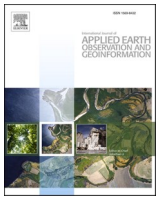
International Journal of Applied Earth Observation and Geoinformation 125 (2023) 103588

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Comparison of 2D and 3D vegetation species mapping in three natural scenarios using UAV-LiDAR point clouds and improved deep learning methods

Liwei Deng [a], Bolin Fu [a,*], Yan Wu [a], Hongchang He [a], Weiwei Sun [b], Mingming Jia [c], Tengfang Deng [a], Donglin Fan [a]

[a] College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China
[b] Department of Geography & Spatial Information Techniques, Ningbo University, Ningbo, China
[c] Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

ARTICLE INFO

Keywords:
Karst wetland
Mangroves
Vegetation classification
Point cloud semantic segmentation
Multi-scale feature selection
Class imbalance
Feature importance

ABSTRACT

Collaboration between Light Detection and Ranging (LiDAR) point clouds and deep learning has been proven to be an effective approach for vegetation mapping. Current studies have predominantly focused on 2D vegetation mapping, whereas 3D mapping, which directly classifies point clouds at point level, offers a more comprehensive understanding of the stratified structural information of vegetation. However, there is a lack of research on 3D vegetation species mapping, and the disparities between 2D and 3D mapping in natural scenarios remain unclear. To resolve these issues, we compared the deep learning performance of 2D and 3D vegetation species mapping across three distinct natural scenes: karst wetland, mangrove forest, and hill forest. In addition, the 2D and 3D mapping in natural scenes are adversely affected by the elevated channel count of LiDAR-derived features and the extreme category imbalance in point cloud. To mitigate these challenges, we propose a novel Multi-resolution Feature Selection Network (MrFSNet) to select optimal feature combinations at different scales for better 2D mapping performance. Additionally, we introduce a novel Dynamic Weighted Sampling (DWS) strategy, which is combined with KPConv to address the extreme category imbalance present in 3D mapping. Results indicate that: (1) 3D vegetation species mapping exhibited the highest performance, achieving an mF1 of 89.78% for karst wetland, 92.25% for mangrove forest, and 92.05% for hill forest. (2) 3D mapping outperformed 2D mapping, improving mF1 by 3.43% to 27.08%. (3) MrFSNet adaptively extracted optimal features at various scales and performed well with the limited training data in 2D vegetation mapping, resulting in a 1.66%–18.46% higher mF1 than that of Swin Transformer. (4) DWS effectively resolved the extreme category imbalance problem and produced 1.28%–2.80% higher mF1 than the non-DWS version in 3D vegetation mapping.

1. Introduction

Accurate vegetation classification and mapping are crucial for biomass measurement (Maxwell et al., 2023), estimation of above-ground carbon stocks (Zhao et al., 2023) and assessment of carbon sequestration capacity (Tong et al., 2020). In natural scenes characterized by intricate vertical structures, Unmanned Aerial Vehicle (UAV) Light Detection and Ranging (LiDAR) offers a rapid and efficient approach for remote sensing of vegetation (Calders et al., 2020; Campbell et al., 2023). Its advantages encompass precise 3D measurements, canopy-penetrating capabilities (Pourshamsi et al., 2021), and independence from light conditions (Yang et al., 2015). High-density 3D point clouds obtained from LiDAR offer intricate insights into the vertical structure of vegetation, making it a popular tool for remote sensing vegetation mapping (Cao et al., 2021).

Deep learning has found extensive application in LiDAR mapping (Adam et al., 2023). Currently, the methodology for deep learning mapping using LiDAR mainly includes two aspects: 2D and 3D mapping (semantic segmentation). In the 2D mapping approach, the 3D point cloud is usually converted into 2D raster LiDAR features (Rana et al., 2022; Shi et al., 2018) and employed raster classifier such as CNN (Li et al., 2021), DeepLabv3+ and HRNet (Liu et al., 2021a) for pixel-wise

* Corresponding author.
E-mail address: fubolin@glut.edu.cn (B. Fu).

https://doi.org/10.1016/j.jag.2023.103588
Received 31 August 2023; Received in revised form 7 November 2023; Accepted 24 November 2023
Available online 30 November 2023

classification. In contrast, 3D mapping directly processes the original unstructured 3D point cloud to obtain classification results on a point-wise basis (Landrieu and Simonovsky, 2018). This approach primarily leverages deep learning methods such as PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), and KPConv (Thomas et al., 2019).

In 2D vegetation mapping, detailed spatial structure information is usually lost during the conversion of 3D point clouds to 2D raster LiDAR features (Mao et al., 2022). Unfortunately, this conversion also introduces a significant amount of data redundancy. The raster LiDAR features derived from point cloud are images with tens or even hundreds of channels, which is significantly more than ordinary RGB and multi-spectral images with a limited number of bands. This makes it challenging for general neural networks, which are usually designed for only RGB three-channel inputs (Chen et al., 2018; Dosovitskiy et al., 2020). The first few layers of such networks only have limited hidden dimensions, which can impede the capture of valuable feature information. In addition, the features in remote sensing images have varying significance at different scales (Liu et al., 2021a). Therefore, for high-channel images, it is possible that some features may offer broad contextual information at a large-scale but may not be suitable for capturing finer details at a local scale, and other features are suitable for extracting local details but cannot provide global contextual information. Based on this conjecture, our hypothesis is that by selectively combining features from different scales, it may be possible to improve the performance of 2D vegetation mapping using high-channel LiDAR-derived images.

In 3D mapping, the utilization of LiDAR point cloud allows for spatial convolution and point-wise classification based on the raw morphology of the point cloud (Guo et al., 2023; Thomas et al., 2019). This effectively avoids information loss that may occur when converting point clouds into structured formats such as 2D raster or 3D voxel representations (Qi et al., 2017b). In natural environments, this subtle spatial structural information of point clouds, such as vegetation canopy morphology and branching structures of tree trunks, plays a crucial role in accurately identifying vegetation types within the natural scene (Dersch et al., 2021; Du et al., 2023). Currently, point cloud classification methods have gained extensive utilization in mapping indoor or urban scenes by leveraging point clouds (Lin et al., 2021; Luo et al., 2020; Wen et al., 2020; Zhu et al., 2017). However, in natural scenes, their application has focused on individual tree segmentation and classification (Qin et al., 2022; Seidel et al., 2021; Zhou et al., 2022a), or predicting vegetation stratum occupancy (Kalinicheva et al., 2022a, b). There has been a notable absence of point-wise vegetation species classification across entire natural scenes. Hence, the suitability of 3D mapping for complex vegetation species in natural environments remains uncertain, requiring further investigation.

Category imbalance problems in 3D vegetation mapping are exacerbated by the high density of point clouds within canopies and the disorderly distribution of vegetation in the natural environment. For instance, in the point cloud dataset of the hill forest scene in this study, *natural arbor* occupies 60.20 % of the total points, while *grass and shrub* account for 25.57 %, leaving the remaining six categories with only 14.23 % of the points. *Building* and *water* occupy only about 0.46 % and 0.67 % of the points, respectively. The effects of category imbalance can be significant. Leading models perform well in the overrepresented categories but poorly in the underrepresented categories. Several techniques have been proposed to address category imbalance, including data augmentation, class weighting, and oversampling (Turkoglu et al., 2021; Waldner et al., 2019). Nevertheless, in point cloud semantic segmentation tasks where the entire scene is directly input into the model, class weighting is typically the only viable option (Han et al., 2021; Zhou et al., 2022b). In practical implementation, it is necessary to incorporate a smoothing constant into the class weights to ensure that rare categories do not yield excessively large weights, which could lead to unstable training. This causes the model to still ignore rare categories in scenes with extremely imbalanced category distributions. In addition, due to the large amount of data per unit area in point clouds and the limitation of computer memory space, models are often trained with smaller sampling ranges and batch sizes (Hu et al., 2021). This constraint may lead to the model being unable to learn rare category samples during multiple iterations, resulting in decreased classification accuracy for these categories and even causing certain categories to be missed altogether. Therefore, addressing the category imbalance problem during model training is crucial for utilizing 3D deep learning for vegetation mapping in natural conditions.

The distinctions between 3D mapping using point clouds and 2D mapping utilizing LiDAR-derived raster features exhibit noteworthy disparities in terms of spatial dimensions, classifier architecture, and outcome presentation. Nevertheless, there exists a notable dearth of endeavors aimed at contrasting the performance of vegetation species mapping within natural scenes through the utilization of these methodologies. To fill this gap, we undertook an exhaustive experiment involving vegetation species mapping in both 2D and 3D contexts across three distinct natural scenarios with the UAV-LiDAR point clouds. Furthermore, we introduced two innovative approaches, namely MrFSNet and DWS-KP-FCNN, specifically designed to mitigate the challenges identified during the course of our research. The primary contributions of this study can be summarized as follows:

(1) This study performs 3D vegetation species mapping using deep learning and LiDAR data in three distinct natural scenarios, including karst wetland, mangrove forest, and hill forest.
(2) We present a multi-scene comparison between 2D and 3D deep learning for vegetation species mapping, utilizing LiDAR point clouds and their derived data.
(3) We proposed a novel Multi-resolution Feature Selection Network (MrFSNet) to improve the accuracy of 2D vegetation mapping, which effectively leverages high-channel raster LiDAR features by adaptively determining the optimal feature combinations at each scale.
(4) We developed a DWS-KP-FCNN for 3D vegetation mapping, which stacked a novel Dynamic Weighted Sampling (DWS) strategy and KPConv to handle the extreme category imbalance by dynamically increasing the sampling frequency of rare categories.

2. Study area and data source

2.1. Study areas

We selected three representative study areas in Guangxi Zhuang Autonomous Region, southern China, which encompass a total area of 101 ha, as illustrated in Fig. 1. Each of these study areas exhibits distinct vegetation, covering a diverse range of scenarios:

(1) **Karst wetland.** Huixian karst wetland, located in Huixian town, Guilin City, is the largest karst wetland in China and is also recognized as a wetland of international importance. Within this karst wetland scene, vegetation and water bodies intricately intertwine, creating a unique ecosystem. The vegetation predominantly consists of aquatic and terrestrial herbaceous plants, complemented by certain water-friendly tree species (Fu et al., 2021). The terrain exhibits a flat topography with minimal variation in height, where the maximum elevation difference is less than 1 m.
(2) **Mangrove forest.** Maowei Sea Mangrove Nature Reserve in Qinzhou City, which is the largest mangrove island in China. Mangrove forests play a vital ecological role (Fu et al., 2023), primarily characterized by the growth of diverse shrubs in coastal mudflats. In this region, the terrain gradually slopes downward from the inland areas towards the coastline, exhibiting an approximate height difference of 3.5 m.

