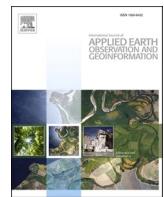




Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Segmentation of individual trees in urban MLS point clouds using a deep learning framework based on cylindrical convolution network

Tengping Jiang^{a,b,c,d}, Shan Liu^{a,b,c}, Qinyu Zhang^{a,b,c}, Xin Xu^{a,b,c}, Jian Sun^{a,b,c,*}, Yongjun Wang^{a,b,c,*}

^a Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

^b Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210093, China

^c State Key Laboratory Cultivation Base of Geographical Environment Evolution, Nanjing 210093, China

^d State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China



ARTICLE INFO

Keywords:

Mobile laser scanning (MLS) point clouds
Individual tree segmentation
Instance segmentation
Cylindrical voxel representation
Cylinder convolution

ABSTRACT

Automatic and accurate instance segmentation of street trees from point clouds is a fundamental task in urban green space research. Previous studies have achieved satisfactory tree segmentation results in simple scenarios. However, for challenging cases, including adjacent overlapping tree crowns, irregular tree shapes, and incompleteness caused by occlusion, most methods show under- or over-segmentation effects. In this study, an automated two-stage framework (tree extraction and individual tree segmentation) using vehicle-mounted mobile laser scanning (MLS) point clouds is developed to robustly detect single roadside trees. In the first stage, the ground points are filtered to reduce the processing time. Subsequently, an improved graph-based semantic segmentation network extracts roadside tree points from the urban scenes. For individual tree segmentation, a segmentation strategy combining cylindrical convolution and dynamic shift detects instance-level roadside trees. A simple road environment and two complex urban areas are used to verify the performance of the individual urban tree segmentation. The proposed method achieves 84–92% overall segmentation accuracy of the roadside tree point clouds and significantly outperforms existing methods in various challenging cases. Some applications can benefit from individual tree segmentation. For instance, the 3D green volume is calculated at the level of individual urban trees. The proposed method provides a practical solution for ecological assessment based on the 3D green volume of urban roads.

1. Introduction

Rapid urbanization causes increasing urban environmental problems. The construction of urban green spaces is one possible solution to urban diseases; thus, knowing the inventory of urban trees is essential (Chen et al., 2019). However, more than 80% of cities have incomplete and inaccurate urban tree inventories owing to urban development funding and strategy problems (Xu et al., 2023). Techniques such as field surveys, video log, and aerial and satellite remote sensing are traditionally regarded as the main methods for collecting roadside tree data. Although these techniques can obtain most of the required urban tree information, their limitations include time consuming and data-quality challenges (Jiang et al., 2022). Recently, light detection and ranging (LiDAR) system has become a conventional means for producing various

types of Geo-information (Jiang et al., 2023a). In particular, mobile laser scanning (MLS) systems with LiDAR devices offer a promising style to collect high-dense point clouds of urban road environments, which are also becoming more prominent in retrieving roadside tree information. Automated segmentation of individual trees in complex street environments has entered a new era, and designing a feasible method for segmenting single trees using urban MLS data is a prerequisite for scientifically evaluating of urban ecosystems (Wang et al., 2023). This provides a significant impetus for urban green space management.

In recent years, extensive methods are developed to perform instance segmentation of trees from point clouds, which are roughly classified into bottom-up and top-down segments. The bottom-up method usually begins with point clustering, then analyzes the structural features to further assemble segments into complete trees. This method

* Corresponding authors at: Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China.

E-mail addresses: 09332@njnu.edu.cn (J. Sun), wangyongjun@njnu.edu.cn (Y. Wang).

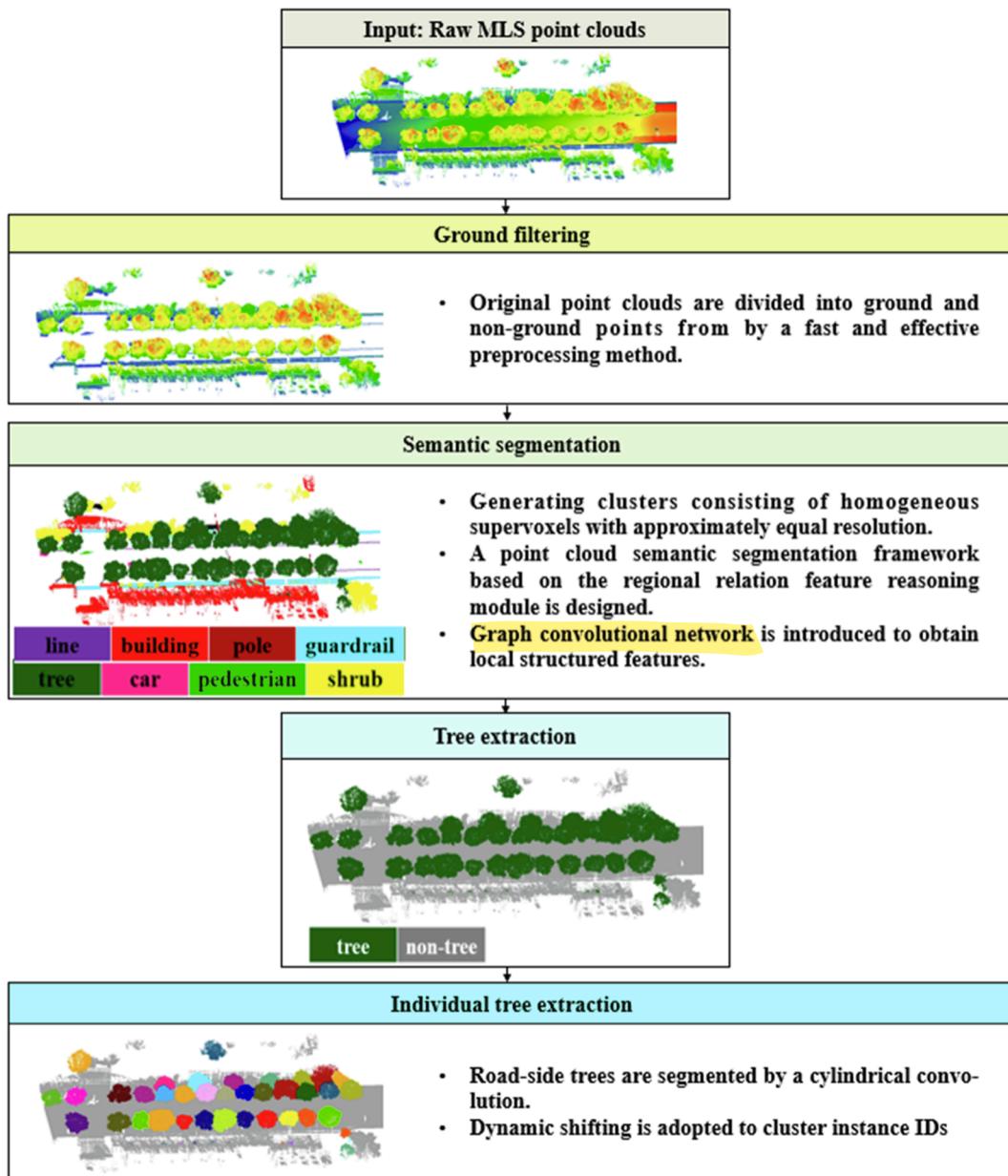


Fig. 1. Pipeline of the proposed framework.

accomplishes individual tree segmentation by calculating useful hand-crafted features from each point segments and applying classic machine learning models (Yu et al., 2015). A three-step individual tree segmentation framework is proposed wherein the mean-shift method is first applied in the spatial domain to isolate individual trees by exploiting the geometric pattern of point distribution of the tree crowns (Dai et al., 2018). Then, a support vector machine classifier organizes the segmentation results and refines the under-segmented trees. An improved mean-shift algorithm is developed based on self-adaptive bandwidth estimation to increase the accuracy (Yan et al., 2020). Similarly, a bottom-up hierarchical clustering approach for MLS point clouds involves minimizing the energy function for grouping target points to optimize cluster combinations (Xu et al., 2018). To reduce the amount of data, Li et al. (2021) first transform the raw point clouds into supervoxel. Then, an uphill clustering algorithm is adopted to extract independent crowns and trunks from potential objects obtained from supervoxel. For touching street trees, Li et al. (2022) develop a branch-trunk-constrained segmentation approach based on tree positions. The

bottom-up methods achieve satisfying individual tree segmentation performance in some areas. However, they require prior manual calculation of features which reduces automation while sensitivity to the choice of features limits their application (Liu et al., 2021). To this end, Wang (2020) proposes an unsupervised method based on the superpoint graph (Landrieu and Simonovsky, 2018), which achieves significant robustness for difficult forest scenes.

In contrast, top-down methods commonly extract tree points and group them into isolated clusters. Generally, clusters contain multiple trees, and the tree clusters are further divided into multiple single trees (Yadav and Lohani, 2020). The classic method is *treeseg* (Burt et al., 2019), which is an automated data-driven approach with a few assumptions of tree structures. To successfully obtain the structural measurements of rainforest fused point clouds, Terryn et al. (2022) use the open-source data-driven approach *treeseg* followed by tree-level ecological indicator quantification. However, post-processing optimization is often required to refine crown boundaries in complicated situations. Crown segmentation using a dual Gaussian filter and energy

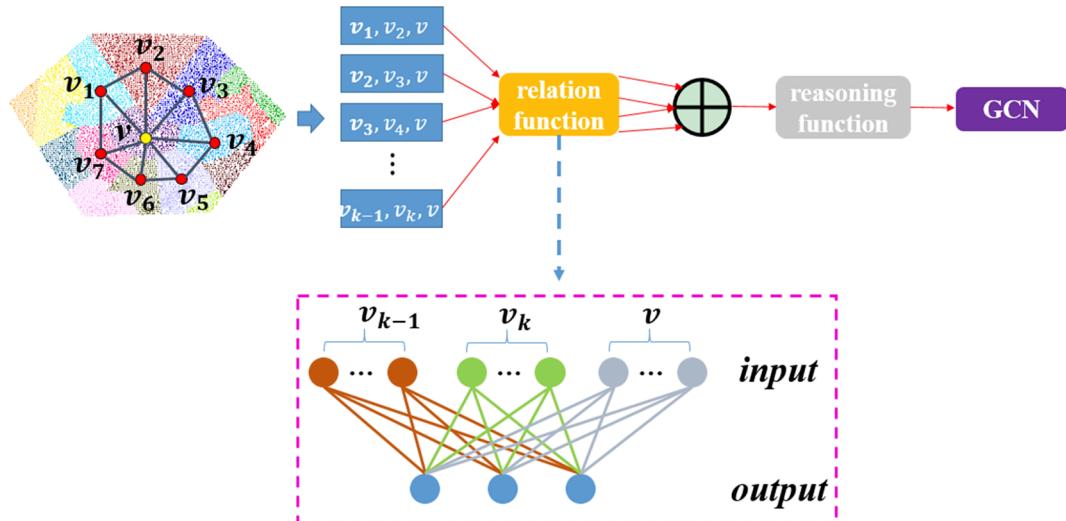


Fig. 2. Illustration of the point cloud semantic segmentation framework based on regional relation feature reasoning. The GCN refers to the graph convolutional network.

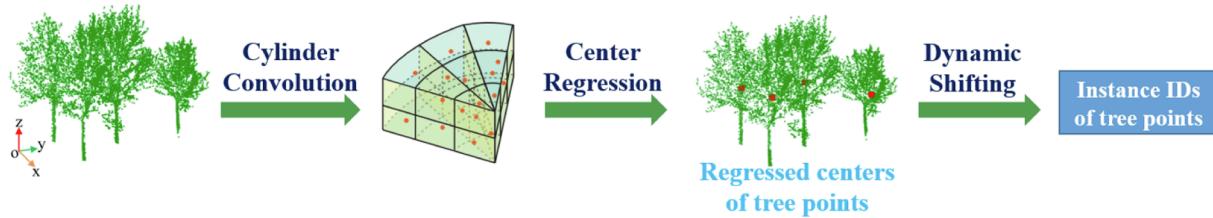


Fig. 3. Pipeline of instance-level tree separation based on instance segmentation that combines cylinder convolution and dynamic shifting.

