

# Multiple Random Observation Strategy for Enhanced ALS Point Cloud Segmentation

Hengming Dai , Zhifang Zhao , Huiwei Jiang , Jiabo Xu , Haihan Duan , *Member, IEEE*, and Xiangyun Hu 

**Abstract**—Random sampling (RS) is widely used in data-driven large-scale point cloud processing due to its high computational efficiency. However, it suffers from two key limitations: first, the randomness introduced during forward propagation can lead to unstable feature extraction, potentially compromising model performance; second, RS does not consider the spatial structure of point clouds, which may result in the loss of critical information. These issues are particularly prominent in airborne laser scanning (ALS) point clouds, which typically exhibit severe class imbalance and substantial variations in object scales. To address these challenges, we propose a multiple random observation (MRO) framework that leverages the efficiency of RS while capturing spatially complementary features. Building upon this, we introduce the MRO-based feature aggregation module, which integrates features from multiple observations to improve feature extraction stability and enhance segmentation accuracy. Furthermore, we propose the MRO-based downsampling strategy, which identifies informative points by evaluating interobservation feature differences during downsampling, thereby boosting overall model performance. The proposed methods are integrated into several RS-based backbones and evaluated on two representative ALS datasets (i.e., ISPRS and LASDU), demonstrating strong competitiveness compared with current leading approaches.

**Index Terms**—Airborne laser scanning (ALS), point cloud, random sampling (RS), semantic segmentation.

## I. INTRODUCTION

WITH the rapid advancement of remote sensing data acquisition and processing technologies, various types

Received 23 May 2025; revised 6 August 2025; accepted 4 September 2025. Date of publication 8 September 2025; date of current version 26 September 2025. This work was supported in part by the Fundamental Research Funds for the Central Universities, China under Grant 2042022dx0001, in part by the National Natural Science Foundation of China under Grant 42161067, and in part by the Young Talent Support Program of the China Association for Science and Technology (CAST). (Corresponding author: Zhifang Zhao.)

Hengming Dai and Zhifang Zhao are with the Institute of International Rivers and Eco-Security, Yunnan University, Kunming 650500, China, also with the School of Earth Sciences, Yunnan University, Kunming 650500, China, also with the Yunnan International Joint Laboratory of China-Laos-Bangladesh-Myanmar Natural Resources Remote Sensing Monitoring, Kunming 650500, China, also with the Research Center of Domestic High-resolution Satellite Remote Sensing Geological Engineering, Universities in Yunnan Province, Kunming 650500, China, and also with the Yunnan Key Laboratory of Sanjiang Metallogeny and Resources Exploration and Utilization, Kunming 650051, China (e-mail: daihm@ynu.edu.cn; zhaozhifang@ynu.edu.cn).

Huiwei Jiang is with the National Geomatics Center of China, Beijing 100830, China (e-mail: jianghw@ngcc.cn).

Jiabo Xu and Xiangyun Hu are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: xujiabo@whu.edu.cn; huxy@whu.edu.cn).

Haihan Duan is with the Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, China (e-mail: duanhaihan@smbu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2025.3607404

of remotely sensed data have demonstrated significant potential across diverse Earth observation applications [1]. Among these, airborne laser scanning (ALS) enables large-scale, high-precision 3-D information acquisition and has been widely applied in urban monitoring [2], biomass estimation [3], and digital elevation model generation [4]. A fundamental preprocessing step for these downstream tasks involves determining the semantic category of each point in the raw ALS point cloud. Traditional ALS data segmentation approaches mainly rely on handcrafted features [5], [6] or classical machine learning classifiers [7], [8]. With the rise of deep learning, data-driven methods have significantly improved classification performance [9], [10]. These methods typically adopt one or more of the following representations: projection-based [11], [12], [13], voxel-based [14], [15], or point-based [16], [17], [18], [19], [20]. While projection and voxel representations convert unstructured point clouds into regular grids to leverage mature convolutional architectures, this conversion inevitably results in the loss of fine-grained geometric details. In contrast, point-based representations operate directly on raw point clouds, preserving spatial fidelity and gaining increasing attention.

Regardless of the chosen representation, most deep learning-based methods rely on downsampling for hierarchical feature extraction. Downsampling is straightforward for grid-based representations, but rather more complicated for unstructured point clouds due to the lack of inherent spatial regularity. Existing point cloud sampling strategies can be broadly categorized into heuristic-based and learning-based approaches [21]. Among them, random sampling (RS) is widely adopted due to its high computational efficiency and scalability to large-scale scenarios [21], [22]. However, RS does not account for the spatial distribution or geometric structure of point clouds, often resulting in oversampling of densely populated regions and undersampling of sparse areas [23]. This issue is particularly critical in ALS data, where significant variation in object scales leads to small-scale features being represented by very few points. In such cases, RS may disproportionately sample these limited points, introducing bias and diminishing the model's ability to effectively represent fine-scale structures. More critically, RS introduces randomness into the feature aggregation process, causing instability in the extracted features, particularly in regions with complex geometry or sharp topological changes, potentially affecting the stability of feature extraction. As visualized in Fig. 1(b), when performing multiple forward propagations on the same input point cloud during the testing phase using RS, the high-dimensional features extracted

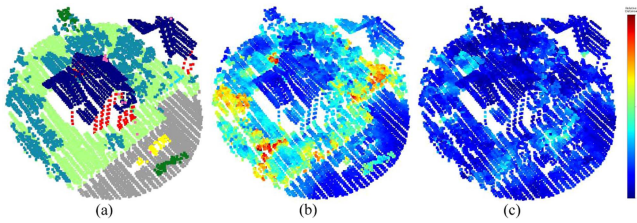


Fig. 1. Feature instability caused by RS. (a) Original point clouds, with colors indicating semantic categories. (b) Relative feature variations caused by RS over ten forward passes. (c) Relative feature variations obtained by our method over ten forward passes. The feature difference of each point was calculated as the L2 distance between its feature and the mean feature across multiple forward passes, which was normalized for visualization.

for classification show noticeable variation across different forward passes, indicating sensitivity to the sampling locations. Such instability may reduce convergence stability during training and negatively impact generalization in cross-scene deployments, thus degrading semantic segmentation accuracy.

To address these challenges, we draw inspiration from ensemble learning and multiview data augmentation strategies and explore a multiple random observation (MRO) framework that builds upon the efficiency of RS while enhancing its spatial awareness. Instead of relying on a single RS pass in each feature extraction stage, MRO features are derived from multiple independently sampled subsets with diverse spatial distributions, thereby capturing complementary spatial information. Based on this, we design an MRO-based feature aggregation (MROA) module that aggregates features extracted from multiple RS using a self-attention mechanism thereby enhancing the stability of RS-based feature extraction and enhancing feature consistency across passes. In addition, we introduce an MRO-based downsampling (MROS) strategy, which performs targeted sampling by evaluating feature differences across MRO, aiming to improve the stability of downsampling process by retaining more informative points. As shown in Fig. 1(c), the proposed MRO framework significantly reduces feature inconsistency across multiple forward passes. We also demonstrate through extensive experiments that the proposed MROA module and MROS strategy effectively improve the semantic segmentation performance of several RS-based baselines on ALS data.

In summary, the main contributions of this article are as follows.

- 1) We explore a MRO framework to capture spatially complementary features and enhance semantic segmentation performance on ALS data. This approach remains underexplored in current literature.
- 2) We design an MROA module that effectively integrates features from diverse spatial distributions, enhancing the accuracy of semantic segmentation.
- 3) We develop an MROS strategy that identifies informative points through interobservation feature difference analysis, introducing a novel paradigm for ALS point cloud downsampling.
- 4) Extensive experiments with various RS-based baselines on representative ALS datasets confirm that our method

delivers competitive, and often superior, performance compared to leading approaches.

## II. RELATED WORK

### A. ALS Point Cloud Semantic Segmentation

Projection-based methods regularize discrete point clouds using a 2-D grid, enabling the application of well-established 2-D convolutional neural networks (CNNs). Since ALS data are typically acquired from a top-down perspective, existing approaches often adopt the bird's-eye view for projecting point clouds into image representations. [11] was the first to apply CNNs to the ground filtering task for ALS data. Their method constructs a three-channel image by computing the elevation differences between each laser point and its neighboring points, with each point serving as the center of its neighborhood. The image is then processed by a CNN for ground point classification. Yang et al. [12] further enhanced this framework by introducing features derived from the covariance matrix of the local neighborhood and expanded the classification task to multiple semantic categories. These methods perform classification at the individual point level, resulting in significant computational redundancy and limited efficiency. Subsequent studies have improved the efficiency by refining feature map generation and adopting fully convolutional networks (FCNs) for semantic segmentation [13]. Despite notable progress, projection-based methods still face inherent limitations, such as susceptibility to occlusions [21] and loss of geometric information [16].

Voxel-based methods organize point clouds with a 3-D grid, thereby enabling the direct application of established convolutional network architectures. However, this approach introduces a substantial increase in computational complexity [21]. To address this challenge, existing methods commonly employ sparse convolution. Schmohl and Sörgel [14] utilized submanifold sparse convolution [24] for semantic segmentation on large-scale ALS data, while Dai et al. [15] extended the MinkoUnet [25] by incorporating sectional partitioning and iterative prediction to capture broader contextual information. Similar to projection-based methods, voxel-based approaches also suffer from quantization errors introduced during the discretization process [26].

