Heliyon 10 (2024) e36814

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

LESA-Net: Semantic segmentation of multi-type road point clouds in complex agroforestry environment

Yijian Duan [a], Danfeng Wu [c,d], Liwen Meng [a], Yanmei Meng [a,*], Jihong Zhu [a,b], Jinlai Zhang [e], Eksan Firkat [f], Hui Liu [a], Hejun Wei [a]

[a] College of Mechanical Engineering, Guangxi University, Naning, 530004, Guangxi, China
[b] Department of Precision Instrument, Tsinghua University, Beijing, 100000, China
[c] College of Robotics, Beijing Union University, Beijing, 100027, Beijing, China
[d] Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, 100000, Beijing, China
[e] College of Automotive and Mechanical Engineering, Changsha University of Science and Technology, Changsha, 410114, Hunan, China
[f] School of Information Science and Engineering, Xinjiang University, Urumqi, 830046, Xinjiang, China

ARTICLE INFO

Dataset link: https://semantic-kitti.org/dataset.html

Dataset link: https://www.nuscenes.org/nuscenes

Keywords:
Road surface segmentation
Deep learning
Agricultural robot
Point cloud

ABSTRACT

Point-cloud semantic segmentation is a visual task essential for agricultural robots to comprehend natural agroforestry environments. However, owing to the extremely large amount of point-cloud data in agroforestry environments, learning effective features for semantic segmentation from large-scale point clouds is challenging. Therefore, to address this issue and achieve accurate semantic segmentation of different types of road-surface point clouds in large-scale agroforestry environments, this study proposes a point-cloud semantic segmentation network framework based on double-distance self-attention. First, a point-cloud local feature enhancement module is proposed. This module primarily extends the receptive field and enhances the generalizability of multidimensional features by incorporating reflection intensity information and a spatial feature-encoding block that is enhanced with contextual semantic information. Second, we introduce a dual-distance attention pooling (DDAPS) block based on the self-attention mechanism. This block initially learns the feature representation of the local neighborhood of each point through the self-attention mechanism. Then, it uses the DDAPS block to aggregate more discriminative local neighborhood point features. Finally, extensive experimental results on large-scale point-cloud datasets, SemanticKITTI and RELLIS-3D, demonstrate that our algorithm outperforms similar algorithms in large-scale agroforestry environments.

1. Introduction

With advancements in sensor technology, the majority of the agroforestry robots now include integrated sensor and control components [1]. The environmental perception technology of robots allows them to perform different conservation tasks such as forest cutting, clearing, tending, lawn trimming, mowing, crop harvesting, and irrigation in complex agroforestry environments [2], [3], [4]. Among these, road-surface-type identification technology in agroforestry environments is a popular research focus nowadays. This technology allows robots to identify road environments while driving, adjust terrain response settings in advance, and accurately

* Corresponding author.
E-mail address: gxu_mengyun@163.com (Y. Meng).

https://doi.org/10.1016/j.heliyon.2024.e36814
Received 20 May 2024; Received in revised form 15 August 2024; Accepted 22 August 2024

Available online 28 August 2024

reach target locations in challenging agroforestry environments. However, agroforestry environments pose difficulties in extracting their characteristics owing to their complex and changeable conditions, their susceptibility to bad weather disturbances, and diverse road materials such as sand, gravel, riverbeds, mud, and rocks, which influence robot safety and passability [5], [6]. Existing semantic segmentation algorithms require improvements regarding accuracy, robustness, and computational complexity. Hence, recognizing diverse road types in agroforestry environments has become a focal point and challenge in environmental perception technology for robots. To address these issues, we propose an efficient semantic segmentation algorithm for agroforestry road types, allowing robots to achieve autonomous navigation, environmental perception, operational planning, fault detection, maintenance, resource management, and disaster prevention, ultimately improving the efficiency, accuracy, and sustainable development of agroforestry conservation [7].

Based on the sensors used, agroforestry road-type classification technologies can be categorized into image-based semantic segmentation, point-cloud semantic segmentation, and image-point-cloud fusion semantic segmentation. Semantic image segmentation has a rapid development along with the availability of numerous successful algorithms [8], [9], [10]. However, camera-based methods lack accurate depth information, resulting in poor performance in road-type recognition tasks. This method is easily influenced by changes in weather and light intensity, especially in the complex environment of agroforestry where the terrain is complex, occlusion is serious, and the anti-interference ability of the camera is poor. In addition, owing to the large-scale of point clouds in agriculture and forestry environments, semantic segmentation technology based on camera-lidar information fusion has slow recognition speed and large computing resource consumption, which influence the real-time performance and high efficiency of agriculture and forestry robots. Semantic segmentation technology based on lidar has low environmental interference, can function under low-light conditions, and maintains high accuracy, making it a commonly adopted solution for agroforestry maintenance robots to address road-type identification challenges.

However, point-cloud maps contain irregularities, disorders, and large data volumes, rendering traditional image semantic segmentation algorithms unsuitable. In recent years, scholars have proposed different point-cloud segmentation networks for agroforestry environments. These can be categorized as follows: (1) point-cloud segmentation networks based on multiple perspectives and voxels, (2) segmentation networks based on graph optimization, and (3) segmentation networks based on point clouds.

1.1. Segmentation network based on multiple views and voxels

Inspired by convolutional neural networks (CNNs) for image processing, Su et al. introduced a multiview CNN (MVCNN) in 2015 [11]. This method generates multiview two-dimensional (2D) representations of objects using 2D images captured from multiple cameras at different angles. These views are then aggregated into three-dimensional (3D) descriptors using a set CNN to achieve classification. SnapNet [12] utilizes depth-segmentation networks to segment an RGB map created using multiple virtual cameras and depth-synthesized images at the pixel level. Improved segmentation is achieved by fusing the features with a residual correction. However, it does not fully leverage spatial context information. Using the SqueezeNet [13] method, SqueezeSeg [14] and SqueezeSegV2 [15] obtain point-cloud information via spherical projection. SqueezeSeg uses the SqueezeNet network to extract features from a projected front view. Conditional random fields are used as recursive layers to further optimize the segmentation results. SqueezeSegV2 mitigates the adverse effects of noise by adding a context-aggregation module to the SqueezeSeg method. An unsupervised domain-adaptive training pipeline is used to solve data migration problems in different fields. The distribution gap between the synthetic and real data is significantly reduced. Moreover, RangeNet++ [16] uses spherical projection to process the input point cloud and uses a 2D CNN to learn the view features. Thus, semantic labels are obtained by combining the distances obtained by the learned laser scanning. Compared with the above methods, this method can effectively improve the adverse effects of information loss inherent in multiple views and is more accurate and faster. Overall, multiview segmentation methods are susceptible to virtual perspective effects and may not fully exploit the geometric structure information in point clouds. Reshaping 3D information from 2D information inevitably leads to information loss, resulting in poor segmentation accuracy. Spherical projection retains more information than conventional multiview methods; however, it continues to face challenges in addressing occlusion and other scenes.

Voxelization transforms irregular 3D point clouds into structured voxel data, thereby addressing the disordered and unstructured nature of point clouds. Similar to 2D image convolution, 3D CNNs can learn from voxelized point-cloud data. Because the voxelization method can retain more original point-cloud structural information than the multiview projection method, numerous scholars have conducted studies in this regard. VV-Net [17] utilizes kernel-based interpolating variational autoencoders and radial basis functions to address the continuous representation of local point clouds. This method captures the distribution of points in each voxel and enriches the local geometric representation. It combines grouping convolution to effectively reduce the computational requirements; however, the lack of correlation between channels can influence the segmentation accuracy. PVCL [18] proposes a comparative learning network between the voxel and primary levels. Voxel-level contrast learning reduces the intraclass distance and increases the interclass distance between samples. Prototype-level contrast learning reduces the dependence of contrast learning on negative sampling and avoids the influence of outliers from the same class. Thus, PVCL can be used more effectively for outdoor point-cloud panoramic segmentation. In another study [19], the original point-cloud coordinates were transformed into cylindrical coordinates and voxels were constructed using these coordinates. Subsequently, asymmetric 3D convolutions are used for feature learning.