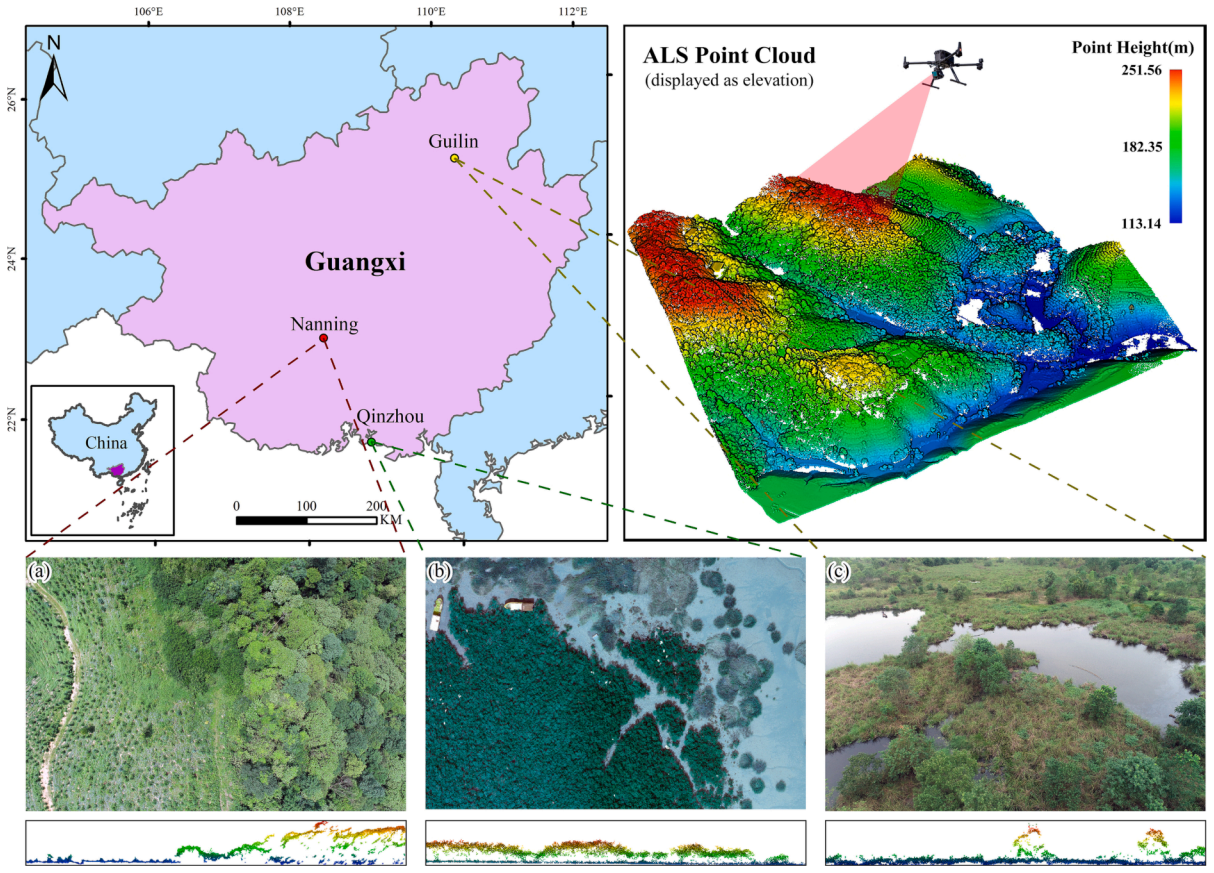


Fig. 1. Study area locations for three natural scenes using UAV-acquired LiDAR point clouds and RGB images: (a) hill forest scene; (b) mangrove forest scene; (c) karst wetland scene.

(3) **Hill forest.** Liuli Forest, located in Nanning City, features undulating terrain and dense vegetation in its landscape. The area encompasses a mix of natural vegetation and planted forests. Within this hill forest scene, tall trees dominate the scenery, thriving on the steep slopes of the hills. The elevation differences within this region can reach up to 100 m (as indicated in the upper right section of Fig. 1).

2.2. Point clouds acquisition and processing

This study utilized an DJI Matrice 300 RTK UAV equipped with a Zenmuse L1 LiDAR scanner to acquire RGB images and LiDAR point clouds. The UAV was flown at a height of approximately 120 m, with a LiDAR maximum scan angle of ± 35°, a sampling interval of 1 ns, and a side overlap rate of 60 %. The specific details regarding data collection for each scene are outlined as follows:

(1) **Karst wetland.** LiDAR data was collected on July 1, 2022, from 10:20 AM to 14:00 PM local time. The resulting point cloud dataset had a horizontal extent of approximately 700 m × 500 m and an average point density of 498 pts/m^2. Based on 597 samples from the field survey, point clouds were classified into eight categories: *grass* (GR), *shrub* (SH), *water* (WT), *cropland* (CL), *aquatic vegetation* (AV), *Salix matsudana* (SM), *Bambusa sinospinosa* (BS), and *other arbor* (AR).

(2) **Mangrove forest.** LiDAR data was collected on October 17, 2022, from 12:00 AM to 16:00 PM local time. The point cloud dataset obtained had an approximately horizontal extent of 500 m × 600 m and an average density of 167 pts/m^2. Six categories, namely *mire* (MR), *artifact* (AR), *Spartina alterniflora* (SA), *Derris trifoliata* (DT), *Aegiceras corniculatum* (AC), and *Avicennia marina* (AM), were classified from the point clouds based on 600 samples obtained from field surveys and visual interpretation.

(3) **Hill forest.** LiDAR data was collected on June 9, 2021, from 10:10 AM to 12:10 PM local time. The point cloud dataset obtained in this study covered an area of approximately 600 m × 600 m and had an average density of around 550 pts/m^2. Field measurements and visual interpretation were conducted to collect a total of 1376 sample data points, which were then categorized into eight categories: *road* (RD), b*uilding* (BD), *grass and shrub* (GS), *bare land* (BL), *natural arbor* (NA), *young artificial eucalyptus* (YE), *adult artificial eucalyptus* (AE), and *water* (WT).

We used DJI SmartMap software (DJI Innovation Technology Co., Ltd.) to process the original LiDAR data and acquire a high-precision true-color point cloud in LAS format. Each point includes 3D coordinates (X, Y, and Z), color (R, G, and B), intensity, scan angle, return number, and GPS time. The categorization of the three scenes based on their distinct characteristics is summarized in Table 1.

Table 1
Summary of categorization for three scenes.

Scene	Categories
Karst wetland	*grass*, *shrub*, *water*, *cropland*, *aquatic vegetation*, *Salix matsudana*, *Bambusa sinospinosa*, *other arbor*
Mangrove forest	*mire*, *artifact*, *Spartina alterniflora*, *Derris trifoliata*, *Aegiceras corniculatum*, *Avicennia marina*
Hill forest	*road*, *building*, *grass and shrub*, *bare land*, *natural arbor*, *young artificial eucalyptus*, *adult artificial eucalyptus*, *water*

3. Methods

In this study, we collected LiDAR point clouds from three distinct natural scenes and constructed 2D and 3D datasets for training vegetation classification models. We proposed two novel deep learning methods: MrFSNet for 2D mapping, and DWS-KP-FCNN for 3D mapping. Then we conducted 2D and 3D vegetation mapping across three scenes. Specifically, we quantitatively evaluated the classification performance of our proposed methods against traditional deep learning methods. Furthermore, based on the experimental results, we investigated the classification performance differences between 2D and 3D mapping. The flowchart of the overall technical route is depicted in Fig. 2.

3.1. Constructing 2D and 3D datasets based on LiDAR point clouds

For each scene, we constructed a point cloud dataset for 3D mapping and a raster dataset using LiDAR-derived raster features for 2D mapping, which were utilized to train deep learning models.

Point cloud datasets were processed using the open-source software CloudCompare v2.12 beta (Girardeau-Montaut, 2022), C++ Point Cloud Library (PCL) v1.12.1 (Rusu and Cousins, 2011), and Python 3.7. The processing for constructing point cloud datasets is summarized in Fig. 3 with the following steps: (1) Remove redundant points to reduce the subsequent computation using uniform subsampling with a minimum spacing of 0.2 m; (2) Remove outliers produced by low-altitude flying objects, LiDAR noise, and so on, using a statistical outlier filter with standard deviation multiplier threshold of 1; (3) Coordinate normalization is performed to ensure the coordinate accuracy in subsequent processing; (4) Manually annotating point-wise category labels using CloudCompare software based on field survey samples; (5) Applying cloth simulation filter (Zhang et al., 2016) and an inverse distance weight algorithm for point cloud elevation normalization to eliminate the impact of terrain undulation while preserving the original elevation as a feature of the point cloud; (6) Computing the normal vector and curvature with a radius of 0.4 m using CloudCompare as additional features; (7) Dividing the total point cloud into several blocks of approximately 100 m × 100 m within the horizontal dimensions. Approximately 20 % of the blocks, containing all categories, were selected as the test set. The remaining blocks were selected as the training set. Finally, the point cloud features used for classification include RGB, intensity, return number, curvature, normal vector, and original elevation.

Raster datasets were made up of LiDAR features derived from point clouds. The original point clouds underwent uniform subsampling with a minimum spacing of 0.1 m, statistical outlier removal, and elevation normalization. LiDAR360 software (Beijing Digital Green Earth Technology Co., Ltd.) was then employed to extract 107 LiDAR features (Wang et al., 2020; GreenValley International, 2021). We experimented with various rasterization resolutions and ultimately selected 0.25 m × 0.25 m as the resolution. This choice aims to achieve the highest possible

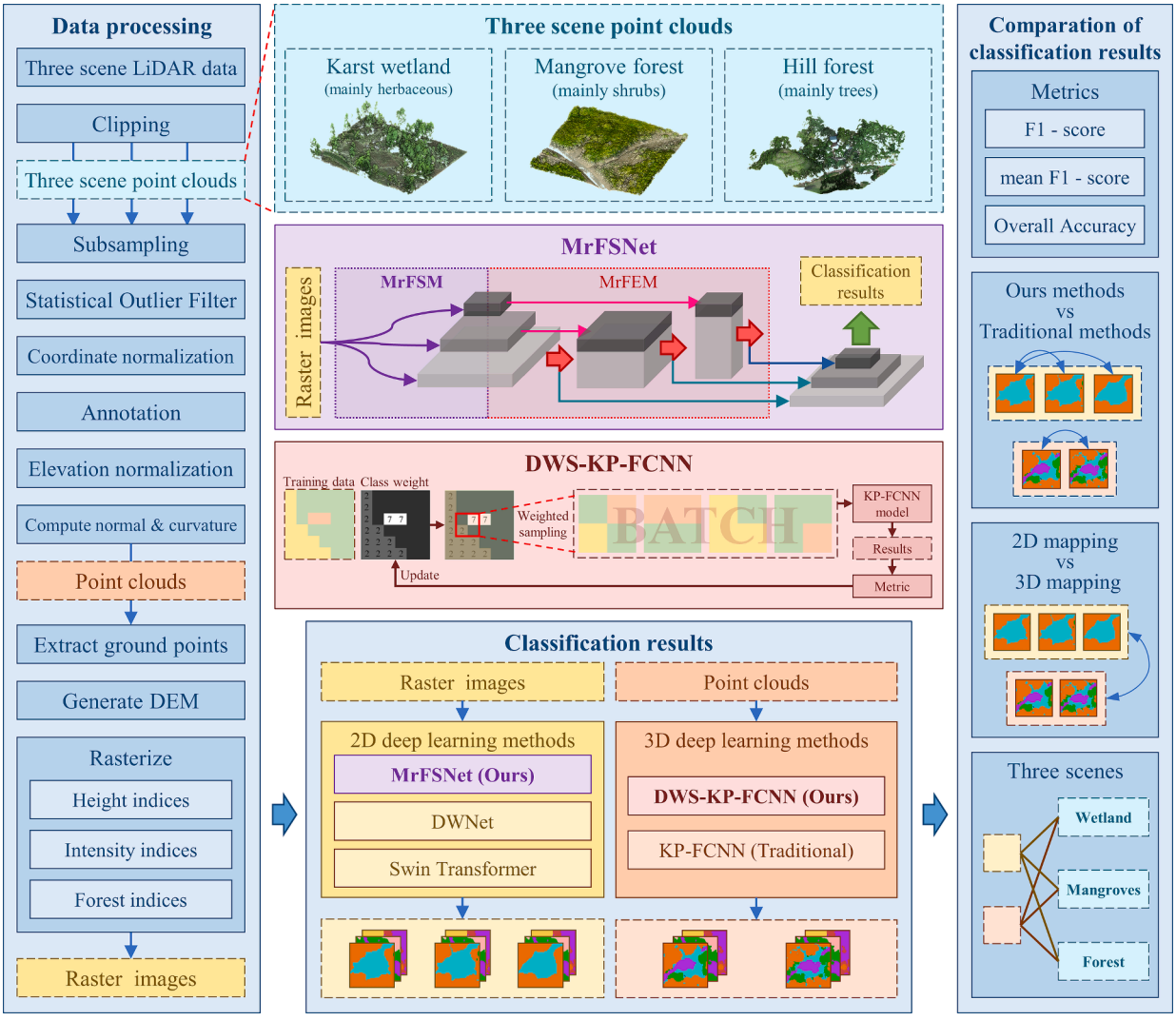


Fig. 2. Flowchart of the overall technical route in this study.

