{-1}\left(\frac{1}{AR}\cdot\tan\frac{\theta_h}{2}\right) \quad (6)$$

Table 4. Validation of diameter estimation.

Tree	Measured diameter (m) d_{meas}	Est. diameter from image (m) d_{est}	% error $\frac{d_{est} - d_{meas}}{d_{meas}}$ %
Left	0.57	0.54	−4%
Right	0.48	0.40	−17%

hence for HFOV of $120°$ the VFOV $\theta_v = 140°$. With an estimated car/camera height $h_c = 2.4m$ and camera angle $\theta_{el.} = 40°$ with $D = 6.925$, and using a different image with the new settings for the same trees as in Fig. 5, vertical heights can be estimated with equation 2 and compared to tree height measurements from local tree survey data [18]. Results are shown in Table 5. Here, the height measurements are highly accurate, particularly considering that the measured heights given in the survey data are only given to the nearest whole number. This could in part be due to the higher number of pixels involved in calculating the tree height; for the trunk diameters, a single pixel of over or under-measurement can cause significant errors comparatively.

Table 5. Validation of height estimation.

Tree	Measured height (m) h_{meas}	Est. height from image (m) h_{est}	% error $\frac{h_{est} - h_{meas}}{h_{meas}}$ %
Left	14	14.1	+0.71%
Right	13	12.7	−2.3%

Fig. 5. Validating tree height calculations by measuring height above centreline. Image capture: Apr 2023 © 2025 Google, Italy.

3.2 Data collection and labelling

To ensure the integrity and accuracy of the dataset, a refinement process was implemented following the initial image acquisition from the Google Street View API. When images are captured in rapid succession along a continuous street segment, some images with the different requested coordinates return identical "actual" coordinate values due to the way the Street View camera takes pictures. By basing the image identifiers on the precise, actual coordinates, duplicated images can be systematically recognized, retaining only one unique instance per location. This careful curation reduces redundancy and enhances the overall quality of the dataset.

A subset of 40 images was manually annotated using the open-source "labelme" application [19]. This manual labelling process involved marking tree boundaries thereby providing reliable ground truth masks for the training and validation of the neural network segmentation models.

The combination of automated deduplication and meticulous manual labelling is needed to enhance the overall data quality, yielding a reliable dataset, ultimately improving the robustness and precision of the tree dimension estimation and following CO_2 sequestration.

The final dataset comprised 130 distinct manually labelled trees: 40 on the east side and 90 on the west side of the street. However, it is important to note that some duplicates may still be present. This is due to overlapping image regions where the

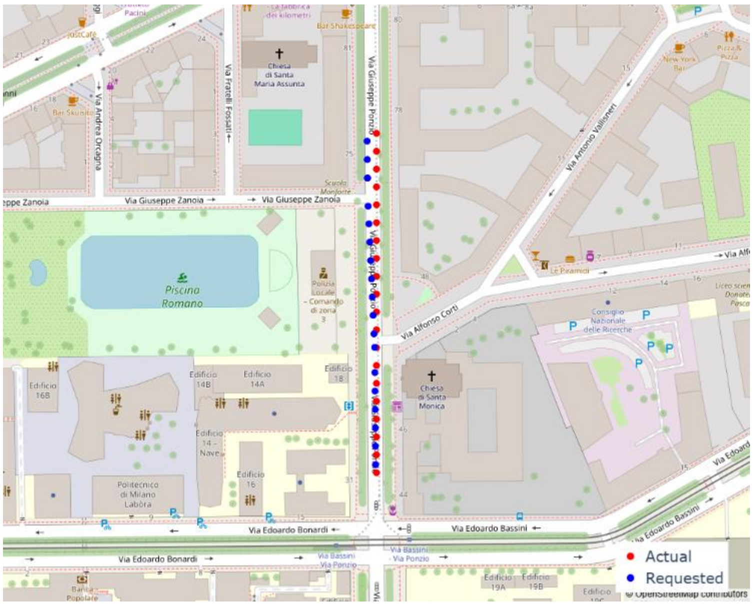

Fig. 6. Requested locations sent to Google Street view, and actual locations returned.

end of one capture coincides with the beginning of the next, resulting in some trees being labelled twice.