This method allows each voxel to contain dense points near to and sparse points far from each other. Subsequently, this significantly increases the proportion of nonempty grids (grids with point clouds), reduces the memory footprint and redundant information in calculations, and achieves the highest level in the SemanticKITTI dataset for the same period. However, the voxel-based approach has limitations such as a lower resolution than the original point cloud, information loss, unnecessary memory consumption, and

computing requirements. Moreover, the voxel size must be controlled within an effective range to maintain the model's accuracy and memory usage.

Conversely, the proposed algorithm operates directly on points, avoiding information loss and memory resource occupation caused by voxel segmentation. It leverages spatial location information to obtain rich local and global features, making it more conducive to feature learning and the recognition of multiple surface point clouds in complex agroforestry environments.

1.2. Segmentation network based on graph optimization

Graph convolution-based point-cloud semantic segmentation considers each point in the point cloud as a vertex of a graph and generates directed edges connecting neighboring points. By learning the features of the points and edges in either the spatial or spectral domains, it captures the local geometric structural features of the 3D point cloud. The spatial-domain graph convolution effectively utilizes the information from each node and its neighboring nodes [20].

To enhance segmentation accuracy in large-scale point-cloud scenes, SPG [21] first divides point clouds into geometrically simple yet meaningful superpoints. These superpoints are connected by superedges to form a superpoint map, and each superpoint is embedded in a PointNet [25] network. The information is refined along the hyperedges of the gated cycle unit to produce the final label. This method provides detailed descriptions of the relationships between adjacent shapes and is suitable for large-scale point-cloud scenarios. However, it faces challenges relative to achieving object division, which can lead to classification errors.

To address the problem of ignoring the correlation between neighborhood points when using PointNet [25], DGCNN [22] introduced edge convolution (EdgeConv) to learn edge features. It constructs local neighborhood graphs and performs EdgeConv operations on each adjacent edge, dynamically updating the graph structure between levels. EdgeConv captures the distance information between each point and its neighboring points; however, it overlooks the direction information of the vector between neighboring points, leading to a loss of structural information. Subsequent graph convolution methods such as LDGCNN [23] and GACNet [24], effectively acquire the local feature information of the point clouds. However, owing to their high computational costs, they are challenging when applied to the semantic segmentation of large-scale point clouds.

In the proposed algorithm, random sampling (RS) is utilized to significantly enhance the sampling speed of the point clouds. Moreover, simple geometric calculations are used to directly extract the spatial features of the key point K neighborhood, significantly reducing the calculation costs and memory consumption. This approach offers an efficient and effective solution for large-scale point-cloud semantic segmentation.

1.3. Point-based segmentation network

PointNet [25] is a pioneering point-cloud segmentation algorithm that directly operates on individual points. It utilizes a multilayer perceptron (MLP) to approximate function h by extracting local features from each point. PointNet concatenates the global features obtained through aggregation with the local features from each point. The MLP then extracts new features from the features of the merged points. Therefore, the semantic labels corresponding to each point can be predicted. However, the original PointNet does not consider the local structure information between adjacent points. Subsequently, scholars have proposed different optimization models based on PointNet, all of which have achieved acceptable results [26], [27]. For instance, Qi et al. [28] proposed PointNet++, which is based on the PointNet model and uses hierarchical neural networks to capture local geometric features. Nevertheless, conducting a K-nearest neighbor (KNN) search in PointNet++ can cause the K-nearest points to be stuck in a particular direction. Therefore, this method cannot efficiently perceive the local feature information of the point cloud, leading to a loss of effective information and low segmentation precision. The literature [30] proposes a novel method to enhance 3D semantic perception by directly processing point cloud data with a convolution-like operator that dynamically attends to local semantic information. The literature [31] Correlation-Based Approach to introduces a kernel correlation learning block (KCB) that enhances network perception by adaptively learning local geometric and global features, integrating them based on kernel correlation, and ensuring end-to-end compatibility with typical point cloud structures. The literature [32] proposes a novel enhanced local semantic learning transformer (LSLPCT) for 3-D point cloud analysis, featuring an efficient local semantic learning self-attention mechanism (LSL-SA) that enhances local semantic feature perception and integrates seamlessly with existing transformers and CNN-based networks for various point cloud tasks. Nowadays, with the expansion of the application field of point-cloud maps, the scale of point-cloud maps has risen to millions of points or even tens of thousands of square meters. The majority of the existing point-cloud segmentation algorithms can only be applied to small-scale point clouds owing to the large amount of computing resources required. To address these issues, RandLA-Net has emerged as a pioneering approach that utilizes RS to address point clouds. This network compensates for the information loss resulting from RS by incorporating a feature aggregation module, delivering excellent results on large-scale point clouds.

The network structures of DLA-Net [33] and RandLA-Net are similar. However, although both involve local feature coding, the two algorithms learn different features. In addition, DLA-Net adds a self-attention weight to the learning features in a manner similar to that of a point transformer. The algorithm then uses a self-attention pool, such as RandLA-Net, to aggregate the functionality. However, the self-attention pool input of DLA-Net also adds a local feature-encoded feature input [29]. In 2021, Fan et al. proposed the SCF-Net model [34], consisting of a local polar-representation block, DDAPS block, and global contextual feature block. The local polar-representation block establishes a local spatial representation of constant rotation along the z-axis. The design of the DDAPS blocks is based on the geometric and feature distances to learn effective local features. The global context feature module uses the spatial location and volume ratio of each neighborhood of the 3D point to the global point cloud to learn its global context. In 2022, Meng et al. [35] introduced a new point-cloud local feature aggregation module to achieve fast semantic segmentation of shrub-point clouds in

large-scale agroforestry environments. By incorporating the SPF features of the point clouds, the local feature aggregation unit gains more effective feature information and better aggregates the shrub-point-cloud features. The attention pooling layer is improved by adopting enhanced ECA-net multichannel attention to achieve more accurate attention allocation and enhance training efficiency. Currently, most studies that improve upon the RandLA-Net baseline algorithm utilize RS methods to reduce computational complexity and enhance point-cloud segmentation efficiency. However, RS can easily lose some important point-pair information, potentially leading to decreased segmentation accuracy and precision. Most current solutions [33], [29] use fixed metrics such as Euclidean distance to represent predefined neighboring points. However, in road semantic segmentation, different types of roads may overlap, such as water, mud, and grass coexisting in a swampy terrain. Therefore, during neighborhood construction, outliers and overlaps between neighborhoods are inevitable, especially when points are densely distributed near the boundaries of different semantic classes. The strict encoding of these methods under fixed constraints in 3D space can weaken the generalizability of high-dimensional feature space. To address these issues and enhance feature generalizability, we propose a local feature enhancement (LFE) module. Unlike the aforementioned methods, we first enhance the geometric information of points by leveraging rich local semantic context information and geometric positional information. Simultaneously, considering the significant differences in scattering/absorption characteristics of point clouds for different types of roads, and the fact that point-cloud reflection intensity is an important indicator reflecting the physical properties of road types, we incorporate point-cloud reflection intensity information to enhance the semantic information of points, thereby improving the discriminative ability for different types of roads. In addition, considering that distance is an important variable to measure the correlation between points, we propose a DDAPS block based on the self-attention mechanism. Unlike SCF-Net [34], to better utilize multidimensional feature information to capture more discriminative geometric features, we first learn the feature representation of local neighborhood of each point through the self-attention mechanism and then utilize the DDAPS block to better aggregate the local neighborhood point features learned by the self-attention block. Finally, we propose a point-cloud semantic segmentation network called LESA-Net. This method not only retains important point-pair information but also fully leverages point-cloud intensity information and local contextual features, thereby improving the accuracy and robustness of point-cloud segmentation.