Point-based methods directly operate on raw point clouds, thereby preserving rich geometric information [27]. To achieve improved classification performance on ALS point clouds, recent approaches have primarily focused on enhancing local feature extraction and contextual modeling capabilities. This has been achieved by incorporating auxiliary local geometric calculation [17], [19] or attention mechanisms [9], [18], [28], [29] to better capture fine-grained structures, and by employing multiscale [10] or multireceptive field [16] fusion strategies, including transformer-based architectures [20], to enrich semantic understanding. These methods have significantly advanced the development of ALS point cloud segmentation. However, hierarchical feature extraction in point-based methods inherently requires point cloud downsampling, which remains another critical yet underexplored aspect in current literature.

Recent studies have also leveraged multiple representation modalities to extract complementary deep features, thereby enhancing overall segmentation performance [30], [31]. Notably, these methods typically incorporate point-based representations, as it is crucial for capturing fine-grained geometric structures, which are especially important in high-precision ALS point cloud processing.

### B. Downsampling Strategies in Point-Based Networks

Downsampling plays a critical role in hierarchical point cloud processing, yet designing efficient and effective sampling strategies remains a major challenge due to the unstructured and irregular nature of point clouds. Unlike downsampling methods based on regular 2-D or 3-D grids, reducing point density in irregular and discrete point clouds is considerably more complex. Existing downsampling techniques can be broadly classified into heuristic-based and learning-based approaches [21]. Learning-based methods often adopt continuous relaxation techniques [32] or reinforcement learning strategies [33] to enable differentiable point selection. However, these methods are computationally expensive and typically rely on supervision or initialization from heuristic methods [34], which restricts their applicability to large-scale data.

In contrast, heuristic methods, such as farthest point sampling (FPS) [10], [16], grid sampling [9], [28], and RS [21], [35], are widely used in practice. Notably, RS provides the highest computational efficiency, making it attractive for large-scale applications. However, its inherent randomness may lead to spatially imbalanced sampling [23] and unstable feature extraction. This limitation serves as a key motivation for our further development of the MRO framework, which builds upon the efficiency of RS while addressing its shortcomings through feature fusion and sampling refinement.

## III. METHODOLOGY

### A. Multiple Random Observations

RS is computationally efficient and therefore widely adopted in large-scale point cloud processing. However, as illustrated in Fig. 1, the introduction of randomness during hierarchical feature extraction can lead to feature instability, which may in turn affect the optimization process of neural networks and compromise the stability of the inference phase. Based on this observation, and considering the high efficiency of RS, we propose a MRO strategy to enhance the overall stability through repeated RS. The difference between our method and existing approaches in the downsampling process is illustrated in Fig. 2. A shared-weight feature extraction module is applied to different downsampled point sets to perceive local geometries. Building on this, an MROA module is introduced to fuse features obtained from multiple observations, thereby enhancing the stability of feature extraction. Furthermore, we propose an MROS method, which measures the importance of sampled points based on feature differences across multiple observations, and utilizes this information to determine the downsampled points for the next

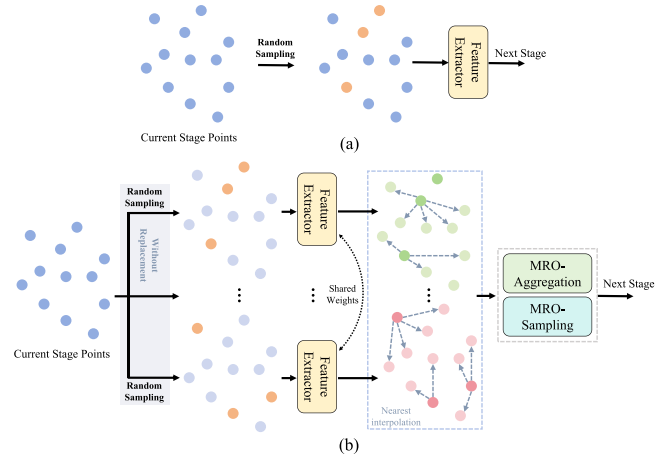


Fig. 2. Primary difference between the proposed method and typical RS-based approaches during feature extraction. (a) RS-based methods, a single RS is performed at each downsampling level. (b) Proposed MRO method, feature aggregation and downsampling are performed based on MRO.

---

#### Algorithm 1: Sampling Process of MRO-Framework.

---

INPUT: Point cloud set  $S$  with  $N$  points, downsampling ratio  $n$   
 OUTPUT: A sequence of  $n$  subsets  $\{S_1, S_2, \dots, S_n\}$ , each containing  $N/n$  points  
 Initialize  $S_{\text{remaining}} \leftarrow S$   
**for**  $i = 1$  **to**  $n$  **do**  
      $S_i \leftarrow \text{RS}(S_{\text{remaining}})$ ,  
      $S_{\text{remaining}} \leftarrow S_{\text{remaining}} \setminus S_i$   
**end for**  
**return**  $\{S_1, S_2, \dots, S_n\}$

---

stage of the network. During sampling iterations, a nonreplacement strategy is utilized to ensure complete coverage of the point cloud as illustrated in Algorithm 1.

As shown in Fig. 2(b), the point subsets obtained through multiple rounds of sampling without replacement are nonoverlapping and spatially complementary. By extracting features from these different subsets, spatially complementary high-dimensional features can be obtained. Subsequently, the features from different subsets are interpolated back to the original point cloud at the current downsampling stage, ensuring that each point receives multiple observations. This facilitates the fusion of spatially complementary features in the subsequent process.

### B. MRO-Based Feature Aggregation

For each point subset  $S_i$  from random observations, we employ the local feature aggregation (LFA) module [22] to extract features. Subsequently, the local features are upsampled and propagated back to the original point set  $S$  through the nearest neighbor interpolation as depicted Fig. 2(b). Feature interpolation addresses the spatial correspondence of multiple observations, ensuring that each point in the input point set  $S$  at the current downsampling stage obtains features from multiple observations, which facilitates subsequent feature fusion. As

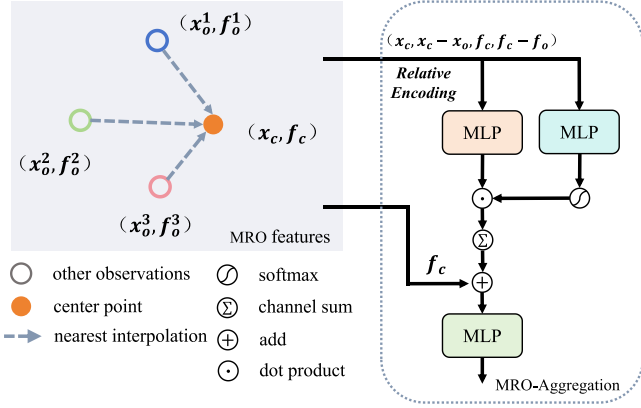


Fig. 3. Illustration of the MROA pipeline. Each random observation is represented by a single feature point for clarity.

illustrated in Fig. 3, after the interpolation process, each point is associated with multiple feature observations. Specifically, the feature extracted from its own subset  $S_i$  is referred to as the central feature  $\mathbf{f}_c$ , while the features interpolated from neighboring observations are denoted as  $\mathbf{f}_o$ . Fig. 3 presents an example corresponding to a four-times downsampling scenario. For these features, we first construct the relative encoding  $\mathcal{F}_r$  based on both the features and spatial coordinates of the central and auxiliary points

$$\mathcal{F}_r = (\mathbf{x}_c, \mathbf{x}_c - \mathbf{x}_o, \mathbf{f}_c, \mathbf{f}_c - \mathbf{f}_o). \quad (1)$$

Here,  $\mathbf{x}_c$  represents the coordinate matrix of the central points and  $\mathbf{x}_o$  denotes the coordinate matrix of the auxiliary observation points. To ensure proper alignment,  $\mathbf{x}_c$  is broadcasted to match the dimensions of  $\mathbf{x}_o$ . This relative encoding provides a unified representation that effectively captures the geometric and feature-level relationships across different observations, facilitating more robust feature aggregation in subsequent stages. The resulting relative encodings are then processed through a sequence of multilayer perceptrons (MLPs) designed to project, refine, and aggregate these features. As shown in Fig. 3, the relative encoding  $\mathcal{F}_r$  is first independently processed by two separate MLPs: one is used for feature projection, while the other is utilized to generate attention weights. The features from multiple observations are then aggregated through an attention-based summation, enabling the network to flexibly model complex interactions across multiple observations. On this basis, the fused features are added to the central feature and further refined through another MLP, in order to prevent the multiple observation features from excessively dominating the central feature representation. The overall computation process can be formulated as follows:

$$\mathbf{f}_{mroa} = \mathcal{M}_3 \left( \mathbf{f}_c + \sum_{c=1}^n \mathcal{M}_1(\mathcal{F}_r) \cdot \sigma(\mathcal{M}_2(\mathcal{F}_r)) \right) \quad (2)$$

where  $\mathcal{M}$  denotes the MLP,  $\sigma(\cdot)$  represents the softmax function, and  $n$  indicates the number of MRO features.

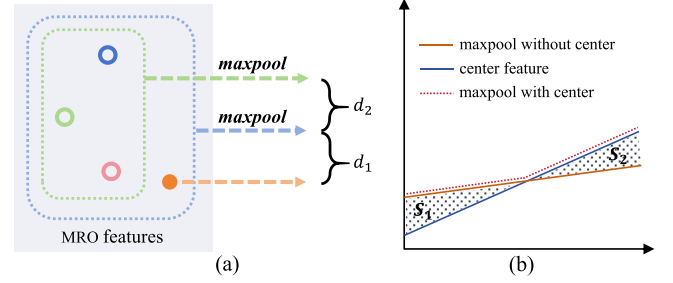


Fig. 4. Illustration of the main calculations in MROS strategy. (a) Feature difference calculation based on MRO features. (b) Geometric interpretations of the feature difference, the max-pooling result that includes the center point (red dashed line) is slightly offset for illustrative purpose.

