International Journal of Digital Earth

ISSN: 1753-8947 (Print) 1753-8955 (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Multi-KPConv: deep learning-based LiDAR point cloud ground point extraction for complex terrains on the Loess Plateau

Chao Zhu, Jingxiang Li, Jonathan Li, Fuquan Tang, Dening Lu, Yuguang Wang, Junlei Xue, Qian Yang & Yu Su

To cite this article: Chao Zhu, Jingxiang Li, Jonathan Li, Fuquan Tang, Dening Lu, Yuguang Wang, Junlei Xue, Qian Yang & Yu Su (2025) Multi-KPConv: deep learning-based LiDAR point cloud ground point extraction for complex terrains on the Loess Plateau, International Journal of Digital Earth, 18:2, 2556235, DOI: 10.1080/17538947.2025.2556235

To link to this article: https://doi.org/10.1080/17538947.2025.2556235

Published online: 14 Sep 2025.

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2025, VOL. 18, NO. 2, 2556235
https://doi.org/10.1080/17538947.2025.2556235

Multi-KPConv: deep learning-based LiDAR point cloud ground point extraction for complex terrains on the Loess Plateau

Chao Zhu[a], Jingxiang Li[a], Jonathan Li[c], Fuquan Tang[a,b], Dening Lu[d], Yuguang Wang[e], Junlei Xue[a], Qian Yang[a] and Yu Su[a]

[a]College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054, People's Republic of China; [b]Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi'an, 710021, People's Republic of China; [c]Departments of Geography and Environmental Management, University of Waterloo, Waterloo, ON, N2L 3G1, Canada; [d]The Department of Systems Design Engineering, University of Waterloo, Waterloo, N2L 3G1, Canada; [e]Jinan Geotechnical Investigation and Surveying Institute, Jinan, Shandong, 250013, People's Republic of China

ABSTRACT
Accurate extraction of ground points from LiDAR point clouds provides important data for understanding terrain changes and supports decision-making in ecological disaster prevention. Recently, deep learning models have been used to process point clouds directly, with a focus on semantic segmentation of urban scenes using RGB features. However, in complex terrains, using point cloud images to generate RGB features often introduces noise, making high-precision ground point extraction a difficult task. This paper presents a new point-based semantic segmentation network, Multi-KPConv, to overcome these challenges. Unlike methods that rely on color point clouds, Multi-KPConv uses shallow features based on domain knowledge as input to the network. The network employs a multi-dimensional kernel point convolutional architecture to extract high-level semantic features, allowing for better data interpretation. Additionally, the SimAM 3D attention mechanism is integrated to adaptively refine feature contributions and highlight important point features. We evaluate Multi-KPConv on datasets from the Dafosi mining area in China's Loess Plateau and the STPLS3D dataset. Experimental results show that Multi-KPConv outperforms current state-of-the-art models in terms of generalization and robustness, effectively extracting ground points in complex terrain, such as the gully areas of the Loess Plateau.

ARTICLE HISTORY
Received 8 November 2024
Accepted 24 August 2025

KEYWORDS
Airborne LiDAR point cloud; Ground filtering; Deep learning; Semantic segmentation; Loess plateau

1. Introduction

Airborne Laser Scanning (ALS) has seen rapid development in recent years, enabling quick, efficient, and accurate acquisition of the Earth's surface information. Employing ALS to obtain reliable ground point clouds under complex geomorphology provides a crucial foundation for geologic disaster monitoring, environmental protection, and related activities (Wang 2017). The Loess Plateau of China, one of the largest loess landform regions in the world, features unique and complex topography and is confronted with severe issues of desertification, soil erosion, and coal mining subsidence (Liu et al. 2020; Wu et al. 2020). Utilizing ALS to acquire ground point clouds of complex terrain can play a pivotal role in monitoring and restoring ecological environments.

