Building and Environment 275 (2025) 112816

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Modelling sunlight and shading distribution on 3D trees and buildings: Deep learning augmented geospatial data construction from street view images

Shu Wang [a], Rui Zhu [b,*], Yifan Pu [a], Man Sing Wong [c,d], Yanqing Xu [e], Zheng Qin [b]

[a] *Department of Geography, National University of Singapore, Singapore 117570, Republic of Singapore*
[b] *Institute of High-Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Republic of Singapore*
[c] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China*
[d] *Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hong Kong SAR, China*
[e] *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China*

ARTICLE INFO

Keywords:
Street view images
Semantic segmentation
Tree Shading
3D Urban model
Building integrated photovoltaics
GIScience

ABSTRACT

In complex urban environments, accurately estimating the shading effects of trees on three-dimensional (3D) building surfaces is crucial to facilitate building design and urban greenery implementation. However, there is a long-unsolved challenge in efficiently and elaborately modelling trees and simulating spatiotemporally heterogeneous shading effects of trees on 3D urban envelopes. To overcome the challenge, this study proposes a research framework that: (i) employs transfer learning to build a deep learning model for accurately segmenting geo-objects in Street View Images (SVIs), (ii) utilizes semantic segmentation results to fit regressions between the pixels of specific geo-objects in the SVIs and the corresponding real-world lengths of standard geo-objects, develops a 3D space geometric projection model for calculating tree coordinates and 3D geometries, and identifies the real spatial relationships between buildings and trees to calibrate errors caused by segmentation inaccuracies for subsequent simulations, and (iii) integrates the calibrated 3D tree models with 3D building models to construct a unified 3D urban model for estimating the spatiotemporal distribution of sunlight and shading. Using Singapore as the study area, we adopted DeepLabV3+, a widely used pre-trained semantic segmentation model, to achieve IoU of 91.51 % for buildings and 76.29 % for trees, with F1-scores of 97.93 % and 88.19 % respectively. Additionally, data calibration optimized initial tree polygons in 39.03 % of the SVIs, reducing outliers and improving modeling accuracy and robustness. The results demonstrate that the proposed framework efficiently and accurately models high-density urban environments, providing a practical solution to complex shading problems and reducing data acquisition and processing costs.

1. Introduction

1.1. Background

As global environmental problems become increasingly serious, the continuous increase in global greenhouse gas emissions has triggered frequent extreme climate events, directly threatening the balance of the earth's ecosystem and the living environment of mankind. In this context, the international community has reached several consensuses and agreements, such as the 2015 Paris Agreement and the 2021 Glasgow Climate Pact focusing on emission reductions and the shift to renewable energy [1,2]. Consequently, promoting renewable energy usage and reducing carbon emissions have thus become central issues in global economic and environmental policies. To further promote global energy transformation, many countries have increased the investment and development in renewable energy with unprecedented efforts. Among them, solar photovoltaic (PV) technology has become an important pillar of global energy transformation with its mature technology, wide application and continuous cost reduction [3]. The installation of PV systems has grown rapidly in the past decade. In 2010, the global total installed PV capacity was about 40 GW, and by 2022, this figure exceeded 770 GW, with an annual growth rate of 18 % [4].

* Corresponding author.
E-mail address: zhur@ihpc.a-star.edu.sg (R. Zhu).

https://doi.org/10.1016/j.buildenv.2025.112816
Received 24 November 2024; Received in revised form 25 February 2025; Accepted 2 March 2025
Available online 4 March 2025

However, with the acceleration of urbanization, land resources are becoming increasingly scarce, and traditional utility-scale PV power stations face land restrictions in many countries. To tackle this issue, Building-Integrated Photovoltaics (BIPV) as an emerging technology that combines PV modules with building structures have been used in a wide range of applications, such as rooftop PV tiles and façade PV curtain walls [5]. Rooftop PV systems benefit from good lighting conditions and optimized tilt angles for maximum solar irradiation, but their scale is limited by the available rooftop space. In contrast, façade PV systems can make full use of the vertical surface of the building, providing more possibilities for achieving energy efficiency and space utilization [6,7]. Nevertheless, in high-density urban environments, sunlight on building façades can be significantly reduced due to shading from surrounding geo-objects, such as trees, and current three-dimensional (3D) city models are difficult to be used for a fine-scale solar distribution estimation. This issue is particularly prominent in cities where trees are densely planted along roads and buildings, such as in some European countries [8] and Singapore [9], making an accurate estimation of solar distribution on 3D urban surfaces more challenging. To address this, shading analysis is crucial, alongside factors such as building orientation and tilt, to ensure accurate assessments of solar potential.

In recent years, LiDAR point clouds have been widely used to produce fine modelling of buildings and trees in 3D urban environments [10]. Although this technique can generate highly accurate results, it presents challenges in large-scale urban applications due to complex data processing, high storage demands, and significant costs [11]. Therefore, there is an urgent need for alternative solutions. The major motivation of this study is to segment trees from Street View Images (SVIs) and accurately obtain their geo-coordinates along the road network, model their 3D geometries, and estimate their areas to facilitate solar distribution estimation particularly on façades. The successful development of the proposed vision will support efficient and systematic solar potential analysis across large urban areas, reducing data costs and improving accuracy in assessing the impact of tree shading on BIPV systems. Moreover, the proposed framework extends beyond solar potential estimation to support the optimization of building energy performance. By incorporating detailed shading analyses, it facilitates dynamic solar gain management, enhancing strategies for improving façade and rooftop solar performance, such as BIPV design and shading device optimization. These capabilities enable more accurate assessments of solar gain impacts at both the building and urban scale, bridging the gap between solar energy modeling and sustainable urban planning.

1.2. Semantic segmentation of street view images

In urban environments, semantic segmentation technology for SVIs has made significant progress across various geo-objects, with broad applications and specific objectives. First, street greening was the most common application of this technology. Numerous studies used indicators such as the Green View Index (GVI) to evaluate urban greening, analyzing its impact on residents' health, walkability, and sense of safety to optimize urban greenery planning [12,13]. Second, street canyon studies played a crucial role in microclimate analysis in high-density cities. They used semantic segmentation of SVIs to calculate the Sky View Factor [14], combined with the morphological characteristics of street canyons and GVI, to assess urban heat island effects and solar irradiation [15,16]. Third, building façade segmentation was important for improving the overall landscape quality of cities [17] and guiding urban poverty. Researchers extracted data on the appearance, structure, and style of buildings to evaluate the impact of urban design and landscape planning [18], as well as the openness of buildings to the streets [19]. Fourth, with the rise of autonomous driving (AD) technology, vehicle and pedestrian detection has become a research hotspot. This technology can accurately identify vehicles and pedestrians in complex traffic environments, ensuring the safety and reliability of AD systems [20]. Additionally, it helped quantify pedestrian and traffic flow through other object detection models, contributing to the development of intelligent transportation systems [21]. Lastly, socioeconomic factor analysis was an emerging application direction. By analyzing vehicle types, building appearances, and other information from SVIs, researchers can infer socioeconomic attributes such as income levels and population distribution in certain areas [22]. In summary, the technology of semantic segmentation of SVIs has achieved broad application and significant progress in multiple fields. Researchers are increasingly focusing on addressing practical real-world application issues, such as how to more effectively apply segmentation results to urban planning and greenery evaluation, improve the management of street spaces and buildings, optimize dynamic environmental monitoring, and expand the use of SVIs in other areas of analysis.

Semantic segmentation technology for SVIs has demonstrated a number of advantages in several areas. First, SVIs typically encompassed multiple perspectives, and by combining with semantic segmentation, they made it possible to analyze the same geo-object from different angles, helping to create a more comprehensive urban model [23]. In addition, semantic segmentation efficiently processed SVIs from large-scale urban environments, automatically identifying and classifying various geo-objects such as buildings and trees, which significantly reduced the time and cost required for manual annotation. Moreover, SVIs offered high real-time capability and frequent updates, allowing them to dynamically analyze constantly changing urban environments, whether for short-term or long-term developments, and helping decision-makers better respond to the needs of urban management [24]. Finally, semantic segmentation technology possessed low environmental dependence, high scalability, and repeatability, requiring no special sensors or expensive hardware. This technology has adapted to various application scenarios and could be reused in different regions, maintaining consistent recognition performance [25]. These advantages made SVIs suitable for a variety of applications in different geographical scenarios worldwide, which can also provide a solid foundation for the development of smart cities, giving it broad application prospects in urban management, transportation systems, and environmental monitoring.

Although semantic segmentation for SVIs has made substantial progress, it still faced several challenges in practical applications. One of the primary issues was the mutual occlusion of geo-objects such as trees, buildings, and vehicles [26], which complicated precise identification and reduced segmentation accuracy. Additionally, while this technique handled large-scale geo-objects like the sky and roads effectively, it struggled with small-scale geo-objects such as road signs and traffic lights [27], as details were frequently lost during down-sampling. The image resolution also posed a limitation, making it difficult to capture the finer details of distant or smaller geo-objects [28]. Finally, geographic positioning errors caused inconsistencies between segmented geo-objects and their actual locations, particularly in complex urban environments where spatial relationships were challenging to represent accurately [29]. These limitations restricted the broader application of SVI-based segmentation, and enhanced data processing and model refinement were still needed in the future to overcome these challenges.

1.3. Deep learning based on semantic segmentation

In recent years, significant progress has been made in the research of semantic segmentation of SVIs using Deep Learning (DL) models. In 2015, Long et al. [30] introduced Fully Convolutional Networks (FCN), marking a breakthrough in pixel-wise prediction methods and laying the foundation for SVI segmentation. Subsequently, Chen et al. [31] developed the DeepLab series, enhancing the ability to segment multi-scale geo-objects in complex scenes. In 2017, Zhao et al. [32] introduced PSPNet, which improved global context capture through pyramid pooling modules. In 2018, Chen et al. [33] further expanded

DeepLabV3+ by adding an encoder-decoder structure, optimizing the segmentation of objects with unclear boundaries. By 2020, Seo et al.'s [34] UPSNet unified instance and semantic segmentation, enhancing the recognition of geo-objects in complex urban scenes. Meanwhile, Laumer et al. [35] and Lumnitz et al. [26] applied CNN-based models to achieve efficient tree detection and improve tree segmentation respectively. By 2023, Transformer-based methods such as Swin Transformers [36] had introduced sliding window mechanisms and multi-scale feature extraction, further improving the ability to handle complex street view scenes. In addition, many other DL models have been proposed to address different challenges, making it crucial to define research objectives clearly and select the most suitable model for specific applications. Numerous SVI datasets have played a key role in training DL models, offering a robust foundation for segmenting complex scenes and various geo-objects across urban environments. Popular datasets such as ADE20 K, and Cityscapes were frequently used for trees, buildings, and sky segmentation, while CamVid and KITTI were widely applied in the segmentation of vehicles, pedestrians, and traffic signs, particularly in autonomous driving and traffic monitoring.

Throughout the development of DL methods in SVI segmentation, several advantages have emerged, offering effective solutions to various challenges. Recent models have been able to address the following issues well. First, they provided high precision in identifying and segmenting geo-objects in complex scenes, ensuring the accurate capture of details [37]. Secondly, their multi-scale processing capability allowed them to handle geo-objects of different sizes, adapting to diverse urban environments. Furthermore, these models excelled at capturing both global and local information, ensuring reliable segmentation even in cases of long-distance dependencies or occlusion [38]. Additionally, they demonstrated strong adaptability and generalization, maintaining high accuracy in different cities and under various weather conditions [39]. Finally, they excelled in edge detection, accurately identifying blurred boundaries in complex backgrounds, and effectively capturing small geo-objects such as traffic signs, thereby improving overall segmentation accuracy [40]. Looking ahead, the next step could focus on translating these segmentation results into practical applications such as traffic monitoring and urban management, where they can provide concrete benefits for smart city development.

Although DL can well address highly complex problems, it still faces several key challenges. First, these models heavily relied on large quantities of labeled data. Despite the availability of numerous public datasets, the generalization ability of models remained constrained when applied to diverse cultural, architectural, and environmental contexts. Second, large models such as Transformers demanded substantial computational resources [41], which raised the cost of training and inference, especially in resource-limited environments. Third, the interpretability of these models was often lacking, making it difficult to diagnose topological relationships of segmented geo-objects and correct spatial mismatch due to their complex internal mechanisms. Fourth, while each model demonstrated its own strengths, no universal model had yet emerged that combined all these advantages, limiting their application in complex urban environments. Thus, it is imperative to develop an adaptive DL model to accurately segment trees from SVIs and effectively identify their measurable geometrics with precise geo-locations. To address these issues, this study will utilize transfer learning to learn the standardization norms of public datasets and determine the optimal DL model capable of segmenting various street-view geo-objects. This approach will enhance model training using the custom SVIs in the study area, improving the recognition accuracy of geo-objects presenting in the complex urban environments.

1.4. Solar potential estimation on 3D urban envelopes

Solar irradiation models were used to estimate solar irradiation that could be collected at a specific location on the Earth's surface. However, traditional models, such as the Perez model and r.sun model, could only run on two-dimensional (2D) raster maps that provided surface elevation data and could not be used to estimate irradiation on vertical surfaces like building façades [42,43]. To improve accuracy, the v.sun model was developed to calculate solar irradiation on a Triangular Irregular Network, which essentially represented the 3D world in a 2.5D form, making it prone to losing 3D geometric information [44]. Building on this, the SOL model computed solar irradiation distribution on vertical surfaces by generating super points [45]. Erdélyi et al. [46] developed the SORAM model, which ignored reflected irradiance but used a high-resolution sky model and ray-tracing methods to detect obstructions. In another approach, Liang et al. [47] employed a novel GPU ray-casting technique to calculate solar irradiation on building envelopes in real time, but due to memory limitations, it could not handle large scenes.

Meanwhile, with advances in LiDAR technology and drone data acquisition, 3D point clouds data of buildings and other structures in urban environments were obtained with higher precision, especially for evaluating shading effects from tall buildings in complex urban environments. Using DSM as the model input, Lindberg et al. [48] developed a raster-based shadow calculation method to determine the shading effect of surrounding buildings on rooftops and façades, but this type of simulation was time-consuming. To address this, Vulkan et al. [49] developed a vector-based modeling approach that used 2.5D polygons to simplify building models, though this method overlooked some building attributes. Moreover, most studies focused only on shading effects between buildings, neglecting obstacles like trees, which could significantly reduce a building PV energy output [50].