The overall design of MROA module is intended to mitigate the feature instability caused by RS and enables the extraction of more robust representations.

### C. MRO-Based Downsampling

As RS neglects the spatial distribution and geometric structure of point clouds, it tends to oversample dense regions while ignoring small-scale objects with few points, which is an issue exacerbated in ALS data with large variations in object scale. To address this, we propose a downsampling strategy based on the MRO mechanism, where each point receives multiple observations from different subsets. The variation among these observations captures local geometric complexity and guides more adaptive sampling. This enables us to preserve points containing richer information during downsampling, resulting in a more representative subset. Specifically, as illustrated in Fig. 4(a), we perform LFA using a max pooling operation, where we first compute the aggregated feature  $\mathbf{f}_o^{\max}$  from the auxiliary observation features, excluding the central point

$$\mathbf{f}_o^{\max} = \text{maxpool}(\mathbf{f}_o^1, \dots, \mathbf{f}_o^n). \quad (3)$$

Meanwhile, the aggregated feature including the central feature is denoted as  $\mathbf{f}_c^{\max}$ , which is computed as follows:

$$\mathbf{f}_c^{\max} = \text{maxpool}(\mathbf{f}_o^1, \dots, \mathbf{f}_o^n, \mathbf{f}_c). \quad (4)$$

On this basis, we compute the distance  $d_1$  between  $\mathbf{f}_c^{\max}$  and  $\mathbf{f}_c$ , as well as the distance  $d_2$  between  $\mathbf{f}_c^{\max}$  and  $\mathbf{f}_o^{\max}$  calculated by  $L_1$ -distance

$$\begin{aligned} d_1 &= \|\mathbf{f}_c^{\max} - \mathbf{f}_c\|_1 \\ d_2 &= \|\mathbf{f}_c^{\max} - \mathbf{f}_o^{\max}\|_1. \end{aligned} \quad (5)$$

Here,  $d_1$  reflects the local importance of the central feature. When the central feature  $\mathbf{f}_c$  exhibits high saliency, the aggregated feature  $\mathbf{f}_c^{\max}$  that includes  $\mathbf{f}_c$  will be closer to  $\mathbf{f}_c$  itself, resulting in a smaller  $d_1$ . Conversely, a larger  $d_1$  indicates that  $\mathbf{f}_c^{\max}$  is predominantly influenced by the auxiliary observation features  $\mathbf{f}_o$ . On the other hand,  $d_2$  measures the contribution of  $\mathbf{f}_c$  to the max pooling operation. A larger  $d_2$  implies that incorporating  $\mathbf{f}_c$  into the aggregation introduces a more significant change in the resulting feature.

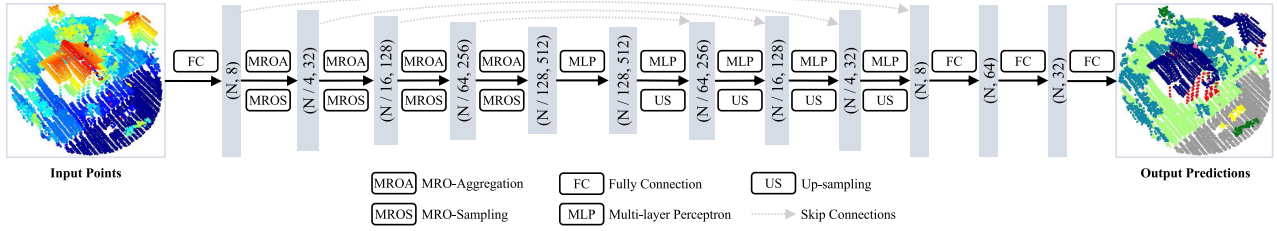


Fig. 5. Overall network architecture of the proposed method. Since our method is compatible with various RS-based frameworks, RandLANet is used here as a representative backbone to illustrate how it integrates with the proposed MRO strategy.

In fact,  $d_1$  and  $d_2$  have clear geometric interpretations, as illustrated in Fig. 4(b). Since the max pooling operation selects only the maximum value across channels,  $d_1$  measures the geometric area  $S_1$  between the part of the feature where the central feature does not participate in the pooling operation and the aggregated feature. The smaller this area, the closer the central feature is to the aggregated feature. On the other hand,  $d_2$  measures the area  $S_2$  between the part where the central feature participates in the pooling operation and the aggregated feature obtained from the auxiliary observations. A larger area indicates a greater contribution of the central feature to the pooling operation. During the sampling process, points with relatively large  $d_2$  and small  $d_1$  are considered to exhibit more complex geometric features. Therefore, we design a sampling score function  $\mathcal{T}(d_1, d_2)$  calculated as

$$\mathcal{T}(d_1, d_2) = \frac{d_2 - d_1}{d_2 + d_1}. \quad (6)$$

This formulation measures the relative magnitude between  $d_2$  and  $d_1$ , with its value range being  $[-1, 1]$ . A larger function value indicates that the corresponding point contains richer information. However, during the sampling process, it is important to not only focus on regions with complex geometric structures but also to preserve relatively flat areas (such as rooftops and ground surfaces) that contribute to the extraction of contextual information. To prevent the sampled points from being overly concentrated in extreme cases, we do not directly perform hard sampling based on the ranking of the  $\mathcal{T}$ -scores. Instead, the  $\mathcal{T}$ -scores are used as probabilistic weights for multinomial sampling.

#### D. Network Architecture

The proposed approach targets feature fusion under multiple RS conditions and the development of an improved sampling strategy, independent of any specific network backbone. To ensure a fair comparison, our network architecture is designed based on several representative RS-based point cloud semantic segmentation networks [22], [36], [37]. As shown in Fig. 5, the architecture follows a symmetric encoder-decoder paradigm. The encoder comprises four hierarchical stages of downsampling and feature extraction, where at each level we incorporate the proposed MROA module for robust feature fusion and apply the MROS strategy to guide downsampling point selection. During decoding, nearest neighbor interpolation is employed for upsampling, and MLPs are utilized for feature propagation.

Furthermore, skip connections are introduced between the corresponding encoder and decoder stages to facilitate effective feature reuse and gradient flow.

## IV. EXPERIMENTS

### A. Datasets

To validate the effectiveness of the proposed method, we conduct extensive experiments on the ISPRS Vaihingen dataset [6] and the LASDU dataset [43]. The ISPRS dataset was acquired using the Leica ALS-50 system at an average flying height of 500 m. This dataset is annotated with nine semantic categories, including power lines, low vegetation, impervious surfaces, cars, fences, roofs, facades, shrubs, and trees. The training set contains a total of 753 876 points, while the test set includes 411 722 points. We provide a visualization of the dataset and the proportion of each semantic category in the upper part of Fig. 6. The LASDU dataset was acquired using the Leica ALS-70 system at an average flying height of approximately 1200 m. It provides annotations for five semantic categories, namely, ground, buildings, trees, low vegetation, and artifacts. Among the four regions included in the dataset, we follow existing partitioning protocols by using Sections 2 and 3 as the training set, and Sections 1 and 4 as the test set [9], [10]. The training and testing data, along with the proportion of each semantic category, are visualized in the lower part of Fig. 6.

### B. Baselines

To validate the effectiveness of the proposed method, we first conduct a comparative analysis against RandLA-Net [21], [22], a pioneering work based on RS. Furthermore, we select two representative methods developed on top of RandLA-Net as baseline approaches. Specifically, BAF-LAC [36] introduces a backward attentive fusing strategy to enhance the robustness of feature extraction, along with a local aggregation classifier to improve the uniformity of predictions. BAAF [37], on the other hand, designs a bilateral feature enhancement strategy that integrates both geometric and semantic features. Our proposed method can be seamlessly incorporated into these network architectures.

### C. Evaluation Metrics

Following established practices in ALS point cloud semantic segmentation [16], [17], we adopt multiple evaluation metrics to comprehensively assess the performance of our method. Specifically, we report overall accuracy (OA), intersection over union