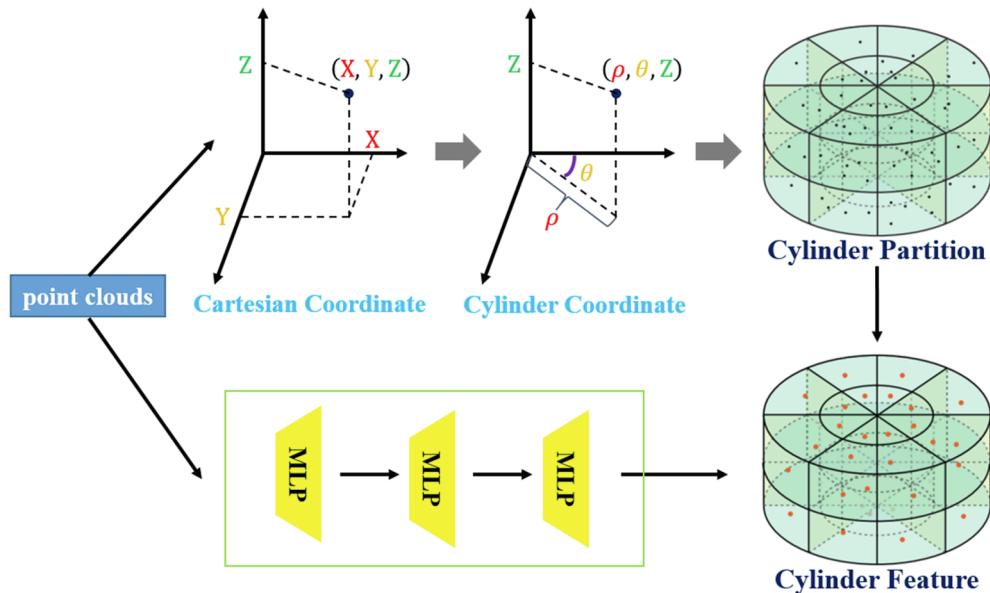


Fig. 4. Pipeline of extracting cylindrical features.

function minimization applied in different scenes preserves their geometric architectures which improves the uncertainty of processing complex scenes using some previous algorithms (Yun et al., 2021). To explore ecosystem structure at the tree-scale, Wang et al. (2021) perform the task of individual tree segmentation from ground-based LiDAR point clouds via graph pathing algorithm. It is rare to ensure a fast computation for point-wise crown segmentation.

Recently, significant progress in deep learning has simplified point cloud analysis and become the dominant method in the task of individual tree segmentation (Chen et al., 2021; Jiang et al., 2023b; Luo et al., 2021; Luo et al., 2022; Mäyrä et al., 2021; Wang et al., 2020a). For example, Chen et al. (2021) develop a PointNLM network that combines point-wise and supervoxel-based representations. By acquiring the long-range relationship among supervoxels, PointNLM may be more helpful

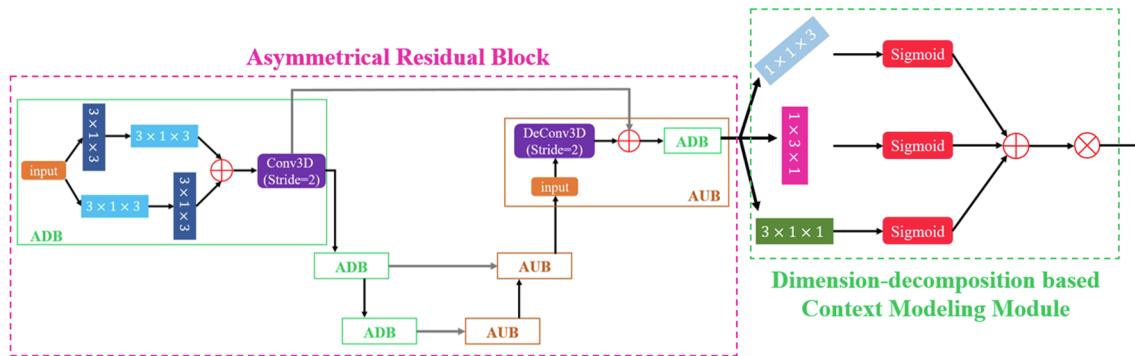
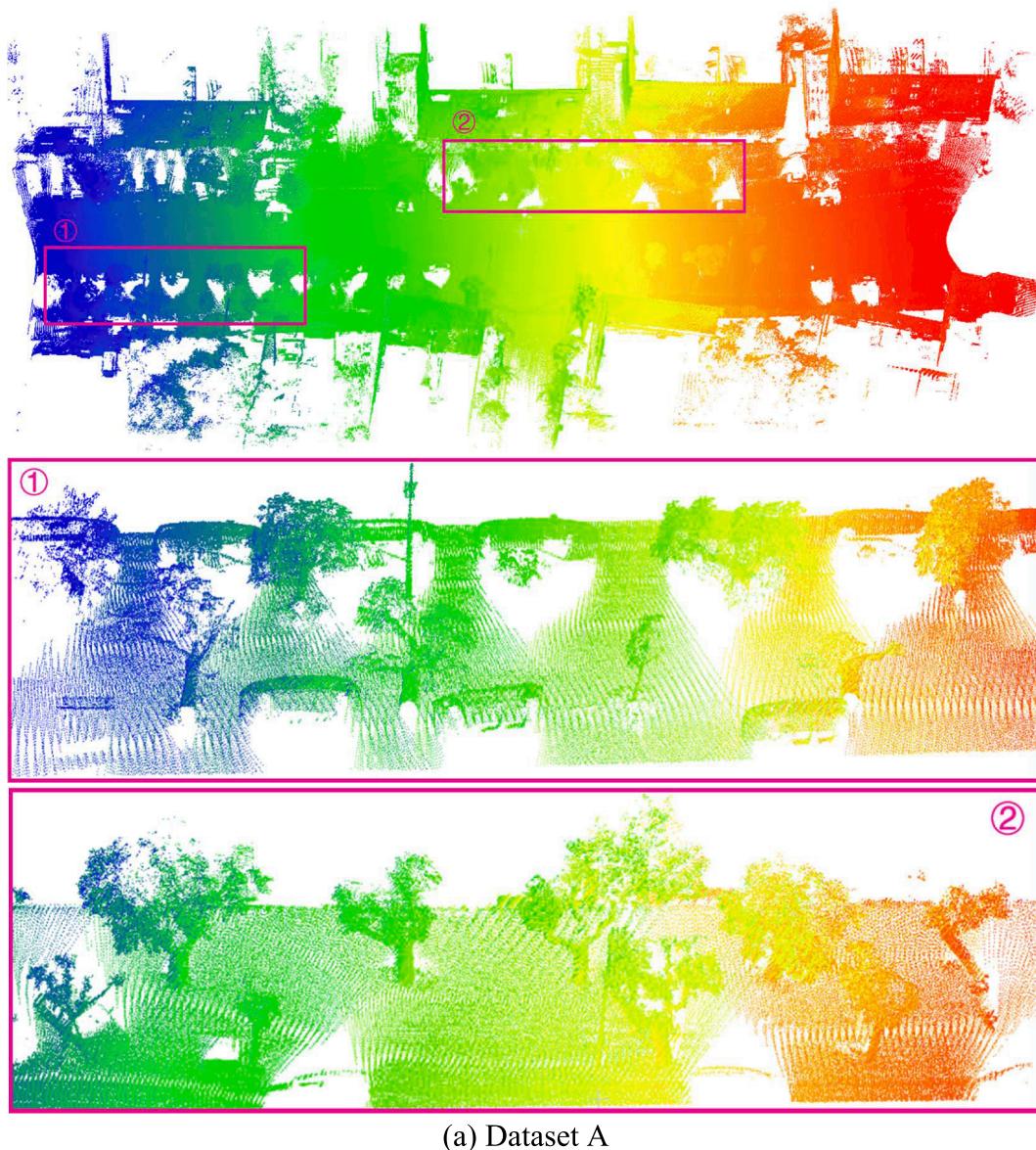
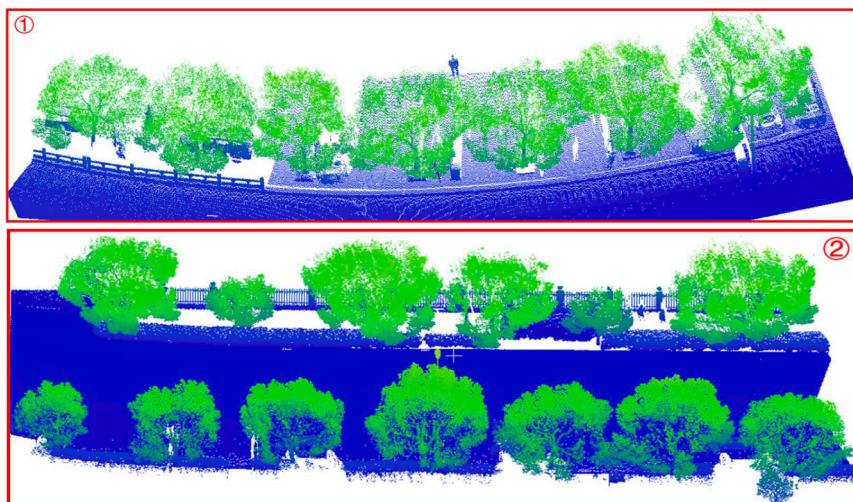
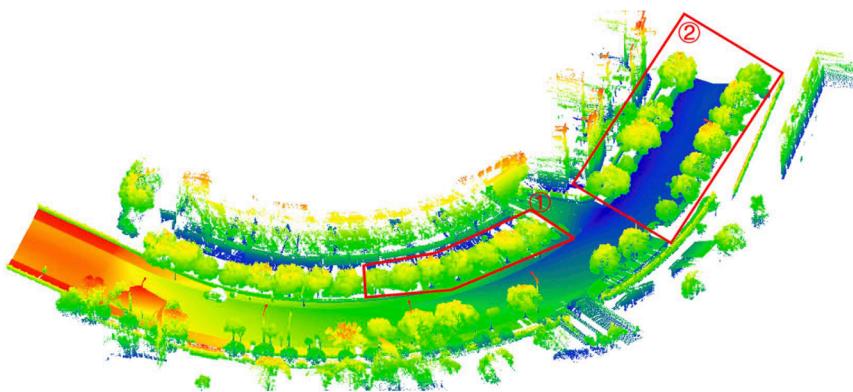


Fig. 5. Overview of the asymmetrical convolution network. ADB denotes an asymmetrical down sample block and AUB refers to an asymmetrical up sample block, respectively.

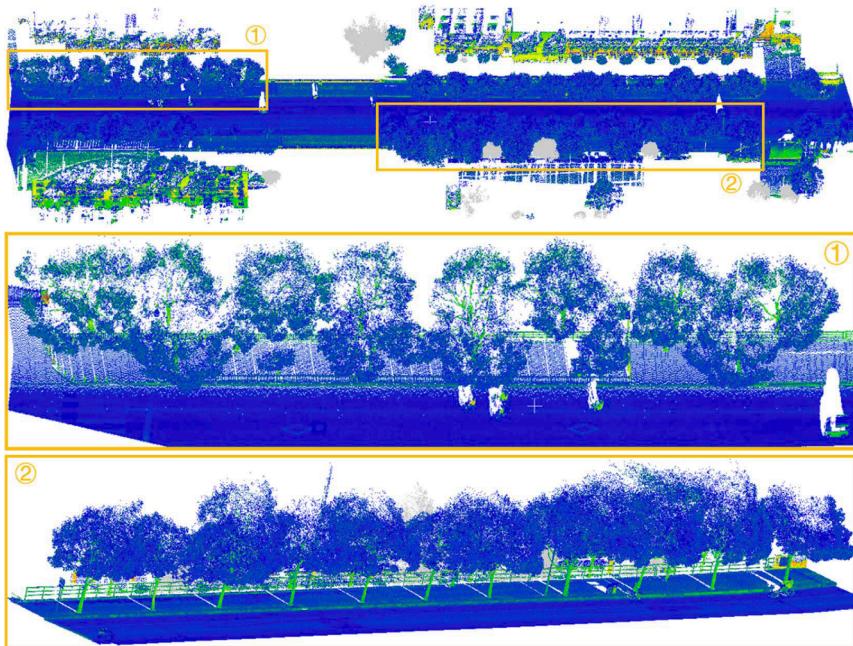


(a) Dataset A

Fig. 6. Illustration of three selected urban MLS point clouds. The selected scene of Dataset A and Dataset B colored by elevation of each point; the selected scene of Dataset C colored by intensity of each point.



(b) Dataset B



(c) Dataset C

Fig. 6. (continued).

in capturing the global geometric features of each single-tree point cloud, improving city-scale vegetation segmentation task. Similar to our idea, [Luo et al. \(2022\)](#) first design an improved graph convolutional

network to filter ground points in various landforms. To obtain more discriminative and detailed structure features, a multi-branch network based on multichannel representation is proposed to detect individual

Table 1

Computational time of each stage (values in s).

Data	Points (million)	Ground Removal	Semantic Segmentation	Cylinder Convolution	Instance Segmentation	Total time
A	206	100.7	7684.6	471.3	77.1	8333.7
B	335	145.1	12640.6	552.6	75.8	13414.1
C	415	195.7	15594.6	757.1	125.4	16672.1

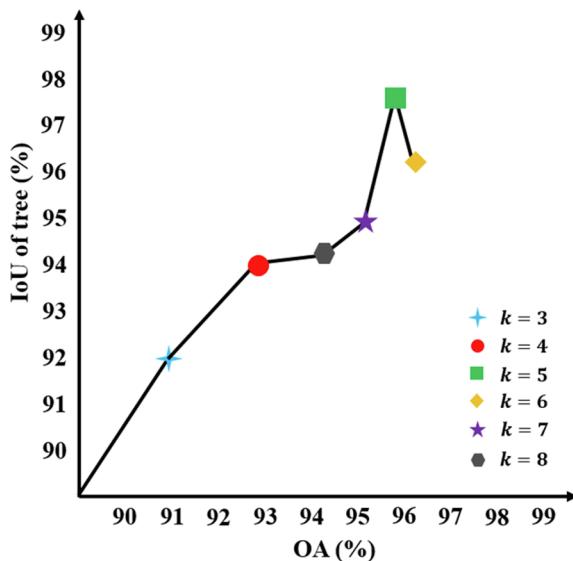


Fig. 7. Overview of the asymmetrical convolution network.

tree from forest airborne laser scanning data. The reliability of deep learning is undeniable in a structurally simple environment. However, outdoor urban MLS data are generally characterized by large volumes, complex structures, and uneven point distribution. Moreover, the boundaries between adjacent trees are usually blurred owing to heavy occlusions and overlaps between the transportation furniture and tree crowns.