In summary, our primary contributions are as follows: (1) By addressing the poor segmentation performance of existing semantic segmentation algorithms on different types of road surfaces in complex agroforestry environments, we propose a new point-cloud semantic segmentation network called LESA-Net. (2) The Local Feature Enhancement (LFE) module is designed, which improves the generalization ability of the network in the high-dimensional feature space, thus improving the segmentation accuracy of various road types. (3) We introduced a DDAPS block based on the self-attention mechanism. The self-attention mechanism is initially used to learn the feature representation of the local neighborhood of each point. Then, the DDAPS block aggregates the local neighborhood point features learned by the self-attention block, allowing better learning of effective local contextual features and achieving more accurate feature learning and attention allocation. (4) Experimental results demonstrate that LESA-Net outperforms other similar algorithms in segmenting various road types such as grassland, water, and soil in complex agroforestry environments, significantly enhancing the efficiency and safety of agricultural and forestry robots.

2. Materials and methods

2.1. LESA-Net overall network architecture

Fig. 1 presents the overall network structure of the LESA-Net algorithm presented in this study. The algorithm structure follows the classic encoder–decoder architecture. The input network for the ambient point cloud initially undergoes a fully connected (FC) layer to increase the feature dimension to (N, 8), where N is the number of point-cloud points. Subsequently, the point-cloud data undergoes four layers of RS and an LFE module, followed by a DDSAP module (LESA) module. The RS ratios of each layer are set as 1/4, 1/4, 1/4, and 1/4. The dimensionality of the obtained depth feature information reaches (N / 256,512), and the feature encoder construction process is completed. Next, the original feature information is recovered through the shared MLP, which continuously performs upsampling to achieve the decoder stage. Finally, the semantically segmented map is obtained through the FC layer and dropout. Subsequently, the semantic segmentation of road surfaces of different types was realized in a large-scale agroforestry environment.

2.2. Architecture of LFE

It has been demonstrated [29] that RS is the fastest method to downsample large-scale point clouds. Other traditional algorithms, such as farthest point sampling [36] and inverse density importance sampling [37], are slower. Hence, for the semantic segmentation of road types in large-scale agroforestry environments, this study first adopts the RS method to reduce computational complexity and enhance the efficiency of the point-cloud segmentation. However, it is worth noting that the RS method can overlook crucial point-pair information, potentially decreasing segmentation precision and accuracy. To address this issue, this paper proposes an LFE module. This module simultaneously enhances the local geometric and feature information of points, thereby expanding the receptive field and improving the segmentation accuracy for various road types. (See Fig. 2.)

Through 3D scanning and sampling of agroforestry environments, lidar sensors can rapidly acquire large-scale, unstructured, and disordered 3D point-cloud data. Each point in the point cloud contains spatial coordinates (x, y, z) and attribute information such as reflection intensity, which is a simple and realistic representation of the real world. Thus, the inputs to the proposed point-cloud LFE module are $P_i(x, y, z)$ of the point-cloud spatial information and f_i of the point-cloud feature information. The reflection intensity of the point cloud serves as a critical source of spectral information related to road properties in agroforestry environments. In addition,

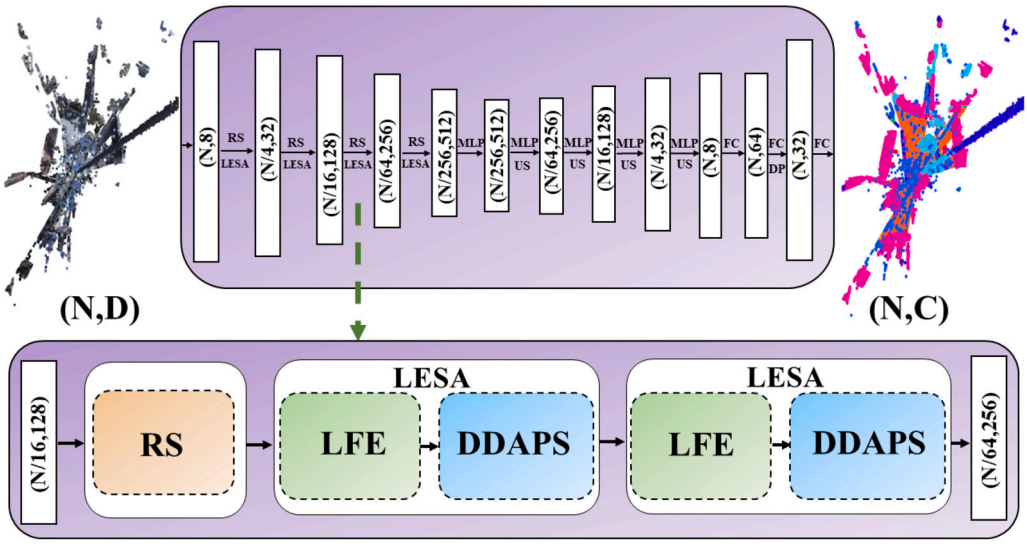

Fig. 1. Structure of the LESA-Net. (N, D) represents the number of input point cloud points and the characteristic dimension of point cloud; (N, C) represents the number of input point cloud points and the number of output point cloud labels; FC is the fully connected layer; RS is random sampling. LESA consists of LFE and DDAPS. RS is the upsample and MLP is the multilayer perceptron.

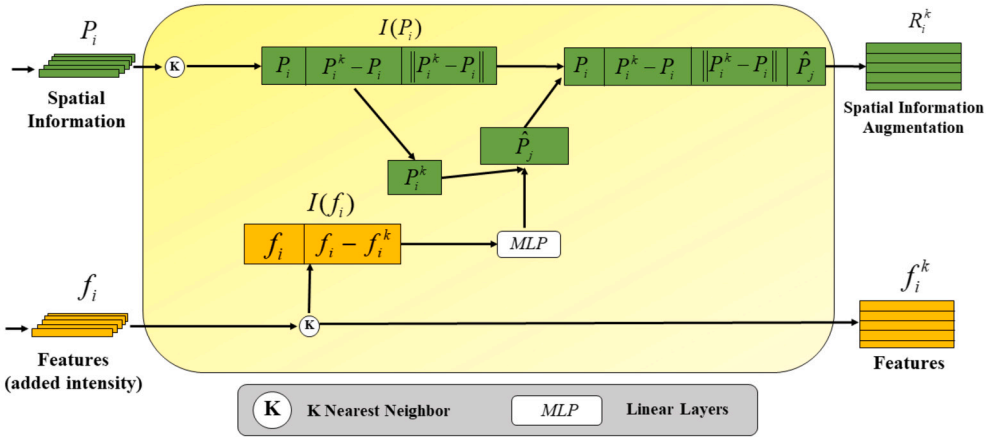

Fig. 2. Structure of the LFE.

it is an important indicator reflecting the physical characteristics of road surfaces. In the design of point-cloud semantic feature enhancement, we first consider that point-cloud reflection intensity information varies significantly among different types of road-surface materials. Herein, the point-cloud intensity is introduced as a semantic feature to enhance the overall semantic representation of the point cloud within the semantic segmentation network. Furthermore, we utilize the KNN method to identify the central point and K adjacent points. Then, the absolute feature of the point and the feature relative to the domain point are combined into the local semantic context information $I(f_i)$. The concrete implementation involves copying the features of point i K times. The eigenvalues of point i are then subtracted from the eigenvalues of the k points around point i to obtain the relative eigenvalues. Finally, the relative feature value is concatenated with the feature value of point i in the feature dimension.