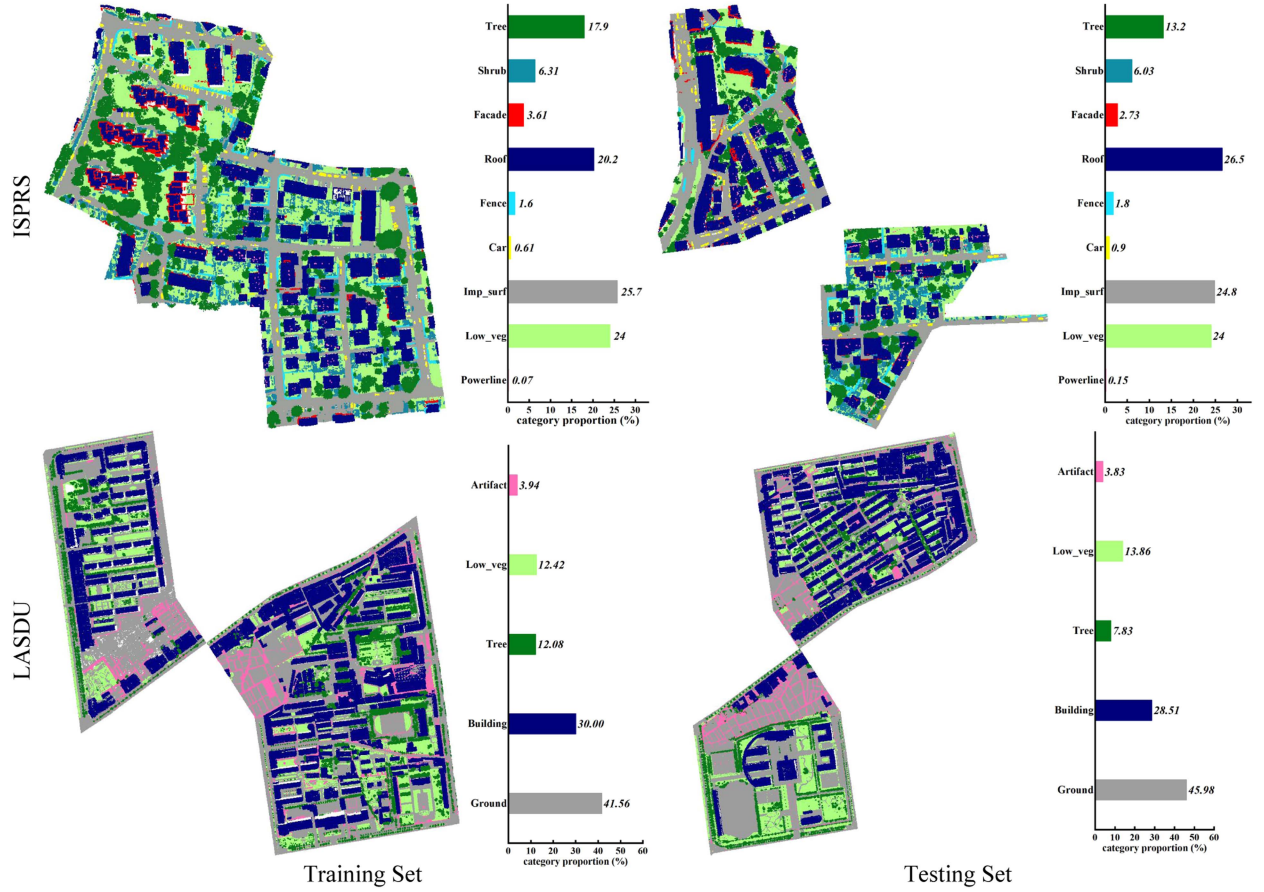


Fig. 6. Visualization of the datasets used in the experiments. Top row: ISPRS training and testing sets along with the corresponding semantic category distribution. Bottom row: LASDU training and testing sets with their respective semantic category statistics.

(IoU), and the  $F_1$ -score as the primary evaluation indicators. OA measures the proportion of correctly classified points relative to the total number of points, offering a general overview of classification effectiveness. Given the inherent class imbalance typically present in ALS datasets, IoU and  $F_1$ -score provide a more nuanced evaluation by focusing on per-class performance. These metrics are computed based on the number of true positives (TP), false positives (FP), and false negatives (FN), according to the following formulations:

$$\begin{aligned}
 \mathcal{P} &= \frac{TP}{TP + FP} \\
 \mathcal{R} &= \frac{TP}{TP + FN} \\
 F_1 &= 2 \times \frac{\mathcal{P} \times \mathcal{R}}{\mathcal{P} + \mathcal{R}} \\
 \text{IoU} &= \frac{TP}{TP + FP + FN}.
 \end{aligned} \tag{7}$$

#### D. Implementation Details

To mitigate the issue of uneven point cloud density caused by strip overlaps, we first applied grid-downsampling to the ISPRS dataset with a fixed grid resolution of 0.25 m. For the LASDU

dataset, we did not apply downsampling and instead used the raw-resolution point cloud directly, as the point distribution in LASDU is relatively uniform. Afterward, we adopted the spatial sampling strategy of RandLANet to generate mini-batches, with a fixed number of 8192 points per sample for both datasets. During the training phase, we employed data augmentations including random rotation around the  $z$ -axis and random scaling. To ensure a fair comparison, all models were optimized using only the weighted cross-entropy loss function. The network was trained using the Adam optimizer with an initial learning rate of 0.01, a batch size of 8, and 50 iterations per epoch, for a total of 100 epochs. An exponential learning rate scheduler with a gamma of 0.95 was used to gradually decay the learning rate after each epoch. All models were implemented using the PyTorch framework and trained on NVIDIA RTX 4090 GPUs.

#### E. Results on ISPRS Dataset

Based on the baseline methods mentioned in Section IV-B, we incorporate the proposed MROA module and MROS strategy to evaluate their effectiveness. The quantitative results on the ISPRS dataset are shown in Table I. It can be observed that the integration of the MROA module leads to performance improvements across all three baseline methods. Specifically,

TABLE I  
QUANTITATIVE COMPARISON WITH OTHER REPRESENTATIVE METHODS ON ISPRS DATASET

Methods	Powerline	Low_veg	Imp_surf	Car	Fence	Roof	Facade	Shrub	Tree	OA	$mF_1$	$mIoU$
D-FCN [38]	70.40	80.20	91.40	78.10	37.00	93.00	60.50	46.00	79.40	82.20	70.70	57.60
DANCE-NET [39]	68.40	81.60	92.80	77.20	38.60	93.90	60.20	47.20	81.40	83.90	71.20	58.30
RFFS-Net [16]	75.50	80.00	90.50	78.50	45.50	92.70	57.90	48.30	75.70	82.10	71.60	58.20
VD-Lab [10]	69.34	80.49	90.35	79.44	38.25	89.51	59.73	47.53	77.19	84.40	72.60	-
GACNN [40]	76.00	81.80	<b>93.00</b>	77.70	37.80	93.10	58.90	46.70	78.90	83.20	71.50	58.70
MCFN [18]	74.50	82.30	91.80	79.00	37.50	94.70	61.70	48.70	83.30	84.40	72.60	56.90
RRDAN [9]	72.20	81.70	91.20	<b>84.60</b>	<b>44.80</b>	94.70	<b>65.20</b>	<b>52.00</b>	<b>85.30</b>	<b>84.90</b>	<b>74.60</b>	<b>62.10</b>
KPConv [28]	73.50	78.70	88.00	79.40	33.00	94.20	61.30	45.70	82.00	81.70	70.60	57.70
DAKAG-Net [28]	71.70	<b>83.10</b>	91.70	81.20	43.10	<b>94.90</b>	64.10	50.00	83.90	85.20	73.70	61.20
Dense-LGEANet [29]	72.20	80.80	92.10	78.70	41.20	93.90	61.50	47.20	80.50	83.20	72.00	58.50
BAFNet [41]	<b>80.86</b>	81.34	91.24	76.91	40.66	93.37	61.50	47.68	81.90	83.69	72.83	60.10
MIA-Net [42]	65.80	79.50	89.70	71.10	26.20	94.00	63.50	48.10	82.80	83.30	69.00	56.01
BAF-LAC	69.40	81.09	<b>91.89</b>	78.56	35.62	94.53	59.69	43.88	81.68	83.70	70.71	58.00
BAF-LAC + MROA	71.81	<b>81.86</b>	91.70	76.55	33.68	94.71	61.20	46.47	82.89	84.23	71.21	58.59
BAF-LAC + MROA + MROS	<b>72.78</b>	81.64	91.73	<b>79.19</b>	<b>39.66</b>	<b>94.77</b>	<b>61.75</b>	<b>47.88</b>	<b>83.11</b>	<b>84.32</b>	<b>72.50</b>	<b>59.83</b>
BAAF	73.60	81.85	<b>92.39</b>	75.90	38.99	93.58	60.04	48.87	80.49	83.78	71.74	58.80
BAAF + MROA	71.53	<b>82.52</b>	91.51	78.32	<b>42.61</b>	93.19	<b>60.25</b>	<b>49.24</b>	80.34	83.68	72.17	59.10
BAAF + MROA + MROS	<b>74.61</b>	81.93	91.68	<b>78.83</b>	39.62	<b>94.23</b>	59.34	48.43	<b>82.93</b>	<b>84.26</b>	<b>72.40</b>	<b>59.71</b>
RandLANet	66.67	81.57	90.94	79.08	38.80	<b>94.79</b>	60.04	43.88	82.02	83.97	70.86	58.04
RandLANet + MROA	71.62	81.79	<b>92.03</b>	78.99	<b>40.13</b>	94.62	<b>61.66</b>	46.21	81.72	84.12	72.09	59.34
RandLANet + MROA + MROS	<b>76.12</b>	<b>81.98</b>	91.70	<b>80.47</b>	39.97	94.53	61.12	<b>47.14</b>	<b>82.45</b>	<b>84.35</b>	<b>72.83</b>	<b>60.28</b>

The  $F_1$ -score (%) is reported for each semantic class. (Best performances are marked in bold.)