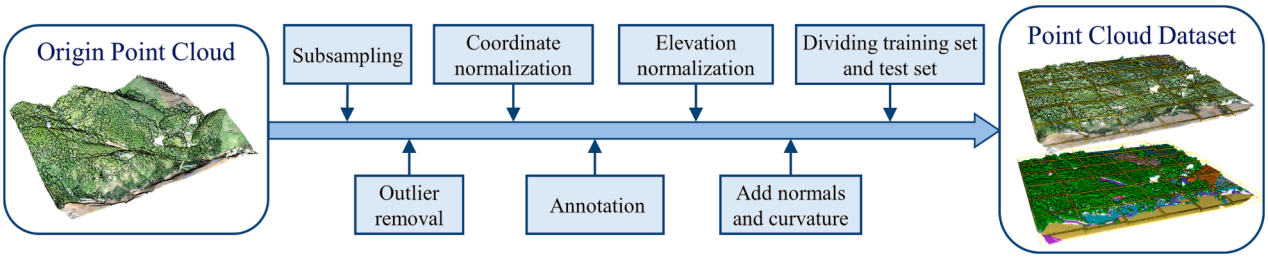


Fig. 3. The main process of constructing the point cloud dataset.

resolution while ensuring that, in most cases, each raster cell contains at least two points for calculating LiDAR metrics. This process resulted in 107-channel images at a 0.25 m × 0.25 m resolution. The correspondence between the point cloud features and the extracted raster features is summarized in Table 2. Moreover, we utilized the CloudCompare software to project point-wise labels in point cloud datasets into pixel-wise labels by selecting the highest semantic label of each pixel. Finally, raster datasets include 107-channel feature images and pixel-wise semantic labels. The division strategy of the training and test sets in raster datasets is consistent with that of point cloud datasets.

3.2. 2D deep learning mapping using multi-resolution feature selection

In this study, we have employed a deep learning approach to implement multi-resolution feature selection and fusion for 2D vegetation mapping using high-channel LiDAR features. Specifically, we proposed a Multi-resolution Feature Selection Network (MrFSNet), whose overall architecture is illustrated in Fig. 4. MrFSNet is composed of novel MrFSM and MrFEM, each containing four FSBlocks and four FEBlocks, respectively, to capture features at four different scales. Commencing from the scale with the highest resolution, MrFSNet dynamically selects an appropriate combination of features for each scale. These selected features are then passed through MrFEM to integrate with the corresponding scale's feature map. Subsequently, spatial features are captured through the utilization of FEBlocks. Global features are further captured using the Pyramid Pooling Module (PPM) (Zhao et al., 2017) from the final output of MrFEM. The global features and output features from each scale of MrFEM are then fed into Feature Pyramid Networks (FPN) (Lin et al., 2017) for multi-scale feature integration and to produce pixel-wise classification results at the original resolution.

For each scale, MrFSM trains a Feature Selection Block (FSBlock) mainly comprising a unary convolution and a Deep-wise patch convolution (DwpConv). DwpConv is a Depth-Wise convolution Block (DWBlock) (Han et al., 2022) with a step size equal to the size of the convolution kernel, which is used for scaling transformation. The features of the input image are selected and combined using unary convolution, and then mapped to the corresponding scale using DwpConv. DwpConv has lower parameters and computational complexity than regular convolution, making it faster and easier to train on images with multiple feature channels. Since the scale after each DwpConv projection is determined, the unary convolution can learn how to select the feature combinations favorable for classification from the original image at the corresponding scale.

MrFEM consists of Feature Encoding Blocks (FEBlocks) equal to the number of scales; each FEBlock mainly consists of a DWBlock and a Patch Merging Layer (Liu et al., 2021b). DWBlock adopts a transformer-like architecture and uses a big convolutional kernel with DWConv instead of multi-head self-attention to learn spatial features at a lower computational cost. The patch merging layer is utilized to gradually reduce the resolution of the feature map while simultaneously increasing its channel dimension. After each downsampling to a new scale, the features selected by MrFSM at that scale will be added to the first d' features of the feature map. d' is the number of features selected by MrFSM at that scale. The innovative encoding structure, which progressively incorporates desired features scale-by-scale, streamlines information flow. This enables the model to synergize global features and local details, resulting in enhanced classification performance.

We implemented MrFSNet using PyTorch and extracted image features from four different scales, with pixel sizes of 0.25 m × 0.25 m (128 × 128 pixels), 0.5 m × 0.5 m (64 × 64 pixels), 1 m × 1 m (32 × 32 pixels), and 2 m × 2 m (16 × 16 pixels). At each scale, 96 feature components were selected by MrFSM from the original images. To avoid overfitting, we employed techniques such as random scaling, rotation, flipping, and cropping of 128 × 128 pixels on the original data during each sampling. We employed the AdamW optimizer with an initial learning rate of 0.0024 to minimize the weighted cross-entropy loss. Category weighting is employed to mitigate the effect of category imbalances. A batch size of 8 was used, and the learning rate was decreased exponentially, divided by 10 every 1000 iterations. Our model converged in 3000 iterations.

Table 2
Raster features produced by the original point clouds data.

Point cloud features	Raster features	Number of features
XYZ	DEM	1
	DSM	1
	CHM	1
	Gap fraction	1
	Height variable	46
XYZ + Return number	Density variable	10
	Canopy cover	1
XYZ + Intensity	Intensity variable	42
XYZ + Scan angle	LAI	1
XYZ + RGB	RGB	3

3.3. 3D KPConv mapping based on dynamic weight sampling

3.3.1. Dynamic weight sampling (DWS)

Our proposed DWS adaptively samples point clouds of different categories based on the real-time prediction performance of each category during model training. In DWS, a weight, denoted as $e_i \in \mathbb{R}$, is assigned to every point p_i within the point cloud $P = \{p_i \in \mathbb{R}^3 | i = 1, 2, \ldots, n\}$ of the training set. The objective of e_i is to determine the weight of each sample, initially randomized with small variance during the initial stages of the training process. In each sampling iteration, we select the point p_m based on the maximum weight, $e_m = \max\{e_i | i = 1, 2, \ldots, n\}$. Subsequently, we extract a set of neighboring points $G = \{p_i \in P | \|p_i - p_m\| < r\}$ centered at p_m with radius r as the sampling result. The indices of points in G are denoted as $I = \{i | p_i \in G\}$. After each sampling, we decrease $\{e_i | i \in I\}$ by Δe_i, which can be calculated by:

$$\Delta e_i = \frac{r \cdot w(d)}{\|p_i - p_m\|} \tag{1}$$

where $w(d)$ is the dynamic weight determined based on the category indicator gap d, which is calculated as follows:

$$w(d) = (1 - \alpha) + \alpha \cdot \max(0, 1 - d) \tag{2}$$

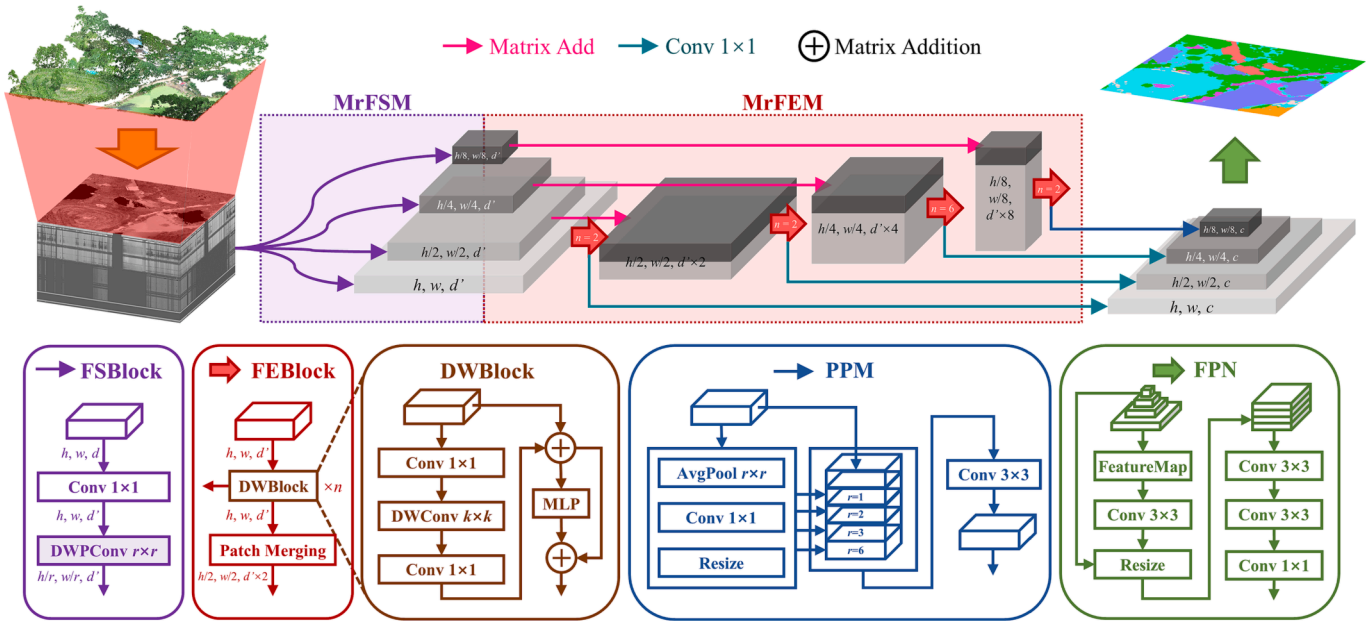


Fig. 4. The overall architecture of MrFSNet proposed in this study. This deep neural network consists of MrFSM, MrFEM, and other modules, which could input high-channel LiDAR-derived images, and output pixel-wise classification results.

where hyperparameter $\alpha \in [0, 1]$ controls the maximum sampling multiplicity of DWS for the areas containing low indicator categories, and the maximum sampling multiplicity is $\frac{\alpha}{(1-\alpha)} + 1$. The category indicator gap d measures the difference between the highest category indicator among all categories and the lowest category indicator within each sample, as shown in the formula below:

$$d = \max(\{c_l | l \in L\}) - \min(\{c_l | l \in L^{'}\}) \tag{3}$$

where L and $L^{'}$ are the set of categories in the dataset and the categories that are present in the current batch G . c_l is the historical average F1-score for each category on the validation set, updated after each epoch.