As DL advanced, recent research began to explore different tree shading modeling methods, including full point clouds models, fully opaque entities, semi-transparent entities and perforated entities, to more accurately simulate the shading effects of trees on buildings. For example, Tian et al. [51] modeled the solid surfaces of buildings and terrain, treated trees as point clouds, and used the Disordered Graph Convolutional Neural Network (DGCNN) model for semantic segmentation of trees. This produced a hybrid model that outperformed other models in accurately predicting the impact of partial tree shading on PV performance. In addition, Kurdi et al. [52] introduced a 3D parameterizable and visualizable mathematical model of individual tree point clouds, using a rotational surface to simulate tree geometry, creating a layered and segmented structure that was more realistic. The relative accuracy of this model ranged from 0.4 % to 17.5 %, making it a simple and effective 3D modeling algorithm for individual trees.

Current 3D solar PV potential estimate models exhibited a high level of advancement. They can relatively accurately calculate solar irradiation distribution on complex geometric shapes, especially on building façades and rooftops, making them more reliable for architectural design and PV system evaluation. Additionally, they effectively simulated building shading effects in high-density urban environments, improving the accuracy of solar PV estimation [53]. Moreover, they can integrate various data sources to provide detailed environmental modeling [54]. However, their drawbacks included high computational resource requirements, complexity, and time-consuming modeling processes. Furthermore, many models tended to overlook factors such as building materials, reflectivity, and the impact of trees and other natural obstacles on irradiation, leading to reduced simulation accuracy. To overcome these limitations, DL-based semantic segmentation can effectively model trees from SVIs and integrated them into 3D solar irradiation models [55,56], which addressed the issue of obstacles obstructing buildings, allowing for a more accurate estimation of PV potential and enhancing the overall accuracy of the simulations.

1.5. Contribution

This study addressed a critical gap in existing solar potential models, which were constrained by limitations in incorporating the complex shading effects of surrounding geo-objects, especially trees. To overcome this challenge, we developed a new technique that uses standard geo-objects in SVIs as references to establish the relationship between pixel length and real-world dimensions. This approach enables accurate calculation of the area and location of trees, which are then systematically integrated into sunlight and shading models, allowing for fast and precise analysis. To the best of our perception, this should be the first study to rapidly construct vertical surface sunlight models for buildings while accounting for tree shading. It significantly reduces the time required to locate and analyze shading effects in urban environments, thereby enhancing the efficiency and accuracy of large-scale sunlight and shading potential assessments.

2. Methodology

We proposed a multi-module research framework for a comprehensive analysis of trees, buildings, and sunlight and shading distribution in the urban environment (Fig. 1). The first module conducts data collection in the study area and generates sampling points on the roads near EV charging stations to collect SVIs, for modelling shading effects of trees and analyzing the capability of solar charging of EVs. The second module: (i) evaluates the performance of datasets and DL models to select the optimal segmentation model for semantic segmentation of SVIs, (ii) compares the segmentation results with real-world data to fit regressions between the pixels of specific geo-objects in the SVIs and the corresponding real-world lengths of standard geo-objects [57], and (iii) determines the nearest points of buildings and trees relative to the camera based on the segmentation results, uses these points as references to establish the geospatial relationships of buildings and trees in the real world, and calibrates inaccurately segmented SVIs to ensure more precise 3D positioning of buildings and trees. The third module projects the polygonal coordinates of trees into 3D space to form 3D tree model, which is integrated with the 3D building model to estimate the spatiotemporal distribution of sunlight and shading on the 3D urban envelopes. The workflow is presented in Fig. 2.

2.1. Study area

Singapore is in southeast Asia, close to the equator, which provides Singapore sufficient sunshine throughout the year, with an abundant annual land-surface solar irradiation about 1580 kWh/m^2 [58]. While due to the limited land resources, dense buildings and a trend towards high-rise buildings in Singapore, the government has actively promoted the development of green buildings with the integration of BIPV technology. An effective penetration of solar energy requires an accurate estimation of PV potential on 3D building envelopes, affected by solar azimuth, elevation, and shadows from surrounding buildings and geo-objects such as trees. Singapore is known for its image as a “garden city” with a green coverage rate of more than 50 %, which presents both challenges and opportunities to the design and application of BIPV systems. Combining BIPV systems with EV charging stations not only provides clean energy for charging EVs, but also significantly improves the energy efficiency of buildings, optimizes space utilization, reduces energy losses, and reduces dependence on the power grid. Therefore, Singapore is an ideal study area for studying sunlight and shading distribution in a complex urban environment.

2.2. Data description and collection

2.2.1. Data sources

Road network information was obtained through OpenStreetMap. Road directions and SVIs were collected from Google Maps Platform via the purchased Directions API and Street View Static API respectively. Building data, including building height, location, and shape, were obtained from a private sector. EV charging station data were from a public data portal.

2.2.2. Generation of sampling points

We selected the vicinity of EV charging stations as the experimental area, drew a 100-meter buffer around the 252 EV charging station locations collected, and retained all road and building data within the buffer to ensure that the analysis covers key areas related to EV charging (Fig. 3). The road network was further refined by removing sections without SVIs, such as viaducts and internal building roads. Only urban streets on the ground were kept for maintaining accuracy and relevance for subsequent analysis. Finally, 3739 sampling points were then generated every 20 meters along the remaining roads to capture precise

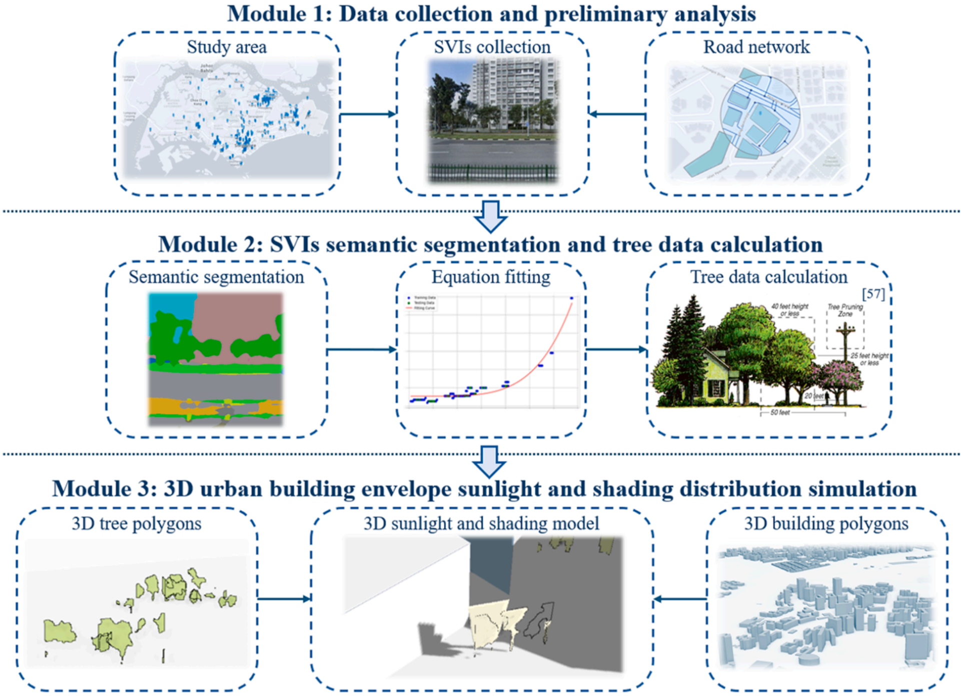

Module 1: Data collection and preliminary analysis
Study area
SVIs collection
Road network
Module 2: SVIs semantic segmentation and tree data calculation
Semantic segmentation
Equation fitting
Tree data calculation
[57]
40 feet height or less
Tree Pruning Zone
25 feet height or less
20 feet
50 feet
Module 3: 3D urban building envelope sunlight and shading distribution simulation
3D tree polygons
3D sunlight and shading model
3D building polygons

Fig. 1. Research framework for 3D urban sunlight and shading distribution simulation.

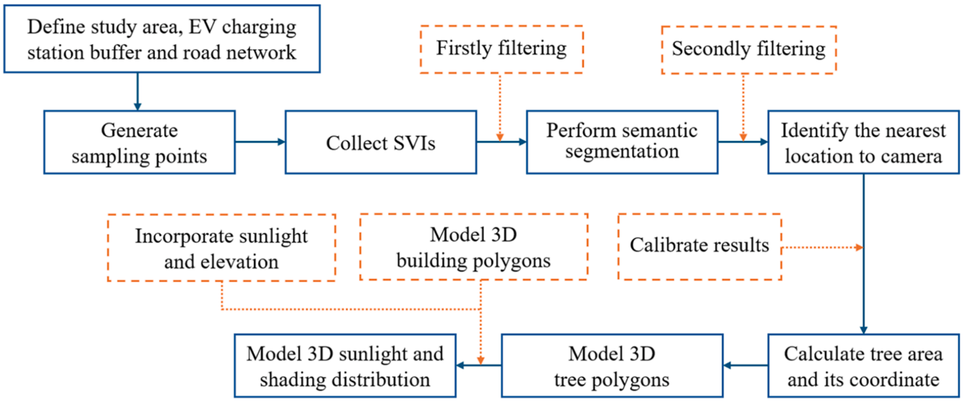

Define study area, EV charging station buffer and road network
Firstly filtering
Secondly filtering
Generate sampling points
Collect SVIs
Perform semantic segmentation
Identify the nearest location to camera
Incorporate sunlight and elevation
Model 3D building polygons
Calibrate results
Model 3D sunlight and shading distribution
Model 3D tree polygons
Calculate tree area and its coordinate

Fig. 2. Workflow for simulating the distribution of sunlight and shading.

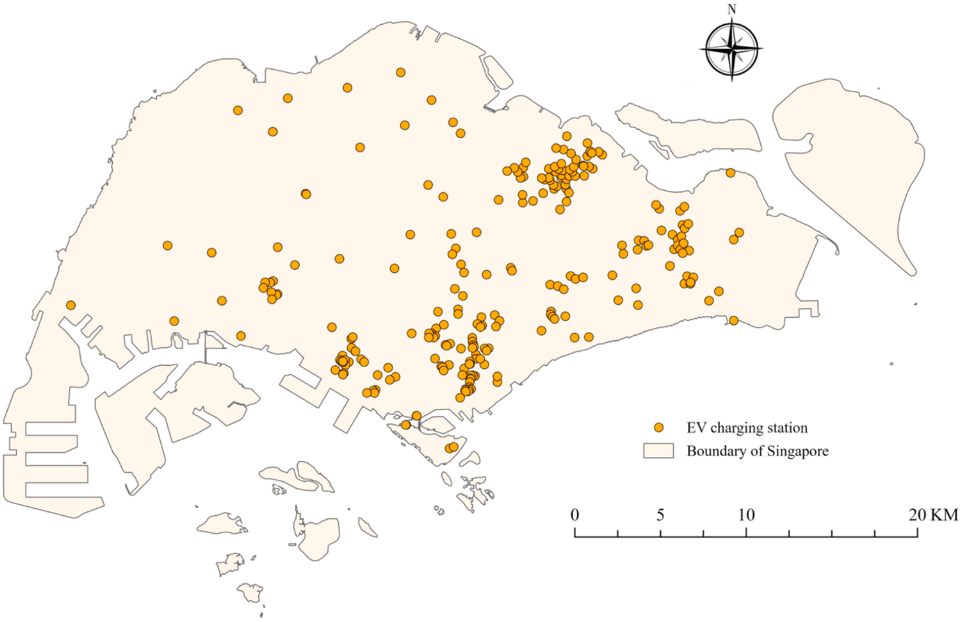


Fig. 3. Map of EV charging station distribution.

coordinates.

2.2.3. *Collection of SVIs*

SVIs were collected on both sides of the road at 90° to the road directions, which can clearly identify the spatial relationship between the position of trees relative to buildings, as well as the shading effect of trees on buildings. The size of them is set to 400 pixels × 400 pixels. The field of view (FOV) is 90°, which enables the image to cover a broader area and capture a wider range of environmental information by encompassing a wide angle across the horizontal plane. The pitch angle is 0°, and the camera is shot in a horizontal position without tilting up or down. Among the 7862 SVIs collected, there are 2 images without data, 126 images collected in the invalid locations, 378 images did not meet the requirements due to significant angel deviation, 414 images were largely obstructed (Fig. 4), and another 646 images were repeatedly collected due to proximity. Finally, 6278 SVIs were retained, accounting

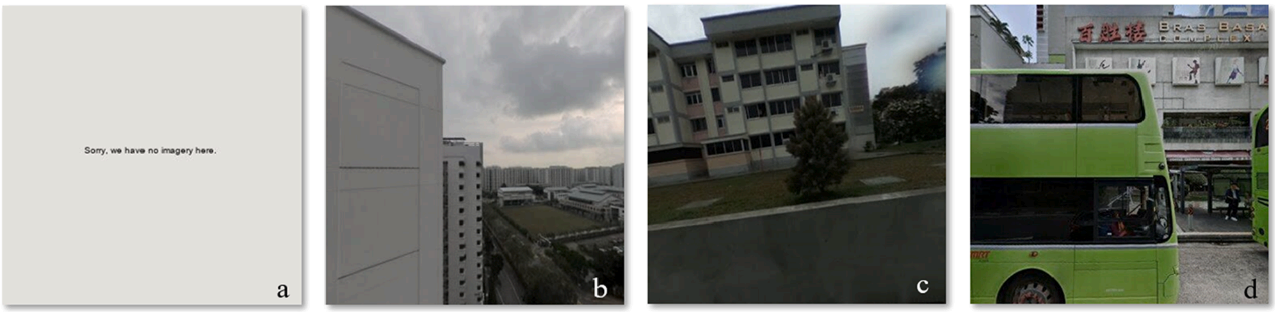


Fig. 4. Sample SVIs that were eliminated for the first time. (a) No data. (b) Invalid location. (c) Angle deviation. (d) Large occlusion.

for 79.85 % of the original images.

2.3. Recognition of geo-objects in SVIs

2.3.1. Semantic segmentation and binarization

We first evaluated the transfer learning effectiveness of the same DL model by using two public datasets, ADE20 K and Cityscapes. The results showed that the ADE20 K dataset performed more accurately in segmenting geo-objects, including buildings and trees, and was therefore selected as the foundational dataset for this study. Subsequently, we evaluated the transfer learning ability of three DL models (FCN [30], PSPNet [32], and DeepLabV3+ [33]) with ADE20K. These three models are based on CNN architectures. We chose to use them instead of more advanced models like Transformer because Transformers excel in capturing global features but lack the sensitivity to local details that CNNs provide, especially in high-resolution segmentation tasks for SVIs [59]. By comparing their performance on different geo-objects in terms of metrics such as Recall, Accuracy, Precision, Intersection over Union (IoU), and F1-Score, DeepLabV3+ was ultimately identified as the best-performing model. To further optimize segmentation performance, we adopted the DeepLabV3+ model pre-trained on the ADE20 K dataset via the GluonCV platform [60], directly utilizing the trained and validated model parameters. These parameters were optimized through a combination of Cross-Entropy Loss and Dice Loss to ensure precision and robustness in pixel-level labeling tasks. Validation results demonstrated the model's high generalization capability and outstanding performance.