Since the original ALS point cloud comprises both ground and non-ground components such as vegetation, buildings, and noise, it is necessary to mitigate the influence of non-ground points when using ALS data to quantify land surface structure. In applications of ALS point clouds across various fields, the critical step involves obtaining ground points from the point cloud (Yang et al. 2023). The traditional ground point extraction methods usually employ various filtering algorithms to remove non-ground points. Based on the algorithm's characteristics, these can be categorized into slope-based methods (Susaki 2012;

CONTACT Fuquan Tang tangfuquan@xust.edu.cn College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054, People's Republic of China; Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi'an, 710021, People's Republic of China

Yang, Zhang, and Li 2017), mathematical morphology methods (Balado et al. 2020; Hui et al. 2016), segmentation-based methods (Chen, Xu, and Gao 2016), and progressive triangulation techniques (Chen et al. 2023; Quan et al. 2017). However, in the Loess Plateau, characterized by its complex terrain, the broken terrain and the surface's intricate vertical and horizontal gully patterns pose significant challenges. The diversity of natural geographical conditions renders traditional filtering methods less adaptive, tends to cause significant misclassifications in challenging areas and necessitating extensive manual refinement.

With the rapid advancement of artificial intelligence technology, deep learning (DL) has been widely applied to computer vision tasks such as image classification (Li et al. 2019), object detection (Luo et al. 2023), and semantic segmentation (Hao, Zhou, and Guo 2020). However, these methods cannot be directly applied to 3D point clouds due to their fundamentally distinct data formats. Various methods have been proposed to solve this problem. Early point cloud classification methods converted irregular point clouds into regular 3D meshes (images, voxels or grids), on which 3DConvNets (Qi et al. 2016) were applied to mitigate the effects of characteristics such as point cloud disorder, uneven density and high data redundancy. However, the sparsity of the point cloud leads to a strictly limited mesh resolution, and the feature extraction of 3D convolutional kernels also results in an exponential growth of computational costs.

In recent years, DL-based models that directly utilize point clouds as inputs have garnered increased attention (Hu et al. 2020; Qi et al. 2017a, 2017b; Thomas et al. 2019), and these approaches more effectively preserve the geometric structural information of point clouds. It is worth noting that most of the current deep learning-based point cloud semantic segmentation studies primarily focus on urban environments, such as indoor scenes and streets, and use RGB as auxiliary features input (Aditya et al. 2024; Cao et al. 2025). In this context, the objective of this paper is to employ DL-based approaches for the intelligent point clouds filtering, specifically to filter ground and non-ground points in sparse and discrete point clouds. However, research on natural regions such as the Loess Plateau remains scarce, which is characterized by complex topography, dramatic geomorphic undulations, and extensive vegetation cover (Tang and Yang 2023). In such areas, UAV-based optical images are often significantly affected by non-uniform illumination and shadows (Kushwaha et al. 2023). During the fusion of point cloud and image, the extraction of RGB features depends heavily on both image quality and point cloud alignment accuracy, which is prone to introducing unstable information due to alignment errors and illumination differences, make it unable to ensure that semantic spatial consistency can be captured through color consistency information (Jing et al. 2021). Although RGB information is often used as an auxiliary feature to enhance semantic segmentation performance in urban scenes, its regional validity is difficult to maintain consistency in complex natural environments. In addition, the current mainstream DL-based segmentation models are supervised learning frameworks, whose core objective is to robustly capture object's spatial consistency and semantic consistency and assign appropriate weight scores to these features (Du et al. 2024; Wang et al. 2019). The effectiveness of this process relies on the differential representation between the geometric features of object's categories and the spatial semantic information, as well as the effective mining of the geometric relationships between the input points and their neighborhoods to provide spatially supervised guidance signals for the model (Engelmann et al. 2019). To this end, in this paper, starting from modeling local spatial geometric relationships in point clouds, we propose replacing the RGB input with a combination of geometric features to enhance the model's representation of fine-grained geometric variations. Additionally, the accuracy of DL-based classification depends on the number and diversity of training samples, which poses a challenge to applied to sparse ALS point cloud filtering, as ALS-acquired point clouds exhibit sparse distributions in regions with steep and complex landscapes. The main contributions of this study are as follows:

(1) We propose a Multi-KPConv model for processing LiDAR point clouds in complex terrain and design a multi-dimensional integrated kernel point convolution for multi-level learning to obtain reliable ground point clouds in geomorphologically complex regions, thereby achieving intelligent point clouds filtering.
(2) For sparse point clouds, we propose an effective shallow features combination derived from a data-driven synthesis of neighbor domain knowledge, to enhance the representation of typical landforms at the input feature level, and integrate these with deep features to enhance the network's capability to extract multi-dimensional integrated semantic features from ALS point clouds.