3.3.2. DWS + KPConv 3D classification model

In order to apply DWS in 3D vegetation mapping, we combined our DWS with KP-FCNN based on KPConv (Thomas et al., 2019) and constructed a point-wise classification method DWS-KP-FCNN, as displayed in Fig. 5. KP-FCNN employs an Encoder-Decoder architecture that utilizes KPConv to extract point cloud features and aggregates them using global average pooling. The decoder is trained to obtain point-wise features using skip connections and neighborhood upsampling. In addition, DWS adjusts the weights of the difficult-to-identify categories based on the model's performance metric, which leads to more frequent sampling of areas containing such categories. This approach effectively improves the model's performance for vegetation classification in point clouds with category imbalance.

We implemented DWS-KP-FCNN using PyTorch and trained it on the point cloud dataset. DWS was applied with α = 0.95, sampling point clouds of 15 m radius for training. For comparison, the original KP-FCNN utilized category weighting to address category imbalances. To mitigate overfitting, training samples were subjected to random horizontal rotation, random horizontal flip, random scaling, and random position noise up to 0.001 m. We optimized the weighted cross-entropy loss using momentum gradient descent with an average batch size of 2. The initial learning rate was set to 0.001, and we employed an exponentially decaying learning rate with a decay factor of 10 every 50,000 iterations. The model achieved convergence after 150,000 iterations.

3.4. Classification scheme

To comprehensively evaluate and compare the vegetation mapping methods, experiments were conducted on MrFSNet, DWS-KP-FCNN, and traditional 2D and 3D deep learning methods in three natural scenes. The classification experiment is comprised of two parts, as outlined in

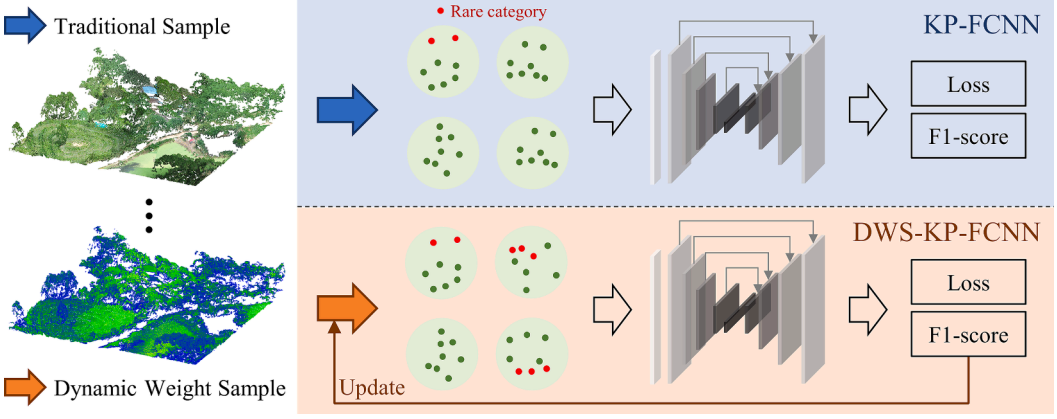


Fig. 5. Diagram of DWS-KP-FCNN and traditional KP-FCNN.

Table 3
Classification scheme of vegetation mapping on each scene in this study.

Data format	Label type	Method type	Method
Point cloud	Point-wise label	3D mapping	KP-FCNN
			DWS-KP-FCNN (Ours)
Raster images	Pixel-wise label	2D mapping	Swin Transformer
			DWNet
			MrFSNet (Ours)

Table 3: (1) The original KP-FCNN (Thomas et al., 2019) and our DWS-KP-FCNN were chosen to perform 3D classification of vegetation using point cloud datasets; (2) Swin Transformer (Liu et al., 2021b), DWNet (Han et al., 2022), and our MrFSNet were chosen to perform 2D classification of vegetation using raster datasets. All models were trained and evaluated on a workstation equipped with two NVIDIA Tesla T4 GPUs with 16 GB of memory each.

3.5. Accuracy assessment

This study evaluates the classification performance of different models using Overall Accuracy (OA), F1-score, and mean F1-score (mF1). F1-score is used to measure the classification performance of each category, which is a combination of precision and recall and is very suitable for natural scenes with category imbalance. OA and mF1 are used to evaluate the overall performance of the model. OA indicates the proportion of overall correctly classified samples, and mF1 is the average F1-score of each category. mF1 is more suitable for tasks with category imbalance. Even if a category only makes up a small portion of the test set, it is obviously reflected in the mF1 when that category performs poorly. The formula for F1-score is as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{F1} - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

where TP, FP and FN are the numbers of true positive, false positive, and false negative sample points for each category in the test dataset, respectively.

4. Results

4.1. 2D vegetation species mapping

Table 4 presents a comparison of three deep learning methods (MrFSNet, DWNet, and Swin Transformer) across three scenes. The proposed MrFSNet outperforms other methods in terms of mF1 and OA metrics. It achieved F1-score for the karst wetland, mangrove forest, and the hill forest scene that were 18.46 %, 5.92 %, and 1.66 % higher than those of the Swin Transformer, and 5.63 %, 1.82 %, and 1.50 % higher than those of DWNet, respectively. The most significant improvement was observed in the SM and BS categories of karst wetlands, where training samples are limited, reflecting the stronger generalization ability of MrFSNet.

Localized visual comparison results of these 3D classification methods are presented in Fig. 6. In the karst wetland and hill forest scenes, the original LiDAR point cloud data were missing in clear water due to laser absorption by the water, leading to Swin Transformer misclassifying WT as AV in the karst wetland scene, and DWNet misclassifying WT as GS in the hill forest scene. In contrast, MrFSNet correctly identified the missing areas in both scenes. Moreover, MrFSNet achieved the highest accuracy in identifying BS in the karst wetland, outperforming other methods by more than 41.95 % of F1-score. In the mangrove forest scene, the three methods had similar vegetation classification performance due to the concentrated vegetation. However, for scattered ARs, MrFSNet showed an F1-score over 10.59 % higher than the other methods. The results presented above demonstrate that our proposed MrFSNet performed well with height-channel raster LiDAR images in 2D vegetation mapping.

4.2. 3D vegetation species mapping

We evaluated the impact of the DWS strategy on 3D vegetation classification accuracy by comparing the performance of DWS-KP-FCNN with KP-FCNN on point cloud datasets. The results, shown in Table 5, compare the performance of both models for each category in three scenes, with the better performance highlighted in bold. DWS-KP-FCNN outperformed original KP-FCNN in terms of mF1 in all scenes, including karst wetland (+1.36 %), mangrove forest (+1.28 %), and hill forest (+2.80 %). The most significant improvement of the DWS-KP-FCNN method was observed in the hill forest scene with the most category imbalance. The BD category, which had the lowest percentage of point clouds (0.46 %), experienced the most significant increase of 7.83 % in F1-score. The RD category, which had the lowest original F1-score, saw an increase of 7.02 %, while the AE category, which was unevenly distributed, experienced an increase of 2.94 %. These results indicate that DWS can significantly enhance the classification performance of rare and challenging categories.

Fig. 7 presents the localized vegetation classification results for the three scenes obtained from both 3D methods based on point clouds. DWS-KP-FCNN showed improved accuracy in identifying BS and produced fewer misclassifications in the karst wetland scene. Similarly, DWS-KP-FCNN demonstrated better classification of AR without confusing AC with DT in the mangrove forest scene. In the hill forest scene, KP-FCNN misclassified parts of the canopy structures of AE as NA, but DWS-KP-FCNN correctly separated the two types of trees. The results demonstrate that DWS-KP-FCNN outperforms KP-FCNN in vegetation classification, providing further evidence of the effectiveness of DWS.

Table 4
Quantitative results (%) of 2D vegetation classification for 2D methods.

Scene	Method	Category F1-score								mF1	OA
Karst wetland		GR	SH	WT	CL	AV	SM	BS	AR		
	Swin Transformer	85.16	73.31	81.74	92.72	94.70	0.00	0.00	79.87	63.44	88.54
	DWNet	89.01	**77.42**	**91.39**	94.30	**96.49**	72.15	7.11	82.28	76.27	91.27
	MrFSNet	**89.35**	77.02	89.85	**94.73**	96.17	**75.72**	**49.06**	**83.27**	**81.90**	**91.49**
Mangrove		MR	AR	SA	DT	AC	AM				
forest	Swin Transformer	98.85	14.40	92.06	90.47	88.36	94.37			79.75	93.47
	DWNet	99.22	32.68	94.35	**91.41**	90.05	**95.37**			83.85	94.64
	MrFSNet	**99.43**	**43.27**	**94.56**	91.34	**90.09**	95.34			**85.67**	**94.75**
Hill		RD	BD	GS	BL	NA	YE	AE	WT		
forest	Swin Transformer	80.00	84.25	88.92	94.30	96.61	81.32	91.98	90.88	88.53	91.38
	DWNet	80.04	**86.36**	88.73	88.43	97.19	85.84	92.33	90.61	88.69	91.55
	MrFSNet	**80.67**	85.68	**90.47**	**94.89**	**97.30**	**87.93**	**92.90**	**91.66**	**90.19**	**92.79**

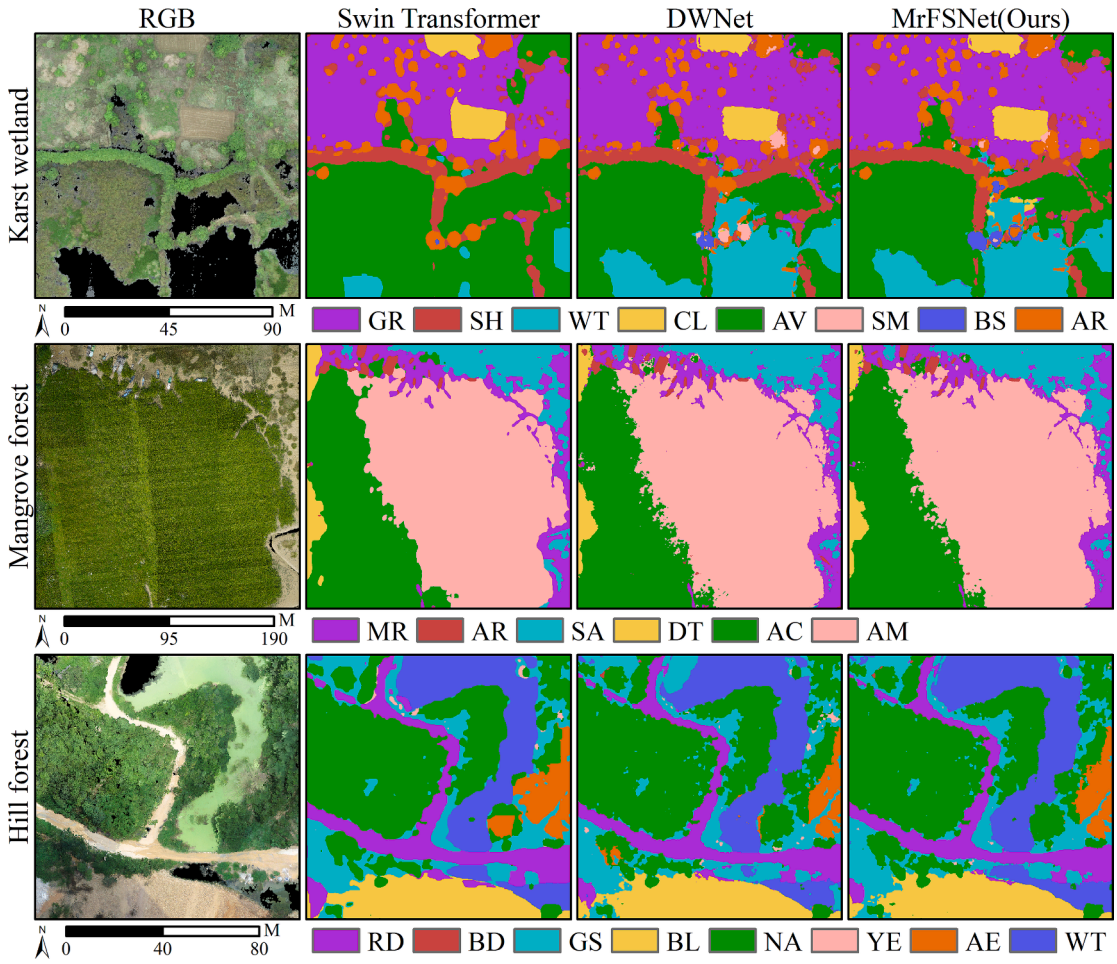


Fig. 6. Localized 2D classification results of Swin Transformer, DWNet, and our proposed MrFSNet model on three scenes.