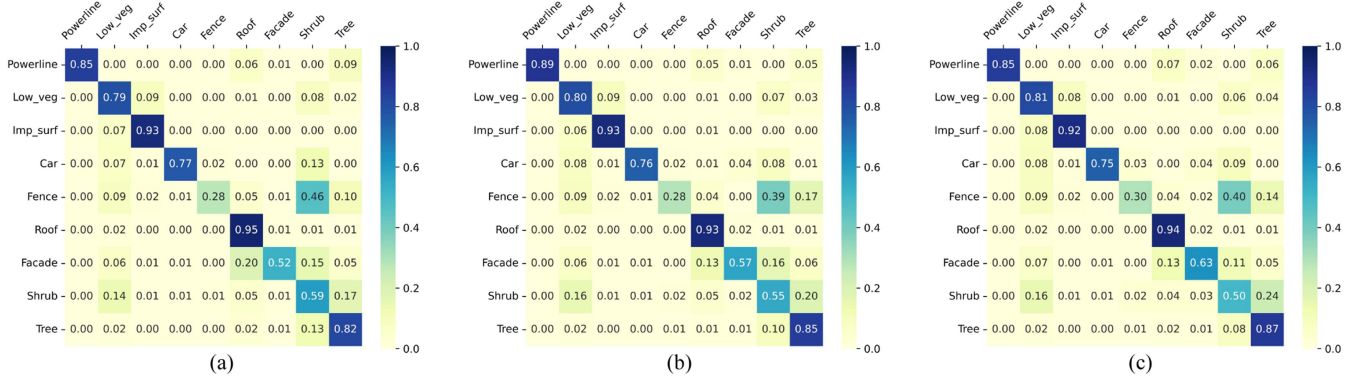


Fig. 7. Normalized confusion matrix of the MRO-based methods on the ISPRS dataset. (a) MRO-BAF-LAC. (b) MRO-BAAF. (c) MRO-RandLANet.

the  $mF_1$  scores of BAF-LAC, BAAF, and RandLANet increased by 0.5%, 0.43%, and 1.23%, respectively, while the  $mIoU$  scores improved by 0.59%, 0.3%, and 1.3%, respectively. These results demonstrate the effectiveness of the MROA module. Building upon this, the addition of the MROS strategy further enhances performance. The  $mF_1$  scores of BAF-LAC, BAAF, and RandLANet were further improved by 1.29%, 0.23%, and 0.74%, respectively, while the  $mIoU$  scores increased by 1.24%, 0.61%, and 0.94%, respectively. These results provide additional evidence supporting the effectiveness of the MROS strategy. Overall, the proposed MRO approach consistently enhances the performance of all three baseline methods. Although it does not achieve state-of-the-art segmentation results on the ISPRS dataset, the best obtained OA score (MRO-RandLANet) is only 0.55% lower than that of the current leading method, indicating strong competitiveness. The classification confusion

matrix of the proposed method on the ISPRS dataset is shown in Fig. 7. As illustrated, MRO-RandLANet achieves the best recognition performance for the categories of low vegetation, fence, facade, and tree. Moreover, it also achieves the highest overall  $mF_1$  and  $mIoU$  scores among the three baseline methods.

In Fig. 8, we further provide a visual comparison between the proposed method and the three baseline methods (top row: RandLANet, middle row: BAF-LAC, bottom row: BAAF). The segmentation results with the integrated MRO strategy are shown in the right column. It can be observed that our method effectively improves the baselines' ability to recognize categories, such as power lines, cars, low vegetation, and trees. Notably, power lines and cars constitute a very small proportion of the ISPRS dataset. The improved recognition of these underrepresented classes contributes significantly to the performance gains in terms of

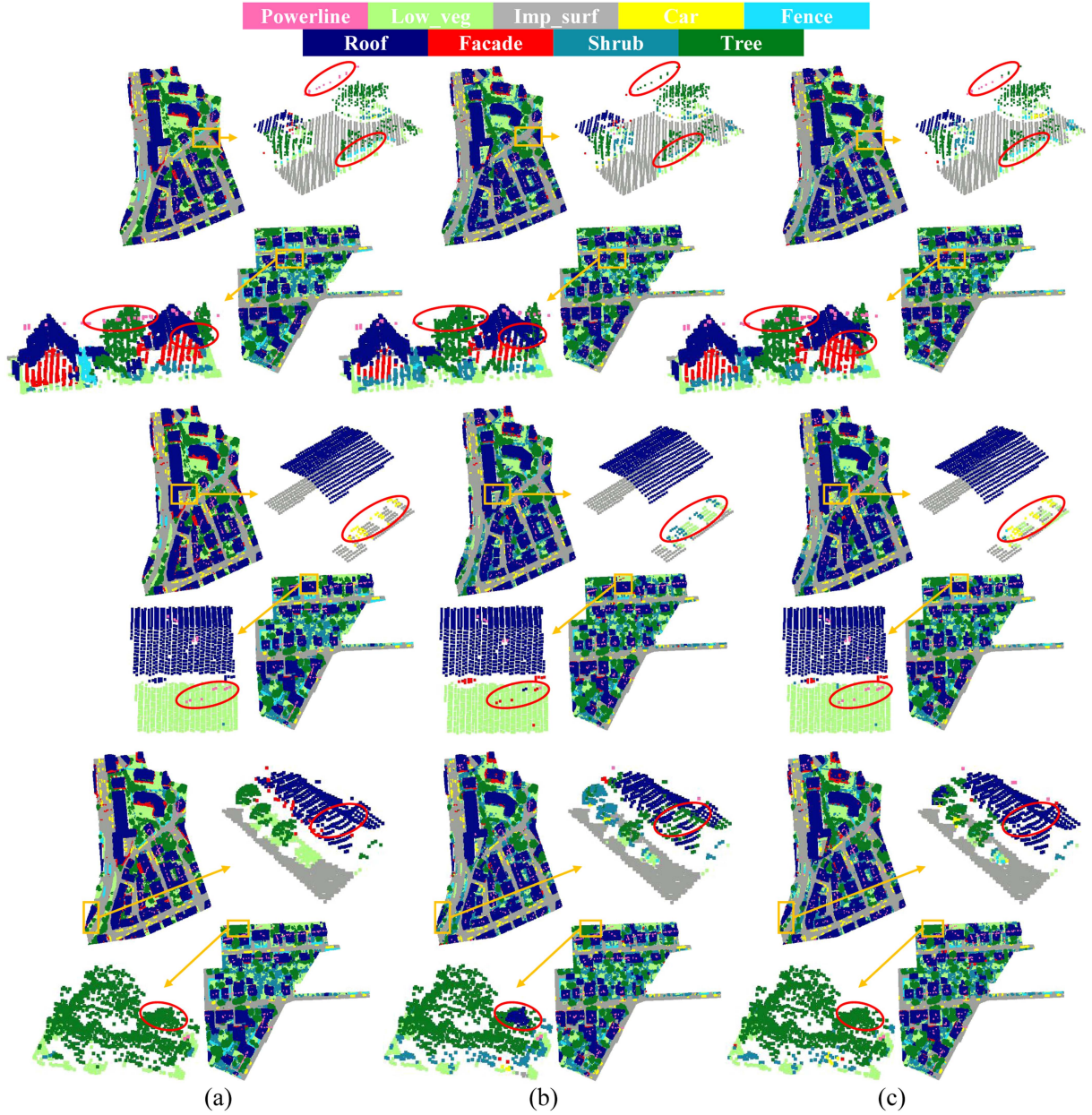


Fig. 8. Visual comparison of classification results on the ISPRS dataset between the proposed method and baseline methods. From top to bottom: BAF-LAC, BAAF and RandLANet. (a) Ground Truth. (b) Baselines. (c) Ours.

both  $mF_1$  and mIoU scores achieved by the proposed MRO method.

#### F. Results on LASDU Dataset

In the experiments conducted on the LASDU dataset, we also evaluated the three baseline methods described in Section IV-B. The quantitative results are presented in Table II. It can be observed that, after incorporating the proposed MRO strategy, all three baseline methods achieved leading segmentation performance, with  $mF_1$  scores exceeding 79%, surpassing the previous best score [41] of 78.95%. Specifically, the introduction of the MROA module led to improvements

in  $mF_1$  scores of 0.69%, 0.46%, and 0.51% for BAF-LAC, BAAF, and RandLANet, respectively. Similarly, mIoU scores increased by 0.64%, 0.44%, and 0.67%, respectively, indicating that the MROA module contributes consistently to these RS-based methods. Furthermore, with the addition of the MROS strategy, the  $mF_1$  scores of BAF-LAC, BAAF, and RandLANet were further improved by 0.78%, 0.87%, and 0.92%, respectively, while the mIoU scores increased by 0.87%, 1.03%, and 1.00%, respectively. These results provide further evidence of the effectiveness of the MROS strategy, which may be attributed to its ability to retain more informative points, thereby facilitating better feature extraction and aggregation after downsampling.