(3) This study introduces a 3D parameter-free attention mechanism, SimAm, to enable the redistribution of feature weights and complementary fusion, thereby highlighting important features, suppressing noise interference, and enhancing the accuracy of ground point extraction.

2. Related work

With the widespread application of deep learning in 2D images, numerous researchers have extended its applications to 3D point cloud semantic segmentation. This section provides a succinct overview of existing shallow feature applications and deep learning methods.

2.1. Classification of point clouds based on shallow features

Given the disorder, varying densities, and non-uniform distribution of point clouds, extracting high-precision ground points from complex scenes remains a significant challenge. The conventional approach involves using point neighborhood information, calculating specific geometric features between points, and leveraging local spatial context information for point cloud classification. In shallow feature-based 3D scene analysis, feature extraction and selection are essential for classification (Xiao et al. 2023). Guo et al. employed 26 diverse features, including intensity information, echo times, and geometric shape, as supplementary attributes for point cloud classification using the JointBoost classifier. However, due to the broad feature set, the process of feature extraction resulted in significant time consumption (Guo et al. 2015). Machine learning-based classifiers independently label each point based on local features, omitting semantic information from neighboring points. Consequently, this method frequently introduces significant classification noise in complex scenes (Weinmann et al. 2015). Kuprowski et al. suggested incorporating geometric features, elevation data, and laser reflection characteristics into a machine learning algorithm to improve the accuracy of classifying power supply line point clouds by distinguishing between ground and building points (Kuprowski and Drozda 2023). The study by Joseph-Rivlin et al. focused on indoor point clouds, introducing the first and second moments as polynomial inputs for a deep learning model. They analyzed the significance of geometric moments in characterizing the shape of 3D point clouds, proposing incorporating polynomial functions into deep learning models to enhance the efficiency and accuracy of point cloud classification (Joseph-Rivlin, Zvirin, and Kimmel 2019). Hsu et al. integrated shallow features into the PointNet model, demonstrating that combining shallow features with deep learning models can improve classification accuracy for urban scene point clouds (Hsu and Zhuang 2020). The selection of effective features plays a crucial role in enhancing classification accuracy. Therefore, we employ high-order moments, surface curvature, and mean elevation as input features for the Multi-KPConv to aggregate high-level features that more effectively capture the local detailed geometric structure of surfaces.

2.2. Point cloud classification based on deep learning

With the increasing adoption of ALS technology, 3D point clouds are increasingly utilized in ecological environment monitoring, urban planning, autonomous vehicle driving, mapping, and other domains. This section provides a succinct overview of existing models for semantic segmentation of point clouds, covering methods based on multi-view methods, voxelization techniques, Transformer networks, and point-based methodologies.

2.2.1. Multi-view-based methods

Multi-view based methods projects the point cloud to a 2D plane, processing it using a 2D convolutional neural network, and subsequently maps the result back to 3D data to achieve the semantic segmentation of the original point cloud (Boulch, Saux, and Audebert 2017; Qin, Hu, and Dai 2018; Wang, Wang, and Cai 2022). Su et al. were the first to propose an image-based MultiView Convolutional Neural Networks, demonstrating that 2D views can provide significant information for 3D shape recognition (Su et al. 2015). However, the method applies a pooling operation on each view, which may result in the loss of local information of each view. To address this issue, Feng et al. proposed to group feature descriptors of different views based on feature similarity, which strengthens the connection of different visual features and improves the segmentation accuracy of the point cloud (Feng et al. 2018). Lawin et al. proposed that different modal features are complementary in semantic segmentation tasks, and that fusing these features through a 2D-CNN network architecture can significantly improve the segmentation performance (Lawin et al. 2017). For scene segmentation, the advantage of the aforementioned methods is that they can directly utilize the developed and mature 2D neural networks, but such methods in the ground point extraction task, the occlusion, undulation and viewpoint changes in the point cloud seriously affect the projection results, resulting in poor semantic consistency. In addition, the multi-view projection is prone to lose the ground continuity information, which affects the classification accuracy, especially in the Loess Plateau region with significant terrain undulations and complex feature structures.