During the testing phase, we collected 100 SVIs in Singapore with as much diversity as possible to annotate and divide them into a test set, using the test set to comprehensively evaluate its performance in handling specific scenarios within real urban environments. This evaluation helped us understand the model's segmentation precision and generalization capability for different geo-objects.

After performing semantic segmentation on the SVIs, buildings and trees were identified as key categories for further investigation. To analyze them more accurately, we applied a binarization process to separate them from other geo-objects. Specifically, we conducted two sets of identifications on the same image: in the first set, trees were assigned a pixel value of 1, and all other geo-objects were assigned a pixel value of 0; in the second set, buildings were assigned a pixel value of 1, with all other geo-objects again assigned a pixel value of 0. Consequently, we generated corresponding black-and-white images, which clearly highlight the target geo-object and provide a solid foundation for subsequent detailed analysis.

2.3.2. Identifying the nearest ground point between the geo-objects and the camera

When processing the buildings and trees in the binary image, we first needed to determine the nearest ground points of buildings and trees to the camera, which typically represented where these geo-objects touch the ground in the image. If the SVI was considered as the *x-y* plane in 3D space, this point demonstrated the start of the geo-object's connection with the ground. Using this as a reference point allowed us to accurately determine the real-world locations of buildings and trees in a 3D urban space and assisted with 3D modelling.

To simplify the calculation and 3D reconstruction process, we define the nearest point for each geo-object as the ground point nearest to the camera, rather than the true nearest point on the geo-object itself (Fig. 5). This choice was made because it provides a clear and fixed reference point, enabling a more consistent and efficient mapping of pixel data to real-world coordinates. Nevertheless, this assumption does not introduce significant errors or deviations because the camera is positioned on the road and is at least 5 meters away from the building. Assuming a camera height of 2 meters, the actual nearest distance error is only about 0.3 meters. As the building moves farther from the camera, this error decreases.

Theoretically, to estimate the shading of a tree on building envelopes from SVIs, the tree had to be closer to the camera than buildings. To ensure the accuracy of the results, we conducted a second round of image filtering. First, we selected the images which contain both buildings and trees. Second, we excluded situations where the arrangement of trees and buildings made it difficult to determine their front-back spatial relationship subject to the viewpoint of the camera, such as at intersections and T-junctions (Fig. 6). As a result, we deleted 669 and 1154 images during the two consecutive filtering, and the remaining 4455 valid images were used for subsequent calculations.

2.3.3. Tree position adjustment after segmentation

After eliminating objective interferences, to ensure accurate differentiation between trees and buildings, we compared the nearest points to the camera for buildings and trees in each SVI as described in Section 2.3.2. Theoretically, the distance value for trees should have been smaller than that for buildings. Thus, if a tree's value was greater than or equal to that of a building, it generally suggested errors in the segmentation results. In this case, we recalculated the tree's new reference point based on urban greening regulations and moved the tree 3 meters in front of the building [61]. It is important to note that this adjustment relied on the process of mapping pixel coordinates to real-world coordinates. We need convert the pixel coordinates to their corresponding geographic locations in the real-world coordinate system, calculate the

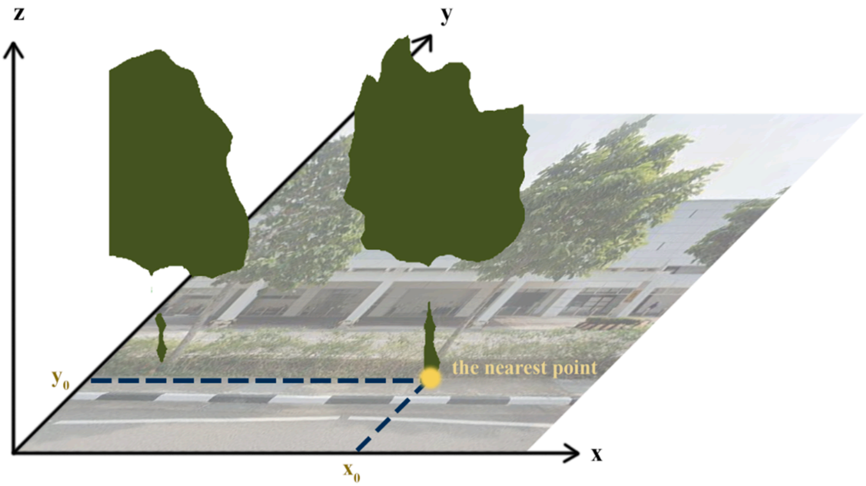


Fig. 5. The nearest ground point in SVIs using a 2D to 3D perspective.

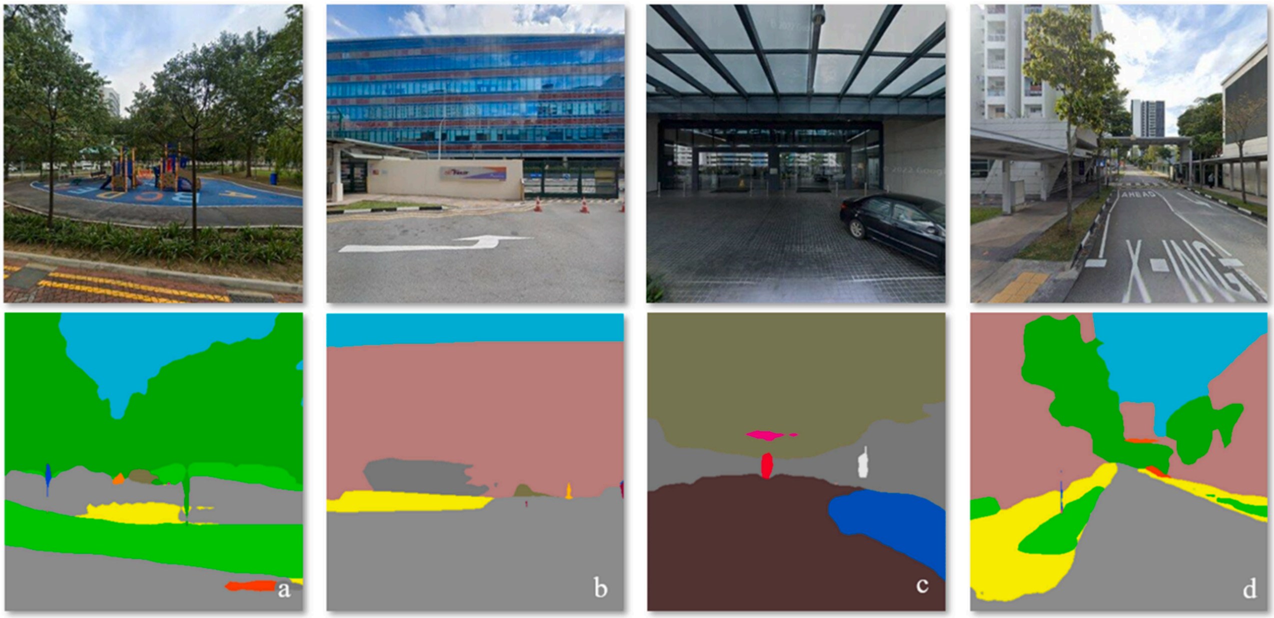


Fig. 6. Sample segmentation images that were eliminated for the second time. (a) (b) (c) No buildings or trees or both. (d) Crossroad.

3-meter offset in the real-world coordinate system, and then convert this offset back to the pixel coordinate system. The specific calculation of this mapping process will be discussed further in Section 2.4. Additionally, in the camera optical system, the vanishing point typically aligns with the camera's optical axis, which is the line from the center of the lens to the center of the image plane, usually pointing toward the image center.

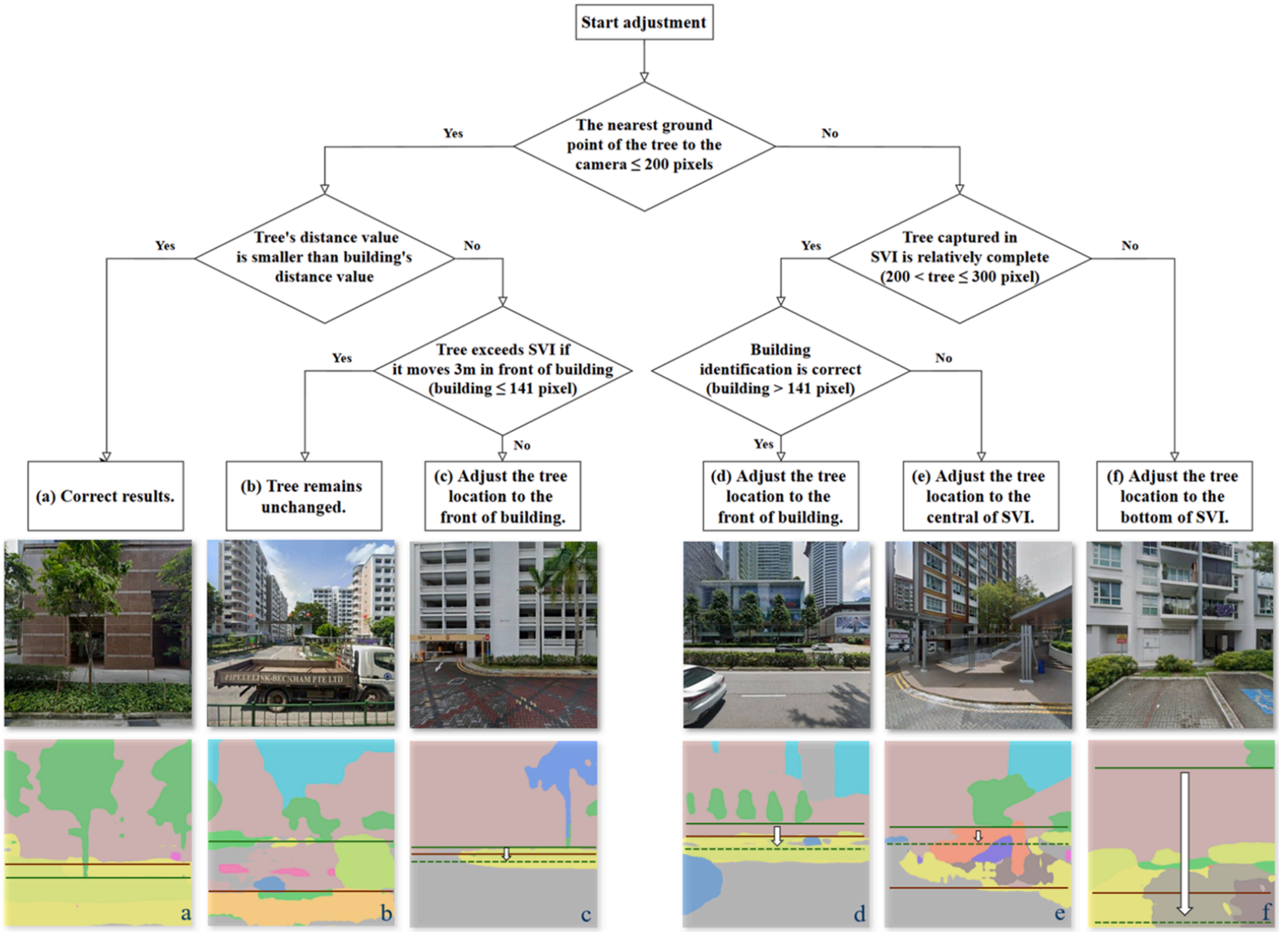

Start adjustment
The nearest ground point of the tree to the camera ≤ 200 pixels
Yes
No
Tree's distance value is smaller than building's distance value
Yes
No
Tree captured in SVI is relatively complete (200 < tree ≤ 300 pixel)
Yes
No
Tree exceeds SVI if it moves 3m in front of building (building ≤ 141 pixel)
Yes
No
Building identification is correct (building > 141 pixel)
Yes
No
(a) Correct results.
(b) Tree remains unchanged.
(c) Adjust the tree location to the front of building.
(d) Adjust the tree location to the front of building.
(e) Adjust the tree location to the central of SVI.
(f) Adjust the tree location to the bottom of SVI.
a
b
c
d
e
f

Fig. 7. Flowchart for tree location calibrations. The brown solid line represents the lowest detected building position, the green solid line represents the lowest detected tree position, and the green dotted line represents the calibrated tree position.

In the SVI collection standards we established, the camera's FOV was perfectly symmetrical, so all perspective projections converged toward the center. As a result, the vanishing point of the horizon should be located at the center of each SVI, and the ground contact points of geo-objects (e.g., buildings or trees), should have appeared below the vanishing point in the lower half of SVI. Although perspective distortion can affect the shape of objects, it does not affect the position of the vanishing point in the perspective projection. The vanishing point generated by the perspective projection is determined by the geometric relationship between the geo-object's spatial position and the camera, independent of distortion. Considering that the size of SVIs was 400 × 400 pixels, the nearest ground points of buildings and trees to the camera should not have exceeded 200 pixels [62].

In the segmentation results, we processed them according to the positional relationship between tree and building and different situations in each SVI, and the specific steps is shown in Fig. 7. When the nearest ground point of the tree to the camera did not exceed 200 pixels, firstly, if the tree's distance value was smaller than the building's distance value, the segmentation result was considered correct (Fig. 7(a)), and it aligned with the expected geometric relationship according to perspective. Secondly, if the tree's distance value was greater than or equal to the building's distance value, and moving the tree 3 meters in front of the building would have caused it to be outside the image frame, this usually indicated that other nearby geo-objects had been incorrectly identified as buildings. In this case, we retained its original value, as shown in Fig. 7(b). Thirdly, if the tree's distance value was greater than or equal to the building's distance value, and moving the tree 3 meters in front of the building was feasible, this indicated that the building was correctly identified, but the tree trunk was recognized incompletely. We moved the tree to the correct position (Fig. 7(c)).