3.3 Neural network performances

The different neural network modifications are compared using the IoU metric. The IoU metric allows a quantitative evaluation of a segmentation's accuracy by comparing predicted mask areas to ground truth mask areas. The equation for the IoU metric is

$$IoU = \frac{Area(gt \cap pd)}{Area(gt \cup pd)} \tag{7}$$

where gt is the ground truth mask of an image, and pd is the predicted mask of an image given by the predictor. For a given image, the IoU is calculated as the intersecting area of the ground truth and prediction masks as a ratio of the union of their areas. A value of 0 means that the two masks do not overlap at all, whilst a value of 1 describes perfect overlap, i.e. matching masks.

For each of the 4 trained neural networks, predictions were run for all test images. IoU values were calculated for each image and averaged to find a mean IoU score. Two test image sets were use - the test set of the USTD dataset (386 images) and the test set of labelled local street view images (40 images). A confidence score of > 0.5 was used, meaning that predictions with lower confidence were discarded.

The IoU scores for each of the networks are displayed in Table 6. With the USTD dataset, it is shown that the additional augmentation of data using the methods described in section 2.4 has improved the neural network performance at segmenting trees. The increase in number of learning iterations from 100 to 500 initially demonstrated a decrease in segmentation precision, shown by the decrease in mean IoU value, however decreasing the initial learning rate by half improves the end-result significantly. This initial poorer behaviour then can be attributed to the additional iterations with the original learning rate running without performing improvements. Further, the use of an initial "warm-up" period in the first 20% of iterations may have caused an excessively high initial learning rate, moving the network further away from the optimal solution in the first 100 iterations.

For the Via Ponzio dataset, the IoU is lower, but shows the highest improvement at the the 500 iterations with higher learning rate. Potentially the use of images with mild distortion (caused by the FOV of the street view images) reduces detection accuracy at the edges of the image. Adding a further augmentation to images to simulate this distortion would then potentially aid in the future improvement of tree segmentation from street view images.

Fig. 7 shows the segmentation of a tree from the different trained networks, making comparisons to the prelabelled ground truth mask. As can be seen, the initial trained network struggles generally with identifying the trunks of trees, which gradually improves with network augmentation and more training. The standard training based on the dataset paper's original parameters fails to identify the full height of the tree crown, whilst augmentation manages to identify it. Larger numbers of training iterations are also shown to improve segmentation accuracy.

Table 6. IOU scores for the tree neural network instance segmentation.

Model	USTD Mean IoU	Via Ponzio Mean IoU
100 iterations, LR 0.001	0.764	0.170
100 iterations, LR 0.001 (aug.)	0.770	0.120
500 iterations, LR 0.001 (aug.)	0.765	0.477
500 iterations, LR 0.0005 (aug.)	0.780	0.347
PercepTree	0.0182	0.0643