2.2.2. Voxel-based methods

The voxel-based approach relies on the representation of image pixels by segmenting the 3D space and introducing spatial dependencies into the point cloud data, transforming the point cloud into a regular voxel representation (Ben-Shabat, Lindenbaum, and Fischer 2018; Maturana and Scherer 2015; Roynard, Deschaud, and Goulette 2018). Finally, feature learning is achieved using 3D convolution to assign all points within each voxel the same category label as the current voxel. Voxelization alleviates the disordered and unstructured problem of point clouds to some extent. MATURANA et al. were the first to propose the use of voxels for semantic segmentation of point clouds and designed the semantic segmentation model VoxNet (Maturana and Scherer 2015). This method voxelizes the point cloud, learns the features using Volumetric CNN network, and outputs the segmentation results. However, the sparsity of voxels and partial loss of spatial information lead to unsatisfactory segmentation results. TCHAPMI et al. proposed the SEG-Cloud model integrated 3D fully convolutional network, trilinear interpolation, and Fully-Connected Conditional Random Fields (FC-CRF), achieving the fine segmentation(Tchapmi et al. 2017). In addition. Some studies have used 3D fully convolutional neural networks to learn multi-scale voxel information as a way to mitigate the problem of partial information loss, but the choice of voxel size affects the final segmentation results (Roynard, Deschaud, and Goulette 2018; Ye, Xu, and Cao 2020). Although these methods achieve better segmentation results, for sparse terrain point clouds, the small number of voxel interior points leads to insufficient information representation, and the voxel averaging process at the boundary points of ground and non-ground is prone to confusion. Furthermore, the choice of voxel size significantly impacts accuracy and computational overhead. Smaller voxels improve the accuracy but at the cost of higher memory and time consumption, which limits the efficiency of their application in large-scale natural terrain.

2.2.3. Transformer-based methods

Transformer utilizes Multi-Head Self-Attention (MSA) to establish model attention, enabling the model to filter and learn the most important information. Additionally, Transformer provides more stable outputs when the input point cloud undergoes substitution or rigid body transformation (Lai et al. 2022; Wang, Chakraborty, and Yu 2022). Engel et al. employed Point Transformer to extract local and global features, linking the two representations by introducing an attention mechanism (Engel, Belagiannis, and Dietmayer 2021). In ground point extraction tasks, it has the advantage of perceiving a broader range of spatial dependencies, which aids in recognizing ground points in occluded or edge features. However, the main issue with Transformer is its high consumption of computational resources, particularly when processing large-scale, high-density point clouds, coupled with low training and inference efficiency. This limitation hinders its engineering deployment for rapid ground point extraction tasks.

2.2.4. Point-based methods

Point-based methods can directly process point cloud data without the need for additional data conversion, preserving the original information of the point cloud (Atzmon, Maron, and Lipman 2018; Chiang et al. 2019; Engelmann, Kontogianni, and Leibe 2020). PointNet uses a multilayer perceptron (MLP) to map low-dimensional features of the point cloud to high-dimensional features, followed by aggregation of the global features through a pooling function (Qi et al. 2017a). A series of methods have emerged based on this approach, such as the point-by-point MLP method(Boulch 2020; Fan et al. 2021; Hermosilla et al. 2018; Li et al. 2018; Wang et al. 2018). However, such methods only capture global information and struggle to handle fine-grained differences in complex landscapes. To address this, Qi et al. sampled and regionally divided the point cloud, utilizing PointNet for feature extraction in each small region (Qi et al. 2017b). However, this widely used framework still suffers from stability issues when handling unevenly distributed and sparse regions, despite introducing local structure modeling. Additionally, it is sensitive to the number of input points, which makes it challenging to adapt to the scale differences across various terrain regions. Moreover, recurrent neural network-based (Ye et al. 2018; Zhao, Liu, and Ramani 2019), point convolution methods (Engelmann, Kontogianni, and Leibe 2020; Hua, Tran, and Yeung 2018; Lei, Akhtar, and Mian 2019; Thomas et al. 2019), and graph-based methods (Phan et al. 2018; Zhang et al. 2020) are widely used in point cloud semantic segmentation. Among these deep learning methods, the KPConv point convolution method, which utilizes a geometric kernel for point-by-point convolution of the original point cloud, achieves excellent results in semantic segmentation of large urban ALS datasets and outperforms traditional deep learning methods such as PointNet++, voxel-based, and multi-view-based approaches. In ground point extraction, it can more effectively distinguish surface and non-surface structures by capturing local undulation features of the terrain, making it especially suitable for scenes with drastic changes in geomorphology, such as the Loess Plateau. Considering the overall performance, we have chosen to use the KPConv kernel function to extract advanced features of the point cloud.