Table 5
Quantitative results (%) of 3D vegetation classification for 3D methods.

Scene	Method	Category F1-score								mF1	OA
Karst wetland		GR	SH	WT	CL	AV	SM	BS	AR		
	KP-FCNN	91.83	79.15	88.55	**95.17**	96.89	**95.23**	72.58	87.95	88.42	92.47
	DWS-KP-FCNN	**92.01**	**81.60**	**94.14**	94.83	**97.33**	95.00	**74.80**	**88.56**	**89.78**	**93.01**
Mangrove		MR	AR	SA	DT	AC	AM				
forest	KP-FCNN	98.87	83.34	91.01	86.37	89.50	**96.73**			90.97	94.80
	DWS-KP-FCNN	**99.11**	**85.13**	**92.93**	**89.79**	**89.92**	96.63			**92.25**	**95.04**
Hill		RD	BD	GS	BL	NA	YE	AE	WT		
forest	KP-FCNN	66.27	86.46	94.77	93.58	98.08	93.83	94.52	86.54	89.25	95.08
	DWS-KP-FCNN	**73.29**	**94.29**	**96.10**	**93.78**	**98.92**	**94.70**	**97.46**	**87.88**	**92.05**	**96.63**

4.3. Comparison of 2D and 3D mapping using deep learning

To effectively evaluate the differences between 2D and 3D classification methods in vegetation classification, we projected the point cloud classification results of 3D methods into the 2D raster and compared them with the raster classification results of 2D methods. Table 6 compares the pixel-based vegetation classification performance of MrFSNet and DWS-KP-FCNN, which are the two methods demonstrating the highest performance in 2D and 3D classification, respectively. The results indicate that 3D classification achieved mF1 3.43 %–8.62 % higher than 2D classification. DWS-KP-FCNN outperformed MrFSNet in most categories (19/22), with the largest increase observed in AR (47.27 %) in the mangrove forest scene, followed by BS (28.68 %) and SM (19.40 %) in the karst wetland scene.

Fig. 8 presents the localized 2D vegetation classification results of MrFSNet and DWS-KP-FCNN for three scenes. The black areas in the figure correspond to missing data in the original point clouds, which is not classified by the 3D classification method since it only operates on the acquired point clouds. However, DWS-KP-FCNN performed better in distinguishing BS and AR in the mangrove forest scene and accurately identifying vegetation distribution at the junction of two vegetation types in the karst wetland scene. Unlike the urban situation (Hu et al., 2022), 3D method has lower accuracy in classifying RD in the hill forest scene because gravel RD has a different morphology than concrete RD in cities, with irregular shapes and interlacing with GS. However, DWS-KP-FCNN more accurately identified the RD under the canopy of trees and precisely determined the canopy boundaries of AE. These results suggest that using direct 3D point cloud classification methods provides better performance than 2D raster classification methods, due to more accurate classification of tree species and category boundaries.

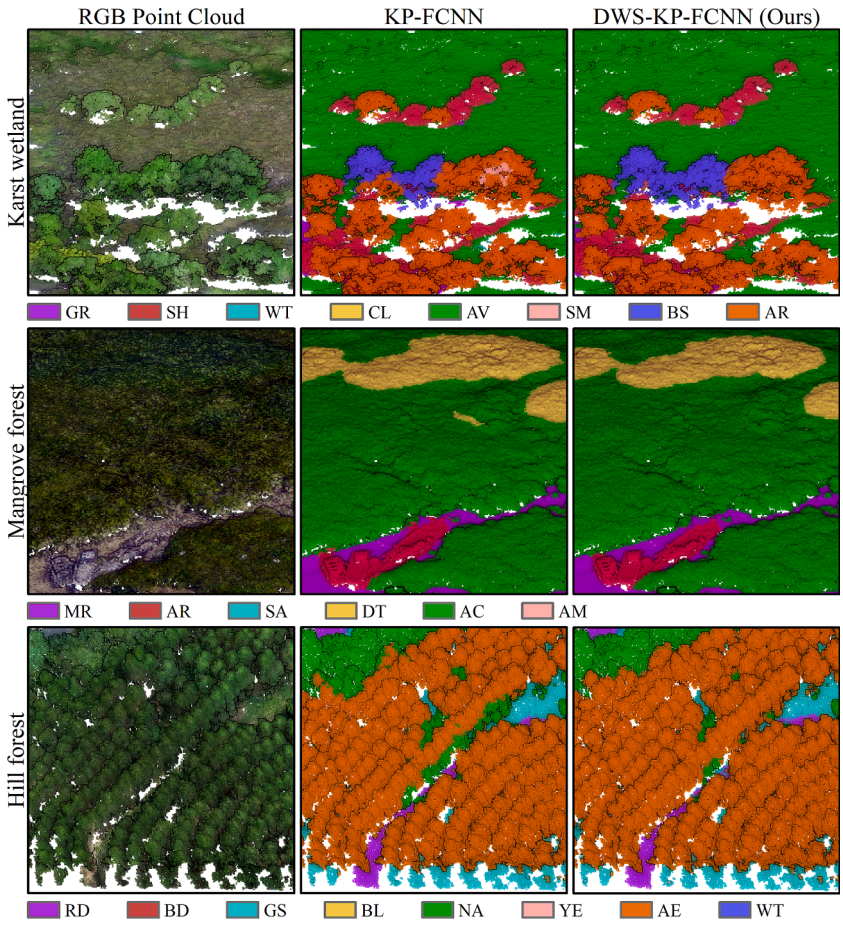


Fig. 7. Localized 3D classification results of KP-FCNN and our proposed DWS-KP-FCNN on three scenes.

Table 6
Performance comparison (%) of 2D vegetation classification for MrFSNet (2D) and DWS-KP-FCNN (3D).

Scene	Method	Category F1-score								mF1	OA
Karst wetland		GR	SH	WT	CL	AV	SM	BS	AR		
	MrFSNet	89.35	77.02	89.85	94.73	96.17	75.72	49.06	83.27	81.90	91.49
	DWS-KP-FCNN	**91.82**	**82.84**	**95.09**	**94.88**	**97.56**	**95.12**	**77.74**	**89.11**	**90.52**	**93.77**
Mangrove		MR	AR	SA	DT	AC	AM				
forest	MrFSNet	99.43	43.27	**94.56**	**91.34**	90.09	95.34			85.67	94.75
	DWS-KP-FCNN	**99.44**	**90.54**	92.10	90.99	**90.81**	**95.99**			**93.31**	**95.28**
Hill		RD	BD	GS	BL	NA	YE	AE	WT		
forest	MrFSNet	**80.67**	85.68	90.47	94.89	97.30	87.93	92.90	91.66	90.19	92.79
	DWS-KP-FCNN	77.70	**95.57**	**95.05**	**95.80**	**99.05**	**96.06**	**97.99**	**91.71**	**93.62**	**95.74**

5. Discussion

5.1. Effect of multi-resolution feature selection on 2D vegetation mapping

To investigate the reasons for the varying classification performance of 2D deep learning methods, we summarize the structural differences between MrFSNet, Swin Transformer, and DWNet in Table 7. Our proposed MrFSNet features the following: (1) employing multi-resolution feature selection instead of patch embedding to encode the original data; (2) incorporating input features from each scale that encompass the features obtained at the previous scale and feature combinations chosen by MrFSM for that scale from the original data; and (3) using lightweight and easily trainable DWConv as the primary feature extractor in the encoder, instead of the multi-head self-attention model.

Given the severe category imbalance in natural vegetation scenes, rare categories often have limited training data and are prone to overfitting. Therefore, using lightweight and easily trainable DWConv as the primary feature extractor in the encoder is preferable to the multi-head self-attention model (Howard et al., 2017). This is evident in our results, which show that DWNet and MrFSNet outperformed Swin Transformer in F1-score for rare categories that accounted for less than 1 % of the raster datasets in all three scenes. Specifically, in the karst wetland scene, SM and BS vegetation species only account for 0.35 % and 0.38 % of the raster dataset, respectively. In this case, Swin Transformer failed to recognize both vegetation species due to limited training samples, resulting in their F1-score of 0. As a result, DWNet and MrFSNet achieve F1-score 7.11 %–75.72 % higher than Swin Transformer. In the mangrove forest scene, the AR category accounts for 0.67 %, and both DWNet and MrFSNet show an improvement of 18.28 %–28.87 % in F1-score. Similarly, in the hill forest scene, the BD category accounts for 0.89 %, and both DWNet and MrFSNet show an improvement of 1.43 %–2.11 % in F1-score. These results indicate that the use of

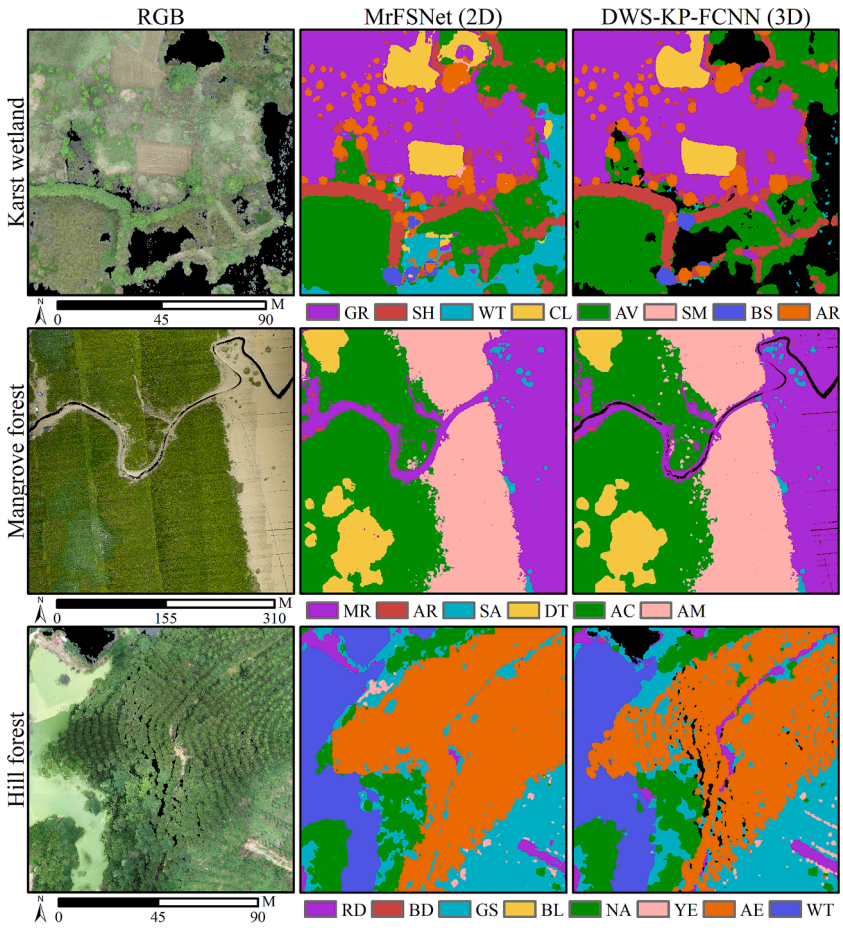


Fig. 8. Raster-based comparison of MrFSNet (2D) and DWS-KP-FCNN (3D) for localized vegetation classification results.