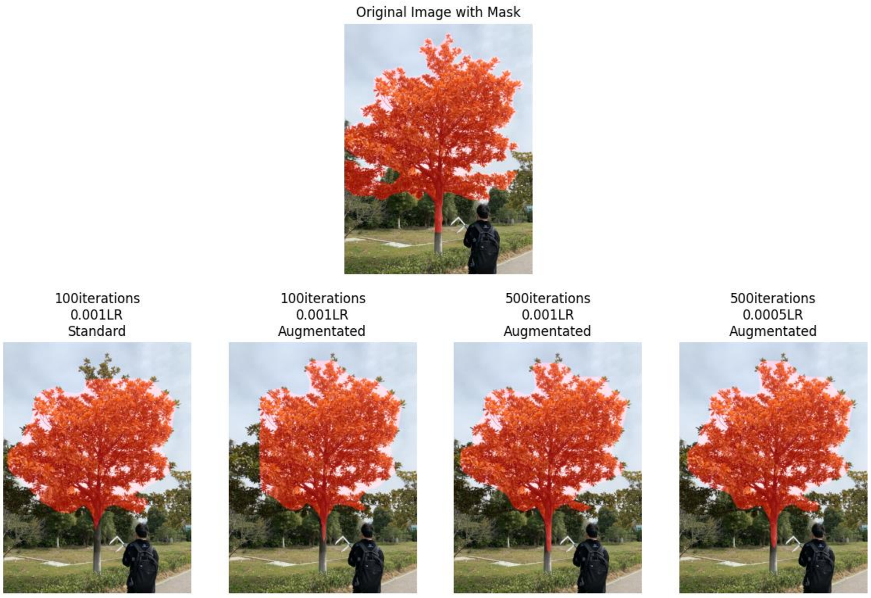

Fig. 7. Visual results of the different trained networks on an image of a tree, showing comparison to ground truth value.

PercepTree appears to perform poorly from the data shown in Table 6, however this is due to the "trunk" masks being predicted by PercepTree being compared to the full-tree masks from the new neural network – the full tree masks have significantly larger areas than is made up by just the trunk, hence the union of the two masks is much larger than the intersection. Additionally, the new neural networks perform poorly for identifying trunks as shown in Fig. 7. The network does however perform better on the local data for Via Ponzio. Improvements could then be made by separately adding "trunk" annotations within the test datasets for more accurate comparison and training, potentially by further tuning of the PercepTree network.

3.4 Accuracy of inferring tree dimensions

Data from the local governmental tree survey [18] was collated for the trial area, including tree co-ordinates (shown in Fig. 8 compared to estimated tree locations from images; locations estimated through adaptations of Eq. 1 [20]), trunk diameter and height, for comparison with segmentation dimension inference data. Tree labels were associated with trunks by finding the trunk with a minimum horizontal distance within a set range, from centre-point to centre-point, with trees without trunks

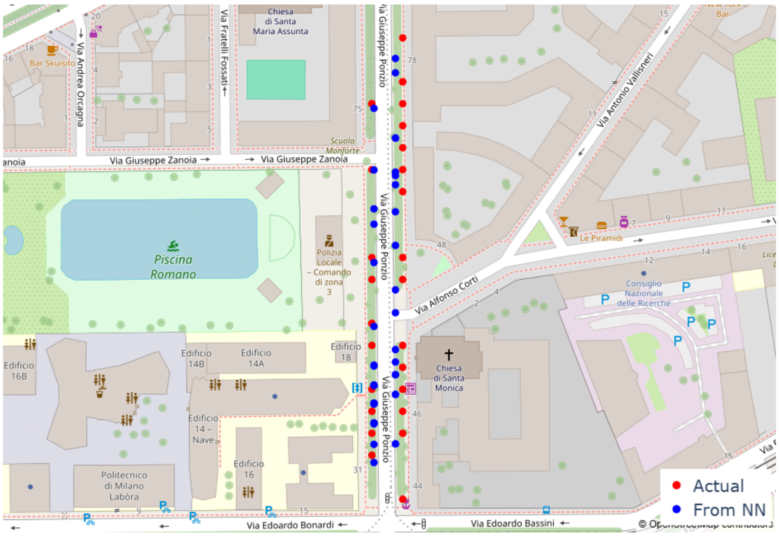

Fig. 8. Estimated tree locations from images, compared to actual tree locations provided by local comune.

being discarded (13 identified tree regions out of 53 were discarded in this way). Latitude and longitude were estimated from the camera location and trunk centre-point positions within the image.

A total of 39 potential trees were found within the survey area. Fig. 9 shows a comparison between the inferred and real dimensions of the tree, showing heights and diameters. For both dimensions, the neural network-based dimension measurement overestimates the data, with errors increasing as the base dimension increases.