Realizing an efficient and robust approach for individual tree segmentation from structurally complex urban MLS point clouds remains a challenge. This study proposes a hierarchical framework for recognizing individual roadside trees from urban MLS data, as shown in Fig. 1. In the tree extraction stage, an existing ground-filtering method is first adopted to separate the original data into the ground and non-ground points. Subsequently, an improved graph-based 3D semantic segmentation framework extracts the tree points from the segmented non-ground points. In the individual tree segmentation stage, a cylindrical convolution-based dynamic shift approach obtains instance-level trees from the extracted tree points. Finally, the 3D green volume (3D GV) of the streets at an individual tree level is estimated.

Concretely, the main contributions of this study are as follows:

- (1) A novel framework based on semantic and instance segmentation is proposed to overcome the challenging task of individual tree segmentation from large-scale urban MLS point clouds.
- (2) The proposed convolution network comprises an asymmetrical residual block, and a dimension-decomposition based context modeling module that can mine the distribution pattern to obtain more discriminative structural information and detailed features for urban tree segmentation.

Following the Introduction, the crucial components of the proposed method to automatically recognize roadside trees are elaborated in Section 2. Thereafter, the experimental results are presented in Section 3 and experimental analysis is presented in Section 4. A discussion of the

3D GV estimation is provided in Section 5. Finally, the conclusion is summarized at the end of this paper.

2. Methods

2.1. Tree point extraction

Generally, an MLS system has a relatively direct scanning angle of the view of the ground. Therefore, the collected 3D points contain massive ground points, which undoubtedly increase algorithm complexity. Therefore, a fast and effective preprocessing method (Zhang et al., 2016) was adopted to separate ground points from original point clouds and reduce the data search range of point cloud processing (see the top of Fig. 1).

The complexity of an urban scene is the main obstacle in semantically extracting tree points, which usually contain many categories of objects and overlapping or closely neighboring objects. This study proposes a graph convolution network that integrates a lightweight representation learning module and a deep context-aware sequential module with embedded residual learning to classify urban scenes into tree and non-tree point clouds.

Unstructured off-ground point clouds were first divided into geometrically homogeneous supervoxels to ameliorate non-uniform distribution and reduce computational complexity (Luo et al., 2020; Jiang et al., 2023c). Specifically, we directly consumes original point clouds generating clusters consisting of homogeneous supervoxels with approximately equal resolution. This process was adopted from Lin et al. (2018), the nearest neighbors based on the clusters generally cover a large enough area to obtain meaningful local information. A sparse auto-encoder is employed to compress and encode high-dimensional information as the embedding to represent the geometric attributes of every patch. Moreover, the spatial position of geometric clusters is concatenated into the final descriptor to increase spatial relationships. An adjacency graph $G = \{V, E\}$ is constructed to model neighboring relationships among the supervoxels to promote the formation of associated areas from geometric supervoxels generated from the off-ground points. The center of precomputed supervoxels acts as the nodes $V = \{v_i\}$ in the G and edges $E = \{e_{ij}\}$ is established between each pair of adjacent supervoxels to allow the network to be relatively robust in handling varying point densities. Specifically, variable adjacency is adopted instead of a fixed-size neighborhood.

Subsequently, the relationship between the central vertex v and its k -nearest neighbors is established to embed regional relation features of v . Inspired by the study of Luo et al. (2020), a point cloud semantic segmentation framework based on the regional relation feature reasoning module was designed, as shown in Fig. 2. Each supervoxel is considered as an independent 3D shape containing n points. Then, PointNet (Qi et al., 2017) is imitated to aggregate global information from all points in these shapes. For building the neighbor-based associated region, the graph structure is adopted to construct the neighbor relations of supervoxels. Finally, our previously proposed graph convolutional network (GCN) (Jiang et al., 2021) is introduced to obtain local structured features. For the local feature extraction, different-order neighboring associated regions for each supervoxel are established, which are represented as 3D tensors. After extracting regional relation features for each scale, a multilayer perceptron (MLP) is used to expand the features into very high dimensions. Then, max pooling is performed for elementwise aggregation. A repeating operation is applied to restore the

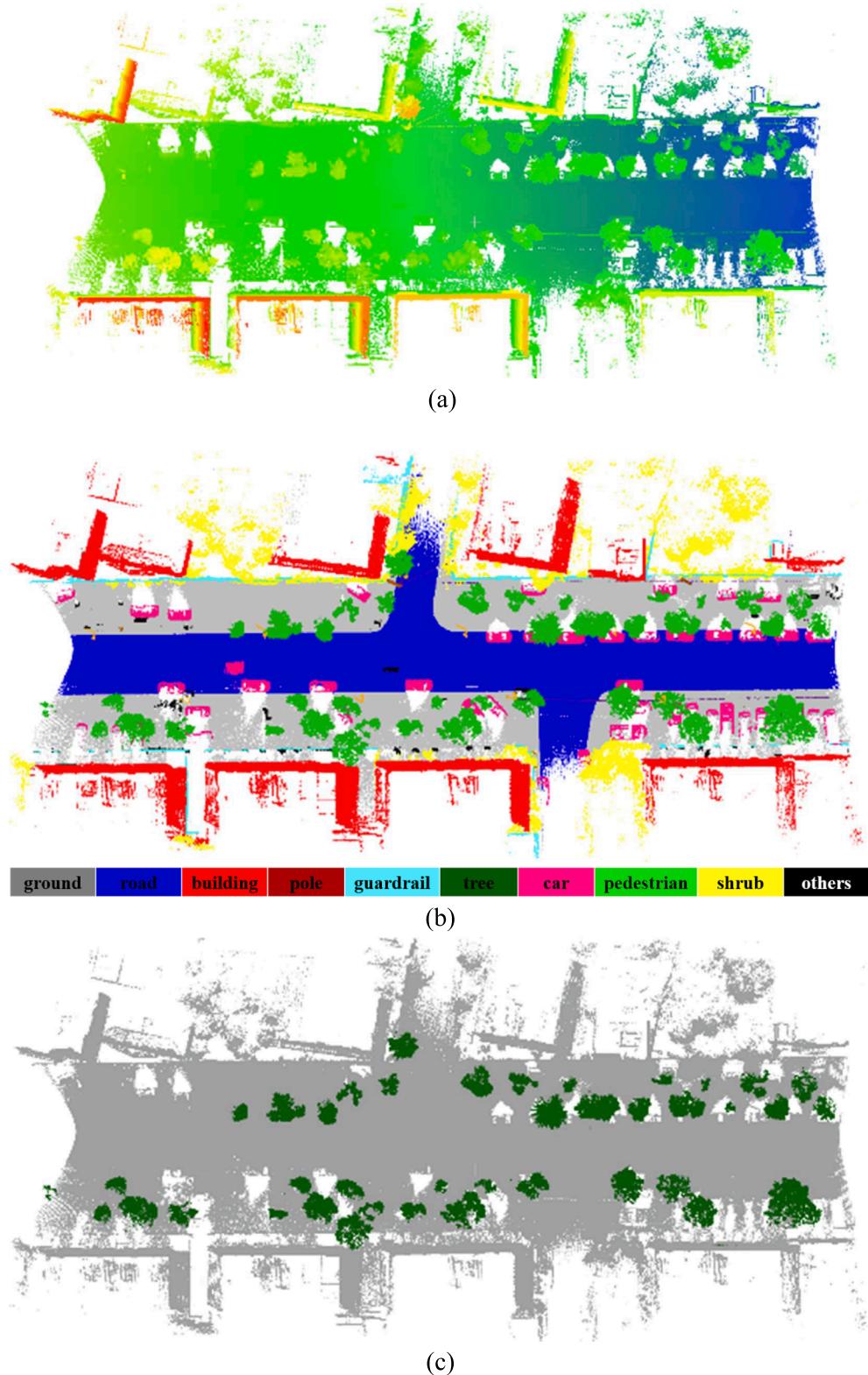


Fig. 8. Individual tree segmentation of dataset A.

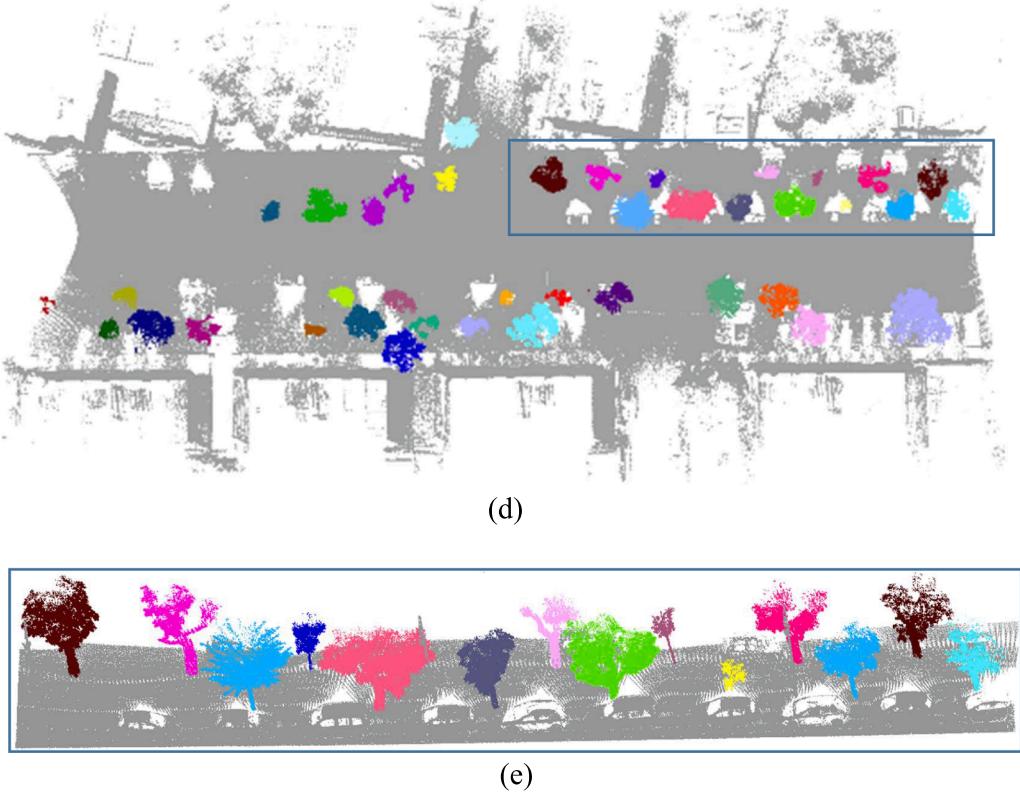


Fig. 8. (continued).

tensor shape to concatenate the global features with the hierarchical local regional relation features by skip-connection. A softmax layer is used to infer the semantic label for each supervoxel. The node labels obtained by classification are $Y = \{y_1, y_2, \dots, y_n\}$, n is the number of supervoxels, $y_i = \{l_1, l_2, \dots, l_k\}$ is the label of node i , and k is the number of categories.

The last step is semantic label optimization based on higher-order conditional random fields (Wang et al., 2020b). The logarithmic model of joint conditional probability of label optimization is defined as

$$\log P_w(y|x) = \sum_{i \in V} \phi_V(y_i, x_i; w_v) + \sum_{(i,j) \in E} \phi_E(y_i, y_j, x_{ij}; w_e) + \sum_{c \in S} \phi_S(y_c, x_c; w_c) - \log Z_w(x) \quad (1)$$

where $Z_w(x) = \sum_y \exp[\sum_{s \in C} \phi_s(y_s, x_s)]$ is a regularization factor and $s \in C$ denotes multi-order cliques. x_i is the observation vector of node i , corresponding to the feature vector of the node, x_{ij} is the observation vector of the edge formed by nodes i and j , corresponding to the eigenvector of edge $e(i, j)$, and x_c is the observation vector of the higher-order clique where node i is located, corresponding to the distributed spatial context of partition c . w_v , w_e , and w_c are the weight vectors associated with nodes, edges, and higher-order cliques, respectively. $\phi_V(y_i, x_i; w_v)$ measures the probability of a node being labeled y_i given an observation vector x_i , which is defined as

$$\phi_V(y_i = l_k, x_i; w_v) = w_v^k \cdot x_i \quad (2)$$

where w_v^k is the weight vector when the class of node i is l_k .