TABLE II  
QUANTITATIVE COMPARISON WITH OTHER REPRESENTATIVE METHODS ON LASDU DATASET

Methods	Ground	Buildings	Trees	Low_veg	Artifacts	OA	$mF_1$	mIoU
DensePoint [10]	89.78	94.77	85.20	65.45	34.17	86.31	73.87	63.00
DGCNN [28]	90.52	93.21	81.55	63.26	37.08	85.51	73.12	61.57
PosPool [28]	88.25	93.67	83.92	61.00	38.34	83.52	73.03	61.39
HDA-PointNet++ [43]	88.74	93.16	82.24	65.24	36.89	84.37	73.25	61.56
VD-Lab [10]	91.19	95.53	87.26	73.49	44.64	88.01	78.42	-
RFFS-Net [16]	90.92	95.35	86.81	71.01	44.36	87.12	77.69	66.94
RRDAN [9]	<b>91.60</b>	96.60	84.10	66.30	<b>48.30</b>	87.70	77.40	66.40
MCFN [18]	<b>91.60</b>	<b>96.70</b>	85.90	67.10	43.80	88.00	77.00	66.38
DAKAG-Net [28]	90.80	96.63	85.85	67.78	47.92	87.38	77.80	66.92
Dense-LGNet [29]	91.40	95.60	<b>88.00</b>	72.50	43.90	87.70	78.30	<b>67.20</b>
MIA-Net [42]	91.00	95.70	86.20	71.60	40.60	87.70	77.00	-
BAFNet [41]	91.36	95.69	86.71	<b>73.36</b>	47.64	<b>88.04</b>	<b>78.95</b>	-
BAF-LAC	<b>92.09</b>	95.92	85.31	73.07	41.60	88.03	77.60	67.15
BAF-LAC + MROA	91.60	95.86	86.47	73.28	44.24	87.82	78.29	67.79
BAF-LAC + MROA + MROS	91.92	<b>96.08</b>	<b>86.75</b>	<b>74.28</b>	<b>46.32</b>	<b>88.28</b>	<b>79.07</b>	<b>68.66</b>
BAAF	91.32	95.63	86.40	73.07	43.01	87.55	77.89	67.34
BAAF + MROA	91.36	96.02	86.26	72.98	45.14	87.75	78.35	67.78
BAAF + MROA + MROS	<b>91.64</b>	<b>96.18</b>	<b>87.54</b>	<b>73.45</b>	<b>47.27</b>	<b>88.22</b>	<b>79.22</b>	<b>68.81</b>
RandLANet	91.38	95.66	85.59	73.82	44.71	87.43	78.23	67.58
RandLANet + MROA	91.71	96.10	86.19	74.04	45.67	87.96	78.74	68.25
RandLANet + MROA + MROS	<b>91.87</b>	<b>96.21</b>	<b>86.91</b>	<b>74.67</b>	<b>48.64</b>	<b>88.44</b>	<b>79.66</b>	<b>69.25</b>

The  $F_1$  score (%) is reported for each semantic class. (Best performances are marked in bold.)

A visual comparison between the proposed method and the three baseline methods is presented in Fig. 9 (top row: RandLANet, middle row: BAF-LAC, bottom row: BAAF), with the segmentation results after applying the MRO strategy shown in the right column. It can be observed that our method facilitates better recognition of low vegetation and artifact. As illustrated in the confusion matrix in Fig. 10, these two classes are prone to confusion, and both share a high degree of similarity with ground points. Although the proposed method outperforms existing approaches in distinguishing low vegetation and artifact, further improving their separability remains an important direction for future research.

### G. Effectiveness of MROS Strategy

In this section, we further analyze the effectiveness of the proposed MROS strategy. Specifically, we employ RandLANet as the backbone network and conduct inference on the ISPRS dataset using both RS and MROS strategies. During the entire inference process, we record the semantic category distribution of the points obtained after each downsampling step. Fig. 11 illustrates the percentage change in the number of sampled points per category when using MROS compared to RS. It can be observed that categories, such as powerline, car, facade, shrub, and tree all experience an increase in the number of sampled points, with the number of powerline points showing a notable increase of over 22%. As shown in Fig. 6, powerline is the least represented category in the dataset, and the others also account for relatively small proportions. Although these categories exhibit distinct geometric characteristics, they are likely to be underrepresented by RS due to their low frequency. In contrast, categories like low vegetation, impervious surface, and roof show a reduction in sampled points. These are the three most dominant categories in the ISPRS dataset and typically contain large, flat regions (roof and ground). These results help

explain the performance improvements observed in Table I for the aforementioned minority classes and also demonstrate the feature selection capability of MROS. Compared with RS, MROS provides a more balanced and task-relevant sampling scheme, which is beneficial for enhancing semantic segmentation performance.

Furthermore, we conducted a visual analysis of the sampling results generated by the proposed MROS strategy and compared them with those of the widely used RS and FPS methods. As shown in Fig. 12, the first row displays the input point clouds, while the second to fourth rows illustrate the results obtained by MROS, RS, and FPS, respectively, under  $4\times$  and  $16\times$  downsampling rates. Regions with noticeable differences are highlighted with red boxes for clearer comparison. It can be observed that RS tends to overlook objects with fewer laser points but distinct structural characteristics, such as powerlines, cars, and facades. As a result, small-scale objects may be discarded during the early stages of downsampling, adversely affecting the model's ability to extract and recognize features from such categories. In contrast, the MROS strategy helps preserve more of these small yet structurally significant objects, thereby enhancing the representativeness of the subsampled points. This advantage can be attributed to MROS's ability to consider local feature variations across MRO, enabling it to retain points that are more informative and better reflect the spatial geometric structures. Regarding sampling uniformity, FPS achieves the most uniform results. However, MROS tends to prioritize geometrically complex regions while preserving fewer points in geometrically simple areas, such as flat ground surfaces. It is noticeable that MROS retains fewer ground points than RS and FPS under both  $4\times$  and  $16\times$  downsampling, yet it still captures the underlying ground geometry effectively. Considering the quantitative comparisons on efficiency (see Table IV) and the analysis in Section IV-I, these observations further support the effectiveness of the proposed MROS strategy. They also highlight the

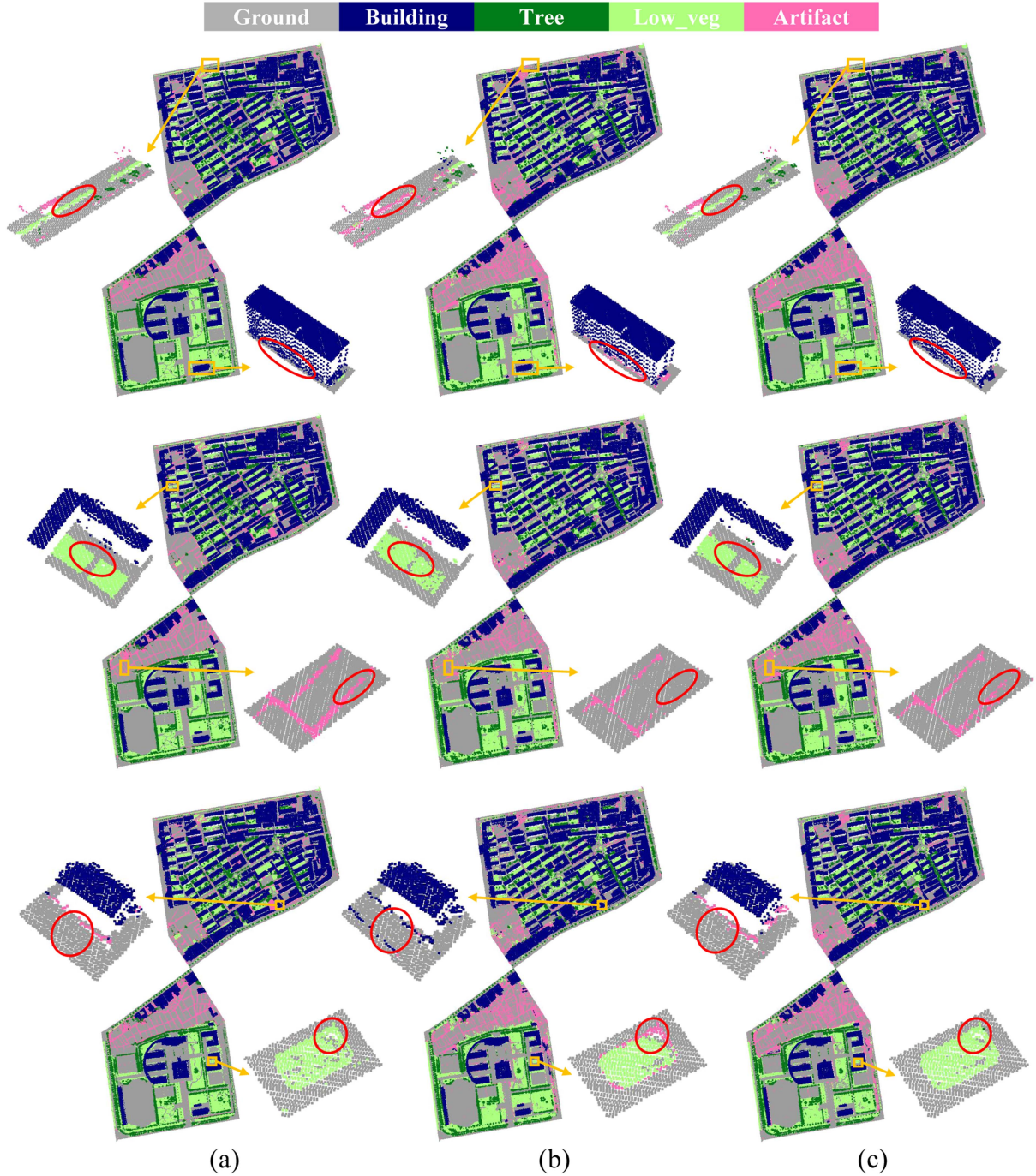


Fig. 9. Visual comparison of classification results on the LASDU dataset between the proposed method and baseline methods. From top to bottom: BAF-LAC, BAAF, and RandLANet. (a) Ground truth. (b) Baselines. (c) Ours.

significance of developing nonuniform sampling methods that selectively retain features from geometrically complex structures.