$$I(f_i) = [f_i; f_i^k - f_i] \tag{1}$$

For the design of the point spatial information enhancement, we combine the point space coordinate P_i, relative distance information $P_i^k - P_i$, and Euclidean distance $\|P_i^k - P_i\|$ of the center point to effectively enhance the spatial information of the center point. Then the local geometric context information expression is ad follows:

$$I(P_i) = [P_i; \|P_i^k - P_i\|; P_i^k - P_i] \tag{2}$$

Subsequently, the spatial information of neighborhood points $I(P_i)$ by utilizing rich context information $I(f_i)$. The main methods include using the MLP on $I(f_i)$ to estimate the semantic offset values of neighboring points. Therefore, the semantic context information offset $\hat{P}_j$ are formulated as,

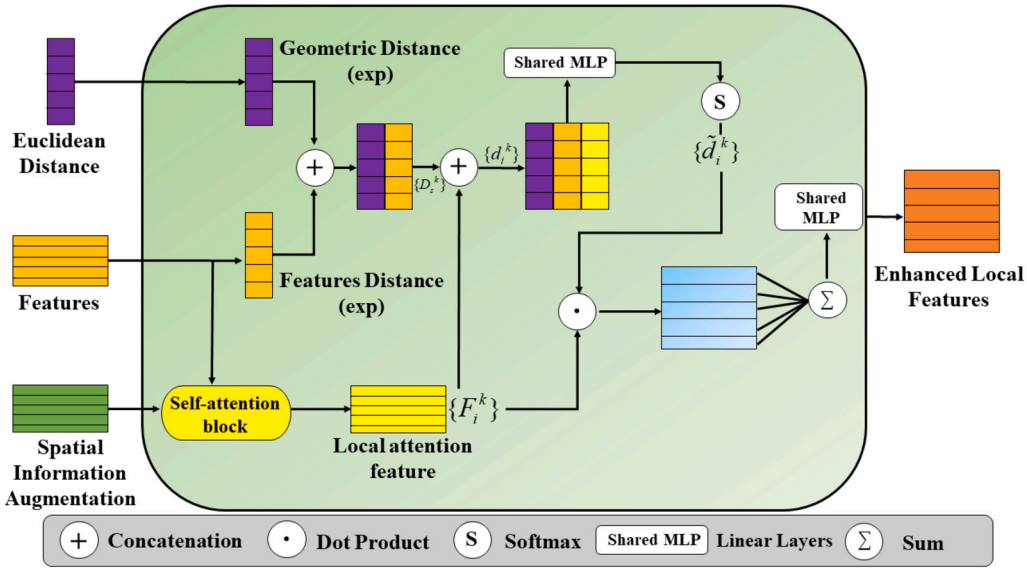

Fig. 3. Structure of the DDAPS.

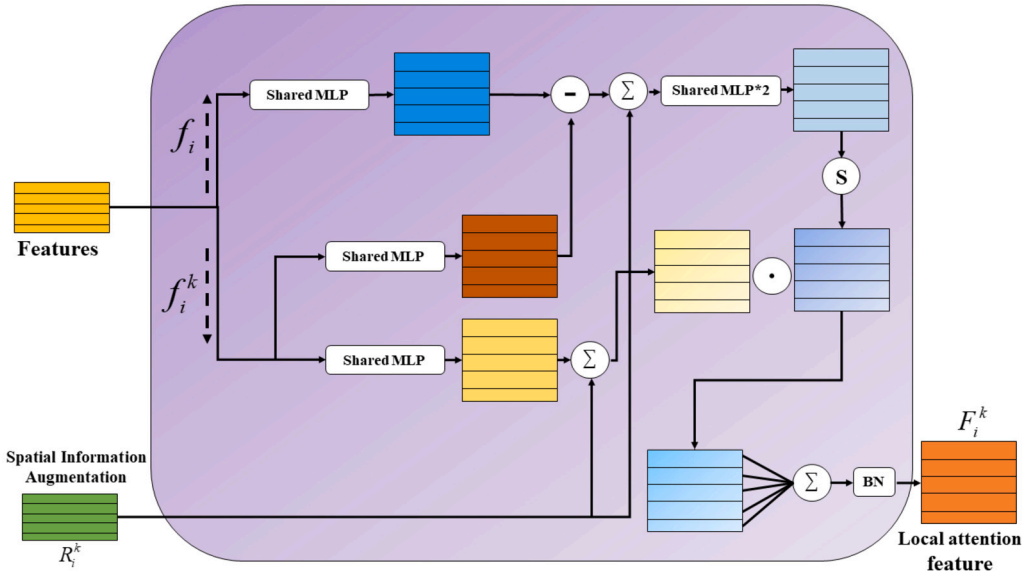

Fig. 4. Structure of the self-attention block.

$$\hat{P}_j = MLP(I(f_i)) + P_i^k, \hat{P}_j \in \mathbb{R}^n \tag{3}$$

Next, we use the semantic context information offset $\hat{P}_j$ to enhance the local geometric context information $I(P_i)$. The expression is as follows,

$$R_i^k = [P_i; \|P_i^k - P_i\|; P_i^k - P_i; \hat{P}_j] \tag{4}$$

2.3. Architecture of DDAPS

Distance is a key variable in measuring the correlation between neighboring points; the smaller the distance, the stronger the correlation. Therefore, we propose a DDAPS. Unlike SCF-Net [34], we do not directly concatenate the spatial encoding information with the point features. Initially, we leverage a self-attention mechanism to learn more effective feature representations of each point's local neighborhood. Subsequently, we employ the DDAPS block to aggregate the local neighborhood point features learned by the self-attention block. This aggregation is based on the Euclidean distance and feature distance between adjacent points to obtain more discriminative local features. Fig. 3 presents the architecture of DDAPS.

We first learn the feature representation of the local neighborhood of each point through the self-attention mechanism. As shown in Fig. 3, DDAPS has three inputs: Euclidean distance, point features, and spatial information augmentation. Then, we input the point features and spatial information augmentation into the self-attention module, which outputs the learned local attention features F_i^k. Fig. 4 presents the structure of the self-attention mechanism. The detailed process is as follows:

The self-attention mechanism utilizes the subtraction between the central point features f_i and the neighborhood point features f_i^K, combined with spatial information R_i^k to form the self-attention weights. These weights are then used to perform a dot-product operation with the neighborhood point features and spatial information features, completing the self-attention weighting process. The self-attention module is expressed as

$$F_i^k = \sum_{k=1}^{k} softmax(\eta(\alpha(f_i) - \beta(f_i^k) + R_i^k)) \odot (\gamma(f_i^K) + R_i^k) \tag{5}$$

where F_i^k denotes the output feature and $\odot$ denotes the Hadamard product. α, β, and γ are MLPS of the linear layer. The mapping function η is an MLP with two linear layers and a rectified linear unit activation. Herein, we used two linear layers to initially convert the feature to a higher dimension (2D) and then revert it to the original dimension (D).

Next, we use the DDAPS block to aggregate the local neighborhood point features learned by the self-attention block. This aggregation exploits the geometric distance and the feature distance between adjacent points to produce more discriminative local features.