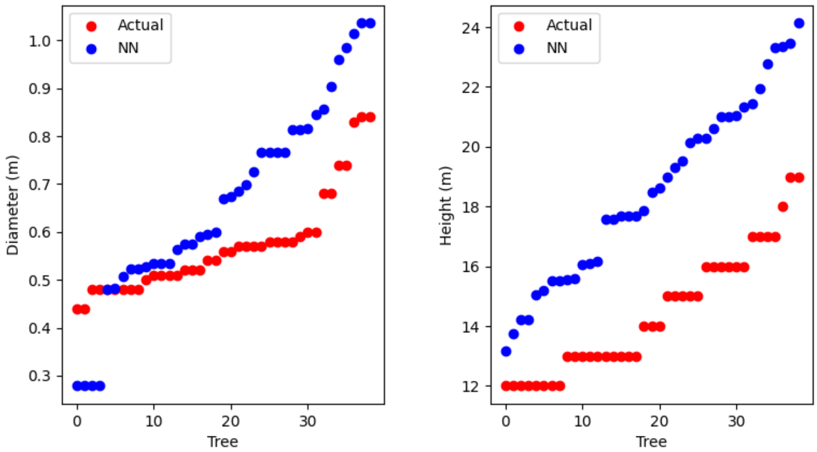

Fig. 9. Plots of actual tree dimensions vs. NN estimated dimensions. Tree diameters on left, height on right.

This indicates a systematic error, likely due to an issue with either the distance value being assumed, or the FOV angles. Additionally, there is no indication of how old the survey data is, and the trees may have simply grown since the data was taken.

3.5 CO_2 sequestration and heat effects

Out of the 130 trees manually segmented in Section 3.2, only 39 were retained in the CSV file after processing through the neural network. Trees were excluded either because the model's confidence level was below the threshold or because their trunks did not fall within the designated segmentation area. For this proof-of-concept, only trees with reliable segmentation were selected.

The smallest tree in the chosen sample is 13.2 meters high when the highest one riches 24.1 meters from the ground level to the top of the canopy. The widest tree has a trunk of 1 meter in diameter whereas the thinnest is 30 cm. An average tree from the subset is 18.5 meters high and has a 70cm-wide trunk.

By applying equations 3 and 4 to the resulting dataset, the following (Table 7) estimates were obtained. A total of 54 tonnes of Carbon is estimated to be stored within the trees' biomass, with an equivalent CO_2 sequestration of 200 tonnes (thus having converted around 146 tonnes of CO_2 to oxygen during their lifetime). Therefore, the average tree on Via Ponzio street stores 1,383kg of Carbon, sequestering around 5t of CO_2.

These estimates align closely with the independent assessments available on the Milano Geoportale [18], where specialized analyses were used to calculate the carbon stored by urban trees. This consistency validates the approach and underscores the robustness of the segmentation and estimation methodology.

While these figures provide a robust snapshot of the carbon sink potential of the sample trees, several factors can affect their accuracy:

- **Measurement Uncertainties:** Errors in measuring tree height or trunk diameter can compound when squared or multiplied.
- **Species Variability:** Different tree species have varying wood densities, growth rates, and carbon fractions. Species-specific parameters can refine the estimates. It is assumed that all entries in the chosen sample belong to European nettle trees (*celtis australis*)

Table 7. Tree survey aggregated results.

Tree Height, m	Trunk Diameter, m	Stored Carbon, kg	CO_2 Sequestration, kg	Aggregate
723.2	25.9	53,972	197,534	SUM
18.5	0.7	1,384	5,065	MEAN
13.2	0.3	173	633	MIN
24.1	1.0	2,827	10,348	MAX

- **Environmental Conditions:** Local growing conditions, tree age, and health also influence biomass accumulation and, hence, carbon sequestration.