When the nearest ground point of the tree to the camera exceeded 200 pixels, fourthly, if the tree was complete but the tree trunk was not recognized and the building was correctly identified, similar to the previous case, we moved the tree to the correct position (Fig. 7(d)). Fifthly, if the tree was complete but the tree trunk was not recognized and the nearby geo-object was incorrectly identified as a building, we could not find the correct reference for the building. In this case, we moved the tree to its furthest possible growth position, which was the center of the image, at the 200th pixel, as shown in Fig. 7(e). Sixthly, for incomplete trees where only the tip appeared in the image, and their planting points were actually outside the image frame, regardless of whether the building was correctly identified, we moved the tree's new reference point uniformly to a position 1 pixel grid from the bottom of the image. This adjustment ensured that the tree remained in the image, but when only a very small part was visible, it was not enough to cast a noticeable shadow (Fig. 7(f)). Through this calibration method, we were able to process and analyze the trees and buildings in the SVI more accurately, thus improving the practical application value of the model.

2.3.4. Pre-processing of tree area and location

Since we aimed to precisely extract the trees in the foreground of buildings from SVIs, including their area and location, additional processing steps were required for the trees in the images. To accurately quantify the area, it is necessary to calculate the number of pixels in the binary image of the trees, providing essential pixel data for estimating the coverage area. Simultaneously, to accurately determine the trees' locations, we needed to identify their polygonal outlines. After detecting the full outlines of the tree areas in the images, these outlines were converted to polygons determined by refined vertices. This process preserved the basic shape and characteristics of the trees while reducing computational complexity, providing a clear and accurate outline for subsequent spatial analysis and 3D modeling (Fig. 8).

2.4. Estimating 3D coordinates of the segmented geo-objects

2.4.1. Fitting of 3D coordinate equations

To project segmented geo-objects from 2D images into 3D space, we can use the real-world distance of known geo-objects from the camera and their pixel dimensions in the image. For example, a curb, a sidewalk, or even a building that can be fully visible in the image. Similar approaches have been conducted in several studies to address the tree dimension estimation measurement problem [63,64]. The 3D space had three axes: x, y, and z. If a geo-object is at the d-th pixel from the bottom edge of the SVI, the horizontal size in the x direction and the vertical size in the z direction for one pixel at this position can be directly calculated using the real-world width or height of the geo-object as displayed in the image based on Eqs. (1) and (2):

$$Pixel_{width} = \frac{w_{real}}{w_{pixel}} \tag{1}$$

$$Pixel_{height} = \frac{h_{real}}{h_{pixel}} \tag{2}$$

where w_{real} and h_{real} represent the real width and height of the geo-object respectively, w_{pixel} and h_{pixel} represent the total number of pixels that the geo-object occupies in the horizontal and vertical direction at the d-th pixel from the bottom edge of the SVI. Therefore, $Pixel_{width}$ and $Pixel_{height}$ represent the real width and height of one pixel at the d-th pixel from the bottom edge of the SVI. In fact, d is a variable, and at different pixel positions d in the same SVI, the total number of pixels occupied by the same geo-object may be different. Thus, we can use $Pixel_{width}$ and $Pixel_{height}$ as dependent variables to establish a fitting relationship with respect to d. Based on both theoretical and practical considerations, we selected non-linear fitting to represent the relationship between pixel positions and real-world dimensions. The inherent perspective projection in a fixed camera setup with a horizontal FOV of 90° leads to non-linear scaling of pixel dimensions across the image. Specifically, as the distance from the image center increases, the real-world dimensions corresponding to each pixel expand non-linearly. Numerous studies have shown that nonlinear distortions in photographs captured by cameras, such as perspective distortion, can be effectively modeled using polynomials. In simpler scenarios, such as with a 90-degree FOV, retaining only the lower-order terms in the polynomial is sufficient to achieve the required accuracy [65,66].

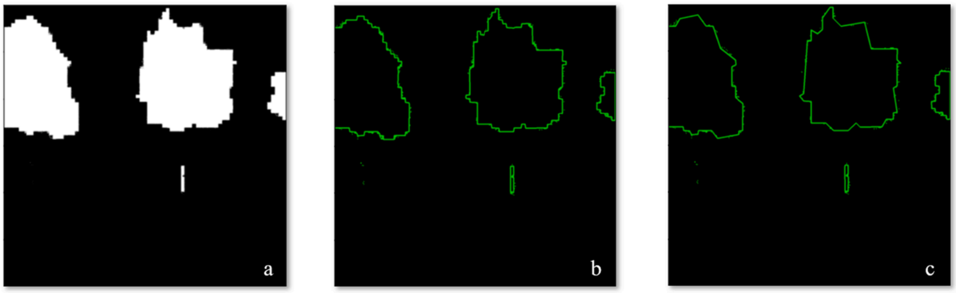


Fig. 8. The process of converting trees into polygonal outlines. (a) Binary image. (b) Complex outlines. (c) Suitable outlines.

Furthermore, compared to polynomials, the power-law function involves fewer parameters, making it equally suitable for simpler scenarios and helping to reduce the risk of overfitting. For this reason, we compared three types of low-order polynomials and a power-law function using the same dataset for fitting in the *x, y*, and *z* directions. Table 1. shows the coefficient of determination (R^2) values for each fitting method.

As demonstrated, the power-law function provides significantly better performance than the low-order polynomial function. On this basis, we can fit them and derive Eqs. (3) and (4):

$$x_d = a_x \times d^{b_x} + c_x \tag{3}$$

$$z_d = a_z \times d^{b_z} + c_z \tag{4}$$

where x_d and z_d represent $Pixel_{width}$ and $Pixel_{height}$ at different pixel positions d respectively. During the formula fitting process, 80 % of the collected data are fitted and 20 % are tested. Due to x increases with the increase of d, when d = 0, $x > 0$, the fitted equations require $a_x > 0$, $b_x > 0$, $c_x > 0$. Likewise, $a_z > 0$, $b_z > 0$, $c_z > 0$. The calculation for the y direction is relatively complicated. The real distance to the camera is determined by the focal length and FOV, which can be calculated in Eq. (5):

$$f_y = \frac{Pixel_{image}}{2 \times \tan\left(\frac{fov}{2}\right)} = 200\ (px) \tag{5}$$

where f_y is the focal length, which helps to determine the proportional relationship of the image (Fig. 9(a)). Assuming we know the real length of a geo-object in the image *l*, and the length of the pixel it occupies in the image w, we can use the principle of similar triangles (Fig. 9(b)) to calculate the real distance from the camera to the geo-object *Y*, as presented in Eq. (6):

$$\frac{l}{Y} = \frac{w}{f_y} \tag{6}$$

In addition to the calculation based on the above method, the real distance *Y* from a known building to the camera, that is, the sampling point, can be measured directly in ArcGIS Pro. Meanwhile, we know the pixel position d, *Y* and d then be fitted in Eq. (7):

$$Y_d = a_y \times d^{b_y} + c_y \tag{7}$$

where Y_d represents the real distance from the camera to the geo-object when it is at d. Note that the increasing rate of Y_d is getting faster nonlinearly with the increase of d. Therefore, the fitting parameter here needs $a_y > 0$, $b_y > 1$, $c_y > 0$. To get the varying distance of a single pixel on the *y*, further calculation is required based on Eqs. (8) and (9):

$$Y_{d+1} = a_y \times (d+1)^{b_y} + c_y \tag{8}$$

$$y_d = Y_{d+1} - Y_d \tag{9}$$

2.4.2. Calculation of tree area

Assume that the coordinates of the nearest location of the tree to the camera in the image is (u, v), the calculation method for the real area of the tree is designed using Eqs. (10)-(12):

$$Pixel_{w_0} = a_x \times v^{b_x} + c_x \tag{10}$$

$$Pixel_{h_0} = a_z \times v^{b_z} + c_z \tag{11}$$

$$Area = Pixel_{w_0} \times Pixel_{h_0} \times N_{pixel} \tag{12}$$

where $Pixel_{w_0}$ and $Pixel_{h_0}$ represent the width and the height of a single pixel of the tree respectively when it is located at v. N_{pixel} is the total number of pixels of trees in the image. Since the tree's scale is standardized based on a fixed reference point in the image (here is v), all pixels of the tree in the same image are calibrated using the same scaling factor, which makes the area of a unit pixel constant for the entire tree, without needing to account for variations in pixel area due to positional changes.

2.4.3. Calculation of tree polygonal coordinates

Assume that the 2D polygonal coordinates of the outline of the tree in an image are $[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), (x_1, y_1)]$ and the 3D polygonal coordinates of the real tree location relative to the camera location in the geographical coordinate system are $[(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \ldots, (X_n, Y_n, Z_n), (X_1, Y_1, Z_1)]$. The calculation of the tree polygonal coordinates is proposed according to Eq. (13):

$$\begin{cases} X_n = (x_n - 200) \times Pixel_{w_0} \\ Y_n = a_y \times v^{b_y} + c_y \quad n \in Z^+,\ n > 2 \\ Z_n = (y_n - v) \times Pixel_{h_0} \end{cases} \tag{13}$$

where X_1 is based on the camera location, that is, the coordinates of the midpoint of the image. When $X_1 < 0$, this point is on the left side of the sample point, and when $X_1 > 0$, this point is on the right side of the sampling point. Y_1 is a fixed value because the trees always lie on the same *y*-axis plane. Z_1 is based on the shortest distance of the tree to the camera, and its height is calculated from the ground v (Fig. 10).

2.5. 3D modeling of sunlight and shading distribution on building envelopes introducing tree shade

We selected the districts of Bishan and Toa Payoh in central Singapore as the study area because of their central location and diverse urban characteristics. These areas feature a wide range of building types, including high-rise public housing (HDB), private condominiums, and commercial buildings, providing diverse scenarios and data for analysis. In addition, the functional zoning in these regions is well-defined, encompassing residential areas, commercial zones, educational institutions, parks, and other open spaces, fully reflecting the complexity of real urban environments. This diversity not only supports multidimensional analysis but also enhances the robustness of the 3D modeling framework, making it more applicable to complex, high-density urban contexts. We conducted detailed 3D modeling of the environment within a 100-meter radius surrounding seven EV charging stations (Fig. 11). Within this selected area, we identified and simulated 49 buildings and 69 sampling points corresponding to trees. These polygons were integrated into our 3D urban model, allowing us to perform precise simulations and analyses of sunlight and shading distribution.

3. Results

3.1. Street view images segmentation results

Table 2. presents the semantic segmentation testing results of the DeepLabV3+ model on the Singapore SVI dataset. The five key metrics collectively evaluate the model's performance across different categories. The results show that the model performs excellently when segmenting large geo-objects such as buildings, houses, cars, and roads. In contrast, Recall declines when segmenting more complex structures like skyscrapers and trees, where the geometry and complexity of these

Table 1
Comparison of R^2 results for different fitting equations.

	x	Y	z
2nd-degree polynomial	0.7447	0.7311	0.8996
4th-degree polynomial	0.9509	0.8502	0.9286
6th-degree polynomial	0.9594	0.8491	0.9117
Power-law	0.9630	0.8610	0.9301

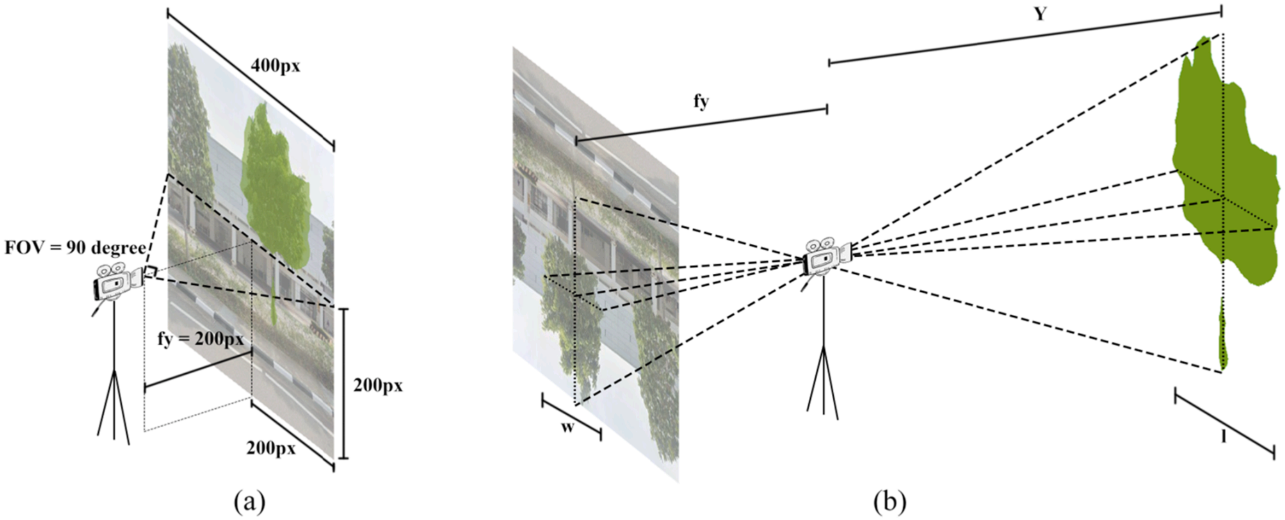


Fig. 9. The process of calculating the real distance from the camera to the point in SVI. (a) FOV calculation diagram. (b) Camera imaging diagram.

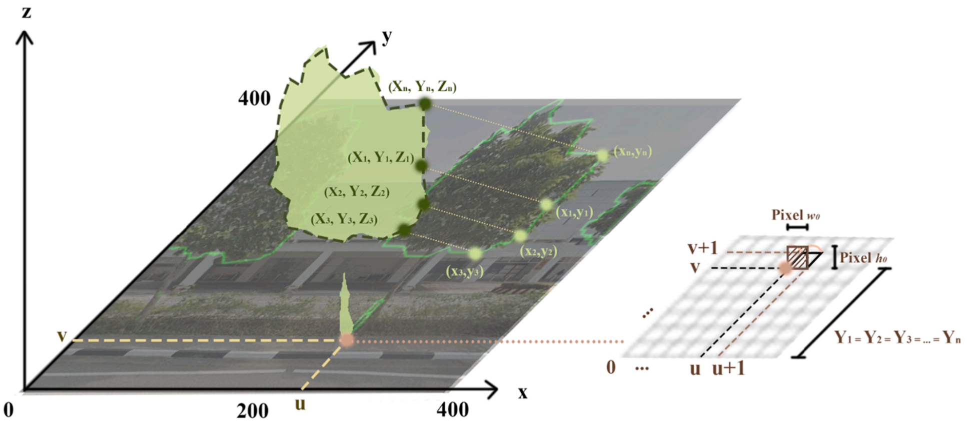


Fig. 10. Illustration of the transformation of tree coordinates from 2D to 3D.

categories pose challenges to the model's segmentation ability. While the model shows weaker performance when handling smaller geo-objects or those with rich details, such as persons, bins and fences, with segmentation results significantly inferior to those of large geo-objects. Overall, the model demonstrates high reliability in segmenting large geo-objects, including buildings and trees, which builds a solid foundation for this study on fitting formulas for calculating real distances from the camera to trees in SVIs.