As shown in Fig. 4, in the DDAPS module, the two distances are the Euclidean distance D_{iR}^k of the point in the world space and the feature distance D_{is}^k of the point in the feature space. It should be noted that the neighborhood of the feature distance D_{is}^k of the point in the feature space is still the neighborhood of the Euclidean space, not the neighborhood of the feature space. Without loss of generality, let $D(i)$ and $D(k)$ denote the input feature vectors of the i-th point and its k-th (k = 1, 2,..., K) neighbor to the DDAPS block, respectively. The feature distance D_{is}^k between $D(i)$ and $D(k)$ is defined as:

$$D_{is}^k = mean(|D(i) - D(k)|) \tag{6}$$

where || denotes the L1 norm and *mean*() denotes the mean function. The negative exponents of both are used to learn the attention-pooling weights. In addition, λ is introduced to adjust D_{is}^k to solve the instability problem. The equation for D_s^k is as follows:

$$D_s^k = exp(-D_{iR}^k) \oplus \lambda exp(-D_{is}^k) \tag{7}$$

where $\oplus$ denotes the concatenation operation. Function exp(-d) monotonically decreases with respect to distance and is always greater than zero. Therefore, the greater the distance D_{is}^k, the smaller the weight. Subsequently, the dual-distance feature D_s^k and feature F_i^k are concatenated; the equation is as follows:

$$d_i^k = D_s^k \oplus F_i^k \tag{8}$$

Then, a shared MLP followed by softmax is applied to d_i^k, and the attentive pooling weight $\widetilde{d_i^K}$ is learnt automatically as,

$$\widetilde{d_i^K} = softmax(MLP(d_i^k)) \tag{9}$$

Finally, the local contextual features are obtained by calculating the weighted-sum of the neighboring point features with the learnt weights $\widetilde{d_i^K}$.

$$\hat{F}_i = \sum_{k=1}^{k} (\widetilde{d_i^K} \cdot F_i^k) \tag{10}$$

3. Results and discussions

Our experiments were conducted on a computer equipped with an AMD Ryzen 7 4800H CPU running at 2.90 GHz with Radon graphics, 16 GB RAM, and an RTX2060-6G GPU. The main parameters of the network were as follows: we employed the Adam optimizer with the default parameters, setting the initial learning rate to 0.01 and decreasing it by 5% after each period. The number of nearest points K was set to 16. For the parallel training of the proposed algorithm, we sampled a fixed number of points from each point cloud as the input.

3.1. Evaluation index

We employed intersection over union (IOU) and mean intersection over union (MIOU)) as the evaluation metrics to compare algorithms. IOU represents the ratio between the intersection of the predictions (Pre) and true values (True) for a certain category and their union. This can be calculated using the following equation:

$$IOU = \frac{Pre \bigcap True}{Pre \bigcup True} \times 100\% \tag{11}$$

Table 1
Semantic segmentation results on Rellis-3D dataset. Note that the original dataset had 14 types of annotations, but we didn't include them because the number of annotations for 3 types was too small.

	MIOU (%)	Inference Time (ms)	Grass	Tree	Vehicle	Log	Person	Bush	Concrete	Barrier	Puddle	Mud	Rubble
KPConv	**27.4**	**313**	**56.3**	**49.3**	**2.1**	**1.0**	**81.3**	**57.3**	**33.2**	**3.5**	**0.8**	**5.3**	**11.2**
GACNet	**41.9**	-	**45.3**	**84.9**	**38.9**	**7.8**	**85.3**	**61.0**	**23.4**	**75.3**	**10.3**	**8.3**	**21.4**
RandLA-Net	**42.2**	**66**	**44.1**	**81.8**	**40.8**	**5.3**	**88.1**	**58.4**	**23.5**	**72.1**	**6.5**	**11.1**	**32.3**
SCF-Net	**44.7**	**62**	**53.3**	**82.3**	**42.1**	**6.0**	**88.7**	**65.2**	**24.4**	**73.5**	**8.5**	**13.2**	**34.7**
BushNet	**44.2**	**81**	**55.6**	**80.7**	**38.5**	**4.6**	**86.4**	**69.8**	**24.7**	**74.2**	**5.8**	**12.6**	**33.1**
Ours	**46.5**	**93**	**58.3**	**86.1**	**41.9**	**5.9**	**90.1**	**66.8**	**26.3**	**75.4**	**13.9**	**13.5**	**33.7**

The MIOU represents the ratio between the intersection and union of the results predicted by the model for each class and true value IOU_i^m, assuming m classes, the result of summing, and then averaging. The MIOU Equation is as follows:

$$MIOU = \frac{1}{m}\sum_{i=1}^{m} IOU_i^m \tag{12}$$

The tow aforementioned indices are widely recognized as important metrics for evaluating the performance of semantic-segmentation algorithms because they effectively reflect the accuracy and robustness of the algorithm's classification results.

3.2. RELLIS 3D dataset experiment

The data used for the evaluation in this study were from the RELLIS 3D dataset [38], which was obtained from Texas A&M University in the United States. This is a multimodal dataset captured in a field setting and contains annotations of 13,556 LiDAR scans and 6,235 images. Among these data, point-cloud labels were available for 14 categories. For our evaluation, we selected the 0004 sequence from the RELLIS 3D dataset. The 0004 sequence provides rich point-cloud information including a substantial number of point clouds representing grassland, water, concrete, and mud, which aligns well with our verification objectives. Therefore, this sequence was selected as a compromise for the evaluation. To ensure fairness, we used a fixed number of points, 24,570, as input. The primary aim of the proposed algorithm is to enhance the segmentation accuracy of different types of road surfaces, such as grasslands, mud, concrete, and water, in large-scale agroforestry environments. We believe that this dataset is well-suited for verifying the segmentation performance of the proposed algorithm on diverse road-surface point clouds in agroforestry environments. We selected several recently proposed algorithms for comparison: KPconv [39], GACNet [24], RandLA-net [29], SCF-Net [34], and BushNet [35]. The main evaluation metrics used were the MIOU, Inference Time, and IOU for the different categories. The experimental results are presented in Table 1.

Table 1 indicates that the proposed algorithm performed optimally in the MIOU, Grass, Tree, Person, Concrete, Barrier, Puddle, and Mud dimensions. Compared with the BushNet algorithm, it improved by 2.3%, 2.7%, 5.4%, 3.7%, 1.6%, 1.2%, 8.1% and 0.9%, respectively. Although our proposed algorithm is only 12 ms slower than BushNet, our MIOU is greatly improved compared with BushNet. The above results prove that the segmentation accuracy of multiple types of road surfaces (grassland, water ground, concrete ground, and mud) in agricultural and forestry environments improved. In addition, this indicates that the proposed algorithm can achieve acceptable semantic segmentation effects on different surfaces in the complex environments of agriculture and forestry.

Figs. 5 (1), (2), and (3) present the comparison of segmentation effects on the RELLIS-3D dataset. The red-marked boxes in Fig. 5 (1) indicate that our algorithm successfully segments the trees, whereas SCF-Net fails to do so. In addition, the SCF-Net algorithm oversegments the barrier, causing edge deformation. However, our algorithm is more accurate for barrier segmentation. As shown in the red-marked box of Fig. 5 (2), our algorithm achieves higher accuracy in the segmentation of grass, bush, mud, and other road surfaces than SCF-Net. Furthermore, the red-marked box in Fig. 5 (3) further highlights that our algorithm outperforms SCF-Net in puddle, grass, bush, and mud segmentation accuracy.

3.3. SemanticKITTI dataset test

In this study, the large-scale point-cloud dataset SemanticKITTI [40] was selected as the experimental data. SemanticKITTI was developed by Behley et al. based on the semantic annotation of LiDAR point-cloud data from the KITTI dataset collected by the Karlsruhe Institute of Technology in Germany. The dataset comprised 22 sequences of point clouds, totalling 43,552 densely annotated LiDAR scans belonging to 21 sequences. Each scan represents a substantial point cloud containing approximately 105 points covering a 3D space of up to 160 m × 160 m × 20 m. The original 3D points contained only spatial coordinates and lacked color information. The evaluation was based on the MIOU scores in more than 19 categories that served as standard indicators. The SemanticKITTI dataset included different scenes such as traffic fields, residential areas, high-speed road scenes, and rural roads in the center of Karlsruhe, Germany. It provided abundant point-cloud information including grass and trees, making it suitable for evaluating the performance of point-cloud semantic-segmentation algorithms applicable to agroforestry environments. We selected several recently proposed algorithms for comparison: RangeNet++ [16], RandLA-Net [29], SqueezeSeg V2 [15], SCF-Net [34], FA-ResNet [41] and GAF-Net [42]. The main evaluation metrics used were the MIOU, Inference Time, and IOU for the different categories. The experimental results are presented in Table 2.