Table 7
Structural differences between Swin Transformer, DWNet and MrFSNet.

Method	Data embedding		Layer feature source		Primary feature extractor		Parameters	Trainingtime
	Padding	Feature selection	Last layer	Feature selection	Self attention	DWConv		
Swin Transformer	√	×	√	×	√	×	59 M	26 h
DWNet	√	×	√	×	×	√	56 M	4 h
MrFSNet	×	√	√	√	×	√	55 M	4 h

lightweight and easily trainable DWConv as the primary feature extractor in the encoder can help to improve classification performance for rare categories with limited training samples in vegetation scenes.

Swin Transformer benefits from the flexibility of the Transformer by combining a moving window and applying self-attention to non-overlapping local windows, allowing the encoder to obtain cross-window information to facilitate the application of global information (Liu et al., 2021b). In contrast, the DWConv (Han et al., 2022) with a small stride and large convolution kernel does not entirely replace this structure and thus suffers from a deficiency in the extraction of global information, particularly in the hill forest scene. The classification of flat ground categories (BL, WT, and GS) in hill forest scenes depends on the spatial distribution of the surrounding environment, as the color, elevation, and curvature features may not be apparent in small areas. Swin Transformer achieved a higher F1-score than DWNet for these categories, due to Swin Transformer's attention mechanism being better suited for capturing spatial dependencies (Fan et al., 2022). However, MrFSNet achieved a 1.05 %–6.46 % increase in F1-score compared to DWNet for these categories and correctly predicted the missing areas in the original point cloud by context. This indicates that the multi-scale sampling capability of MrFSNet compensates for the deficiencies of DWNet in capturing spatial dependencies and context, resulting in a higher F1-score for these categories.

To capture information across multiple scales, MrFSNet employs a set of FSBlocks, with each block dedicated to a particular scale. Within each FSBlock, a unary convolution is used to select and combine the most important image features, which can be estimated by the absolute values of the corresponding weights in the convolution layer (Ye et al., 2021). This approach enables MrFSNet to estimate the relative importance of features for vegetation classification at different scales. The top 3 most important features at each scale of MrFSNet in the hill forest scene are summarized in Table 8. The top 3 most important features at the first scale (0.25 m resolution) are spectral features; at the second and third scales (0.5 m and 1 m resolution), they include spectral features and forest metrics; at the fourth scale (2 m resolution), they are mainly intensity features and height features. It reveals that different features perform different functions. MrFSNet mainly used spectral features for classified local details, while intensity and height features were used for recognized global information.

Table 8
Summary of the top 3 features with the highest importance at each scale of MrFSNet.

Resolution	Top 1 feature	Top 2 feature	Top 3 feature
0.25 m × 0.25 *m*	Green band	Red band	Blue band
0.5 m × 0.5 *m*	Blue band	Canopy cover	Gap fraction
1 m × 1 *m*	Blue band	Canopy cover	Leaf area index
2 m × 2 *m*	Intensity kurtosis	Elevation canopy relief ratio	Elevation percentile 90th

5.2. Effects of sampling methods on 3D vegetation mapping

In the point cloud of natural scene, category imbalance significantly affects vegetation mapping. The classification accuracy of categories with a relatively small proportion is typically poor. Thus, the average logarithm of the proportion of each category is employed as a Category Imbalance Metric (CIM) to quantify the degree of category imbalance in each scene:

$$\mathrm{CIM} = \frac{1}{n}\sum_{i=1}^{n}\log_{1/n}(p_i) \tag{7}$$

where n be the number of categories,p_i be the percentage of labels in category i. CIM value is from 1 to infinity. CIM = 1 represents a completely balanced distribution, which each category displays an equal proportion. The larger the CIM, the more significant the category imbalance of the scene. In the karst wetland, mangrove forest, and hill forest scenes, the CIM values are 1.42, 1.40, and 1.57, respectively. DWS-KP-FCNN outperforms KP-FCNN by 1.36 %, 1.28 %, and 2.80 % in these scenes. Our analysis reveals that the improvement of DWS-KP-FCNN over KP-FCNN is positively correlated with the CIM value, which reflects the degree of category imbalance. This implies that the performance improvement of DWS is more significant in highly imbalanced scenes.

We have conducted an in-depth analysis of the DWS in the hill forest scene with the highest CIM. Table 9 displays the distribution of sampled points for each category with and without the DWS. The traditional sampling method resulted in the largest proportion of sampling point clouds belonging to the NA category, accounting for 67.68 % of the total sample points. Conversely, the BD and WT categories had the smallest proportions, representing only 0.41 % and 0.34 %, respectively. After applying DWS, the decrease in the largest sampling proportion was limited to 4.83 %, while the increase in the smallest sampling proportions was merely 0.15 % and 0.35 %. Despite this adjustment, the category distribution still exhibits significant disparity. Unlike traditional over-sampling or under-sampling techniques used in sample-by-sample classification tasks (Azadbakht et al., 2018; Roy et al., 2022), our DWS approach does not aim to equalize the distribution of sampled points for each category. This is because in the semantic segmentation task, multiple nearby points are classified simultaneously. Consequently, DWS retains a similar category distribution to that of the traditional sampling strategy.

Due to the high information density of the point cloud and the limitation of GPU memory size, the 3D model can only input a maximum of 80,000 points for each training iteration, which accounts for only 0.47 % of the training set. We define the average training interval (ATI) as the number of times a category participates in training divided by the total number of training steps. A high ATI indicates the model may fail to learn certain category samples during multiple training iterations, leading to unstable training and slow convergence. Table 10 summarizes the ATI for each category in the first 25,000 iterations trained by the model, both before and after applying DWS. Initially, the average ATI was 7.27 using the original sampling method. The category with the lowest proportion in the dataset, BD, had an ATI of approximately 11, while the category with an extremely uneven distribution in the study area, AE, had an ATI of approximately 20. However, after applying DWS, we observed a decrease in ATI for all categories except NA (with the lowest original ATI). The extent of the reduction in ATI varied across categories, with a more substantial decrease observed for those with higher original ATI. As a result, the average ATI was reduced by 49.5 % to 3.67. In addition, we observed an inverse relationship between ATIs and final F1-scores for each category. Excluding the atypical RD category, a strong negative correlation was found with a correlation coefficient of −0.83 for KP-FCNN, −0.89 for DWS-KP-FCNN, and −0.82 in total (p-value < 0.001), as displayed in Fig. 9. It indicates that DWS can improve classification accuracy by reducing the ATI during training, and this approach has strong potential for application in point clouds with imbalanced and unevenly distributed categories.

5.3. Comparative analysis of 3D and 2D vegetation species mapping

Previous studies have shown that including LiDAR features can enhance vegetation classification performance (Anderson et al., 2018). This study further demonstrates that 3D mapping using point clouds directly yield superior classification performance in vegetation scenes compared to 2D mapping using raster LiDAR features derived from point clouds. The 3D methods achieved an average 10.18 % higher mF1 score in vegetation classification compared to the 2D methods. This approach maximizes the utilization of LiDAR data and is particularly well-suited for identifying various types of vegetation in scenarios with stratified canopy structures.

Converting point clouds into LiDAR metrics presents a scaling dilemma akin to the challenge of segmenting rasters into objects (Fu et al., 2022), where larger object scales may result in different categories

Table 9
The proportion (%) of sampling point clouds for each category using DWS and traditional sampling strategies.

Sampling strategy	RD	BD	GS	BL	NA	YE	AE	WT
Traditional	1.45	0.41	22.25	4.36	67.68	2.19	1.32	0.34
DWS	2.01	0.56	24.88	4.07	62.84	3.33	1.84	0.69

Table 10
Average training interval (ATI) for each category using DWS and traditional sampling strategy (average of the first 25,000 iterations).

Sampling strategy	RD	BD	GS	BL	NA	YE	AE	WT	Average
Traditional	2.81	11.01	1.02	5.63	**1.01**	6.40	9.90	20.41	7.27
DWS	**1.80**	**5.42**	**1.01**	**5.00**	1.02	**3.22**	**3.50**	**8.37**	**3.67**

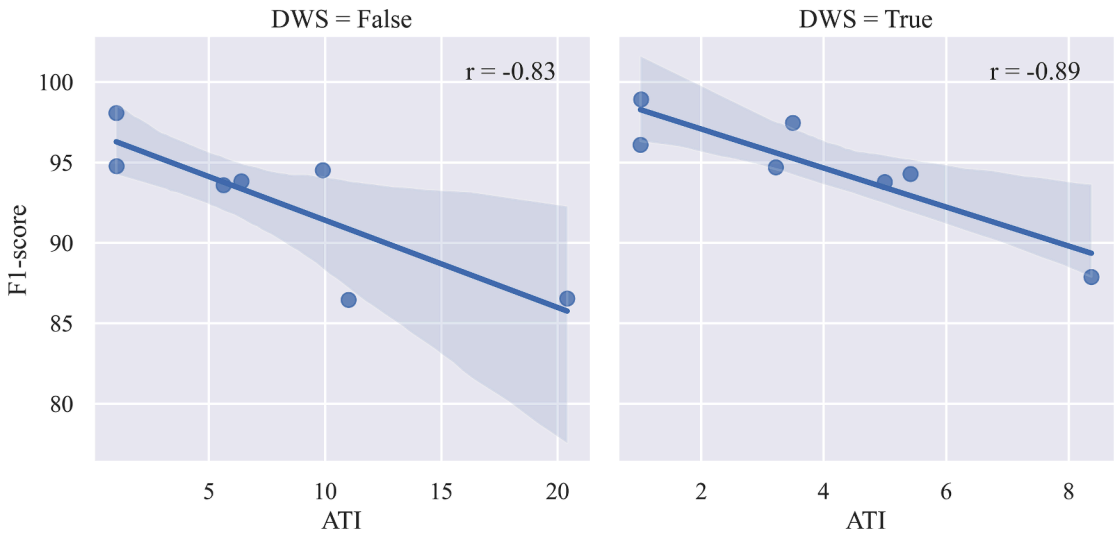


Fig. 9. Scatterplots between F1-score and ATI.

being combined, while smaller scales may cause feature heterogeneity. Although deep learning can overcome the heterogeneity of some features by learning to combine multiple pixels, statistical features such as kurtosis and skewness cannot be directly computed, resulting in the loss of accurate statistical information. However, this issue can be circumvented by directly inputting the original point cloud into the neural network for classification, enabling accurate statistical information to be preserved.