3.2. Equation fitting results

3.2.1. Estimation of the distance in the x and z direction

In the *x*-direction, we selected 100 SVIs to calculate the horizontal distance represented by a pixel at different vertical pixel positions. These images included 50 pedestrian walkway grids and 50 roadway edge grids. Here, we randomly selected 80 images (80 %) for training and the remaining 20 images (20 %) for testing, which were used to build an empirically non-linear regression, as presented in Eq. (14):

$$x_d = 7.6400 \times 10^{-14} \times d^{5.2955} + 0.0149 \ (m) \tag{14}$$

The final fitting and test results are shown in Fig. 12. The model has a mean absolute error (MAE) of 0.0020 m, a mean absolute percentage error (MAPE) of 0.11, and an R^2 of 0.96. These results indicate a high accuracy of fitting, demonstrating strong explanatory power of the model, which can be used to effectively predict the relationship between pixels and real distances.

Similarly, in the *z*-direction, we also used 100 SVIs to estimate the vertical distance represented by a pixel at a given position. These images included 50 ones showing the heights of roadway edges and another 50 ones showing the heights of buildings that were fully visible within the image. From these, we randomly selected 80 % for training and the rest 20 % for testing and fitted a model expressed by Eq. (15):

$$z_d = 2.4754 \times 10^{-14} \times d^{5.5840} + 0.0131 \ (m) \tag{15}$$

The final fitting and test results are shown in Fig. 13. The MAE is 0.0102 m, MAPE is 0.15, and R^2 is 0.94, showing a strong predictive capability similar to that observed in the *x*-direction.

3.2.2. Estimation of the distance in the y direction

The *y*-direction proved more complex. We utilized 180 SVIs featuring fully identifiable building bases. Each image's building base pixel position was paired with its real-world distance from the camera, as measured in ArcGIS Pro. To ensure precise curve fitting near zero, 60 additional datasets from pedestrian walkways were included for fitting purposes only. From the 180 building datasets, 144 (80 %) were used for training and 36 (20 %) for testing, so we can get the Eq. (16):

$$Y_d = 1.6199 \times 10^{-15} \times d^{7.1093} + 3.10 \ (m) \tag{16}$$

To address instances where the building base edges were inaccurately identified, we also employed a secondary set of 36 SVIs where the

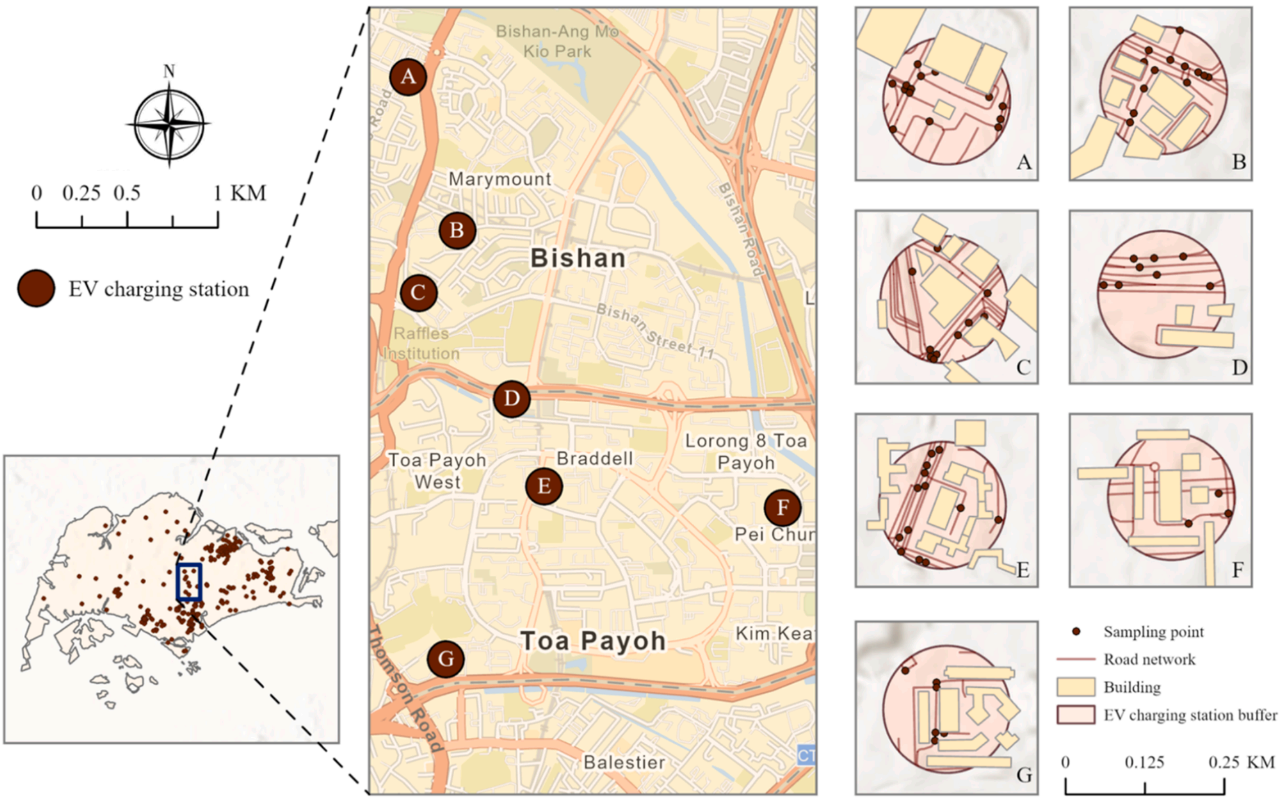


Fig. 11. Map of experimental region.

Table 2
Evaluation of the SVIs semantic segmentation using DeepLabV3+.

Category	Recall (%)	Accuracy (%)	Precision (%)	IoU (%)	F1-Score (%)
Building	**98.16**	**90.64**	**97.71**	**91.51**	**97.93**
House	97.84	89.55	94.1	89.89	95.93
Car	97.75	87.52	92.26	88.71	94.93
Road	96.46	87.12	92.89	87.86	94.64
Path	95.41	86.17	91.35	87.25	93.34
Sidewalk	93.43	82.52	88.94	85.32	91.13
Skyscraper	91.59	79.36	86.23	78.88	88.83
Tree	**91.24**	**78.86**	**85.33**	**76.29**	**88.19**
Streetlight	84.55	74.32	80.84	73.03	82.65
Person	79.79	71.52	78.36	70.46	79.07
Bus	76.73	67.55	73.05	68.77	74.84
Bin	76.85	65.28	72.33	68.23	74.52
Grass	75.25	64.82	72.05	67.36	73.62
Wall	71.84	58.01	65.32	62.36	68.43
Fence	71.5	55.27	62.5	61.45	66.70

building bases were obscured by shrubbery. In these cases, the new base position was estimated as the average of the current base and shrubbery base to verify the fitting results. The fitting and test results are displayed in Fig. 14. The first dataset achieved an MAE of 2.39 m, an MAPE of 0.14, and an R^2 of 0.86. The second dataset showed an MAE of 3.47 m, an MAPE of 0.29, and an R^2 of 0.76. The deviations observed in the y-direction fitting results are larger compared to the x and z directions, mainly due to the inherent scale differences, amplifying the perception of error in the y-direction. However, the accuracy remains high. The use of the same equation form for regression across all three directions ensures consistency, simplicity, and computational efficiency, especially for large-scale urban applications. The first dataset's test results are more accurate compared to the second, as the second dataset experiences visual obstructions, which introduce greater inaccuracies and variability.

3.3. *Tree location calibration results*

Based on the equation fitting results, the distance from the bottom of the SVI to the camera is 3.10 m. When a geo-object is located at the 140th pixel vertically from the bottom of the SVI, its distance to the camera is 6.03 m. When a building identified in the SVI is at the 140th pixel, if trees are moved forward by 3 m, they will not appear in the SVI. Therefore, the 140th pixel is used as a judgment line. Among 4455 SVIs, when the building location exceeded the 140th pixel, trees in 1595 SVIs were moved because they were further from the camera than the buildings. When the building location was equal to or less than the 140th pixel, trees in 144 images were determined to be out of the SVI because the planting point was too far from the ground and were moved to the 1st pixel from the bottom. Additionally, trees in 370 images remained unchanged because their positions are reasonable. A total of 1739 images, accounting for 39.03 %, were calibrated.

Table 3. summarizes the statistical data of pixel position and tree area before and after calibration. For pixel position, both the average and median values have decreased, and the standard deviation and interquartile range have also slightly decreased, indicating an increased concentration trend of the data. Moreover, changes in skewness and kurtosis suggest that the data distribution has become more left-skewed, with more prominent extreme values. In terms of tree area, there has been a significant reduction after calibration, the distribution has become more concentrated, and the significant reductions in skewness and kurtosis indicate a trend towards a more symmetrical and normal distribution, with a noticeable decrease in extreme values.

The box plot in Fig. 15(a) clearly shows that there were many outliers before calibration, with values significantly higher than other data points and a wide distribution range. After calibration, the number of outliers is significantly reduced, and the data distribution tends to be concentrated in a lower range of values. To quantify the changes in area before and after calibration, we calculate the relative change for each sampling point using the following formula:

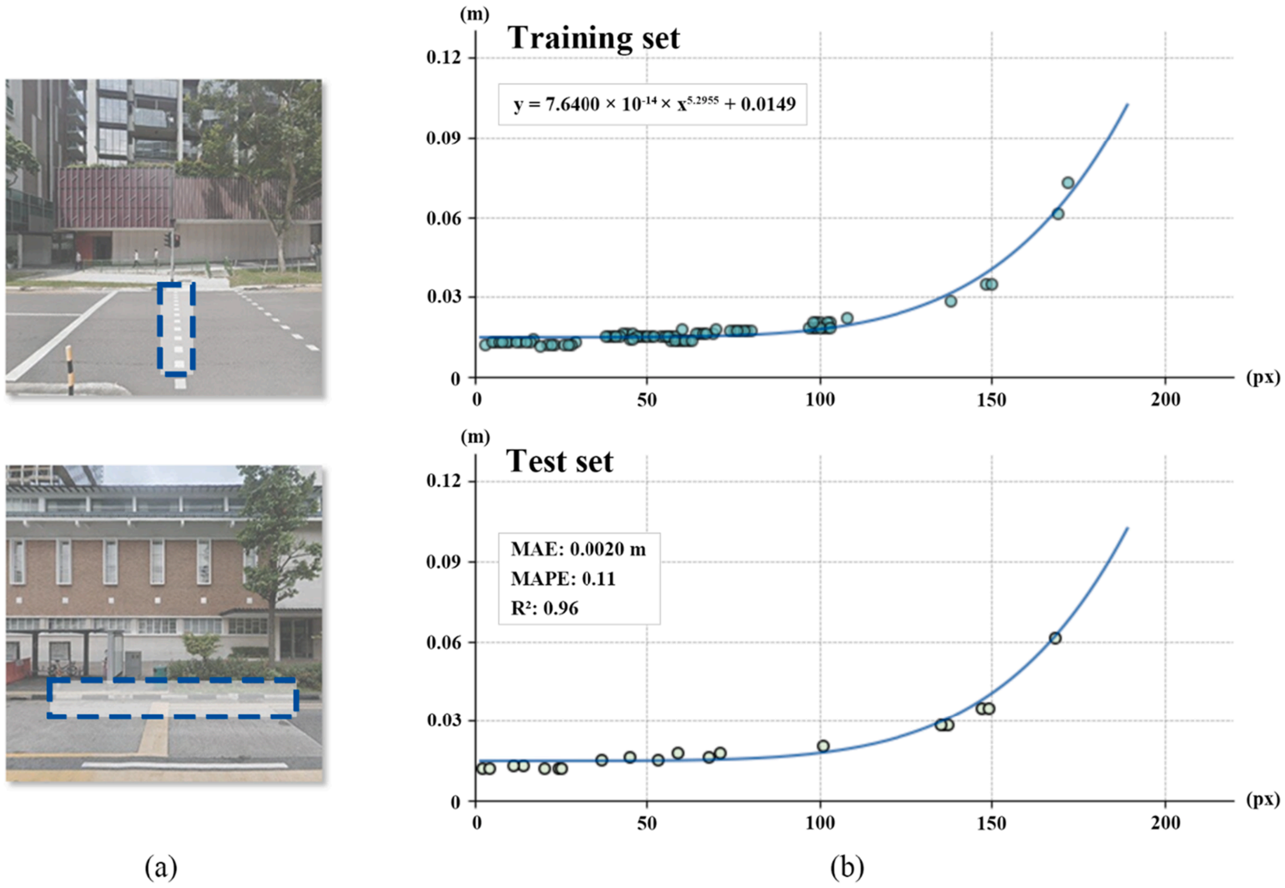


Fig. 12. Equation fitting in the *x* axis of the 3D space (referring to Fig. 10). (a) The standard geo-objects (pedestrian walkway grids and roadway edge grids) used for fitting spatial width equation. (b) Results of training and test set in fitting.

$$\textit{Relative change} = \frac{|A_{modified} - A_{original}|}{A_{original}} \tag{17}$$

Its value ranges from 0 to 1, with values closer to 0 indicating smaller changes and values closer to 1 indicating larger adjustments during the calibration process. As shown in Fig. 15(b), the distribution of relative changes is mostly concentrated in the lower range (0–0.2), indicating that the original area calculations were relatively accurate. In contrast, more outliers are found in the higher relative changes (0.8–1), suggesting they underwent relatively large changes during calibration. This indicates an improvement in data quality, aligning more closely with the expected distribution range.

3.4. Tree and building information for EV charging station buffer

Based on the recalibrated data for tree area calculations, we allocated these data to each sampling point. Each buffer area centralized at the locations of EV charging stations contained approximately 10 sampling points. By summing up tree areas calculated from the sampling points within each buffer, we obtained a rough estimate of the totally projected 2D tree area for each region, as illustrated in Fig. 16. The eastern and southwestern coastal areas of Singapore had more tree cover, whereas the city center and other highly urbanized areas had relatively less tree coverage.

Additionally, we calculated the total building surface area and average building height within each EV charging station buffer using known building data, as shown in Fig. 17. Areas with more EV charging stations tend to have higher urbanization, greater building density, and larger total building surface areas. However, it's also important to consider that high-density buildings may limit the solar irradiation reception area for some buildings. This means that even though there is a large building surface area, the actual effective area available for solar power generation might be restricted. Comparing buildings and trees, in rapidly urbanizing areas, building development generally takes precedence over green space conservation. Therefore, building area and height are more indicative of urban characteristics, while tree coverage is relatively random and much smaller than the building area, with tree shading having a greater impact on lower buildings.