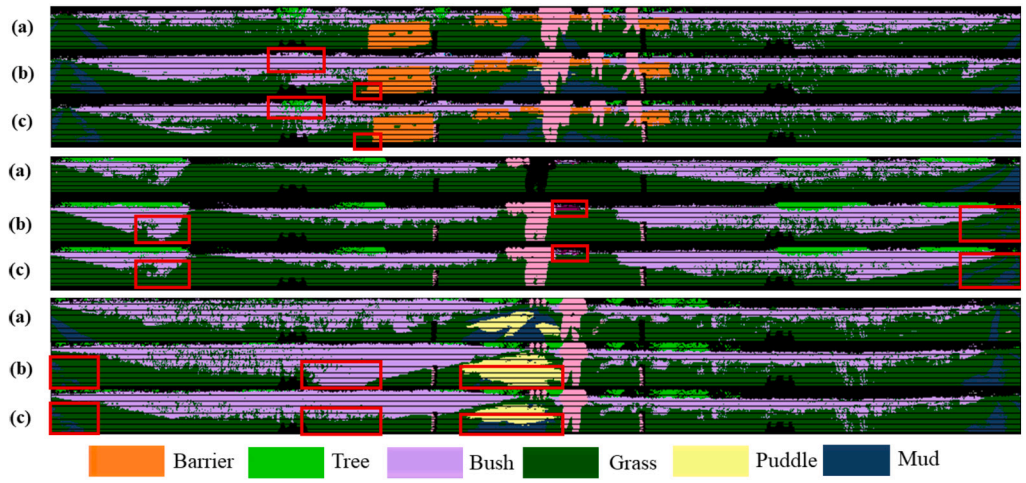

Fig. 5. Comparison of segmentation performance on the Rellis-3D dataset. (a) is the Ground truth, (b) is the SCF-Net segmentation result, and (c) is the LESA-Net segmentation result.

Table 2
Semantic segmentation results on the SemanticKITTI dataset.

	MIOU (%)	Car	Bicycle	Motorcycle	Truck	Other-vehicle	Person	Bicyclist	Motorcyclist	Road	Parking	Sidewalk	Other-ground	Buliding	Fence	Vegetation	Trunk	Terrain	Pole	Traffic-sign
RangeNet++	**52.2**	**91.4**	**25.7**	**34.4**	**25.7**	**23.0**	**38.3**	**38.8**	**4.8**	**91.8**	**65.0**	**75.2**	**27.8**	**87.4**	**58.6**	**80.5**	**55.1**	**64.6**	**47.9**	**55.9**
RandLA-Net	**53.9**	**94.2**	**26.0**	**25.8**	**40.1**	**38.9**	**49.2**	**48.2**	**7.2**	**90.9**	**60.3**	**73.7**	**20.4**	**86.9**	**56.3**	**81.4**	**61.3**	**66.8**	**49.2**	**47.7**
SqueezeSeg V2	**55.9**	**92.5**	**38.7**	**36.5**	**29.6**	**33.0**	**45.6**	**46.2**	**20.1**	**91.7**	**63.4**	**74.8**	**26.4**	**89.0**	**59.4**	**82.0**	**58.7**	**65.4**	**49.6**	**58.9**
SCF-Net	**55.8**	**92.7**	**25.5**	**37.1**	**58.4**	**31.7**	**59.4**	**71.9**	**0.0**	**91.3**	**41.8**	**76.8**	**23.7**	**88.0**	**45.1**	**84.2**	**61.8**	**74.7**	**53.0**	**42.9**
FA-ResNet	**54.9**	**93.7**	**30.6**	**33.2**	**33.9**	**27.2**	**51.6**	**45.3**	**18.7**	**91.3**	**65.2**	**75.5**	**26.0**	**90.6**	**62.7**	**83.1**	**63.5**	**66.3**	**48.4**	**53.3**
GAF-Net	**56.1**	**94.7**	**33.5**	**33.6**	**34.2**	**37.9**	**48.8**	**50.5**	**6.0**	**91.0**	**61.9**	**74.6**	**24.2**	**89.5**	**61.3**	**84.2**	**65.3**	**68.4**	**52.2**	**53.3**
Ours	**58.0**	**93.6**	**17.5**	**48.1**	**70.4**	**38.5**	**58.1**	**66.0**	**0.0**	**92.6**	**41.6**	**80.0**	**24.2**	**89.4**	**51.8**	**86.8**	**64.9**	**76.1**	**57.6**	**44.2**

As indicated in Table 2, the proposed algorithm performed best in the dimensions for MIOU, Motorcycle, Truck, Road, Sidewalk, Vegetation, Terrain, and Pole. Compared with the GAF-Net algorithm, the proposed algorithm improved by 1.9%, 14.5%, 36.2%, 1.6%, 5.4%, 2.6%, 7.7%, and 5.4%, respectively. The segmentation accuracy of trees, grass, and roads in agroforestry environments was significantly improved. These results indicate that the proposed algorithm exhibited strong semantic segmentation capabilities for different types of road surfaces in complex agroforestry environments.

Fig. 6 presents a comparison plot of the segmentation effect of the proposed algorithm on the SemanticKITTI dataset. In the red box in Fig. 6, the SCF-Net algorithm failed to accurately segment the tree dimension, whereas the proposed algorithm achieved successful tree segmentation. In addition, the SCF-Net algorithm incorrectly classified the train dimension in the red box as vegetation, whereas the proposed algorithm demonstrated a significant improvement in this aspect. These results verify that the proposed algorithm is more suitable than the other algorithms for point-cloud semantic segmentation of different road surface types in agroforestry environments.

4. Ablation experiment

All experiments in this section were conducted using point-cloud data from sequences 00000, 00001, 00002, and 00003 of the RELLIS-3D dataset for training; testing was performed on sequence 00004.

4.1. Ablation experiment of LESA

To investigate the effects of the two proposed modules, we conducted the following ablation studies.

The ablation study results of LESA-Net demonstrate the effectiveness of our proposed LFE and DDAPS modules. As shown in Table 3, the inclusion of the LFE module resulted in improvements of 2.1%, 9.4%, 1.3%, 2.6%, and 2% in MIOU, Grass, Concrete, Puddle, and Mud, respectively. Similarly, the addition of the DDAPS module led to enhancements of 2.9%, 7.1%, 1.9%, 5.4%, and 1.5% in MIOU, Grass, Concrete, Puddle, and Mud, respectively. These results confirm that incorporating LFE and DDAPS significantly enhances the segmentation accuracy of roads of various types in complex agroforestry environments.

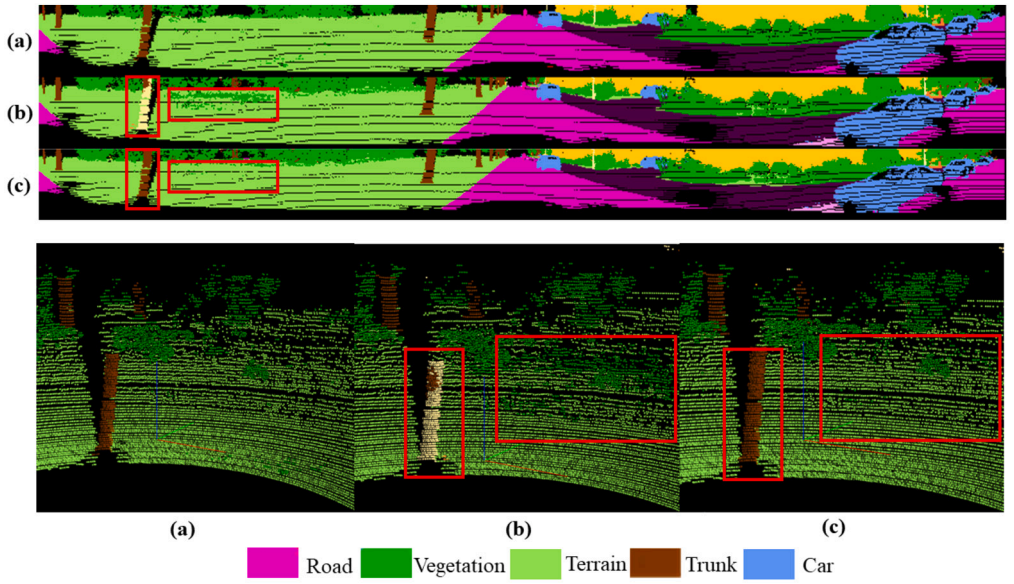

Fig. 6. Comparison of segmentation performance on the Semantickitti dataset. (a) is the Ground truth, (b) is the SCF-Net segmentation result, and (c) is the LESA-Net segmentation result.