Furthermore, in 2D mapping, height is considered as a feature. The coverage of convolution and pooling operations in deep learning model

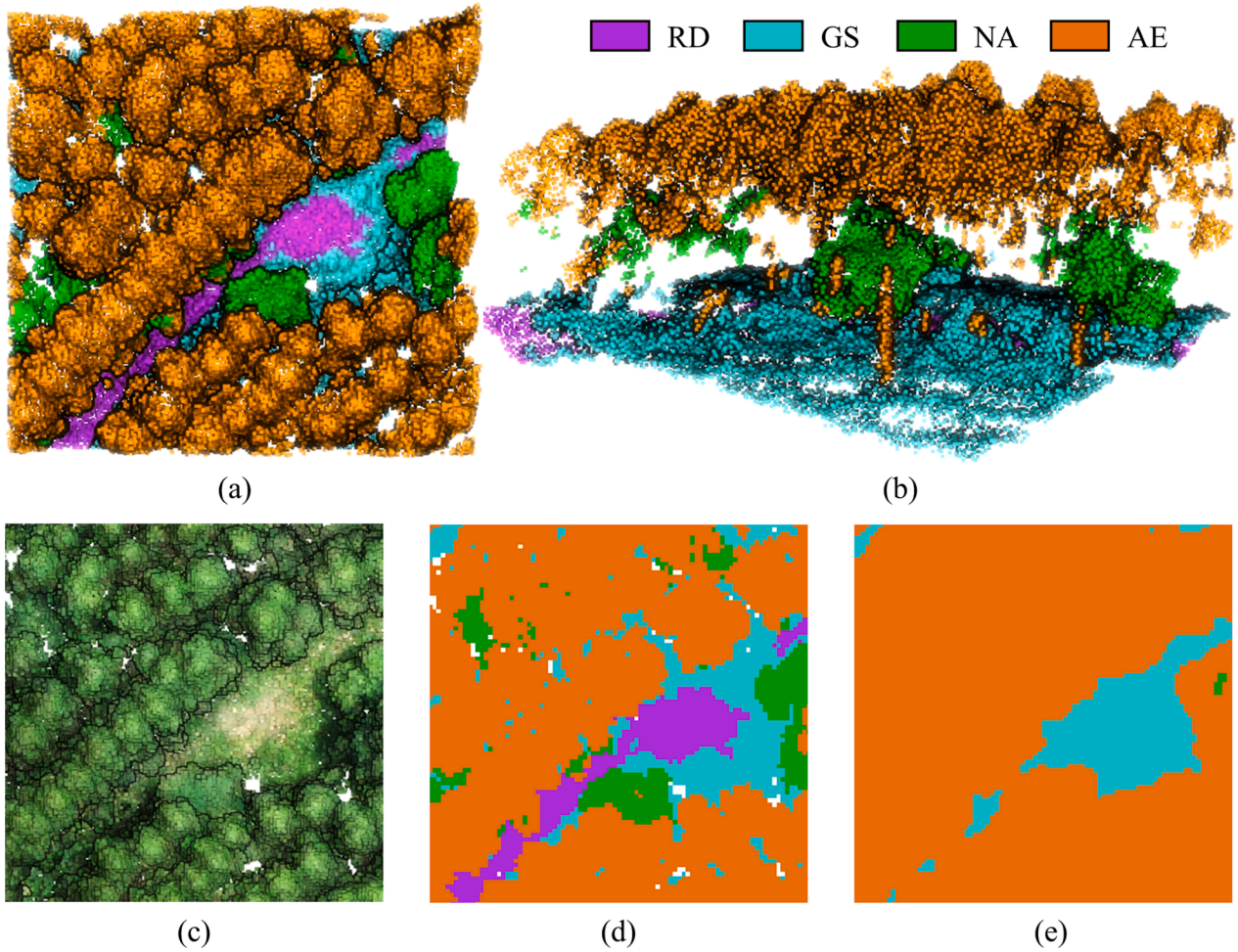


Fig. 10. Different classification results in same region. (a) and (b) are top view and side view of 3D point-wise result of DWS-KP-FCNN (3D). (c) is the RGB view of this region. (d) is 2D pixel-wise result projected from 3D point-wise result of DWS-KP-FCNN. (e) is 2D pixel-wise result of MrFSNet (2D).

is contingent solely upon the 2D pixel distances. In contrast, 3D deep learning with KPConv treats height as a dimension and computes convolution ranges in 3D space. Therefore, 2D convolution uniformly applies initial convolution across various heights, failing to distinguish between different heights at the feature dimension level, thus constraining its ability to utilize height information. In contrast, KPConv performs initial local convolutions individually for different height levels to generate multi-level features. Then KPConv gradually extracts and fuses features from different heights. This approach reduces interference from varying heights during initial feature extraction, enabling the model to distinguish category boundaries at different heights. This also facilitates the consideration of interactions between different heights, enhancing the effectiveness of the 3D classification method.

To better understand the differences between 2D and 3D vegetation classification, the 3D method DWS-KP-FCNN and the 2D method MrFSNet were used to predict the point cloud of a section of an artificial eucalyptus forest, as shown in Fig. 10. The forest comprises three vegetation categories, namely AE, NA, and GS, arranged in a complex structure from top to bottom. 3D methods can effectively classify the complex, multi-layered vegetation in the forest, achieving high F1-scores of 97.46 %, 98.92 %, and 96.10 % for these three categories, respectively. The point-wise classification results allow for observation of the stratified structure within the forest, even beneath the dense canopy. Projecting this stratified 3D structure into 2D, vegetation with varying heights is distinctly segmented, resulting in the 3D classification method having clearer and more precise classification boundaries for BD, NA, and AE in the hill forest scene.

In this study, 3D mapping also has several limitations when compared to 2D mapping. 3D mapping methods often depend on the expensive LiDAR point cloud. Annotating point clouds at a point-wise level is not only time-consuming but also need to cost a lot of labor. In the training process, the 2D models presented in this paper include over 55 million parameters, which is twice the count of the 3D model, but the GPU memory requirements for 3D mapping are higher due to the intermediate state of the point cloud. Furthermore, the 2D method of DWConv demonstrates a sixfold faster convergence rate compared to the 3D methods. Consequently, 3D mapping is suitable for high-precision, multi-layer, and fine-grained mapping of natural scenes, enabling the acquisition of detailed 3D point-wise classification results. Nonetheless, it's crucial to acknowledge that 3D mapping is associated with higher costs and increased time investment.

6. Conclusion

Vegetation species 2D and 3D mapping is a crucial technical task for presenting an inventory of existing plant communities, their locations, and their geographical distribution in the landscape. In this study, our work pioneers the demonstration of the applicability of 3D deep learning vegetation species mapping in natural scenes and the comparison of classification performance in vegetation species 2D and 3D mapping. Our experiments in three distinct natural scenarios reveal significant advantages of 3D mapping in comparison to 2D mapping. These advantages are primarily evident in higher classification metrics, more accurate edge classification, and more valuable 3D point-wise classification outcomes. In addition, the multi-resolution feature selection method in MrFSNet proposed in this study could adaptively select optimal features at four different scales, fuse the multi-scale information by adding desired features scale-by-scale, and perform high-accuracy 2D mapping with the limited training data. The DWS method effectively promotes model training by reducing the average training interval for rare categories, thereby improving 3D mapping performance in three natural scenes with category imbalance.

This study exclusively relies on UAV-LiDAR point clouds as the sole data source, which imposes limitations on accurately classifying finer vegetation species. To overcome this limitation, our future research aims to explore the integration of hyperspectral data and LiDAR point clouds within a 3D framework to enable 3D mapping of ten or more vegetation categories within natural scenes. Furthermore, the successful point-wise classification achieved through 3D mapping of natural scenes inspires us to utilize the 3D mapping results for quantitative analysis of large-scale vegetation structure parameters.

CRediT authorship contribution statement

Liwei Deng: Conceptualization, Methodology, Software, Validation, Writing – original draft. **Bolin Fu:** Conceptualization, Writing – review & editing, Funding acquisition, Supervision. **Yan Wu:** Data curation, Investigation. **Hongchang He:** Data curation, Resources. **Weiwei Sun:** Data curation, Investigation. **Mingming Jia:** Data curation, Investigation. **Tengfang Deng:** Data curation, Investigation. **Donglin Fan:** Data curation, Investigation.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant number 42371341, 42004006, 21976043), the Innovation Project of Guangxi Graduate Education (Grant Number YCSW2022328), and the "BaGui Scholars" program of the provincial government of Guangxi (Grant Number 2019A30), and in part by Zhejiang Province "Pioneering Soldier" and "Leading Goose" R&D Project under Grant 2023C01027, the Guilin University of Technology Foundation (Grant number GUTQDJJ2017096).

References

Adam, J.M., Liu, W., Zang, Y., Afzal, M.K., Bello, S.A., Muhammad, A.U., Wang, C., Li, J., 2023. Deep learning-based semantic segmentation of urban-scale 3D meshes in remote sensing: A survey. Int. J. Appl. Earth Obs. Geoinf. 121, 103365 https://doi.org/10.1016/j.jag.2023.103365.

Anderson, K.E., Glenn, N.F., Spaete, L.P., Shinneman, D.J., Pilliod, D.S., Arkle, R.S., McIlroy, S.K., Derryberry, D.R., 2018. Estimating vegetation biomass and cover across large plots in shrub and grass dominated drylands using terrestrial lidar and machine learning. Ecol. Indic. 84, 793–802. https://doi.org/10.1016/j.ecolind.2017.09.034.

Azadbakht, M., Fraser, C.S., Khoshelham, K., 2018. Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform LiDAR data. Int. J. Appl. Earth Obs. Geoinf. 73, 277–291. https://doi.org/10.1016/j.jag.2018.06.009.

Calders, K., Adams, J., Armston, J., Bartholomeus, H., Bauwens, S., Bentley, L.P., Chave, J., Danson, F.M., Demol, M., Disney, M., Gaulton, R., Krishna Moorthy, S.M., Levick, S.R., Saarinen, N., Schaaf, C., Stovall, A., Terryn, L., Wilkes, P., Verbeeck, H., 2020. Terrestrial laser scanning in forest ecology: Expanding the horizon. Remote Sens. Environ. 251, 112102 https://doi.org/10.1016/j.rse.2020.112102.

Campbell, M.J., Eastburn, J.F., Mistick, K.A., Smith, A.M., Stovall, A.E.L., 2023. Mapping individual tree and plot-level biomass using airborne and mobile lidar in piñon-juniper woodlands. Int. J. Appl. Earth Obs. Geoinf. 118, 103232 https://doi.org/10.1016/j.jag.2023.103232.

Cao, J., Liu, K., Zhuo, L., Liu, L., Zhu, Y., Peng, L., 2021. Combining UAV-based hyperspectral and LiDAR data for mangrove species classification using the rotation forest algorithm. Int. J. Appl. Earth Obs. Geoinf. 102, 102414 https://doi.org/10.1016/j.jag.2021.102414.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp. 801-818. https://doi.org/10.1007/978-3-030-01234-2_49.

Dersch, S., Heurich, M., Krueger, N., Krzystek, P., 2021. Combining graph-cut clustering with object-based stem detection for tree segmentation in highly dense airborne lidar point clouds. ISPRS J. Photogramm. Remote Sens. 172, 207–222. https://doi.org/10.1016/j.isprsjprs.2020.11.016.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929. https://doi.org/10.48550/arXiv.2010.11929.

Du, L., Pang, Y., Wang, Q., Huang, C., Bai, Y., Chen, D., Lu, W., Kong, D., 2023. A LiDAR biomass index-based approach for tree- and plot-level biomass mapping over forest farms using 3D point clouds. Remote Sens. Environ. 290, 113543 https://doi.org/10.1016/j.rse.2023.113543.