3.5. Projecting 2D trees into 3D urban model

First, the latitude and longitude data of the sampling points are converted from the geographic coordinate system WGS84 to the projected coordinate system UTM Zone 48 N, representing the *x* and *y* coordinates in meters. This conversion ensures consistency with the *z*-direction height data, thereby avoiding unit discrepancies between different coordinate systems that could affect model accuracy. Next, a rotation matrix is constructed using the directional angle and relative offset data to calculate the absolute coordinates of tree polygons in 3D space, achieving synchronized positioning and height alignment. This process ensures that the position and height of trees within the study area accurately reflect actual conditions (Fig. 18).

Following this, the 3D tree polygons can be imported into QGIS for 3D visualization. Concurrently, 3D building models can be generated based on location and height data, integrating both tree and building models into a unified urban scene. Based on this integrated 3D model, sunlight and shading projections of trees on buildings at various times are simulated (Fig. 19). The trees appear as 2D planes because we define the position of each tree based on the nearest ground point to the camera. This means that all the trees in the same image share the same y-

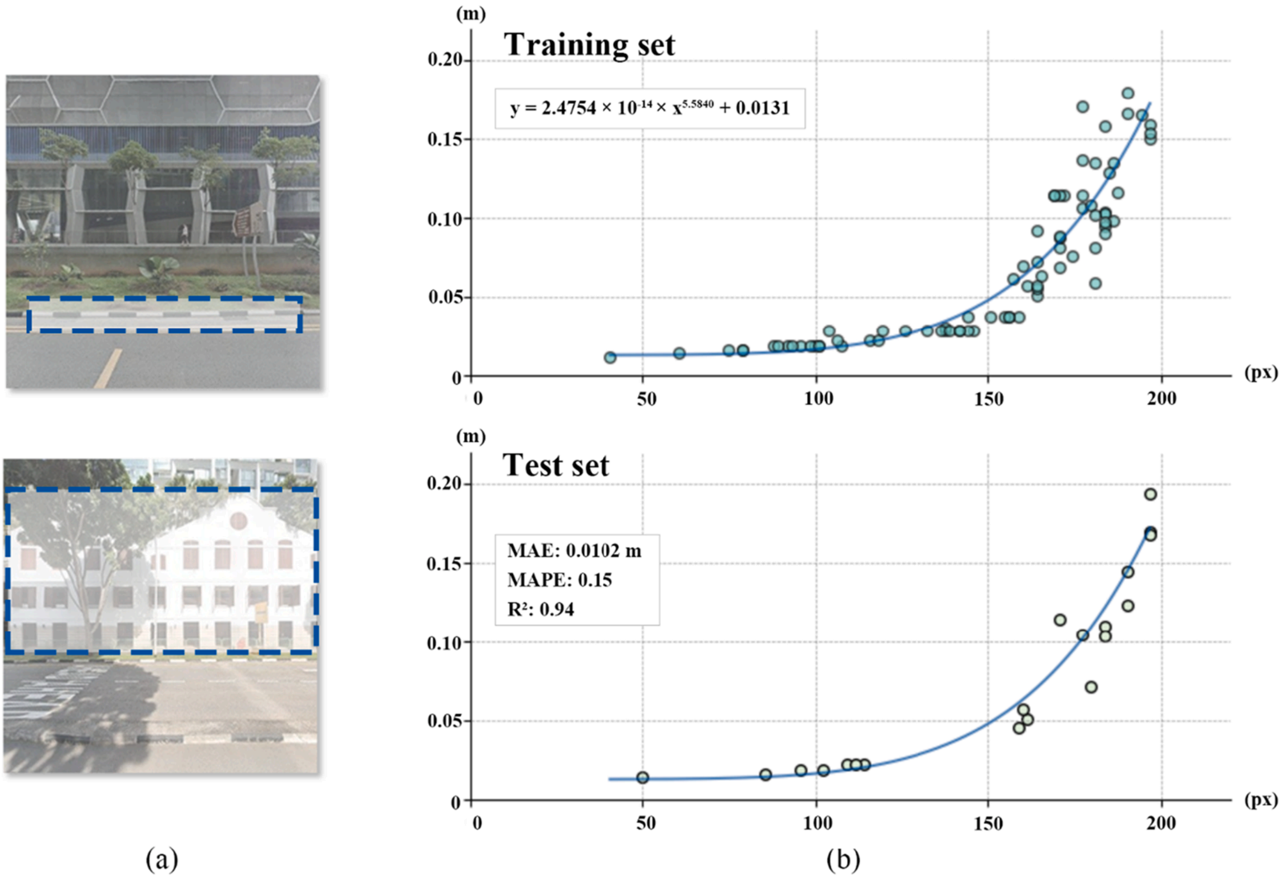


Fig. 13. Equation fitting in the *z* axis of the 3D space (referring to Fig. 10). (a) The standard geo-objects (roadway edges and buildings) used for fitting spatial height equation. (b) Results of training and test set in fitting.

axis value, which gives the impression that they are positioned on a 2D plane, despite being modeled in 3D.

4. Discussion

This study proposes a novel framework that performs DL-based segmentation of SVIs, which efficiently constructs 3D geometries of trees by fitting regressions between pixel lengths and real-world lengths based on semantic segmentation results, achieving seamless integration of 3D tree models and 3D building models on a large urban scale. We selected the DeepLabV3+ model for semantic segmentation and optimized the adaptability of the original pre-trained model to the SVI of the study area through transfer learning. The model demonstrated excellent performance, with IoU of 91.51 % for buildings and 76.29 % for trees, with F1-scores of 97.93 % and 88.19 %, respectively. Using segmentation results, we identified the nearest ground points of buildings and trees relative to the camera to determine their spatial relationships in the SVI, and calibrated inaccurate segmentations by adjusting initial tree polygons in 39.03 % of the SVIs. Finally, the calibrated tree polygons were projected into 3D space to construct tree models, which were integrated with 3D building models to form a unified 3D urban environment and precisely simulate sunlight and shading distribution on building envelopes.

By processing SVIs efficiently, this study achieves precise extraction of 3D geometric features of trees on a city scale, providing an innovative pathway for constructing fundamental geospatial datasets. Unlike conventional methods reliant on LiDAR point clouds or field measurements, which often involve computationally intensive processes for depth estimation and 3D modeling, our approach simplifies the workflow by employing deep learning solely for semantic segmentation. This enables efficient extraction of trees, while lightweight regression models are used to estimate 3D geometric properties, resulting in a comprehensive large-scale database of tree distribution and morphology. By minimizing computational overhead without sacrificing accuracy, the framework is particularly well-suited for large-scale urban applications. Moreover, the widespread availability and periodic updates of SVIs further enhance the timeliness and adaptability of urban tree monitoring and modeling. This ensures the framework remains accessible and scalable, even for cities with limited computational infrastructure. By integrating SVI with DL-based feature extraction, the study significantly lowers the barriers to geospatial data acquisition, offering a more economical, flexible, and scalable solution for urban applications. This advancement presents a substantial breakthrough from traditional geospatial dataset construction methods and lays a robust foundation for the future design and management of smart cities.

Furthermore, our method addresses key challenges in projecting tree pixels segmented from 2D SVIs into measurable 3D space through equation fitting and calibration. This capability overcomes the limitations of traditional remote sensing methods in dense urban environments with complex obstructions and restricted perspectives. It enables detailed studies of tree spatial distribution, morphology, and ecological impact, which are essential for addressing urban sustainability challenges. For instance, in 3D solar PV potential estimation, the method optimizes envelope PV design, enhancing energy utilization efficiency in buildings. In urban heat island effect analysis, it enables precise simulations of tree shading impacts, providing a scientific basis for microclimate regulation. In urban greening and spatial planning, precise tree coverage data inform policy decisions and promote ecological urban development. Overall, through a systematic framework and efficient methodological innovations, this study expands the scope of urban

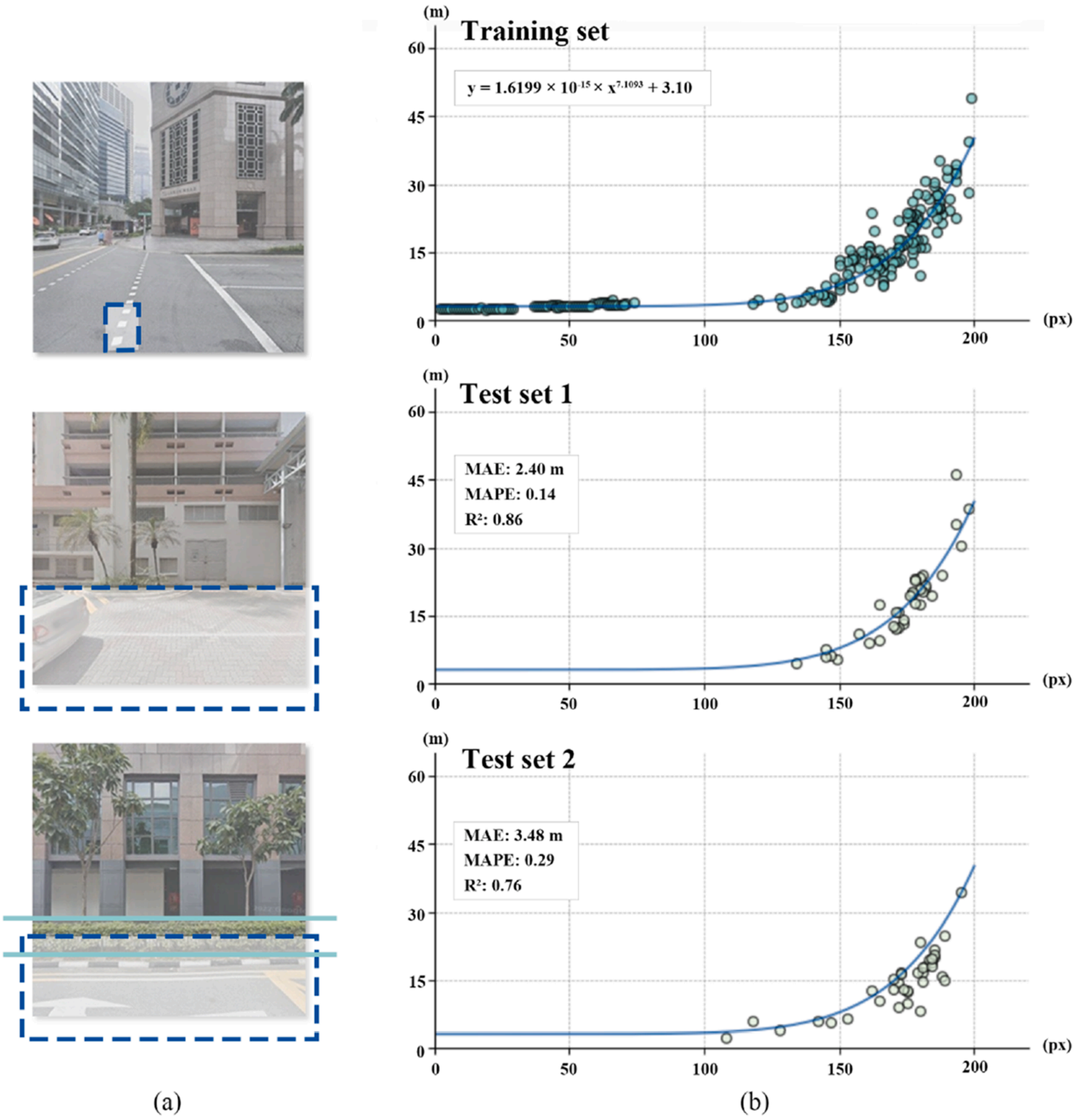


Fig. 14. Equation fitting in the *y* axis of the 3D space (referring to Fig. 10). (a) The standard geo-objects (the distance between camera and pedestrian walkway grids, buildings not obscured by shrubbery and buildings obscured by shrubbery) used for fitting spatial depth equation. (b) Results of training and test set in fitting.

Table 3
Comparison of statistical data before and after calibration.

	Pixel position		Tree area	
	Before	After	Before	After
Mean	164.46	145.43	314.62	92.26
Median	170	155	90.27	45.89
Std Dev	41.98	40.96	2282.1	141.12
IQR	47	43	190.66	106.56
Skewness	−0.11	−1.79	28.73	9.01
Kurtosis	2.83	3.63	1016.49	211.72

spatial analysis, opening new possibilities for addressing complex urban challenges and supporting sustainability goals.

Nevertheless, this study has certain limitations and uncertainties that warrant discussion. First, various types of errors exist in SVIs, which lead us to exclude a portion of the data, retaining only the more accurate and research-valuable images. While this filtering process enhanced data quality to some extent, it remains constrained by the current coverage and quality of available images. If future updates to SVIs could ensure higher frequency, improved quality, and comprehensive regional coverage, it would significantly strengthen similar research and enable further refinement in precision. Second, the estimation of the 3D features of trees relies solely on the perspective provided by SVIs, rather than a full 360-degree 3D reconstruction. This inherent limitation of data perspective may affect the completeness of the results, although its

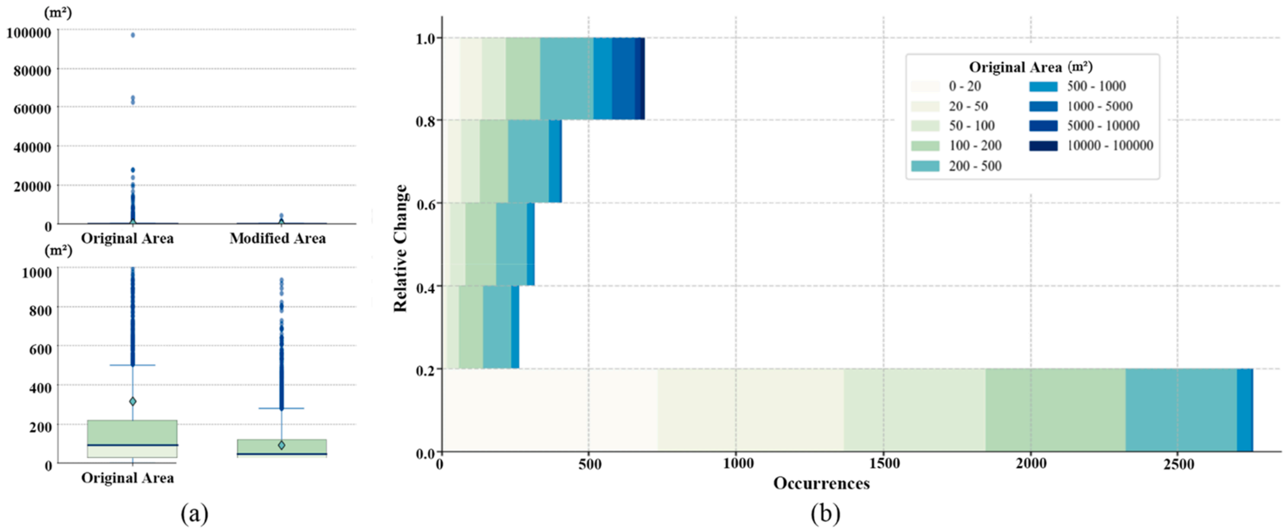


Fig. 15. Comparison of before and after calibration. (a) Boxplots of different area ranges. (b) Relative change of tree area for all sampling points.