4 Discussion & Conclusions

The study demonstrates the feasibility of integrating neural network-based segmentation with geometric modelling to estimate urban tree dimensions and their associated CO_2 sequestration potential using imagery. By employing a dual image acquisition strategy and leveraging a fine-tuned Faster R-CNN/ResNet-50-FPN model, it was possible to extract tree crown and trunk dimensions, which were then converted into real-world measurements using pinhole camera models and trigonometric equations.

Despite initially labelling 130 trees, the pipeline retained 39 trees after filtering based on model confidence and segmentation criteria. This proof-of-concept approach intentionally focused on trees with reasonable confidence scores for segmentation to minimize uncertainties. The resulting dataset, when processed with standard forestry equations produced a total estimated CO_2 sequestration of approximately 200t and stored carbon of 54t. These figures align closely with independent assessments from the governmental portal [18], thereby validating the methodology.

Some inaccuracies were demonstrated in the estimation of tree dimensions, particularly in diameter estimation. This is in part due to the smaller number of pixels involved in measuring the trunk diameter as compared to the height of trees, as single pixel errors in measurement represent a far larger proportion of the dimension and cause greater measurement error up to 20%. This causes further errors in CO_2 sequestration evaluation, as the mass of the tree is proportional to the square of diameter. Height measurements are initially shown however to be highly accurate using the presented technique. Improving distance estimates from trees to the camera plays a part in reducing dimensional measurement errors,

Crown segmentation using the R-CNN networks trained on the USTD dataset provides a high degree of accuracy, shown through over 75% IoU rates on the original dataset's training data and ~45% IoU on newly labelled local data. Trunk segmentation displays lower IoU scores, partially due to comparisons being made between trunk masks and whole-tree masks, and additional labelling of trunk data would certainly show a quantitative improvement in IoU scores. Further improvements could additionally be made through increased augmentation of training data to include local images.

The low success rate of tree identification in this study is an area which would require improvement before scaling. The computational effort of using neural networks is significant, however can be made faster by moving the analysis process to cloud computing for a large-scale inspection of a region. The use of Google Street View imagery is of usefulness due to the high availability of data, and the reduction of operator hours required. In comparison to local government-level data surveys being taken manually, such use of street view data does not require employing operatives or tree specialists and could be undertaken in remote areas without governmental co-ordination.

The study underscores broader environmental implications, demonstrating the significant CO_2 sequestration effects that can be achieved by increasing urban tree inventories. Trees contribute indirectly to urban heat mitigation by providing shade, which reduces ambient temperatures to improve comfort levels within the city. Although the focus was on carbon sequestration, these urban heat island mitigation benefits further emphasize the value of urban trees as a nature-based solution for sustainability. Future studies could investigate the additional benefits to urban heating effects from crown coverage and expand the survey area for larger scale carbon sequestration investigations.

References

[1] D. Perrotti and O. Iuorio, ‘Green Infrastructure in the Space of Flows: An Urban Metabolism Approach to Bridge Environmental Performance and User’s Wellbeing’, in *Planning Cities with Nature*, F. Lemes De Oliveira and I. Mell, Eds., in Cities and Nature. , Cham: Springer International Publishing, 2019, pp. 265–277. doi: 10.1007/978-3-030-01866-5_18.

[2] O. Iuorio and L. A. Harbich, ‘Green Networks as a Key of Urban Planning with Thermal Comfort and Well-being’, in *Planning Cities with Nature*, F. Lemes De Oliveira and I. Mell, Eds., in Cities and Nature. , Cham: Springer International Publishing, 2019, pp. 97–109. doi: 10.1007/978-3-030-01866-5_7.

[3] M. Lindner *et al.*, ‘Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems’, *Forest Ecology and Management*, vol. 259, no. 4, pp. 698–709, Feb. 2010, doi: 10.1016/j.foreco.2009.09.023.

[4] J. C. Minx *et al.*, ‘A comprehensive and synthetic dataset for global, regional, and national greenhouse gas emissions by sector 1970–2018 with an extension to 2019’, *Earth Syst. Sci. Data*, vol. 13, no. 11, pp. 5213–5252, Nov. 2021, doi: 10.5194/essd-13-5213-2021.