Table 3
Effects of LESA.

Method	MIOU (%)	Grass (%)	Concreate (%)	Puddle (%)	Mud (%)
LESA-Net	**46.5**	**58.3**	**26.3**	**13.9**	**13.5**
Removing LFE	**45.1**	**51.2**	**25.4**	**11.9**	**12.6**
Removing DDAPS	**44.3**	**53.5**	**24.8**	**9.1**	**13.1**
Removing all (RandLA-Net)	**42.2**	**44.1**	**23.5**	**6.5**	**11.1**

Table 4
Effect of the local position feature-encoding module.

Method	MIOU (%)
$\|(P_i^k - P_i)\| \oplus (P_i^k - P_i) \oplus P_i^k \oplus P_i$	**42.2**
$\|(P_i^k - P_i)\| \oplus \hat{P}_j \oplus P_i$	**42.9**
$(P_i^k - P_i) \oplus \hat{P}_j \oplus P_i$	**43.4**
$\|(P_i^k - P_i)\| \oplus (P_i^k - P_i) \oplus P_i \oplus \hat{P}_j$	**43.8**

4.2. Ablation experiment of LFE

The ablation experiment of the proposed LFE module consisted of tow parts: an ablation experiment of the local position feature-encoding module and an ablation experiment of the point-cloud reflection intensity.

4.2.1. Ablation experiment of the local position feature-encoding module

To verify the effectiveness of the proposed local position feature- enhancement module, we conducted four sets of ablation experiments.

(1) Position encoding comprises the location coordinate P_i, neighboring point location coordinate P_i^k, relative position $(P_i^k - P_i)$ and Euclidean distance $\|P_i^k - P_i\|$.

(2) Position encoding comprises the relative position $(P_i^k - P_i)$, the location coordinate P_i, and semantically enhanced location-encoded information $\hat{P}_j$.

(3) Position encoding comprises the relative position $(P_i^k - P_i)$, the location coordinate P_i, and semantically enhanced location-encoded information $\hat{P}_j$.

(4) Position encoding comprises the location coordinate P_i, semantically enhanced location-encoded information $\hat{P}_j$, neighboring point location coordinate P_i^k, relative position $P_i^k - P_i$ and Euclidean distance $\|P_i^k - P_i\|$.

As shown in Table 4, our designed spatial information encoding module performs the best in terms of MIOU, with a value of 43.8%. This demonstrates that our designed local positional encoding module is a key factor influencing semantic segmentation accuracy.

Table 5
Effects of point cloud reflection intensity.

Method	MIOU (%)
Removing intensity	**43.8**
Adding intensity	**44.3**

Table 6
Effects of DDAPS.

Method	MIOU (%)
DDAPS	**45.1**
Dual-distance attention pooling block	**43.8**
Attention pooling block	**42.2**

4.2.2. Ablation experiment of point cloud reflection intensity

The reflection intensity of point clouds exhibits significant variations among different road materials in agroforestry environments. To validate whether the added point-cloud intensity information can effectively characterize the road type, we conducted an ablation experiment on the LFE module. The performances of adding and not adding point-cloud intensity were compared to evaluate its influence on MIOU. (See Table 5.)

After introducing the point-cloud reflection intensity feature, the algorithm exhibited a 0.6% improvement in MIOU. This result indicates that the reflection intensity information can better characterize different types of road surfaces in agroforestry environments.

4.3. Ablation experiment of DDAPS

Distance is an important variable in measuring the correlation between points; the smaller the distance, the stronger the correlation. Meanwhile, attention pooling blocks with self-attention mechanisms can achieve more accurate feature learning and attention allocation. Therefore, we aim to validate the positive significance of using the proposed DDAPS for point cloud semantic segmentation. We use MIOU as the evaluation metric and conduct three sets of ablation experiments on the Dual-Distance Attention Pooling (DDAPS) block based on the self-attention mechanism. The ablation experiment design is as follows:

(1) The DDAPS module.
(2) The Dual-distance attention pooling block.
(3) The attention pooling block.

From the data presented in Table 6, we can observe that compared to the algorithms using the Dual-Distance Attention Pooling block and the Attention Pooling block, the algorithm using DDAPS improves the MIOU by 2.3% and 2.9%, respectively. The above experimental results confirm that the introduction of DDAPS further improves the accuracy of road-surface-type segmentation of the proposed algorithm in large-scale environments.

5. Conclusion

Herein, we propose the LESA-Net algorithm for semantic segmentation of various types of roads in agroforestry environments. Initially, we introduce an LFE module that effectively expands the receptive field and enhances the generalizability of multidimensional features by incorporating reflection intensity information and spatial feature-encoding blocks enhanced with contextual semantic information. Subsequently, we introduce the DDAPS mechanism to achieve more accurate attention allocation and key feature learning, thereby improving segmentation accuracy. Finally, after multiple ablation experiments and experimental verification on large public datasets, the experimental results confirmed the superiority of the proposed algorithm. Unfortunately, there are only limited open-source point-cloud datasets with composite annotations for water, grassland, mud, concrete, and other road surfaces. In the future, we will strive to establish a rich point-cloud dataset with multiple road-surface annotations in an agroforestry environment. The semantic segmentation performance of LESA-Net will be further optimized to improve the efficiency and accuracy of the segmentation algorithm. We believe that the proposed algorithm can advance the automated processes for agriculture and forestry conservation. Furthermore, it can serve as a valuable reference for researchers in related fields.

CRediT authorship contribution statement

Yijian Duan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Danfeng Wu:** Formal analysis, Conceptualization. **Liwen Meng:** Formal analysis, Conceptualization. **Yanmei Meng:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Jihong Zhu:** Writing – review & editing, Supervision, Resources, Conceptualization. **Jinlai Zhang:** Formal analysis, Conceptualization. **Eksan Firkat:** Formal analysis, Conceptualization. **Hui Liu:** Methodology. **Hejun Wei:** Formal analysis.

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

The data that support the findings of this study are openly available in https://semantic-kitti.org/dataset.html and https://www.nuscenes.org/nuscenes.

Acknowledgements

This work was supported by the National Natural Science Foundation Project Approval No. 52365001 and Guangxi innovation-driven development special fund project Medium and heavy commercial vehicle power equipment supporting development and industrialization No. Guike AA23062040 and National college Student Innovation and Entrepreneurship Training Program project No. 202310593082.