#### H. Effectiveness of MROA Module

As shown by the quantitative comparison results on the ISPRS dataset (see Table I) and the LASDU dataset (see Table II), the MROA module brings varying degrees of performance improvement to three representative RS-based semantic segmentation

models: RandLANet, BAF-LAC, and BAAF. In this section, we further analyze the effectiveness of the MROA module, with quantitative results summarized in Table III. First, we removed the MROA module from the MRO-RandLANet model. The experimental results on the ISPRS dataset indicate a noticeable degradation in semantic segmentation performance, with declines observed in OA,  $mF_1$ , and mIoU-scores, demonstrating the positive contribution of MROA to the model's effectiveness. Next, we attempted to fuse features from multiple downsampled point subsets using simple max-pooling or average-pooling

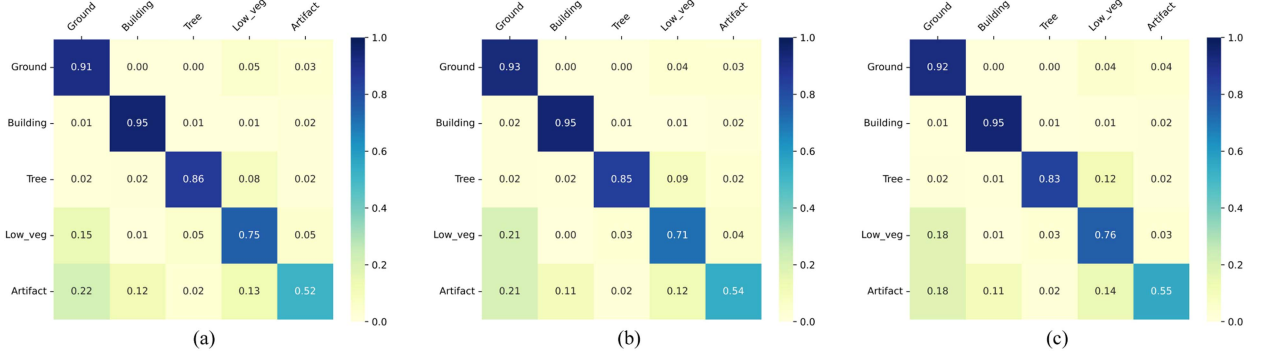


Fig. 10. Normalized confusion matrix of the MRO-based methods on the LASDU dataset. (a) MRO-BAF-LAC. (b) MRO-BAAF. (c) MRO-RandLANet.

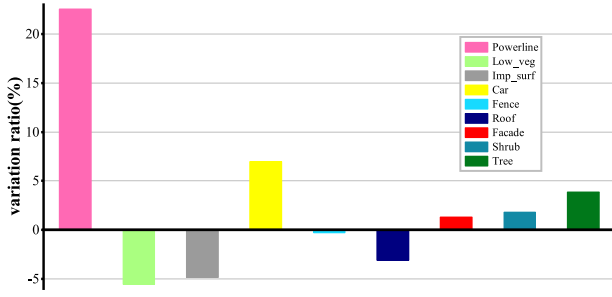


Fig. 11. Variation ratio of subsampled point counts for each semantic category using MROS sampling compared to RS on the ISPRS dataset during inference phase. A positive value indicates an increased presence of the class in the downsampled points, whereas a negative value reflects a reduced presence.

TABLE III  
QUANTITATIVE ANALYSIS OF MROA MODULE EFFECTIVENESS

Fusion strategy	OA(%)	$mF_1$ (%)	mIoU (%)
None	84.07	72.01	59.23
Max-pooling	83.52	72.11	59.19
Average-pooling	83.76	72.03	59.12
MROA	<b>84.35</b>	<b>72.83</b>	<b>60.28</b>

(Best performances are marked in bold.)

TABLE IV  
QUANTITATIVE COMPARISON OF INFERENCE EFFICIENCY AND MODEL ACCURACY OF RANDLANET WITH DIFFERENT SETTINGS ON THE ISPRS DATASET

Methods	Forward pass time (ms)	OA(%)	$mF_1$ (%)
RandLANet	<b>16.14</b>	83.97	70.86
RandLANet+FPS	356.28	84.03	71.43
MRO-RandLANet	78.44	<b>84.35</b>	<b>72.83</b>

(Best performances are marked in bold.)

strategies. However, these approaches failed to produce significant performance gains and even led to a decrease in OA in some cases. This suggests that feature fusion from multiple observations is not straightforward, and that an appropriate fusion strategy is essential for enhancing segmentation performance. These experiments further validate the effectiveness of the MROA module.

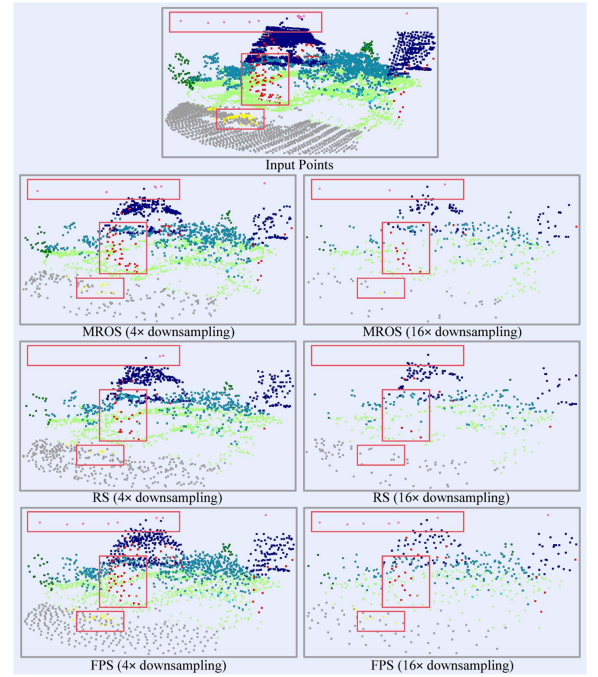


Fig. 12. Visualized comparison of downsampling results of MROS, RS, and FPS sampling methods on the ISPRS dataset.

### I. Efficiency Analysis

In this section, we further analyze the efficiency of the proposed MRO framework. Specifically, we compare the inference efficiency and semantic segmentation performance between RandLANet and MRO-RandLANet during the testing phase. The experiments are conducted on the ISPRS dataset. In addition, we replace the RS strategy in RandLANet with the widely adopted FPS sampling method to perform a comparative study. The quantitative results are presented in Table IV. As observed, the inference efficiency of our method is lower than that of RandLANet. This is primarily because the MRO framework requires multiple forward passes at each feature extraction stage to process different downsampled point subsets, which inevitably impacts the overall efficiency. Nevertheless, the MRO framework brings noticeable improvements in semantic segmentation performance. Specifically, MRO-RandLANet achieves a 1.97% gain in  $mF_1$ -score compared to RandLANet. On the other hand,

when the RS strategy in RandLANet is replaced by FPS, we observe an improvement in semantic segmentation performance. However, this comes at the cost of a significant reduction in runtime efficiency, with the inference speed being approximately 4.5 times slower than that of MRO-RandLANet. In comparison, the MRO framework achieves better runtime efficiency than FPS sampling while also delivering superior semantic segmentation performance. Overall, these results indicate that although the MRO framework incurs some efficiency loss, such tradeoff may be acceptable in application scenarios where runtime efficiency is not the primary concern.

## V. CONCLUSION

In this article, we propose a MRO framework to enhance the semantic segmentation performance of RS-based networks on ALS point clouds. The framework leverages the efficiency of RS by performing multiple rounds of nonreplacement random downsampling during the feature extraction stage, thereby covering the target point set with diverse and spatially complementary subsets. High-dimensional features extracted from these observations are then fused to improve spatial completeness. To this end, we design the MROA module, which aggregates features from multiple observations using local spatial attention, thereby enhancing the robustness of local feature extraction. Furthermore, we introduce the MROS sampling strategy to retain more informative points by computing local feature differences that reflect the relative importance of each center point. These scores are used as probabilistic sampling weights for guided downsampling. We integrate the proposed MRO method with several representative RS-based networks and evaluate its effectiveness on the ISPRS and LASDU datasets. Experimental results demonstrate that both the MROA and MROS components contribute positively to performance improvements. The proposed method achieves competitive results on the ISPRS dataset and outperforms current state-of-the-art methods on the LASDU dataset, with all three baseline networks achieving  $mF_1$  scores above 79% on the LASDU dataset after incorporating the MRO-based components. However, due to the need for multiple forward passes during the feature extraction process to obtain multiobservation features, the MRO framework still has a certain degree of negative impact on model efficiency. In future work, we plan to design lightweight feature extraction modules and more efficient feature fusion strategies to further improve the overall efficiency.