$\phi_E(y_i, y_j, x_{ij}; w_e)$ describes the interaction between neighboring nodes, which is defined as

$$\phi_E(y_i, y_j, x_{ij}; w_e) = \begin{cases} w_e^k \cdot x_{ij}, & y_i = l_k \text{ and } y_j = l_k \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where w_e^k is the weight vector when nodes i and j of edge $e(i, j)$ are labeled as l_k simultaneously.

$\phi_S(y_c, x_c; w_c)$ promotes the consistency of node labeling within higher-order cliques S and is given by

$$\phi_S(y_c, x_c; w_c) = \begin{cases} w_c \cdot x_c, & \forall i \in c, y_c = l_k \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Given the ground truth (x, \hat{y}) , the sub-gradient optimization algorithm (Muñoz et al., 2009) is used to obtain the weight parameters $w = [w_v, w_e, w_c]$. Then, the graph-cuts algorithm (Boykov et al., 2001) is adopted to infer the optimal class label of the node, which is used as the optimal result of point cloud semantic segmentation.

Finally, the supervoxel-level inferences are transformed into point-level predictions based on the homogeneity assumption of the points in each supervoxel. Thus, some supervoxels whose attributes contain multi-dimensional embedding representation and 3D spatial coordinates in sample scenes are taken as the input of the proposed semantic segmentation framework. Only the tree points are used for further individual urban tree segmentation from the various classes of obtained point clouds (see Figs. 8–10 (b) and (c)).

2.2. Instance-level tree separation

Traditional methods that rely on treetop point detection or stems recognition are insufficiently robust for urban scenarios since the boundaries between adjacent urban trees are usually indistinct. Instance segmentation is performed by generalizing the 3D cylinder convolution network to the instance-level roadside tree separation task (Fig. 3).

Sparsity and varying density (nearby area often has greater density than farther away areas) are tricky problems for urban tree point clouds. A cylindrical coordinate system offers a balanced point distribution by increasing the voxel size to cover the farther area (Zhu et al., 2022). Thus, the points on the Cartesian coordinate system (x, y, z) are first transformed into a cylindrical coordinate system (ρ, θ, z) . Next, the

cylindrical partition method is adopted to divide the cylinder coordinates (Fig. 4). The cylindrical features $F_c \in D \times R \times A \times H$ (where D , R , A , and H represents the feature dimension, radius, azimuth, and height, respectively) are obtained by reassigning point-wise features produced by a simplified PointNet (Zhou et al., 2020). Point-wise features within the same cylinder are mapped to cylinders using a point-to-cylinder mapping function according to the index of the point-by-point feature. Finally, a 3D cylindrical representation is obtained via max-pooling.

An asymmetrical convolution network (ACN) is applied to learn the geometric properties while maintaining these inherent characteristics based on the cylindrical representation obtained by unrolling the cylinder from 0-degrees. Specifically, the asymmetrical convolution network includes an asymmetrical residual block (ARB) and a context modeling module (CMM) to suit the properties of roadside trees (Fig. 5). Trees conform to a cylindrical shape distribution according to the designed ARB because they enhance the vertical and horizontal kernels and align the tree point distribution. Moreover, ARB uses a $3 \times 1 \times 3$ convolution kernel followed by a $1 \times 3 \times 3$ convolution instead of the traditional $3 \times 3 \times 3$ convolution kernel, and the effects of these two 3D convolutional kernels are nearly identical. Surprisingly, this pattern has approximately 35% less computation and memory cost. The ARB also contains an asymmetrical down sample block (ADB) and an asymmetrical up sample block (AUB). In the up sampled part, low-level features

are incorporated, and an ARB processed the fused features. The down sampling operation is performed by an ARB and a 3D convolution with a stride of 2. The construction of context feature is complicated and requires considerable 3D space costs due to the large varieties and high-rank context properties. Therefore, the high-rank contextual information is transformed into a combination of three low-rank fragments (height, width, and depth) according to tensor decomposition theory (Chen et al., 2020). Then, three kernels ($3 \times 1 \times 1$, $1 \times 3 \times 1$ and $1 \times 1 \times 3$) are adopted to obtain low-rank features in all three fragments. The weights of each dimension tensor are generated by modulating the convolution results using sigmoid function where the context is mined according to low-rank tensors of different views. Overall, there is sufficient capacity to make use of high-rank contexts in the decomposition-aggregation means. Three low-rank activations are finally aggregated to gain full global context.

Grid-level features learned using specifically designed cylinder convolution from the tree point clouds are input into the instance segmentation architecture. This resulted in instance-level spatially overlapping tree separation.

In the instance segmentation network, the center regression module is first adopted to adapt cylindrical features and regress all points to each tree center by predicting the offset vectors $O_v \in \mathbb{R}^{M \times 3}$ from the points $P \in \mathbb{R}^{M \times 3}$ to the centers $C_{gt} \in \mathbb{R}^{M \times 3}$ of individual trees (Hong et al., 2021). Next, we further cluster regressed centers $O_v + P$ to get instance

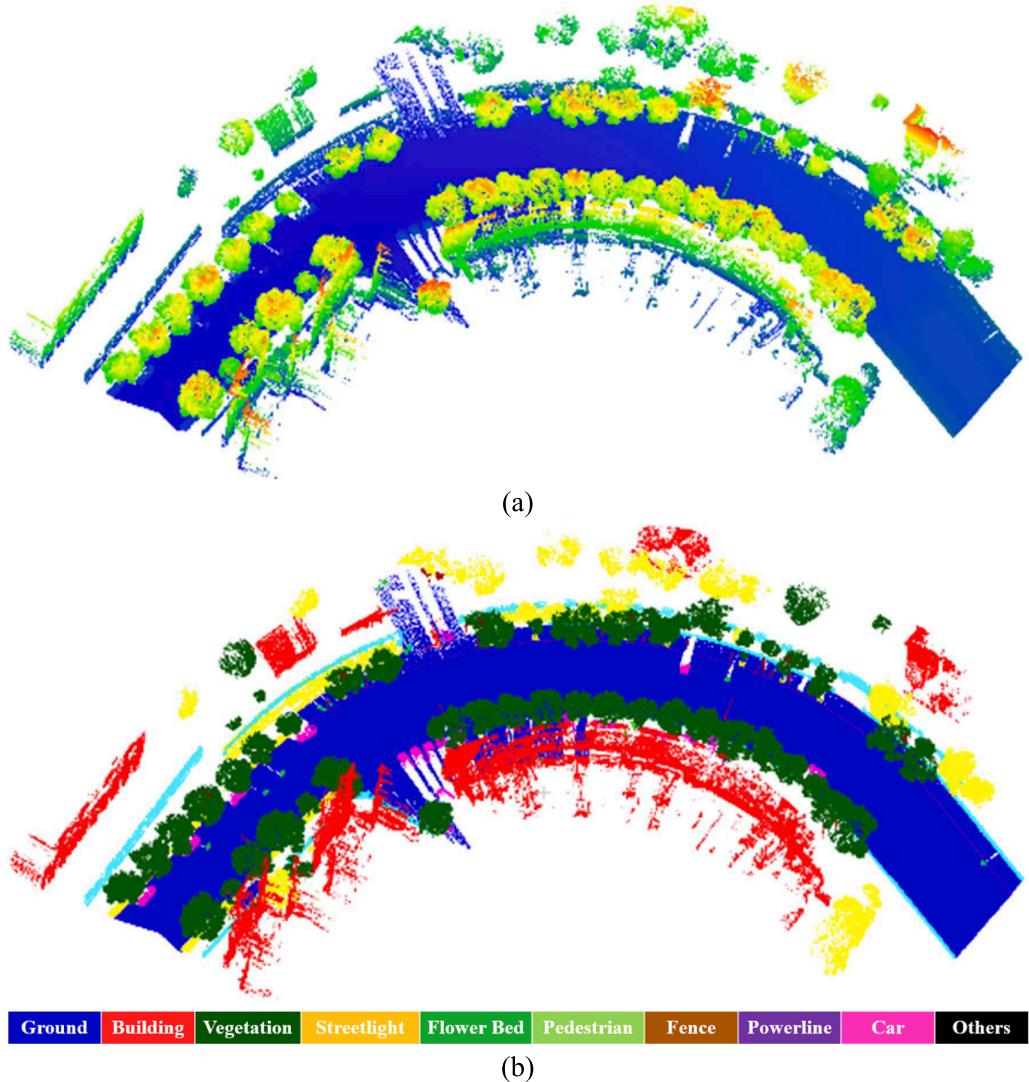


Fig. 9. Individual tree segmentation of dataset B.

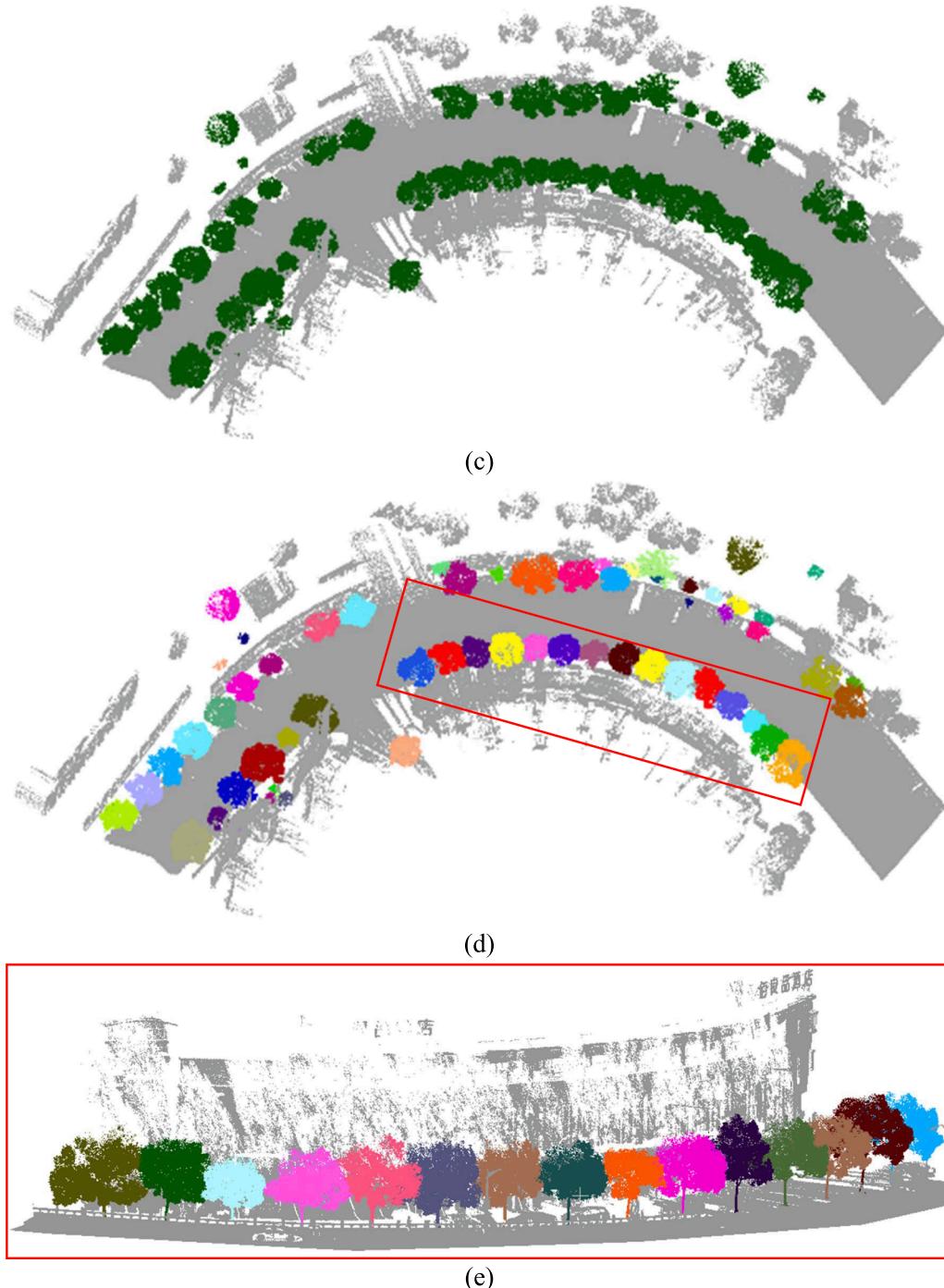


Fig. 9. (continued).

IDs; however, the most commonly used clustering algorithms are incapable of dealing with the urban environment with varying tree sizes. Therefore, a simple learnable clustering method based on dynamic shifting is introduced to divide instance-level urban trees to dynamically adapt kernel functions for various roadside trees. The dynamic shifting module adapts the kernel function for varying tree sizes so that the cluster centers of regressed centers are capable to be efficiently estimated.