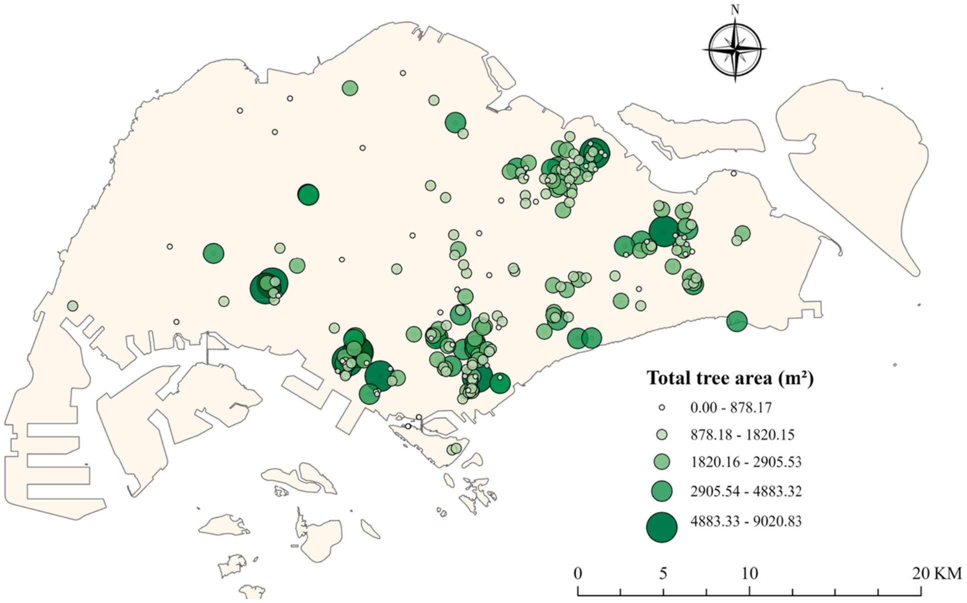


Fig. 16. Summarized information of total projected 2D tree area in each EV charging station buffer.

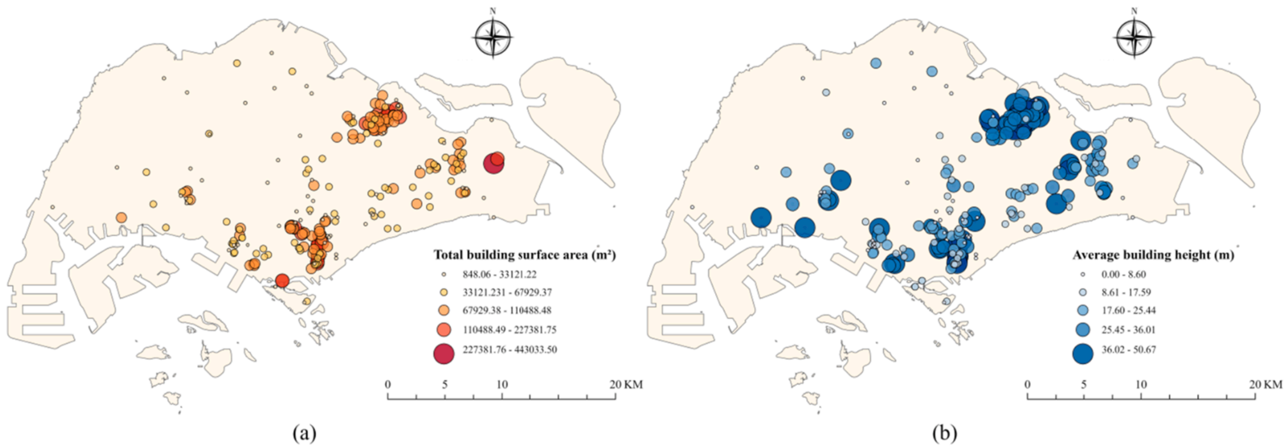
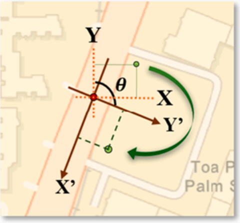


Fig. 17. Summarized information of building in each EV charging station buffer. (a) The total building rooftop and façade area. (b) The average building height.



$$\begin{cases} X'_a = lat + X_a \cdot \cos\theta + Y_a \cdot \sin\theta \\ Y'_a = lon - X_a \cdot \sin\theta + Y_a \cdot \cos\theta \end{cases}$$

Latitude	Longitude	Rotation	Relative Position of Tree Polygon
lat_geo	*lon_geo*	θ	PolygonZ (($X_a\ Y_a\ Z_a, X_b\ Y_b\ Z_b, \dots\ \dots, X_a\ Y_a\ Z_a$))
103.8470	1.3400	110	PolygonZ ((9.03 9.42 9.3, 9.03 9.42 9.25, … …, 9.03 9.42 9.3))

Latitude	Longitude	Absolute Position of Tree Polygon
lat	*lon*	PolygonZ (($X'_a\ Y'_a\ Z_a, X'_b\ Y'_b\ Z_b, \dots\ \dots, X'_a\ Y'_a\ Z_a$))
371730.50	148147.65	PolygonZ ((371736.26 148135.94 9.30, 371736.26 148135.94 9.25, … …, 371736.26 148135.94 9.30))

Fig. 18. Calculation of the absolute coordinates of tree polygons.

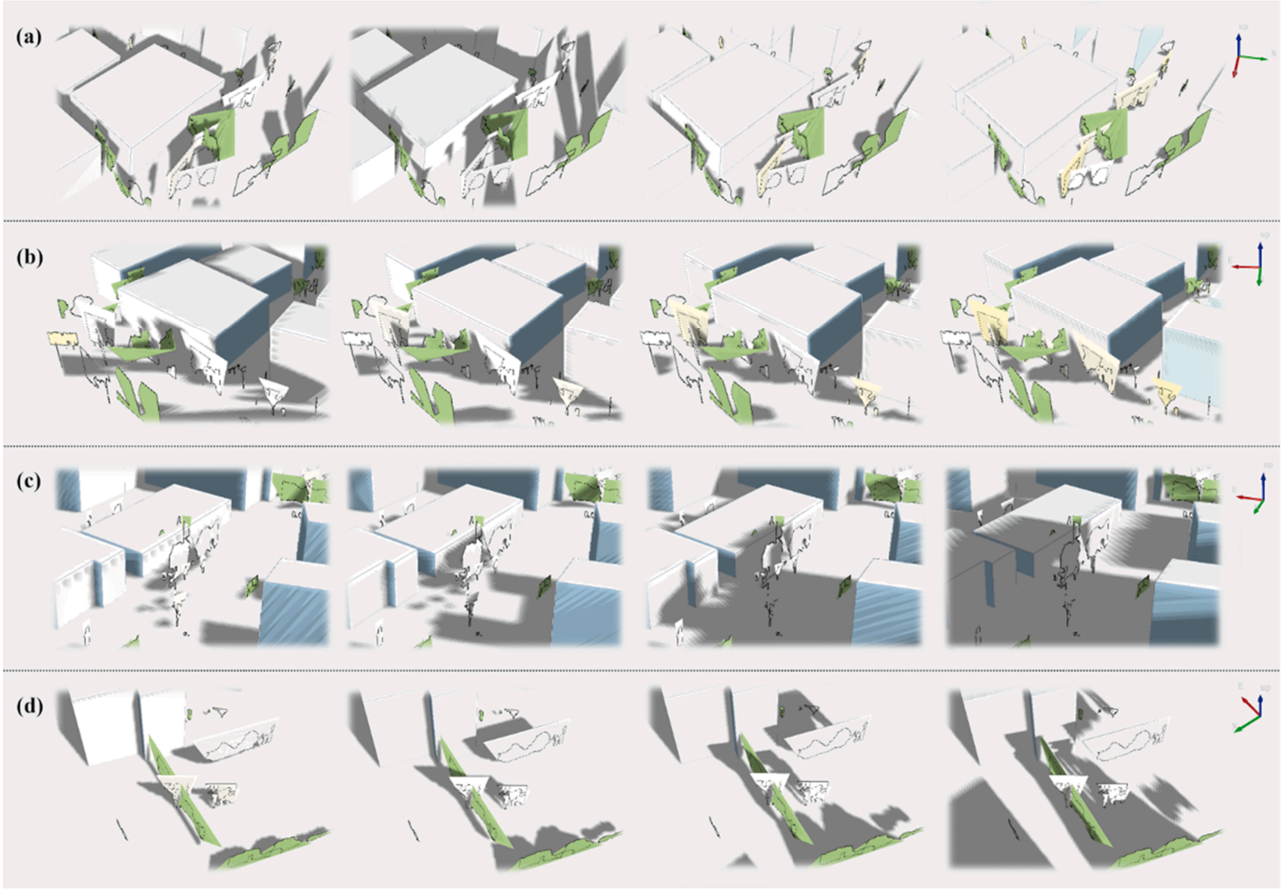


Fig. 19. Spatiotemporal distributions of sunlight and shading at different times of one day. (a) (b) 7am - 10am. (c) (d) 2pm - 5pm.

overall impact on accuracy is minimal, primarily in scenarios with substantial spatial occlusions. The method of assuming the nearest ground point as the reference point for the tree's position, while effective, could be optimized in future work. More accurate camera positioning and perspective correction could refine the location estimation and further improve the precision of the 3D reconstruction. Lastly, our proposed method cannot be directly applied to certain special scenarios when vegetations are not rooted on the ground while planted on building façades (e.g., vertically greening trees), which will produce notable errors due to the mismatch between the referenced spatial scales subject to the complexity of 3D model structure. This limitation hampers the model's ability to accurately estimate the spatial configuration and height of such trees. Nevertheless, vertically greening trees account for only a negligible portion of our dataset (0.34 %), resulting in a minimal impact on the overall outcomes. Therefore, despite these limitations, considering the robustness of the study's framework and results, these constraints have a limited effect on the findings and conclusions. The proposed methodology remains highly credible and holds substantial value for practical applications.

5. Conclusion

In conclusion, this study presents an innovative framework that integrates 3D tree models with urban building datasets to simulate sunlight and shading distribution on buildings. The framework uses advanced DL techniques and geospatial data to bridge the gap between 2D imaging and 3D modeling, offering a solution that is accurate, efficient, and cost-effective. With its adaptability, the model can be applied to various urban contexts globally, contributing to smart city development and sustainable goals. This study offers valuable perspectives on addressing urban sustainability challenges and leveraging technology for energy-efficient urban systems.

CRediT authorship contribution statement

Shu Wang: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Rui Zhu:** Writing – original draft, Supervision, Methodology. **Yifan Pu:** Resources, Methodology. **Man Sing Wong:** Writing – review & editing, Supervision. **Yanqing Xu:** Writing – review & editing, Supervision. **Zheng Qin:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

M.S. Wong thanks the funding support from the General Research Fund (Grant No. 15603920 and 15609421), and the Collaborative Research Fund (Grant No. C5062–21GF) from the Research Grants Council, Hong Kong, China; and the funding support from the Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hong Kong, China (Grant No. 1-BBG2).

Data availability

Data will be made available on request.

References

[1] J. Rogelj, M. Den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, M. Meinshausen, The Paris Agreement climate proposals need a boost to keep warming well below 2 °C, Nature 534 (2016) 631–639, https://doi.org/10.1038/nature18307.

[2] R. Cohen, P.C. Eames, G.P. Hammond, M. Newborough, B. Norton, Briefing: the 2021 Glasgow Climate Pact: steps on the transition pathway towards a low carbon world, Proc. Inst. Civ. Eng.-Energy. 175 (2022) 97–102, https://doi.org/10.1680/jener.22.00011.

[3] C.J. Ramanan, K.H. Lim, J.C. Kurnia, S. Roy, B.J. Bora, B.J. Medhi, Towards sustainable power generation: recent advancements in floating photovoltaic technologies, Renew. Sustain. Energy Rev. 194 (2024) 114322, https://doi.org/10.1016/j.rser.2024.114322.

[4] Q. Hassan, S. Algburi, A.Z. Sameen, J. Tariq, A.K. Al-Jiboory, H.M. Salman, B. M. Ali, M. Jaszczur, A comprehensive review of international renewable energy growth, Energy Built Environ (2024), https://doi.org/10.1016/j.enbenv.2023.12.002 e202312002.

[5] E.D. Rounis, A.K. Athienitis, T. Stathopoulos, BIPV/T curtain wall systems: design, development and testing, J. Build. Eng. 42 (2021) 103019, https://doi.org/10.1016/j.jobe.2021.103019.

[6] K. Fath, J. Stengel, W. Sprenger, H.R. Wilson, F. Schultmann, T.E. Kuhn, A method for predicting the economic potential of (building-integrated) photovoltaics in urban areas based on hourly radiance simulations, Sol. Energy 116 (2015) 357–370, https://doi.org/10.1016/j.solener.2015.03.023.

[7] G. Yu, H. Yang, Z. Yan, M.K. Ansah, A review of designs and performance of façade-based building integrated photovoltaic-thermal (BIPVT) systems, Appl. Therm. Eng. 182 (2021) 116081, https://doi.org/10.1016/j.applthermaleng.2020.116081.

[8] M. Fogl, V. Moudrý, Influence of vegetation canopies on solar potential in urban environments, Appl. Geogr. 66 (2016) 73–80, https://doi.org/10.1016/j.apgeog.2015.11.011.

[9] P.Y. Tan, M.R.B. Ismail, Building shade affects light environment and urban greenery in high-density residential estates in Singapore, Urban For. Urban Green 13 (2014) 771–784, https://doi.org/10.1016/j.ufug.2014.05.011.

[10] M. Münzinger, N. Prechtel, M. Behnisch, Mapping the urban forest in detail: from LiDAR point clouds to 3D tree models, Urban For. Urban Green 74 (2022) 127637, https://doi.org/10.1016/j.ufug.2022.127637.