## REFERENCES

- [1] Z. Xiong, F. Zhang, Y. Wang, Y. Shi, and X. X. Zhu, "Earth-Nets: Empowering artificial intelligence for Earth observation," *IEEE Geosci. Remote Sens. Mag.*, early access, Oct. 23, 2024, doi: [10.1109/MGRS.2024.3466998](https://doi.org/10.1109/MGRS.2024.3466998).
- [2] U. Stilla and Y. Xu, "Change detection of urban objects using 3 d point clouds: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 197, pp. 228–255, 2023.
- [3] L. Du et al., "A LiDAR biomass index-based approach for tree-and plot-level biomass mapping over forest farms using 3D point clouds," *Remote Sens. Environ.*, vol. 290, 2023, Art. no. 113543.
- [4] N. Qin, W. Tan, L. Ma, D. Zhang, H. Guan, and J. Li, "Deep learning for filtering the ground from ALS point clouds: A dataset, evaluations and issues," *ISPRS J. Photogrammetry Remote Sens.*, vol. 202, pp. 246–261, 2023.
- [5] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogrammetry Remote Sens.*, vol. 105, pp. 286–304, 2015.
- [6] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Contextual classification of LiDAR data and building object detection in urban areas," *ISPRS J. Photogrammetry Remote Sens.*, vol. 87, pp. 152–165, 2014.
- [7] L. Landrieu, H. Ragué, B. Vallet, C. Mallet, and M. Weinmann, "A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds," *ISPRS J. Photogrammetry Remote Sens.*, vol. 132, pp. 102–118, 2017.
- [8] T. Hackel, J. D. Wegner, and K. Schindler, "Fast semantic segmentation of 3 d point clouds with strongly varying density," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 177–184, 2016.
- [9] T. Zeng, F. Luo, T. Guo, X. Gong, J. Xue, and H. Li, "Recurrent residual dual attention network for airborne laser scanning point cloud semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5702614.
- [10] J. Li et al., "Vd-lab: A view-decoupled network with local-global aggregation bridge for airborne laser scanning point cloud classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 186, pp. 19–33, 2022.
- [11] X. Hu and Y. Yuan, "Deep-learning-based classification for DTM extraction from ALS point cloud," *Remote Sens.*, vol. 8, no. 9, 2016, Art. no. 730.
- [12] Z. Yang, W. Jiang, B. Xu, Q. Zhu, S. Jiang, and W. Huang, "A convolutional neural network-based 3D semantic labeling method for als point clouds," *Remote Sens.*, vol. 9, no. 9, 2017, Art. no. 936.
- [13] A. Rizaldy, C. Persello, C. Gevaert, S. Oude Elberink, and G. Vosselman, "Ground and multi-class classification of airborne laser scanner point clouds using fully convolutional networks," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1723.
- [14] S. Schmohl and U. Soergel, "Submanifold sparse convolutional networks for semantic segmentation of large-scale ALS point clouds," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 77–84, 2019.
- [15] H. Dai, X. Hu, Z. Shu, N. Qin, and J. Zhang, "Deep ground filtering of large-scale ALS point clouds via iterative sequential ground prediction," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 961.
- [16] Y. Mao et al., "Beyond single receptive field: A receptive field fusion-and-stratification network for airborne laser scanning point cloud classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 188, pp. 45–61, 2022.
- [17] T. Jiang, Y. Wang, S. Liu, Y. Cong, L. Dai, and J. Sun, "Local and global structure for urban ALS point cloud semantic segmentation with ground-aware attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5702615.
- [18] T. Zeng, F. Luo, T. Guo, X. Gong, J. Xue, and H. Li, "Multi-level context feature fusion for semantic segmentation of ALS point cloud," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5506605.
- [19] R. Huang, Y. Xu, and U. Stilla, "Granet: Global relation-aware attentional network for semantic segmentation of ALS point clouds," *ISPRS J. photogrammetry remote Sens.*, vol. 177, pp. 1–20, 2021.
- [20] B. Guo et al., "MCTNet: Multiscale cross-attention-based transformer network for semantic segmentation of large-scale point cloud," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5704720.
- [21] Q. Hu et al., "Learning semantic segmentation of large-scale point clouds with random sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8338–8354, Nov. 2021.
- [22] Q. Hu et al., "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11108–11117.
- [23] H. Cheng, X. Han, H. Jiang, D. He, and G. Xiao, "PCB-RandNet: Rethinking random sampling for LiDAR semantic segmentation in autonomous driving scene," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2022, pp. 4435–4441. [Online]. Available: <https://api.semanticscholar.org/CorpusID:252567968>
- [24] B. Graham, M. Engelcke, and L. Van Der Maaten, "3D semantic segmentation with submanifold sparse convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9224–9232.
- [25] C. Choy, J. Gwak, and S. Savarese, "4 d spatio-temporal convnets: Minkowski convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3075–3084.
- [26] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet : Fast and accurate LiDAR semantic segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2019, pp. 4213–4220.
- [27] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.

- [28] J. Zhao, H. Zhou, and F. Pan, "A dual attention KPCConv network combined with attention gates for semantic segmentation of ALS point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5107914.
- [29] S. Huang, Q. Hu, P. Zhao, J. Li, M. Ai, and S. Wang, "ALS point cloud semantic segmentation based on graph convolution and transformer with elevation attention," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2877–2889, 2024.
- [30] H. Dai, X. Hu, J. Zhang, Z. Shu, J. Xu, and J. Du, "Large-scale ALS point clouds segmentation via projection-based context embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5704216.
- [31] Y. Yang et al., "DQFormer: Towards unified LiDAR panoptic segmentation with decoupled queries for large-scale outdoor scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5702515.
- [32] A. Abid, M. F. Balin, and J. Zou, "Concrete autoencoders for differentiable feature selection and reconstruction," 2019. [Online]. Available: <https://arxiv.org/abs/1901.09346>
- [33] K. Xu et al., "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. 32nd Int. Conf. Mach. Learn.*, Jul. 2015, pp. 2048–2057. [Online]. Available: <https://proceedings.mlr.press/v37/xuc15.html>
- [34] O. Dovrat, I. Lang, and S. Avidan, "Learning to sample," in *Proc. 2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2755–2764. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.00287>
- [35] P. He et al., "Offs-Net: Optimal feature fusion-based spectral information network for airborne point cloud classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 141–152, 2023.
- [36] H. Shuai, X. Xu, and Q. Liu, "Backward attentive fusing network with local aggregation classifier for 3D point cloud semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 4973–4984, 2021.
- [37] S. Qiu, S. Anwar, and N. Barnes, "Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1757–1767.
- [38] R. Huang, Y. Xu, D. Hong, W. Yao, P. Ghamisi, and U. Stilla, "Deep point embedding for urban classification using ALS point clouds: A new perspective from local to global," *ISPRS J. Photogrammetry Remote Sens.*, vol. 163, pp. 62–81, 2020.
- [39] X. Li, L. Wang, M. Wang, C. Wen, and Y. Fang, "Dance-net: Density-aware convolution networks with context encoding for airborne lidar point cloud classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 128–139, 2020.
- [40] C. Wen, X. Li, X. Yao, L. Peng, and T. Chi, "Airborne LiDAR point cloud classification with global-local graph attention convolution neural network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 181–194, 2021.
- [41] C. Liu, Z. Liu, and X. Wang, "Bidirectional feature aggregation and adaptive fusion network for ALS point cloud semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, 2025, Art. no. 6500705.
- [42] Z. Wang, H. Chen, J. Liu, J. Qin, Y. Sheng, and L. Yang, "Multilevel intuitive attention neural network for airborne LiDAR point cloud semantic segmentation," *Int. J. Appl. Earth Observation Geoinformation*, vol. 132, 2024, Art. no. 104020.
- [43] Z. Ye et al., "LASDU: A large-scale aerial LiDAR dataset for semantic labeling in dense urban areas," *ISPRS Int. J. Geo- Inf.*, vol. 9, no. 7, 2020, Art. no. 450.



**Hengming Dai** received the B.E. degree in remote sensing science and technology and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017 and 2023, respectively.

He is currently a Postdoctoral Research Fellow with the Institute of International Rivers and Eco-Security, Yunnan University, and the School of Earth Sciences, Yunnan University, Kunming, China. His research interests include computer vision and point cloud processing.



**Zhifang Zhao** received the B.S. degree in geomorphology and quaternary geology from Nanjing University, Nanjing, China, in 1992, and the Ph.D. degree in geological resources and geological engineering, specializing in resource and environment remote sensing, from the China University of Geosciences, Beijing, China, in 2009.

Since 2009, she has been a Professor with the Department of Earth Sciences, Yunnan University, Kunming, China, since 2009. Her research interests include remote sensing data processing and relevant applications.



**Huiwei Jiang** received the M.Sc. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016 and 2021, respectively.

She is currently an Engineer with the National Geomatics Center of China, Beijing, China. Her research interests include artificial intelligence, semantic segmentation, and change detection.



**Jiabo Xu** received the B.S. degree in software engineering from Nanchang Hangkong University, Nanchang, China, in 2019, the M.S. degree in computer technologies from the School of Mathematics and Computer Sciences, Nanchang University, Nanchang, China, in 2023. He is currently working toward the doctoral degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University.

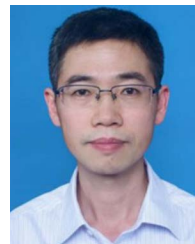
His research interests include computer vision, point cloud processing, and artificial intelligence generated content techniques.



**Haihan Duan** (Member, IEEE) received the B.Eng. degree in computer science and technology from East China Normal University, Shanghai, China, in 2017, the M.Eng. degree in software engineering from Sichuan University, Chengdu, China, in 2020, and the Ph.D. degree in computer and information engineering from The Chinese University of Hong Kong, Shenzhen, China, in 2023.

He is currently an Associate Professor with Artificial Intelligence Research Institute, Shenzhen MSU-BIT University (SMBU), Shenzhen, and also with

Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen. Before joining SMBU, he worked as a Postdoctoral Research Fellow with the University of Ottawa and Mohamed bin Zayed University of Artificial Intelligence, located in Abu Dhabi, UAE. His research interests include multimedia, blockchain and Web3, metaverse, human-centered computing, and computer vision.



**Xiangyun Hu** received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2001.

From 2002 to 2005, he was a Postdoctoral Research Fellow with the Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, Toronto, ON, Canada. He has developed a feature extraction technology SmartDigitizer acquired by PCI Geomatics, Leica Geosystems, and Microsoft. From 2005 to 2010, he was a Senior Software Engineer with ERDAS Inc.,

Atlanta, GA, USA. He is currently a Professor and Head of the Department of Photogrammetry, School of Remote Sensing and Information Engineering, Wuhan University. He is also an Adjunct Professor with Hubei LuoJia Laboratory, Wuhan. He has been leading a team developing an open source deep learning framework-LuoJiaNET. He has authored or coauthored more than 100 papers in journals and conferences on intelligent feature extraction of remotely sensed data.