Specifically, the shift operation on the points to be clustered ($X \in \mathbb{R}^{M \times 3}$, which represents M seeding points) are expressed as a matrix operation:

$$X \leftarrow X + S \quad (5)$$

where X is updated once by the shift vector $S \in \mathbb{R}^{M \times 3}$, $S = f(X) - X$ is calculated by applying a flat kernel function $f(X) = D^{-1}KX$ ($D = \text{diag}(K)$) which represents the diagonal matrix that calculates the number of points within bandwidth δ of each seeding, $K = (XX^T \leq \delta)$ represents the kernel matrix that masks off the points within a certain bandwidth for every seeding point on X .

Next, n shift object candidates for each seeding point are estimated by n kernels with the corresponding candidate bandwidths $B = \{\delta_1, \delta_2, \dots, \delta_n\}$. By learning $W \in \mathbb{R}^{M \times l}$ to weight n candidate objects, the shift objects closest to the cluster centers are dynamically determined for each seeding point. This learnable procedure can be represented as follows:

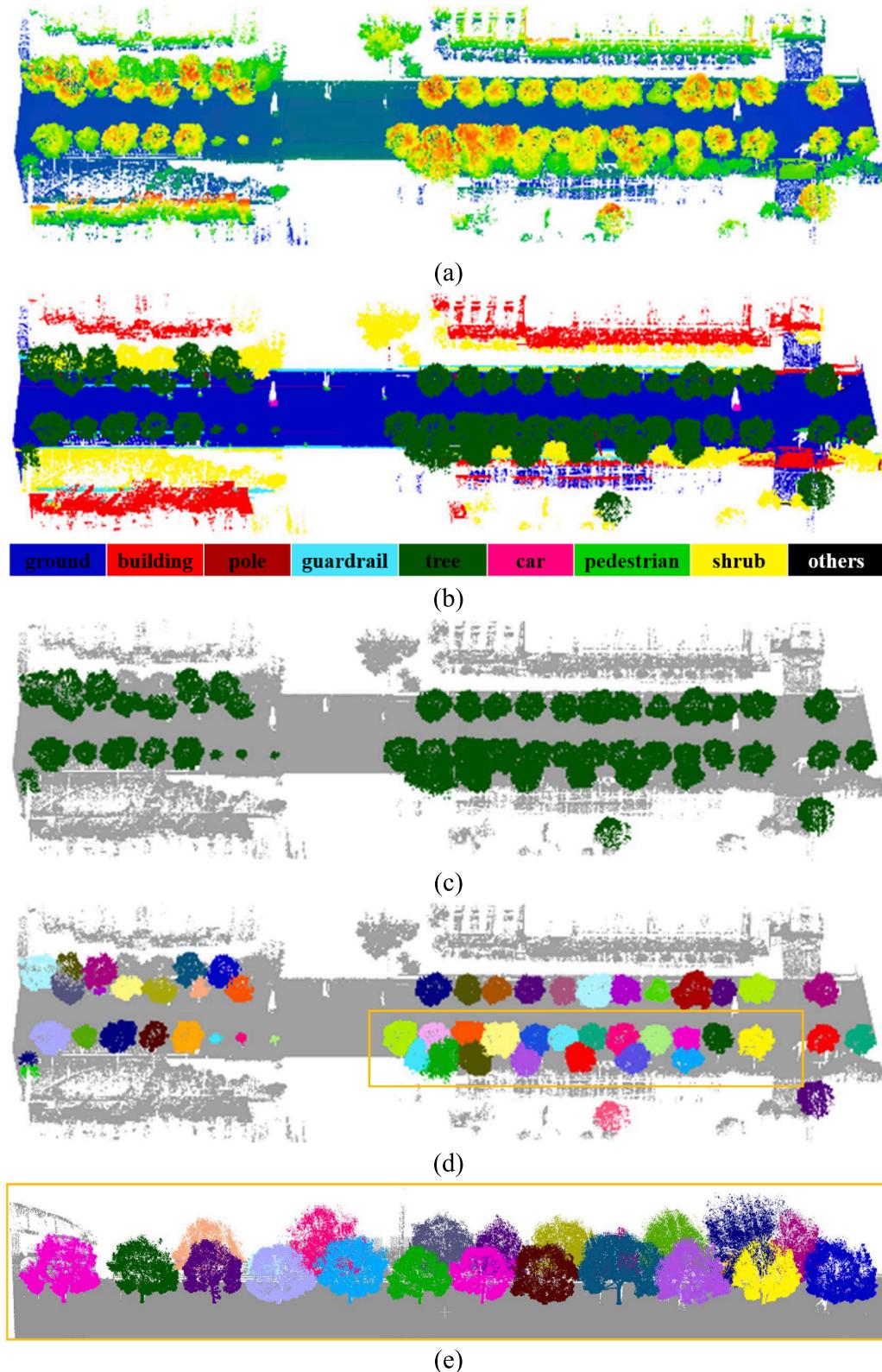


Fig. 10. Individual tree segmentation of dataset C.

Table 2

Influence of the size of cylindrical representation on the individual tree segmentation results.

Size of Representation	SAC	OME	COE
320 × 240 × 24	0.79	0.21	0.26
480 × 360 × 32	0.81	0.19	0.23
640 × 480 × 40	0.84	0.16	0.20
800 × 600 × 48	0.82	0.18	0.21
1000 × 1000 × 60 (cubic)	0.74	0.26	0.31

$$\hat{f}(X) = \sum_{j=1}^l W[:, j] \odot D_j^{-1} K_j X \quad (6)$$

where $\sum_{j=1}^l W[:, j] = 1$ when the weight W is obtained by MLP and Softmax, $K_j = (XX^T \leq \delta_j)$ and $D_j = \text{diag}(K_j)$.

It is critical to choose an appropriate loss function to shift the seed points to their cluster centers to optimize the dynamic shifting module. However, it is unrealistic to acquire the cluster center ground truths. Therefore, the ground truth centers of clusters are approximated by the ground truth centers of trees $C'_{gt} \in \mathbb{R}^{M \times 3}$ (M seeding points are provided for the dynamic shifting) to simplify the work. The loss function, L_{ds} (Eq. (3)) is then expressed by the Manhattan distance between the shift object X and ground truth center C'_{gt} .

$$L_{ds} = \frac{1}{M} \sum_{x=1}^M \|X[x] - C'_{gt}[x]\|_1 \quad (7)$$

The seeding points are traversed until they regress to the cluster centers after several dynamic shifting iterations. To get the instance IDs for every seeding point, a simple cluster algorithm called accessible region growing (Luo et al., 2021) is used to cluster the regressed seeding points. The instance IDs of the remaining tree points are finally assigned by finding nearest seeding points.

3. Experimental setup and results

3.1. Experimental datasets

The performance of instance segmentation of urban trees using the proposed method was verified using a test experiment on three representative MLS point cloud datasets (shown in Fig. 6). Dataset A is a public MLS dataset, named Paris-Lille-3D (Roynard et al., 2018), acquired from several French cities using a truck-based MLS system with a Velodyne HDL-32e laser scanning sensor. Dataset A covered well-planned road scenes of two typical Western European cities, including a large number of roadside objects. Dataset B was collected using a HiScan-Z MLS system from approximately 10.0 km of a bustling street in Wuhan, China. As for Dataset C, it was acquired using a Trimble MX8 MLS system from a downtown street in Shanghai, China containing various significantly overlapping trees. Dataset B and Dataset C are provided by Dynamic Mapping Laboratory at the Wuhan University. Dataset B trees overlapped more compared with Dataset A with no obvious boundary between the canopies and trunks of street trees in Dataset B which differs from regular roadside trees. Dataset C is also an environmentally complex urban scene including numerous road facilities with differing appearances.

It needs to point out that the above three datasets only provide semantic labels (see Figs. 8–10 (b)), and there are no instance labels of trees. For training the network and evaluating the results, the ground truth for individual tree segmentation were obtained by careful manual segmentation. To make the instance label as accurate as possible, they were finally verified by four people.

3.2. Implementation and evaluation metrics

All experiments are carried out on a workstation equipped with two 16 GB NVIDIA GTX 1080 Ti GPUs. The main parts of the proposed approach are completed using C++, except for the tree extraction implemented in Python. The point cloud semantic segmentation framework is built using Tensorflow. The PCL 1.13 (<https://pointclouds.org/>), OpenCV 4.5 (<https://opencv.org/>), and CGAL 5.0 (<https://www.cgal.org/>) are used for ground removal and tree segmentation. Table 1 shows the running time of individual step of the proposed algorithm for several road scenes. Approximately 90% of the total time on each dataset is spent in the point cloud semantic segmentation stage because graph convolution operation and label optimization processing for large-scale point clouds are time-consuming.

If the height difference (HD) between the extracted roadside tree and ground truth is minor than 0.5 m and over 85% points are recognized, this tree is considered to be correctly segmented. The method is evaluated using three evaluation measures: segmentation accuracy (SAC), omission error (OME), and commission error (COE) which are commonly adopted for evaluating tree segmentation in previous studies (Wang et al., 2020a). SAC is the rate of correctly segmented roadside trees; OME is the rate of unsegmented roadside trees; and COE refers to the rate of falsely segmented street trees. These metrics are calculated at the point-level instead of the cluster-level.

$$SAC = \frac{v_{cs}}{v_{gt}} \quad (8)$$

$$OME = \frac{v_{us}}{v_{gt}} \quad (9)$$

$$COE = \frac{v_{fs}}{v_{gt}} \quad (10)$$

where v_{cs} represents the value of correctly segmented roadside trees, v_{us} refers to the value of unsegmented roadside trees, v_{fs} denotes the value of falsely segmented roadside trees, and v_{gt} is the value of ground truth roadside trees.

3.3. Experimental parameter configuration

Several experiments are performed on Dataset C to determine the optimal experimental parameters.

In the semantic segmentation part, the nearest k supervoxels directly affected the precision of semantic segmentation for off-ground points. According to Luo et al. (2020), the value of k determines the associated region in each neighbor. A set of comparison tests are conducted with six different values of k , which are 3, 4, 5, 6, 7, and 8, to evaluate how sensitive semantic segmentation result is to this parameter. Fig. 7 shows the influence of k on semantic segmentation results in terms of overall accuracy (OA) and intersection over union (IoU) of tree points. According to our observation, the selection of k had a significant impact on the semantic segmentation results. The performance presents a fluctuation trend with increasing k . Both large k (7, 8) and small k (3, 4) cause semantic segmentation performance degradation. Meanwhile, moderate k parameters (5 and 6) produce better performance. Comparing the results of these two k parameters, $k = 5$ gives better IoU of tree. In addition, the computational cost increases as k increases. Therefore, k is set to 5 to achieve the optimal results.

For individual tree segmentation based on cylindrical convolution network, the tree point clouds are divided into cylindrical blocks to fulfill the input requirement. To investigate the influence of the size of cylindrical representation on the individual tree segmentation performance, we add some parameters to cross-validate these sizes, including 320 × 240 × 24, 480 × 360 × 32, 640 × 480 × 40, and 800 × 600 × 48. Moreover, a cubic partition with a size of 1000 × 1000 × 60 is also added for comparison to prove the effectiveness of cylindrical partition.

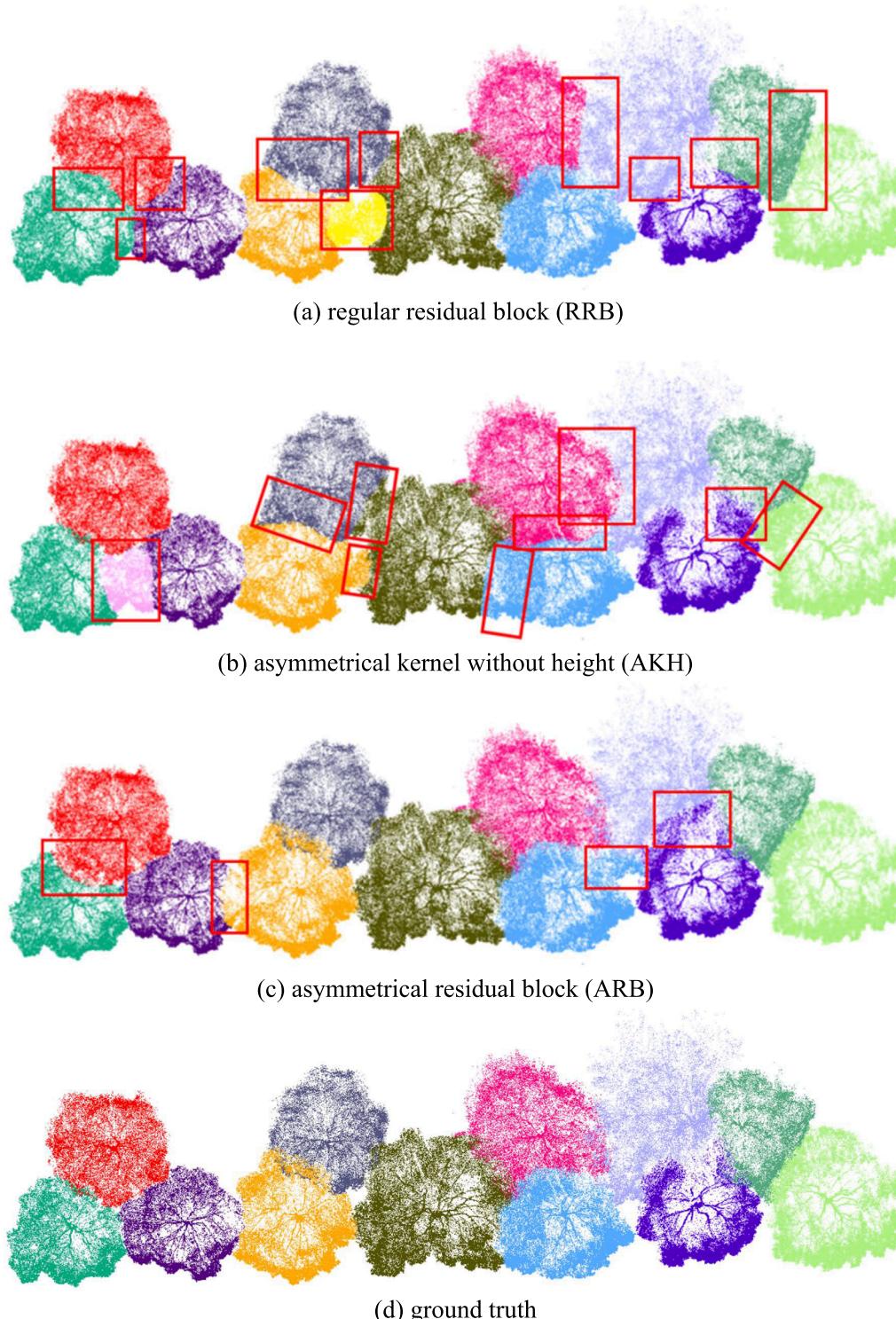


Fig. 11. Performance comparison between the cylindrical convolution networks with different variants of residual block. The red boxes highlight the differences between different residual block variants. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