Fan, Z., Hu, G., Sun, X., Wang, G., Dong, J., Su, C., 2022. Self-attention neural architecture search for semantic image segmentation. Knowledge-Based Systems 239, 107968. https://doi.org/10.1016/j.knosys.2021.107968.

Fu, B., Liu, M., He, H., Lan, F., He, X., Liu, L., Huang, L., Fan, D., Zhao, M., Jia, Z., 2021. Comparison of optimized object-based RF-DT algorithm and SegNet algorithm for classifying Karst wetland vegetation communities using ultra-high spatial resolution UAV data. Int. J. Appl. Earth Obs. Geoinf. 104, 102553 https://doi.org/10.1016/j.jag.2021.102553.

Fu, B., He, X., Yao, H., Liang, Y., Deng, T., He, H., Fan, D., Lan, G., He, W., 2022. Comparison of RFE-DL and stacking ensemble learning algorithms for classifying mangrove species on UAV multispectral images. Int. J. Appl. Earth Obs. Geoinf. 112, 102890 https://doi.org/10.1016/j.jag.2022.102890.

Fu, B., Liang, Y., Lao, Z., Sun, X., Li, S., He, H., Sun, W., Fan, D., 2023. Quantifying scattering characteristics of mangrove species from Optuna-based optimal machine learning classification using multi-scale feature selection and SAR image time series. Int. J. Appl. Earth Obs. Geoinf. 122, 103446 https://doi.org/10.1016/j.jag.2023.103446.

Girardeau-Montaut, D., 2022. CloudCompare - Open Source Project.

Guo, B., Deng, L., Wang, R., Guo, W., Ng, A.-H.-M., Bai, W., 2023. MCTNet: Multiscale Cross-Attention-Based Transformer Network for Semantic Segmentation of Large-Scale Point Cloud. IEEE Trans. Geosci. Remote Sens. 61, 1–20. https://doi.org/10.1109/TGRS.2023.3322579.

Han, X., Dong, Z., Yang, B., 2021. A point-based deep learning network for semantic segmentation of MLS point clouds. ISPRS J. Photogramm. Remote Sens. 175, 199–214. https://doi.org/10.1016/j.isprsjprs.2021.03.001.

GreenValley International, 2021. LiDAR360 V5.0 User Guide. https://greenvalleyintl.com/LiDAR360/ (accessed 23 October 2023).

Han, Q., Fan, Z., Dai, Q., Sun, L., Cheng, M.-M., Liu, J., Wang, J., 2022. On the Connection between Local Attention and Dynamic Depth-wise Convolution. In: International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2106.04263.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N., Markham, A., 2021. Towards semantic segmentation of urban-scale 3D point clouds: A dataset, benchmarks and challenges. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4977-4987. https://doi.org/10.1109/cvpr46437.2021.00494.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N., Markham, A., 2022. SensatUrban: Learning Semantics from Urban-Scale Photogrammetric Point Clouds. Int. J. Comput. Vision 130, 316–343. https://doi.org/10.1007/s11263-021-01554-9.

Kalinicheva, E., Landrieu, L., Mallet, C., Chehata, N., 2022a. Multi-Layer Modeling of Dense Vegetation from Aerial LiDAR Scans. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1341-1350. https://doi.org/10.1109/CVPRW56347.2022.00140.

Kalinicheva, E., Landrieu, L., Mallet, C., Chehata, N., 2022b. Predicting Vegetation Stratum Occupancy from Airborne LiDAR Data with Deep Learning. Int. J. Appl. Earth Obs. Geoinf. 112, 102863 https://doi.org/10.1016/j.jag.2022.102863.

Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 4558-4567. https://doi.org/10.1109/cvpr.2018.00479.

Li, Q., Wong, F.K.K., Fung, T., 2021. Mapping multi-layered mangroves from multispectral, hyperspectral, and LiDAR data. Remote Sens. Environ. 258, 112403 https://doi.org/10.1016/j.rse.2021.112403.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2117-2125. https://doi.org/10.1109/access.2021.3100369.

Lin, H., Wu, S., Chen, Y., Li, W., Luo, Z., Guo, Y., Wang, C., Li, J., 2021. Semantic segmentation of 3D indoor LiDAR point clouds through feature pyramid architecture search. ISPRS J. Photogramm. Remote Sens. 177, 279–290. https://doi.org/10.1016/j.isprsjprs.2021.05.009.

Liu, M., Fu, B., Fan, D., Zuo, P., Xie, S., He, H., Liu, L., Huang, L., Gao, E., Zhao, M., 2021a. Study on transfer learning ability for classifying marsh vegetation with multi-sensor images using DeepLabV3+ and HRNet deep learning algorithms. Int. J. Appl. Earth Obs. Geoinf. 103, 102531 https://doi.org/10.1016/j.jag.2021.102531.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021b. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 10012-10022. https://doi.org/10.1109/iccv48922.2021.00986.

Luo, H., Khoshelham, K., Fang, L., Chen, C., 2020. Unsupervised scene adaptation for semantic segmentation of urban mobile laser scanning point clouds. ISPRS J. Photogramm. Remote Sens. 169, 253–267. https://doi.org/10.1016/j.isprsjprs.2020.10.002.

Mao, Y., Chen, K., Diao, W., Sun, X., Lu, X., Fu, K., Weinmann, M., 2022. Beyond single receptive field: A receptive field fusion-and-stratification network for airborne laser scanning point cloud classification. ISPRS J. Photogramm. Remote Sens. 188, 45–61. https://doi.org/10.1016/j.isprsjprs.2022.03.019.

Maxwell, A.E., Wilson, B.T., Holgerson, J.J., Bester, M.S., 2023. Comparing harmonic regression and GLAD Phenology metrics for estimation of forest community types and aboveground live biomass within forest inventory and analysis plots. Int. J. Appl. Earth Obs. Geoinf. 122, 103435 https://doi.org/10.1016/j.jag.2023.103435.

Pourshamsi, M., Xia, J., Yokoya, N., Garcia, M., Lavalle, M., Pottier, E., Balzter, H., 2021. Tropical forest canopy height estimation from combined polarimetric SAR and LiDAR using machine-learning. ISPRS J. Photogramm. Remote Sens. 172, 79–94. https://doi.org/10.1016/j.isprsjprs.2020.11.008.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Adv. Neural Inf. Process. Syst., pp. 5105–5114. https://dl.acm.org/doi/10.5555/3295222.3295263.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 652-660. https://doi.org/10.1109/cvpr.2017.16.

Qin, H., Zhou, W., Yao, Y., Wang, W., 2022. Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data. Remote Sens. Environ. 280, 113143 https://doi.org/10.1016/j.rse.2022.113143.

Rana, P., St-Onge, B., Prieur, J.-F., Budei, B.C., Tolvanen, A., Tokola, T., 2022. Effect of feature standardization on reducing the requirements of field samples for individual tree species classification using ALS data. ISPRS J. Photogramm. Remote Sens. 184, 189–202. https://doi.org/10.1016/j.isprsjprs.2022.01.003.

Roy, S.K., Haut, J.M., Paoletti, M.E., Dubey, S.R., Plaza, A., 2022. Generative Adversarial Minority Oversampling for Spectral-Spatial Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2021.3052048.

Rusu, R.B., Cousins, S., 2011. 3D is here: Point Cloud Library (PCL). In: IEEE Int. Conf. Robot. Autom. 1–4. https://doi.org/10.1109/ICRA.2011.5980567.

Seidel, D., Annighöfer, P., Thielman, A., Seifert, Q.E., Thauer, J.-H., Glatthorn, J., Ehbrecht, M., Kneib, T., Ammer, C., 2021. Predicting Tree Species From 3D Laser Scanning Point Clouds Using Deep Learning. Front. Plant Sci. 12 https://doi.org/10.3389/fpls.2021.635440.

Shi, Y., Wang, T., Skidmore, A.K., Heurich, M., 2018. Important LiDAR metrics for discriminating forest tree species in Central Europe. ISPRS J. Photogramm. Remote Sens. 137, 163–174. https://doi.org/10.1016/j.isprsjprs.2018.02.002.

Thomas, H., Qi, C.R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. Kpconv: Flexible and deformable convolution for point clouds. In: Proc. IEEE/CVF Int. Conf. Comput. vis. 6411–6420. https://doi.org/10.1109/iccv.2019.00651.

Tong, X., Brandt, M., Yue, Y., Ciais, P., Rudbeck Jepsen, M., Penuelas, J., Wigneron, J.-P., Xiao, X., Song, X.-P., Horion, S., 2020. Forest management in southern China generates short term extensive carbon sequestration. Nat. Commun. 11, 1–10. https://doi.org/10.1038/s41467-019-13798-8.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., Wegner, J. D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. Remote Sens. Environ. 264, 112603 https://doi.org/10.1016/j.rse.2021.112603.

Waldner, F., Chen, Y., Lawes, R., Hochman, Z., 2019. Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. Remote Sens. Environ. 233, 111375 https://doi.org/10.1016/j.rse.2019.111375.

Wang, D., Wan, B., Liu, J., Su, Y., Guo, Q., Qiu, P., Wu, X., 2020. Estimating aboveground biomass of the mangrove forests on northeast Hainan Island in China using an upscaling method from field plots, UAV-LiDAR data and Sentinel-2 imagery. Int. J. Appl. Earth Obs. Geoinf. 85, 101986 https://doi.org/10.1016/j.jag.2019.101986.

Wen, C., Yang, L., Li, X., Peng, L., Chi, T., 2020. Directionally constrained fully convolutional neural network for airborne LiDAR point cloud classification. ISPRS J. Photogramm. Remote Sens. 162, 50–62. https://doi.org/10.1016/j.isprsjprs.2020.02.004.

Yang, H., Chen, W., Qian, T., Shen, D., Wang, J., 2015. The extraction of vegetation points from LiDAR using 3D fractal dimension analyses. Remote Sens. 7, 10815–10831. https://doi.org/10.3390/rs70810815.

Ye, N., Morgenroth, J., Xu, C., Chen, N., 2021. Indigenous forest classification in New Zealand – A comparison of classifiers and sensors. Int. J. Appl. Earth Obs. Geoinf. 102, 102395 https://doi.org/10.1016/j.jag.2021.102395.

Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., Yan, G., 2016. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. Remote Sens. 8, 501. https://doi.org/10.3390/rs8060501.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2881-2890. https://doi.org/10.1109/cvpr.2017.660.

Zhao, C., Jia, M., Wang, Z., Mao, D., Wang, Y., 2023. Toward a better understanding of coastal salt marsh mapping: A case from China using dual-temporal images. Remote Sens. Environ. 295, 113664 https://doi.org/10.1016/j.rse.2023.113664.

Zhou, Y., Ji, A., Zhang, L., 2022b. Sewer defect detection from 3D point clouds using a transformer-based deep learning model. Autom. Constr. 136, 104163 https://doi.org/10.1016/j.autcon.2022.104163.

Zhou, C., Ye, H., Sun, D., Yue, J., Yang, G., Hu, J., 2022a. An automated, high-performance approach for detecting and characterizing broccoli based on UAV remote-sensing and transformers: A case study from Haining, China. Int. J. Appl. Earth Obs. Geoinf. 114, 103055 https://doi.org/10.1016/j.jag.2022.103055.

Zhu, Q., Li, Y., Hu, H., Wu, B., 2017. Robust point cloud classification based on multi-level semantic relationships for urban scenes. ISPRS J. Photogramm. Remote Sens. 129, 86–102. https://doi.org/10.1016/j.isprsjprs.2017.04.022.