References

[1] L. Droukas, Z. Doulgeri, N.L. Tsakiridis, et al., A survey of robotic harvesting systems and enabling technologies, J. Intell. Robot. Syst. 21 (2023) 107.
[2] S. Zhe, H. Shujie, X. Hao, L. Hongyu, Z. Jinchuan, C. Bo, Fuzzy adaptive recursive terminal sliding mode control for an agricultural omnidirectional mobile robot, Comput. Electr. Eng. 105 (2023) 108529.
[3] Y. Zhixin, Z. Chunjiang, Z. Taihong, Agricultural machinery automatic navigation technology, iScience 27 (2024) 108714.
[4] J. Chanyoung, K. Jeongeun, S. Jaehwi, S. Hyoung, A review on multirobot systems in agriculture, Comput. Electron. Agric. 202 (2022) 107336.
[5] Jan Weyler, Thomas Laebe, Federico Magistri, Jens Behley, Cyrill Stachniss, Towards domain generalization in crop and weed segmentation for precision farming robots, IEEE Robot. Autom. Lett. 8 (2023) 3310–3317.
[6] Y. Chen, G. Li, X. Zhang, J. Jia, K. Zhou, C. Wu, Identifying field and road modes of agricultural machinery based on GNSS recordings: a graph convolutional neural network approach, Comput. Electron. Agric. 198 (2022) 107082.
[7] M. Aki, T. Rojanaarpa, K. Nakano, Y. Suda, N. Takasuka, T. Isogai, T. Kawai, Road surface recognition using laser radar for automatic platooning, IEEE Trans. Intell. Transp. Syst. 17 (2016) 2800–2810.
[8] L. Yunzhe, C. Meixu, W. Meihui, H. Jing, T. Fisher, R. Kazem, M. Mohammad, An interpretable machine learning framework for measuring urban perceptions from panoramic street view images, iScience 26 (2023) 106132.
[9] J. Wang, Z. Zhang, L. Luo, H. Wei, W. Wang, M. Chen, S. Luo, DualSeg: fusing transformer and CNN structure for image segmentation in complex vineyard environment, Comput. Electron. Agric. 206 (2023) 107682.
[10] G. Tianrui, K. Divya, C. Rohan, S. Adarsh, W. Kasun, M. Dinesh, GANav: efficient terrain segmentation for robot navigation in unstructured outdoor environments, arXiv:2103.04233, 2022.
[11] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view convolutional neural networks for 3D shape recognition, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
[12] A. Boulch, Y. Guerry, B. Le, N. Audebert, SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks, Comput. Graph. 71 (2018) 189–198.
[13] I. Forrest, H. Song, M. Matthew, A. Khalid, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size, arXiv:1602.07360, 2016.
[14] W. Bichen, W. Alvin, Y. Xiangyu, K. Kurt, SqueezeSeg: convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud, arXiv:1710.07368, 2017.
[15] W. Bichen, Z. Xuanyu, Z. Sicheng, Y. Xiangyu, K. Kurt, SqueezeSegV2: improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud, arXiv:1809.08495, 2018.
[16] Milioto An, I. Vizzo, J. Chley, C. Stachniss, RangeNet plus plus: fast and accurate LiDAR semantic segmentation, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 4213–4220.
[17] M. Hsien, G. Lin, L. YuKun, M. Dinesh, VV-Net: voxel VAE net with group convolutions for point cloud segmentation, arXiv:1811.04337, 2019.
[18] M. Liu, Q. Zhou, H. Zhao, J. Li, Y. Du, K. Keutzer, L. Du, S. Zhang, Prototype-voxel contrastive learning for LiDAR point cloud panoptic segmentation, in: 2022 International Conference on Robotics and Automation (ICRA), 2022, pp. 9243–9250.
[19] X. Zhu, H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, H. Li, D. Lin, Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021, 2021, pp. 9934–9943.
[20] L. Loic, S. Martin, Large-scale point cloud semantic segmentation with superpoint graphs, arXiv:1711.09869, 2018.
[21] Y. Song, C. Yang, Y. Shen, P. Wang, Q. Huang, C.J. Kuo, SPG-net: segmentation prediction and guidance network for image inpainting, in: British Machine Vision Conference, 2018.
[22] Y. Wang, Y. Sun, Z. Liu, E. Sarma, M.M. Bronstein, J. Solomon, Dynamic graph CNN for learning on point clouds, ACM Trans. Graph. 38 (2019) 146.
[23] K. Zhang, Hao Mi, J. Wang, X. Chen, Y. Leng, W. de, C. Fu, Linked dynamic graph CNN: learning through point cloud by linking hierarchical features, in: 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2021.
[24] J. Li, H. Li, G. Cul, Y. Kang, Y. Hu, Y. Zhou, GACNet: a generative adversarial capsule network for regional epitaxial traffic flow prediction, Comput. Mater. Continua 64 (2020) 925–940.
[25] C.R. Qi, H. Su, K. Mo, L.J. Guibas, PointNet: deep learning on point sets for 3D classification and segmentation, in: 30TH IEEE Conference on Computer Vision and Pattern Recognition (CVPR2017), 2017, pp. 77–85.
[26] C. Wang, B. Samari, K. Siddiqi, Local spectral graph convolution for point set feature learning, in: Computer Vision - ECCV 2018, PT IV, vol. 11208, 2018, pp. 56–71.
[27] Y. Shen, Feng Ch, Y. Yang, D. Tian, Mining point cloud local structures by kernel correlation and graph pooling, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4548–4557.
[28] Charles R. Qi, Y. Li, S. Hao, Leonidas J. G, PointNet++: deep hierarchical feature learning on point sets in a metric space, arXiv:1706.02413, 2017.
[29] H. Qingyong, Y. Bo, X. Linhai, R. Stefano, G. Yulan, W. Zhihua, T. Niki, M. Andrew, RandLA-Net: efficient semantic segmentation of large-scale point clouds, arXiv:1911.11236, 2020.
[30] S. Yupeng, H. Fazhi, F. Linkun, D. Jicheng, G. Qing, DSACNN: dynamically local self-attention CNN for 3D point cloud analysis, Adv. Eng. Inform. 54 (2022) 101803.

[31] S. Yupeng, H. Fazhi, D. Yansong, L. Yaqian, Y. Xiaohu, A kernel correlation-based approach to adaptively acquire local features for learning 3D point clouds, Comput. Aided Des. 146 (2022) 103196.
[32] Y. Song, F. He, Y. Duan, T. Si, J. Bai, LSLPCT: an enhanced local semantic learning transformer for 3-D point cloud analysis, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13.
[33] Y. Su, W. Liu, Z. Yuan, M. Cheng, Z. Zhang, X. Shen, C. Wang, DLA-Net: learning dual local attention features for semantic segmentation of large-scale building facade point clouds, Pattern Recognit. 123 (2022) 108372.
[34] S. Fan, Q. Dong, F. Zhu, Y. Lv, P. Ye, F. Wang, SCF-Net: learning spatial contextual features for large-scale point cloud segmentation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021, 2021, pp. 14499–14508.
[35] H. Wei, E. Xu, J. Zhang, Y. Meng, J. Wei, Z. Dong, Z. Li, BushNet: effective semantic segmentation of bush in large-scale point clouds, Comput. Electron. Agric. 193 (2022) 106653.
[36] M. Carsten, Neil A. D, Fast marching farthest point sampling, Eurographics (2023).
[37] G. Fabian, W. Patrick, Hendrik P.A. L, Flex-convolution (million-scale point-cloud learning beyond grid-worlds), arXiv:1803.07289, 2020.
[38] J. Peng, O. Philip, W. Maggie, S. Srikanth, RELLIS-3D dataset: data, benchmarks and analysis, arXiv:2011.12954, 2022.
[39] H. Thomas, C. Qi, J. Deschaud, B. Marcotegui, F. Goulette, L.J. Guibas, KPConv: flexible and deformable convolution for point clouds, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6410–6419.
[40] B. Jens, G. Martin, M. Andres, Q. Jan, B. Sven, S. Cyrill, G. Juergen, SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences, arXiv: 1904.01416, 2019.
[41] L. Zhan, W. Li, W. Min, FA-ResNet: feature affine residual network for large-scale point cloud segmentation, Int. J. Appl. Earth Obs. Geoinf. 118 (2023) 103259.
[42] C. Zhou, Q. Ling, GAF-Net: geometric contextual feature aggregation and adaptive fusion for large-scale point cloud semantic segmentation, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–15.