All experiments are performed on the same settings for a fair comparison. Table 2 lists the numerical comparisons. The SAC of the representation size with $640 \times 480 \times 40$ is the highest and considerably superior to others with a gap of approximately 0.02. The main reason is that compacter representation ($320 \times 240 \times 24$ and $480 \times 360 \times 32$) may mis-split the points across different trees into same cell and a larger representation ($800 \times 600 \times 48$) might burden the training. Note that whether it is a large representation or a small representation, the

proposed method has achieved better performance than the cubic partition with $1000 \times 1000 \times 60$. This fully proves that the performance can be improved by using cylindrical partition.

3.4. Overall performance of individual tree segmentation

Qualitative evaluation of the proposed method in Dataset A, B, and C show three examples of detailed individual tree segmentation outcomes

corresponding to urban scenes with complex trees (Figs. 8–10). The three selected road scenes are colored by the elevation values of every point (Figs. 8–10(a)). The roadside objects are labeled after ground and non-ground point separation, and colored by the label values of every point (Figs. 8–10(b)). The tree point extraction represents points from the tree (green) and non-tree objects (gray), respectively (Figs. 8–10(c)). Street-tree segmentation shows each tree drawn in one color (Figs. 8–10(d)). Some details of the individual roadside segmentation results are shown in Figs. 8–10(e). The proposed method achieves satisfactory segmentation results for multi-tree groups with various shapes and complex position distributions, such as heavily spatially overlapping roadside trees (Figs. 8–10(d)).

Reference trees are obtained by careful segmentation of all tree points using CloudCompare to quantitatively evaluate the performance of individual tree segmentation (Table 3). The proposed method achieved good performance in segmenting roadside trees according to SAC which indicates the quality of instance segmentation of urban tree points. Because the trees of Dataset A are relatively simpler, the corresponding SACs in Dataset A are higher than those of Dataset B and Dataset C. The precision of our method for individual tree segmentation is generally quite high, and some segmentation errors are mainly caused by false positives. Thus, our method is robust to different road scenarios with various roadside trees.

4. Experimental analysis

4.1. Ablation studies

A set of ablation studies are conducted on Dataset B to support the rationality of the main component of the proposed instance-level tree separation.

We evaluate the performance of different variants of the residual block to further investigate the effectiveness of ARB. We first obtain the result by directly using the regular residual block (RRB) without any asymmetrical structure to replace the ARB. Then, another result is obtained from the 1D-ARB without the height, which can strengthen the horizontal or vertical kernels in 1D (named asymmetrical kernel without height, AKH). The results of the two ablation experiments and our ARB that strengthens both horizontal and vertical kernels are shown in Fig. 11. Comparing Fig. 11 (a)–(c) and Fig. 11 (d), shows that ARB significantly boosts the performance. This because ARB effectively matches the object distribution and enhances the robustness to the sparsity.

CMM is a plug-and-play component and can be ported to other models. Ablation experiments are conducted to compare models with and without CMM to verify the importance of CMM in instance-level tree separation framework. Table 4 shows the individual tree segmentation results of two models. Evidently, the validation SAC of full pipeline is approximately 0.03 higher than the version without CMM. It can be concluded that CMM can deliver the effective global context features, contributing to achieve better performance.

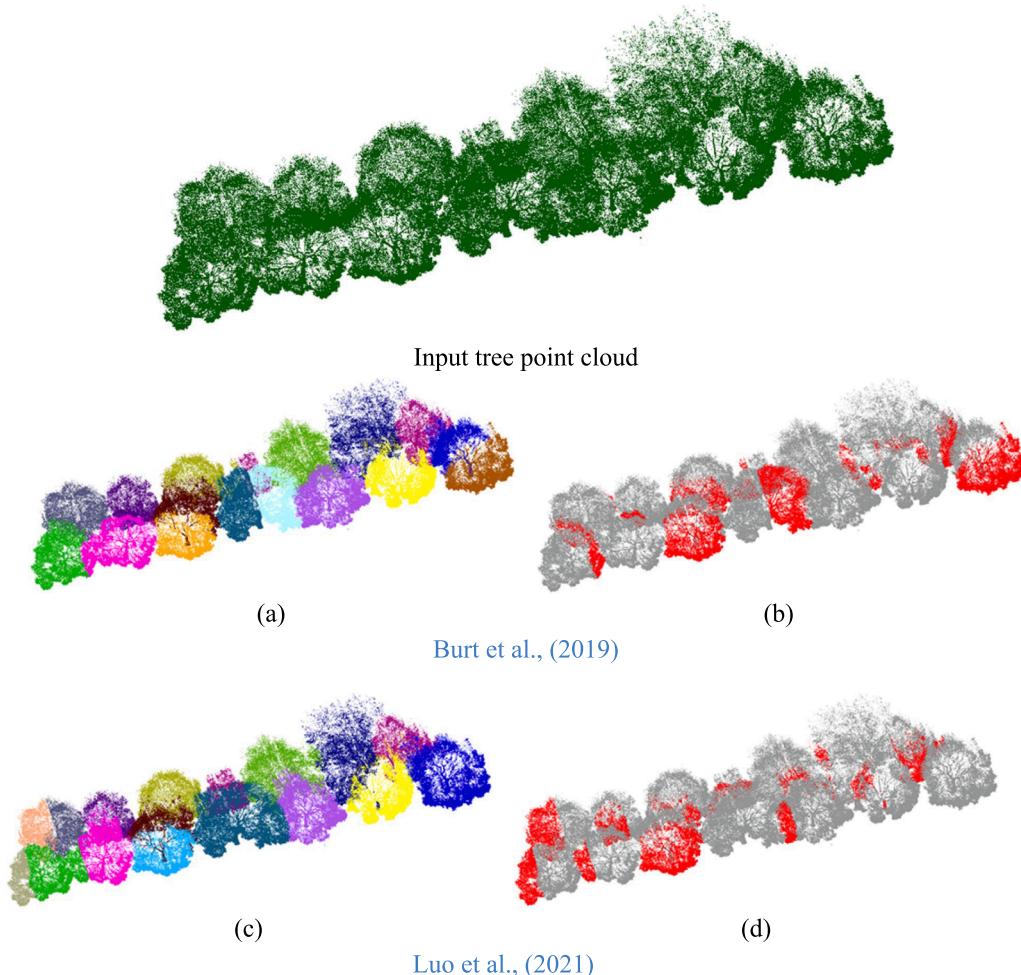


Fig. 12. Individual tree segmentation results with different methods. (a, c, e, g): individual tree segmentation results; (b, d, f, h): the error map of different results.

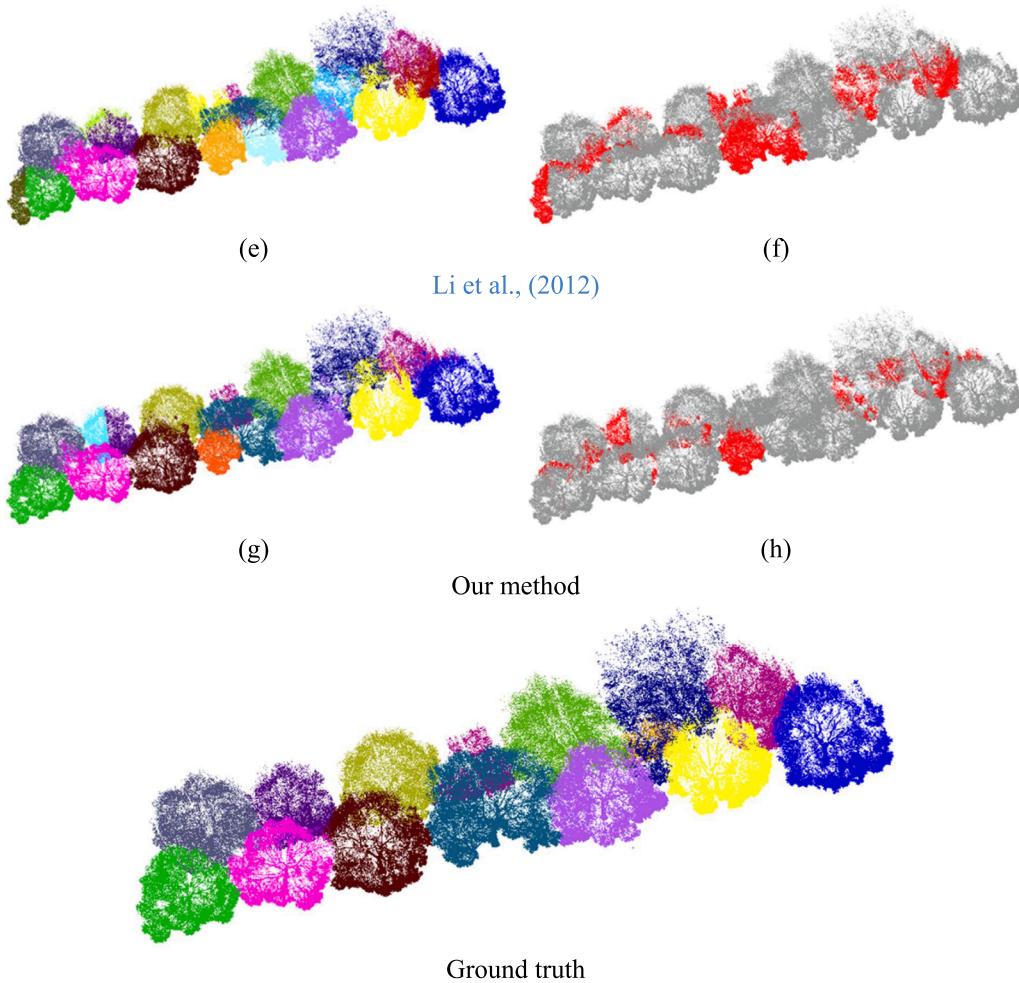


Fig. 12. (continued).

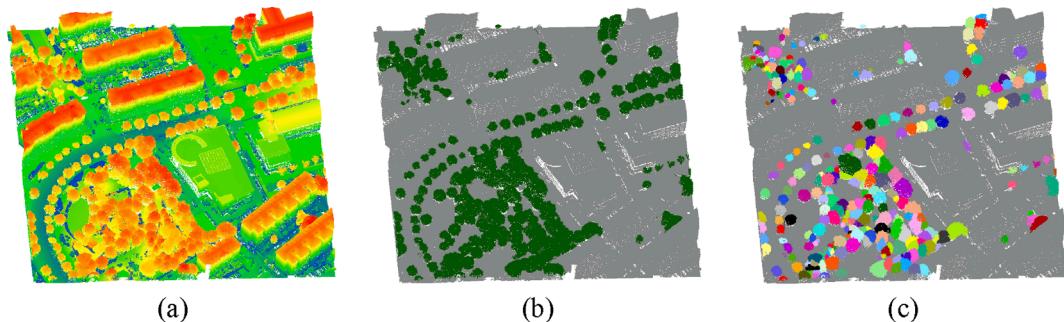


Fig. 13. Individual tree segmentation results by testing with ALS dataset. (a): raw ALS point cloud colored by elevation; (b): the tree point cloud extraction result, green points refer to roadside tree; (c) the individual tree segmentation result for ALS point cloud. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Comparative studies

The proposed tree segmentation method shows competitive results with some published methods based on the same configurations (Burt et al., 2019; Li et al., 2012; Luo et al., 2021) in terms of SAC, OME and COE for individual tree segmentation (Table 3). The average SAC of our method is the highest currently reported, followed by Luo et al. (2021) and Burt et al. (2019) by approximately 2% and 3%, respectively. Furthermore, the classical watershed method (Li et al., 2012) yields the worst SAC. This strategy uses a top-to-bottom marker-controlled watershed segmentation algorithm that cannot distinguish the delicate

boundaries of urban trees with extended and irregular shapes because it considers only height variation. Moreover, some visual comparison examples of segmentation results are shown in Fig. 12. It can be seen that the proposed method has performed well on individual tree segmentation and is superior to comparative methods.