[5] IPCC, *Good practice guidance for land use, land-use change and forestry /The Intergovernmental Panel on Climate Change. Ed. by Jim Penman*. Hayama, Kanagawa, 2003.

[6] A. Tiwary *et al.*, ‘An integrated tool to assess the role of new planting in PM10 capture and the human health benefits: A case study in London’, *Environmental Pollution*, vol. 157, no. 10, pp. 2645–2653, Oct. 2009, doi: 10.1016/j.envpol.2009.05.005.

[7] P. Tsirtsakis, G. Zacharis, G. S. Maraslidis, and G. F. Fragulis, ‘Deep learning for object recognition: A comprehensive review of models and algorithms’, *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 298–312, Dec. 2025, doi: 10.1016/j.ijcce.2025.01.004.

[8] S. Shadman *et al.*, ‘The carbon sequestration potential of urban public parks of densely populated cities to improve environmental sustainability’, *Sustainable Energy*

Technologies and Assessments, vol. 52, p. 102064, Aug. 2022, doi: 10.1016/j.seta.2022.102064.

[9] A. E. Zanne *et al.*, ‘Data from: Towards a worldwide wood economics spectrum’. Dryad, p. 2123413 bytes, Feb. 04, 2009. doi: 10.5061/DRYAD.234.

[10] C. Orwa, A. Mutua, R. Jamnadass, and S. Anthony, ‘Celtis australis’, Agroforestree Database: A tree reference and selection guide version 4.0. Accessed: Mar. 10, 2025. [Online]. Available: https://apps.worldagroforestry.org/treedb/AFTPDFS/Celtis_australis.PDF

[11] A. Kirillov *et al.*, ‘Segment anything’, *arXiv:2304.02643*, 2023.

[12] V. Grondin, J.-M. Fortin, F. Pomerleau, and P. Giguère, ‘Tree detection and diameter estimation based on deep learning’, *Forestry: An International Journal of Forest Research*, Oct. 2022.

[13] M. Cordts *et al.*, ‘The cityscapes dataset for semantic urban scene understanding’, in *Proc. of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016.

[14] T. Yang, S. Zhou, Z. Huang, A. Xu, J. Ye, and J. Yin, ‘Urban street tree dataset for image classification and instance segmentation’, *Computers and Electronics in Agriculture*, vol. 209, p. 107852, Jun. 2023, doi: 10.1016/j.compag.2023.107852.

[15] E. Shelhamer, J. Long, and T. Darrell, ‘Fully Convolutional Networks for Semantic Segmentation’, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.

[16] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, ‘Detectron2’. 2019. [Online]. Available: https://github.com/facebookresearch/detectron2

[17] J. Fang, T. Dong, and J. Fan, ‘A Copy-Paste Data Augmentation Method for Urban Tree Detection’, in *PRICAI 2024: Trends in Artificial Intelligence*, vol. 15283, R. Hadfi, P. Anthony, A. Sharma, T. Ito, and Q. Bai, Eds., in Lecture Notes in Computer Science, vol. 15283. , Singapore: Springer Nature Singapore, 2025, pp. 122–133. doi: 10.1007/978-981-96-0122-6_12.

[18] Comune di Milano, ‘Geoviewer’, Geoportale SIT | Comune di Milano. Accessed: Jan. 30, 2025. [Online]. Available: https://geoportale.comune.milano.it/geoviewer/

[19] K. Wada, ‘labelme: Image polygonal annotation with python’. GitHub, 2018. [Online]. Available: https://github.com/wkentaro/labelme

[20] S. Branson, J. D. Wegner, D. Hall, N. Lang, K. Schindler, and P. Perona, ‘From Google Maps to a fine-grained catalog of street trees’, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 135, pp. 13–30, Jan. 2018, doi: 10.1016/j.isprsjprs.2017.11.008.