[11] R. Wang, J. Peethambaran, D. Chen, Lidar point clouds to 3-D urban models: a review, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (2018) 606–627, https://doi.org/10.1109/jstars.2017.2781132.

[12] A. Khan, W. Asim, A. Ulhaq, R.W. Robinson, A multiview semantic vegetation index for robust estimation of urban vegetation cover, Remote Sens 14 (2022) 228, https://doi.org/10.3390/rs14010228.

[13] D. Ki, S. Lee, Analyzing the effects of green view index of neighborhood streets on walking time using Google Street View and deep learning, Landsc. Urban Plan. 205 (2021) 103920, https://doi.org/10.1016/j.landurbplan.2020.103920.

[14] Y. Xia, N. Yabuki, T. Fukuda, Sky view factor estimation from street view images based on semantic segmentation, Urban Clim 40 (2021) 100999, https://doi.org/10.1016/j.uclim.2021.100999.

[15] F.Y. Gong, Z.C. Zeng, F. Zhang, X. Li, E. Ng, L.K. Norford, Mapping sky, tree, and building view factors of street canyons in a high-density urban environment, Build. Environ. 134 (2018) 155–167, https://doi.org/10.1016/j.buildenv.2018.02.042.

[16] K. Chen, M. Tian, J. Zhang, X. Xu, L. Yuan, Evaluating the seasonal effects of building form and street view indicators on street-level land surface temperature using random forest regression, Build. Environ. 245 (2023) 110884, https://doi.org/10.1016/j.buildenv.2023.110884.

[17] X. Liang, J.H. Chang, S. Gao, T. Zhao, F. Biljecki, Evaluating human perception of building exteriors using street view imagery, Build. Environ. 263 (2024) 111875, https://doi.org/10.1016/j.buildenv.2024.111875.

[18] F. Xu, M.S. Wong, R. Zhu, J. Heo, G. Shi, Semantic segmentation of urban building surface materials using multi-scale contextual attention network, ISPRS J. Photogramm. Remote Sens. 202 (2023) 158–168, https://doi.org/10.1016/j.isprsjprs.2023.06.001.

[19] Q. Rui, H. Cheng, Quantifying the spatial quality of urban streets with open street view images: a case study of the main urban area of Fuzhou, Ecol. Indic 156 (2023) 111204, https://doi.org/10.1016/j.ecolind.2023.111204.

[20] K. Muhammad, T. Hussain, H. Ullah, J. Del Ser, M. Rezaei, N. Kumar, M. Hijji, P. Bellavista, V.H.C. de Albuquerque, Vision-based semantic segmentation in scene understanding for autonomous driving: recent achievements, challenges, and outlooks, IEEE Trans. Intell. Transp. Syst. 23 (2022) 22694–22715, https://doi.org/10.1109/tits.2022.3207665.

[21] C. Zhang, W. Ding, G. Peng, F. Fu, W. Wang, Street view text recognition with deep learning for urban scene understanding in intelligent transportation systems, IEEE Trans. Intell. Transp. Syst. 22 (2020) 4727–4743, https://doi.org/10.1109/tits.2020.3017632.

[22] F. Zhang, L. Wu, D. Zhu, Y. Liu, Social sensing from street-level imagery: a case study in learning spatio-temporal urban mobility patterns, ISPRS J. Photogramm. Remote Sens. 153 (2019) 48–58, https://doi.org/10.1016/j.isprsjprs.2019.04.017.

[23] L. Wang, C. Hou, Y. Zhang, J. He, Measuring solar radiation and spatio-temporal distribution in different street network direction through solar trajectories and street view images, Int. J. Appl. Earth Obs. Geoinf. 132 (2024) 104058, https://doi.org/10.1016/j.jag.2023.104058.

[24] N. He, G. Li, Urban neighbourhood environment assessment based on street view image processing: a review of research trends, Environ. Challenges. 4 (2021) 100090, https://doi.org/10.1016/j.envc.2021.100090.

[25] H. Yue, Investigating the influence of streetscape environmental characteristics on pedestrian crashes at intersections using street view images and explainable machine learning, Accid. Anal. Prev. 205 (2024) 107693, https://doi.org/10.1016/j.aap.2024.107693.

[26] S. Lumnitz, T. Devisscher, J.R. Mayaud, V. Radic, N.C. Coops, V.C. Griess, Mapping trees along urban street networks with deep learning and street-level imagery, ISPRS J. Photogramm. Remote Sens. 175 (2021) 144–157, https://doi.org/10.1016/j.isprsjprs.2021.01.016.

[27] L. Rita, M. Peliteiro, T.C. Bostan, T. Tamagusko, A. Ferreira, Using deep learning and Google Street View imagery to assess and improve cyclist safety in London, Sustainability 15 (2023) 10270, https://doi.org/10.3390/su151310270.

[28] F. Biljecki, K. Ito, Street view imagery in urban analytics and GIS: a review, Landsc. Urban Plan. 215 (2021) 104217, https://doi.org/10.1016/j.landurbplan.2021.104217.

[29] H.E. Pang, F. Biljecki, 3D building reconstruction from single street view images using deep learning, Int. J. Appl. Earth Obs. Geoinf. 112 (2022) 102859, https://doi.org/10.1016/j.jag.2022.102859.

[30] J. Long, E. Shelhamer, T. Darrell, 2015. Fully convolutional networks for semantic segmentation. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2015) 3431–3440. https://doi.org/10.1109/cvpr.2015.7298965.

[31] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2017) 834–848, https://doi.org/10.1109/tpami.2017.2699184.

[32] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, In, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2881–2890, https://doi.org/10.48550/arXiv.1612.01105.

[33] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, In, in: Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 801–818, https://doi.org/10.1007/978-3-030-01234-2_49.

[34] S. Seo, J.S. Choi, J. Lee, H.H. Kim, D. Seo, J. Jeong, M. Kim, UPSNet: unsupervised pan-sharpening network with registration learning between panchromatic and multi-spectral images, IEEE Access 8 (2020) 201199–201217, https://doi.org/10.1109/access.2020.3035802.

[35] D. Laumer, N. Lang, N. van Doorn, O. Mac Aodha, P. Perona, J.D. Wegner, Geocoding of trees from street addresses and street-level images, ISPRS J. Photogramm. Remote Sens. 162 (2020) 125–136, https://doi.org/10.1016/j.isprsjprs.2020.02.004.

[36] C. Zhao, Y. Ogawa, S. Chen, T. Oki, Y. Sekimoto, Quantitative land price analysis via computer vision from street view images, Eng. Appl. Artif. Intell. 123 (2023) 106294, https://doi.org/10.1016/j.engappai.2023.106294.

[37] T. Aikoh, R. Homma, Y. Abe, Comparing conventional manual measurement of the green view index with modern automatic methods using google street view and

semantic segmentation, Urban For. Urban Green 80 (2023) 127845, https://doi.org/10.1016/j.ufug.2023.127845.

[38] T. Zhang, J. Dai, W. Song, R. Zhao, B. Zhang, OSLPNet: a neural network model for street lamp post extraction from street view imagery, Expert Syst. Appl. 231 (2023) 120764, https://doi.org/10.1016/j.eswa.2023.120764.

[39] Y. Hou, M. Quintana, M. Khomiakov, W. Yap, J. Ouyang, K. Ito, Z. Wang, T. Zhao, F. Biljecki, Global Streetscapes—A comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics, ISPRS J. Photogramm. Remote Sens. 215 (2024) 216–238, https://doi.org/10.1016/j.isprsjprs.2024.06.023.

[40] X. Liu, X. Zhang, R. Wang, Y. Liu, H. Hadiatullah, Y. Xu, T. Wang, J. Bendl, T. Adam, J. Schnelle-Kreis, X. Querol, High-precision microscale particulate matter prediction in diverse environments using a long short-term memory neural network and street view imagery, Environ. Sci. Technol. 58 (2024) 3869–3882, https://doi.org/10.1021/acs.est.3c06511.

[41] Y. Chuang, S. Zhang, X. Zhao, Deep learning-based panoptic segmentation: recent advances and perspectives, IET Image Process 17 (2023) 2807–2828, https://doi.org/10.1049/ipr2.12853.

[42] S. Freitas, C. Catita, P. Redweik, M.C. Brito, Modelling solar potential in the urban environment: state-of-the-art review, Renew. Sustain. Energy Rev. 41 (2015) 915–931, https://doi.org/10.1016/j.rser.2014.08.060.

[43] Š. Kolečanský, J. Hofierka, J. Bogľarský, J. Šupinský, Comparing 2D and 3D solar radiation modeling in urban areas, Energies 14 (2021) 8364, https://doi.org/10.3390/en14248364.

[44] J. Liang, J. Gong, X. Xie, J. Sun, Solar3D: an open-source tool for estimating solar radiation in urban environments, ISPRS Int. J. Geo-Inf. 9 (2020) 524, https://doi.org/10.3390/ijgi9090524.

[45] C. Catita, P. Redweik, J. Pereira, M.C. Brito, Extending solar potential analysis in buildings to vertical facades, Comput. Geosci. 66 (2014) 1–12, https://doi.org/10.1016/j.cageo.2014.01.002.

[46] R. Erdélyi, Y. Wang, W. Guo, E. Hanna, G. Colantuono, Three-dimensional SOlar RAdiation Model (SORAM) and its application to 3-D urban planning, Sol. Energy. 101 (2014) 63–73, https://doi.org/10.1016/j.solener.2013.12.023.

[47] J. Liang, J. Gong, W. Li, A.N. Ibrahim, Visualizing 3D atmospheric data with spherical volume texture on virtual globes, Comput. Geosci. 68 (2014) 81–91, https://doi.org/10.1016/j.cageo.2014.03.015.

[48] F. Lindberg, P. Jonsson, T. Honjo, D. Wästberg, Solar energy on building envelopes –3D modelling in a 2D environment, Sol. Energy. 115 (2015) 369–378, https://doi.org/10.1016/j.solener.2015.03.001.

[49] A. Vulkan, I. Kloog, M. Dorman, E. Erell, Modeling the potential for PV installation in residential buildings in dense urban areas, Energy Build 169 (2018) 97–109, https://doi.org/10.1016/j.enbuild.2018.03.052.

[50] M. Abuseif, K. Dupre, R.N. Michael, Trees on buildings: opportunities, challenges, and recommendations, Build. Environ. 225 (2022) 109628, https://doi.org/10.1016/j.buildenv.2022.109628.

[51] B. Tian, R.C.G.M. Loonen, J.L.M. Hensen, Combining point cloud and surface methods for modeling partial shading impacts of trees on urban solar irradiance, Energy Build 298 (2023) 113420, https://doi.org/10.1016/j.enbuild.2023.113420.

[52] F.T. Kurdi, E. Lewandowicz, J. Shan, Z. Gharineiat, 3D Modeling and Visualization of Single Tree Lidar point Cloud Using Matrixial Form, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens, 2024, https://doi.org/10.1109/jstars.2024.3349549.

[53] J. Guo, B. Sun, Z. Qin, M.S. Wong, S.W. Wong, C.W. Yeung, H. Wang, S. Abba, G. Q. Shen, Analysing the effects for different scenarios on surrounding environment in a high-density city, Cities 99 (2020) 102585, https://doi.org/10.1016/j.cities.2019.102585.

[54] A. Katal, M. Mortezazadeh, L.L. Wang, H. Yu, Urban building energy and microclimate modeling–From 3D city generation to dynamic simulations, Energy 251 (2022) 123817, https://doi.org/10.1016/j.energy.2022.123817.

[55] R. Zhu, L. You, P. Santi, M.S. Wong, C. Ratti, Solar accessibility in developing cities: a case study in Kowloon East, Hong Kong. Sustain. Cities Soc. 51 (2019) 101738, https://doi.org/10.1016/j.scs.2019.101738.

[56] R. Zhu, M.S. Wong, L. You, P. Santi, J. Nichol, H.C. Ho, L. Lu, C. Ratti, The effect of urban morphology on the solar capacity of three-dimensional cities, Renew. Energy. 153 (2020) 1111–1126, https://doi.org/10.1016/j.renene.2020.02.050.

[57] W.B. Most, S. Weissman, Trees and Power Lines: Minimizing Conflicts Between Electric Power Infrastructure and the Urban Forest, City Streets, Berkeley Law, 2012.

[58] T. Fujii, R. Ray, Singapore As a Sustainable City: Past, Present and the Future, SMU Economics & Statistics Working Paper No, 2019, https://doi.org/10.2139/ssrn.3480894, 18-2019.

[59] M. Jiang, S. Khorram, L. Fuxin, Comparing the decision-making mechanisms by transformers and CNNs via explanation methods, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit, 2024, pp. 9546–9555, https://doi.org/10.48550/arXiv.2212.06872.

[60] J. Guo, H. He, T. He, L. Lausen, M. Li, H. Lin, X. Shi, C. Wang, J. Xie, S. Zha, A. Zhang, H. Zhang, Z. Zhang, Z. Zhang, S. Zheng, Y. Zhu, GluonCV and GluonNLP: deep learning in computer vision and natural language processing, J. Mach. Learn. Res. 21 (2020) 1–7, https://doi.org/10.48550/arXiv.1907.04433.

[61] C.M. Hsieh, J.J. Li, L. Zhang, B. Schwegler, Effects of tree shading and transpiration on building cooling energy use, Energy Build 159 (2018) 382–397, https://doi.org/10.1016/j.enbuild.2017.10.045.

[62] Y. Yan, B. Huang, Estimation of building height using a single street view image via deep neural networks, ISPRS J. Photogramm. Remote Sens. 192 (2022) 83–98, https://doi.org/10.1016/j.isprsjprs.2022.08.006.

[63] W. Wang, L. Xiao, J. Zhang, Y. Yang, P. Tian, H. Wang, X. He, Potential of internet street-view images for measuring tree sizes in roadside forests, Urban For. Urban Green 35 (2018) 211–220, https://doi.org/10.1016/j.ufug.2018.08.009.

[64] L. Cheng, Y. Yuan, N. Xia, S. Chen, Y. Chen, K. Yang, L. Ma, M. Li, Crowd-sourced pictures geo-localization method based on street view images and 3D reconstruction, ISPRS J. Photogramm. Remote Sens. 141 (2018) 72–85, https://doi.org/10.1016/j.isprsjprs.2018.04.006.

[65] F. Devernay, O. Faugeras, Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured environments, Mach. Vis. Appl. 13 (1) (2001) 14–24, https://doi.org/10.1007/PL00013269.

[66] J. Kannala, S.S. Brandt, A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses, IEEE Trans. Pattern Anal. Mach. Intell. 28 (8) (2006) 1335–1340, https://doi.org/10.1109/TPAMI.2006.153.