In contrast, there are fewer errors at the boundaries in the proposed method because we iteratively perform shift operations based on a kernel density function. Burt et al. (2019) started that stem detection is critical for individual tree segmentation implying that their method overly depends on the assumptions of stationary tree structures and the quality of input. This may lead to segmentation errors for urban trees

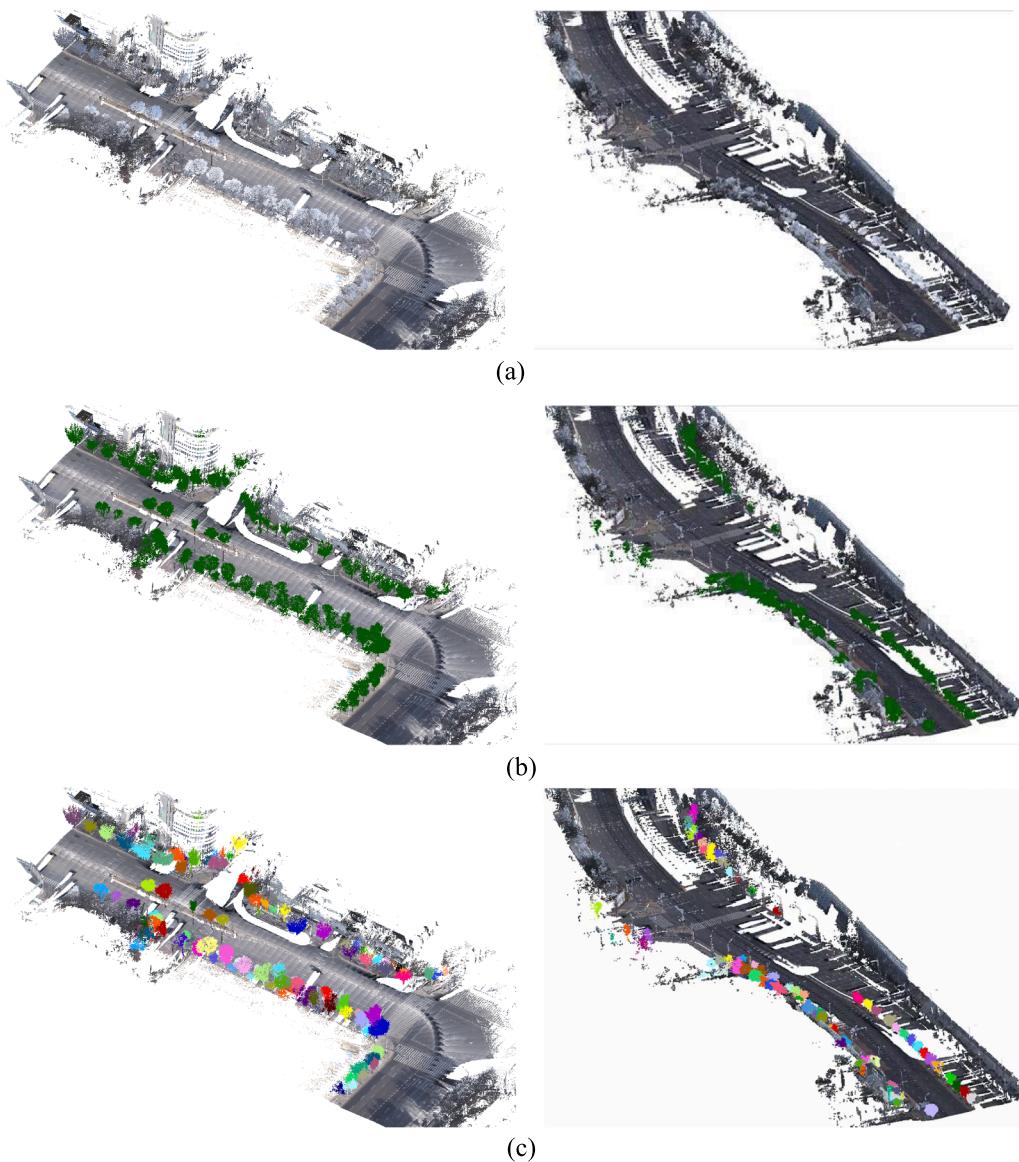


Fig. 14. Individual tree segmentation results by testing with Urban_SGPCM dataset. (a): raw MLS point cloud colored by RGB information; (b): the tree point cloud extraction result, green points refer to roadside tree; (c) the individual tree segmentation result for MLS point cloud. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

without stem points. Unfortunately, the point clouds of tree stems are usually incomplete because of occlusion in environmentally complex urban scenes. Luo et al. (2021) used an end-to-end strategy based on instance segmentation to detect individual tree, which effectively eliminated the requirement for high-quality input. This study performs slightly better in some areas by combining point-wise direction embedding and tree center detection. Although precise individual tree segmentation was realized in theory, Luo et al. (2021) may have significant separation errors in some extreme cases. Moreover, there is an empirical threshold which require time-consuming manual adjustment and limit its generalization. From the runtime comparison in the Table 3, we can see that the proposed method requires a little more running time. The main reason is that our work combines cylindrical convolution and dynamic shift, significantly improves the computational efficiency. In conclusion, the proposed method has consistently steep gains for all testing datasets whether it is a complex or straightforward urban scene. These tree segmentation results confirm that the proposed method is robust and significantly better than other competitive methods.

4.3. Generalization capability

To demonstrate the superiority of the proposed solution using cylindrical voxels, some generalization experiments are performed on the data with different point density and the data coming from different platforms. Specifically, two additional tests are performed on an ALS urban point cloud and a MLS highway point cloud with color information, respectively. The ALS data consists of large regions from three Chinese cities, which is collected using a drone over three days and spanned approximately 5.0 km^2 of the city landscape. By contrast, the density and the size of the point cloud of MLS data are more variable due to occlusions, complicated and incomplete structures. The MLS point cloud dataset, named Urban_SGPCM (Song et al., 2022), is collected in the Pangyo city, South Korea using an MLS system equipped with a ZF 9012 laser scanner and a 360° panoramic camera.

Fig. 13 shows the segmentation result by testing with ALS dataset, which demonstrates that the proposed method is able to achieve promising solutions for individual tree segmentation tasks in urban regions. Although ALS point clouds collected in urban environments are

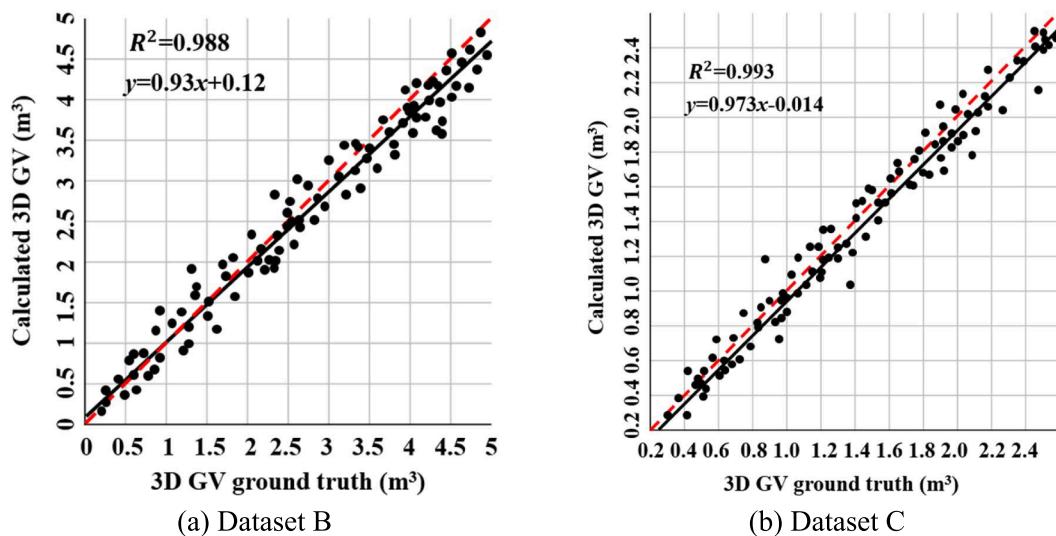


Fig. 15. Comparison of estimated 3D GV with ground truths for the two datasets.

Table 3
Performance comparison between our method and others.

	A	B	C	average SAC	Runtime (s)
Burt et al., (2019)	SAC	0.89	0.88	0.80	0.85
	OME	0.11	0.12	0.20	
	COE	0.11	0.12	0.23	
Luo et al., (2021)	SAC	0.91	0.80	0.81	0.86
	OME	0.09	0.20	0.19	
	COE	0.11	0.18	0.22	
Li et al., (2012)	SAC	0.89	0.81	0.67	0.80
	OME	0.11	0.19	0.33	
	COE	0.20	0.19	0.35	
Ours	SAC	0.92	0.87	0.84	0.88
	OME	0.08	0.13	0.16	
	COE	0.05	0.13	0.20	

different from large-scale MLS data, the results indicate the majority of road-side trees were effectively segmented. Based on the zoom-in visual inspection, several small tree were also completely extracted. In addition, based on the same testing settings, the proposed method is compared with two existing baselines (i.e., SGE_Net (Wang et al., 2020a) and DAE_Net (Luo et al., 2021)). Table 5 provides the performance comparison results by calculating three quantitative evaluation indicators mentioned in Section 3.2. As can be perceived, the proposed method outperforms two segmentation baselines, which are pioneers that segment tree point clouds using deep learning models.

As for the Urban_SGPM, the qualitative results are presented in Fig. 14, which shows that the proposed method can segment the majority of trees from various scenes correctly. We also list the numerical results in Table 6 according to the instance-level evaluation metrics. It can be seen that the SAC of the proposed method is the highest at present and is followed by SGE_Net (Wang et al., 2020a) with a gap of roughly 0.05, while the DAE_Net (Luo et al., 2021) is slightly inferior to SGE_Net by approximately 0.04. In general, our solution can achieve promising individual tree segmentation results for more complex situations (i.e.,

Table 5
Generalization performance comparison on ALS point clouds.

Methods	SAC	OME	COE
SGE_Net (Wang et al., 2020a)	0.81	0.19	0.21
DAE_Net (Luo et al., 2021)	0.83	0.17	0.20
The proposed method	0.85	0.15	0.16

multiple trees distributed in a queue with slight spatial overlap). The main reason is that high quality grid-level features are obtained by the cylindrical convolution from point clouds, which is efficient enough for providing important contextual information.

5. Application: 3D GV calculation

Competitive individual tree segmentation results support many tree-related applications, i.e., 3D GV calculation and leaf-wood separation (Jiang et al., 2023d). We introduce a differential method for the calculation of 3D GV of a single tree to further verify the performance of roadside tree segmentation. The 3D GV calculation is a classical problem of urban ecology (Handayani et al., 2018). Benefited from these high precision instance segmentation results of roadside trees, we provide a satisfactory solution to calculate 3D GV.

First, three parameters, i.e. crown width (CW), crown height (CH), and tree height (TH), are estimated. Subsequently, the maximum and minimum values of 3D geometric coordinates for every tree are obtained to construct the minimum bounding volume. The cell length is determined by the minimum value of the length and width of the leaves of a tested tree. Bounding cuboids shaped as $L \times W \times H$ are divided into small differential cells with a fixed cell length d . A total of $M = \frac{L \times W \times H}{d^3}$ spatial differentiated cells can be obtained. Then, the 3D points are placed into the corresponding spatial cells one by one. The point is labeled as a valid grid point if there is a 3D point within a cell. After obtaining all valid grid points, the model is used to calculate the 3D GV G of a single tree using the differential method, which is expressed as

Table 4
Performance comparison between the cylindrical convolution network with CMM and without CMM. Note that ‘w/o’ refers to ‘without’ and ‘w/’ refers to ‘with’, respectively.

Methods	SAC	OME	COE
w/o CMM	0.84	0.16	0.18
w/ CMM	0.87	0.13	0.13

Table 6
Generalization performance comparison on Urban_SGPM dataset.

Methods	SAC	OME	COE
SGE_Net (Wang et al., 2020a)	0.84	0.16	0.17
DAE_Net (Luo et al., 2021)	0.80	0.20	0.23
The proposed method	0.89	0.11	0.15

Table 7

Calculation results of 3D green volume based on single tree point clouds.

Dataset	ID	DBH/m	CH/m	CW/m	3D GV/m ³
A	#1	0.308	11.701	11.705	2.581
	#2	0.572	12.737	15.901	3.385
	#3	0.618	12.072	16.120	3.648
	#4	0.611	12.554	15.781	3.509
	#5	0.485	11.979	13.450	2.783
	#6	0.777	18.880	11.141	5.500
	#7	0.786	19.161	12.892	5.828
	#8	0.793	19.118	12.302	6.029
	#9	0.799	19.207	13.381	6.220
	#10	0.806	19.218	13.651	6.582
B	#1	0.151	3.647	4.312	0.420
	#2	0.196	3.811	4.469	0.579
	#3	0.187	3.774	4.353	0.501
	#4	0.169	3.610	4.275	0.465
	#5	0.156	3.665	4.419	0.431
	#6	0.319	9.102	11.215	3.336
	#7	0.351	9.826	10.993	3.787
	#8	0.341	9.238	11.002	3.528
	#9	0.445	10.227	11.360	4.018
	#10	0.452	10.440	11.211	4.153
C	#1	0.109	4.666	3.912	0.224
	#2	0.115	4.884	4.083	0.332
	#3	0.121	4.999	4.128	0.355
	#4	0.125	5.020	4.218	0.392
	#5	0.130	5.116	5.003	0.410
	#6	0.359	8.381	5.879	1.857
	#7	0.373	8.541	5.996	1.998
	#8	0.389	8.528	5.981	1.892
	#9	0.405	8.998	6.321	2.222
	#10	0.388	8.450	5.234	1.676

Table 8

Estimation errors of calculated 3D GVs for datasets B and C.

	Average 3D GV (m ³)	Average GT (m ³)	Average errors (%)
Dataset B	2.13	2.21	3.61
Dataset C	1.38	1.43	3.50

$$G = \frac{N}{M} \times V_{box} \quad (11)$$

where N is the total number of valid space cell and V_{box} is the volume of the smallest circumscribed cuboid formed by the space units.

Several trees are selected from each of the three sets of experimental data to calculate 3D GV (Table 7). The 3D GV size is directly related to the tree diameter breast height (DBH), CH, and CW. The larger the CH and CW, the greater the 3D GV, and the smaller the DBH and 3D GV. In other words, the amount of greenness is related to the growth period of the tree.

To evaluate the accuracy of the computed 3D GV based on our individual tree segmentation results, the computed 3D GV results are compared with manual measurements. The estimation errors (the ratio of the absolute difference between the calculated value and the manual measurement to the manual measurement in percentage) on datasets B and C are listed in Table 8. Because dataset A does not provide ground-truth of 3D GVs, there is no evaluation error on dataset A. The average estimation errors are less than 4%, which satisfies the accuracy requirements of the simulation calculation. The linear correlations between the calculated 3D GVs and the ground truths of the two datasets are shown in Fig. 15. The definition of the linear correlation is as follows:

$$R^2 = 1 - \frac{\sum_{k=1}^K (m_k - \bar{m}_k)^2}{\sum_{k=1}^K (m_k - \bar{m}_k)^2} \quad (12)$$

where K denotes the number of trees; m_k is the value of the manual measured 3D GV; \bar{m}_k is the value of 3D GV determined from the segmented tree point clouds; and \bar{m}_k is the mean value of the manually

measured 3D GV.

It is not difficult to find two fitted lines that are close to $y = x$, and the accuracy of the estimated 3D GVs results for datasets B and C were both greater than 0.9. Overall, the proposed individual tree segmentation algorithm supports the calculation of tree 3D GV and lays a solid foundation for the quantitative study of the 3D GV in the entire urban forest.

6. Conclusion

The accurate instance segmentation of roadside trees is a requirement for ecological construction. In this study, a multi-scale method is proposed by integrating semantic and instance segmentation for individual tree segmentation, improving the segmentation accuracy and computational efficiency for the urban MLS point clouds. Overall accuracies of 0.92, 0.87 and 0.84, respectively, on three different MLS datasets demonstrate the effectiveness of the proposed tree segmentation method. These results are superior to that of other methods using the same configurations. Individual tree segmentation results provide support for future eco-city analysis, such as calculating the 3D GV of urban roads. In conclusion, our study offers a potential advance in the instance segmenting roadside trees. However, the proposed method assumes that most roadside trees conform to a cylindrical shape distribution. In the further, the method will be extended to segment trees with non-cylindrical crown morphologies.

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.

Data availability

Data will be made available on request.

Acknowledgement

This study was supported by the National Natural Science Foundation of China under grant No. 42171446 and No. 41771439.

References

- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (11), 1222–1239.
- Burt, A., Disney, M., Calders, K., 2019. Extracting individual trees from LiDAR point clouds using treeseg. *Methods Ecol. Evol.* 10, 438–445.
- Chen, W., Zhu, X., Sun, R., et al., 2020. Tensor low-rank reconstruction for semantic segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, pp. 4558–4567.
- Chen, Y., Wang, S., Li, J., Ma, L., Wu, R., Luo, Z., Wang, C., 2019. Rapid urban roadside tree inventory using a mobile laser scanning system. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (9), 3690–3700.
- Chen, Y., Wu, R., Yang, C., Lin, Y., 2021. Urban vegetation segmentation using terrestrial LiDAR point clouds based on point non-local means network. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102580.
- Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. *ISPRS J. Photogramm. Remote Sens.* 144, 400–411.
- Handayani, H., Estoqe, R., Murayama, Y., 2018. Estimation of built-up and green volume using geospatial techniques: A case study of Surabaya, Indonesia. *Sustain. Cities Soc.* 37, 581–593.
- Hong, F., Zhou, H., Zhu, X., et al., 2021. LiDAR-based panoptic segmentation via dynamic shifting network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 19–25 June, pp. 13085–13094.
- Jiang, T., Zhang, Q., Liu, S., et al., 2023d. LWSNet: A point-based segmentation network for leaf-wood separation of individual trees. *Forests* 14, 1303.
- Jiang, T., Sun, J., Liu, S., Zhang, X., Wu, Q., Wang, Y., 2021. Hierarchical semantic segmentation of urban scene point clouds via group proposal and graph attention network. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102626.
- Jiang, T., Wang, Y., Liu, S., Cong, Y., Dai, L., Sun, J., 2022. Local and global structure for urban ALS point cloud semantic segmentation with ground-aware attention. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15.

- Jiang, T., Wang, Y., Zhang, Z., Liu, S., Dai, L., Yang, Y., Jin, X., Zeng, W., 2023a. Extracting 3D structural lines of building from ALS point clouds using graph neural network embedded with corner information. *IEEE Trans. Geosci. Remote Sens.* 61, 1–28.
- Jiang, T., Wang, Y., Liu, S., Zhang, Q., Zhao, L., Sun, J., 2023b. Instance recognition of street trees from urban point clouds using a three-stage neural network. *ISPRS J. Photogramm. Remote Sens.* 199, 305–334.
- Jiang, T., Liu, S., Zhang, Q., et al., 2023c. ShrimpSeg: a local-global structure for mantis shrimp point cloud segmentation network with contextual reasoning. *Appl. Opt.* 62 (10), 97–103.
- Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, pp. 4558–4567.
- Li, J., Cheng, X., Wu, Z., Guo, W., 2021. An over-segmentation-based uphill clustering method for individual trees extraction in urban street areas from MLS data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 2206–2221.
- Li, J., Cheng, X., Xiao, Z., 2022. A branch-trunk-constrained hierarchical clustering method for street trees individual extraction from mobile laser scanning point clouds. *Measurement* 189, 110440.
- Li, W., Guo, Q., Jakubowski, M.K., Kelly, M., 2012. A new method for segmenting individual trees from the LiDAR point cloud. *Photogramm. Eng. Remote Sens.* 78 (1), 75–84.
- Lin, Y., Wang, C., Zhai, D., Li, W., Li, J., 2018. Toward better boundary preserved supervoxel segmentation for 3D point clouds. *ISPRS J. Photogramm. Remote Sens.* 143, 39–47.
- Liu, H., Dong, P., Wu, C., Wang, P., Fang, M., 2021. Individual tree identification using a new cluster-based approach with discrete-return airborne LiDAR data. *Remote Sens. Environ.* 258, 112382.
- Luo, H., Chen, C., Fang, L., Khoshelham, K., Shen, G., 2020. MS-RFFSegNet: Multiscale regional relation feature segmentation network for semantic segmentation of urban scene point clouds. *IEEE Trans. Geosci. Remote Sens.* 58 (12), 8301–8315.
- Luo, H., Khoshelham, K., Chen, C., He, H., 2021. Individual tree extraction from urban mobile laser scanning point clouds using deep pointwise direction embedding. *ISPRS J. Photogramm. Remote Sens.* 175, 326–339.
- Luo, Z., Zhang, Z., Li, W., Chen, Y., Wang, C., Nurunnabi, A.A.M., Li, J., 2022. Detection of individual trees in UAV LiDAR point clouds using a deep learning framework based on multichannel representation. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15.
- Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuapää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., Vihervaara, P., 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.* 256, 112322.
- Munoz, D., Vandapel, N., Hebert, M., 2009. Onboard contextual classification of 3D point clouds with learned high-order Markov Random Fields. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, pp. 2009–2016.
- Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, pp. 652–660.
- Roynard, X., Deschaud, J., Goulette, F., 2018. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *Int. J. Rob. Res.* 37 (6), 545–557.
- Song, H., Jo, K., Cho, J., Son, Y., Kim, C., Han, K., 2022. A training dataset for semantic segmentation of urban point cloud map for intelligent vehicles. *ISPRS J. Photogramm. Remote Sens.* 187, 159–170.
- Terryn, L., Calders, K., Bartholomeus, H., Bartolo, R.E., Brede, B., D'hont, B., Disney, M., Herold, M., Lau, A., Shenkin, A., Whiteside, T.G., Wilkes, P., Verbeeck, H., 2022. Quantifying tropical forest structure through terrestrial and UAV laser scanning fusion in Australian rainforests. *Remote Sens. Environ.* 271, 112912.
- Wang, D., 2020. Unsupervised semantic and instance segmentation of forest point clouds. *ISPRS J. Photogramm. Remote Sens.* 165, 86–97.
- Wang, P., Tang, Y., Liao, Z., et al., 2023. Road-side individual tree segmentation from urban MLS point clouds using metric learning. *Remote Sens.* 15: 1992.
- Wang, Y., Jiang, T., Liu, J., Li, X., Liang, C., 2020a. Hierarchical instance recognition of individual roadside trees in environmentally complex urban areas from UAV laser scanning point clouds. *ISPRS Int. J. Geoinf.* 9 (10), 595.
- Wang, Y., Jiang, T., Yu, M., Tao, S., Sun, J., Liu, S., 2020b. Semantic-based building extraction from LiDAR point clouds using contexts and optimization in complex environment. *Sensors* 20 (12), 3386.
- Wang, D.i., Liang, X., Mofack, G.I., Martin-Ducup, O., 2021. Individual tree extraction from terrestrial laser scanning data via graph pathing. *For. Ecosyst.* 8, 67.
- Xu, X., Juricich, F., Calders, K., Armston, J., De Floriani, L., 2023. Topology-based individual tree segmentation for automated processing of terrestrial laser scanning point clouds. *Int. J. Appl. Earth Obs. Geoinf.* 116, 103145.
- Xu, S., Xu, S.S., Ye, N., Zhu, F., 2018. Automatic extraction of street trees' nonphotosynthetic components from MLS data. *Int. J. Appl. Earth Obs. Geoinf.* 69, 64–77.
- Yadav, M., Lohani, B., 2020. Identification of trees and their trunks from mobile laser scanning data of roadway scenes. *Int. J. Remote Sens.* 41 (4), 1–26.
- Yan, W., Guan, H., Cao, L., Yu, Y., Li, C., Lu, JianYong, 2020. A self-adaptive mean shift tree-segmentation method using UAV LiDAR data. *Remote Sens.* 12 (3), 515.
- Yu, Y., Li, J., Guan, H., Wang, C., 2015. Automated extraction of urban road facilities using mobile laser scanning data. *IEEE Trans. Intell. Transp. Syst.* 16 (4), 2167–2181.
- Yun, T., Jiang, K., Li, G., Eichhorn, M.P., Fan, J., Liu, F., Chen, B., An, F., Cao, L., 2021. Individual tree crown segmentation from airborne LiDAR data using a novel Gaussian filter and energy function minimization-based approach. *Remote Sens. Environ.* 256, 112307.
- Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., Yan, G., 2016. An easy-to-use airborne LiDAR data filtering method based on cloth simulation. *Remote Sens.* 8 (6), 501.
- Zhou, H., Zhu, X., Song, X., et al., 2020. Cylinder3D: An effective 3D framework for driving-scene LiDAR semantic segmentation. In arXiv preprint arXiv: 2008.01550. <https://arxiv.org/abs/2008.01550>.
- Zhu, X., Zhou, H., Wang, T., Hong, F., Li, W., Ma, Y., Li, H., Yang, R., Lin, D., 2022. Cylindrical and asymmetrical 3D convolution networks for LiDAR-based perception. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (10), 6807–6822.