3. Method

3.1. Shallow features extraction

In point cloud deep learning models, RGB features are commonly employed as additional input. To obtain RGB information, ALS point clouds require overlaying with images. However, in environments like the Loess Plateau of China, where the terrain is furrowed and geometrically complex, overlaying point clouds with images from various data sources can lead to significant noise due to object occlusion and issues with geometrical alignment. Therefore, this paper proposes a suite of shallow point cloud features specifically designed for complex terrain, conceived from domain-specific expertise. By fully utilizing the information in the dataset, we augment the feature input to the convolutional layer, enabling the network to learn comprehensive and context-aware information for performance improvement. The following subsections will detail the proposed combination of shallow features, including average elevation, surface curvature, and higher-order moments.

3.1.1. Average elevation

ALS data contains a significant number of non-ground points, where points at the interface with the ground are usually prone to misclassification. To address this issue, we utilize the mean elevation within a local region as the classification feature, thereby enhancing the spatial relationships among neighboring point clouds. Typically, the weighted average method is employed to calculate the elevation mean, assigning weights according to the elevation disparity between laser points and nadir points. Points exhibiting greater deviation from the minimum elevation are assigned higher weights, whereas those with lesser deviation receive lower weights. The weight function is calculated using Equation (1), while the mean elevation is derived using Equation (2). In these equations, P_i represents the weight assigned to each point, Z_{ave} denotes the average elevation of all point clouds in the local area, $Z_{\max}$ represents the maximum elevation value within that area, and $Z_{\min}$ represents the minimum elevation value within that area.

$$P_i = \frac{Z_i - Z_{\min}}{Z_{\max} - Z_{\min}} \tag{1}$$

$$Z_{ave} = \frac{\sum_{i=1}^{n} (Z_i \cdot P_i)}{\sum_{i=1}^{n} P_i} \tag{2}$$

3.1.2. Surface curvature

For steep landforms in complex terrain, this paper utilizes the surface curvature metrics of the point cloud to characterize the undulation conditions of the points on the surface. Surface curvature is a significant attribute of point clouds, with numerous studies demonstrating its positive influence on point cloud classification (Demantké et al. 2012; Lalonde et al. 2006; Yang and Dong 2013). For each point P_i in the point cloud, identify its nearest k points, and compute the local plane **P** using the least squares method for these points; the plane **P** can be expressed as equation (3).

$$P\left(\vec{n}, d\right) = \underset{\left(\vec{n}, d\right)}{\arg\min} \sum_{i=1}^{k} \left(\vec{n} \cdot P_i - d\right)^2 \tag{3}$$

where **n** represents the normal vector of plane **P**, and d denotes the distance from plane **P** to the coordinate origin, the covariance matrix **M** is computed from the points on surface **P**, as delineated in Equation (4).

$$M = \frac{1}{k} \sum_{i=1}^{k} (p_i - p_0)(p_i - p_0)^T \tag{4}$$

Eigenvalue decomposition of the covariance matrix **M** is performed to determine each eigenvalue of **M**. If the eigenvalues satisfy the condition $\lambda_0 < \lambda_1 < \lambda_2$, then the surface curvature of point P is given by Equation (5).

$$\delta = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \tag{5}$$

where a smaller δ signifies a flatter neighborhood, while a larger δ suggests a more varied neighborhood characterized by undulations.

3.1.3. Higher-order moment

For complex landscapes such as landslides and terraces, this paper utilizes geometric moments as additional inputs to enable the network to take into account higher-order moments of complex terrain. Previous studies have demonstrated the advantages of higher-order moment features in addressing the problem of insufficient feature extraction, enabling the network to automatically learn additional effective features (Benouini et al. 2019; Li and Shen 1991). Joseph-Rivlin et al. demonstrated that using higher-order moments as inputs to individual hidden units enables the network to extract higher-order features without necessitating large memory resources (Joseph-Rivlin, Zvirin, and Kimmel 2019). Neural networks require numerous weights and network layers to compute geometric moments, necessitating substantial computational resources. Therefore, we approximate the polynomial function of the input point cloud using the powers (x^2, y^2, z^2, xy, xz, yz) as additional features, leveraging the model's ability to process the point cloud and reduce time complexity.

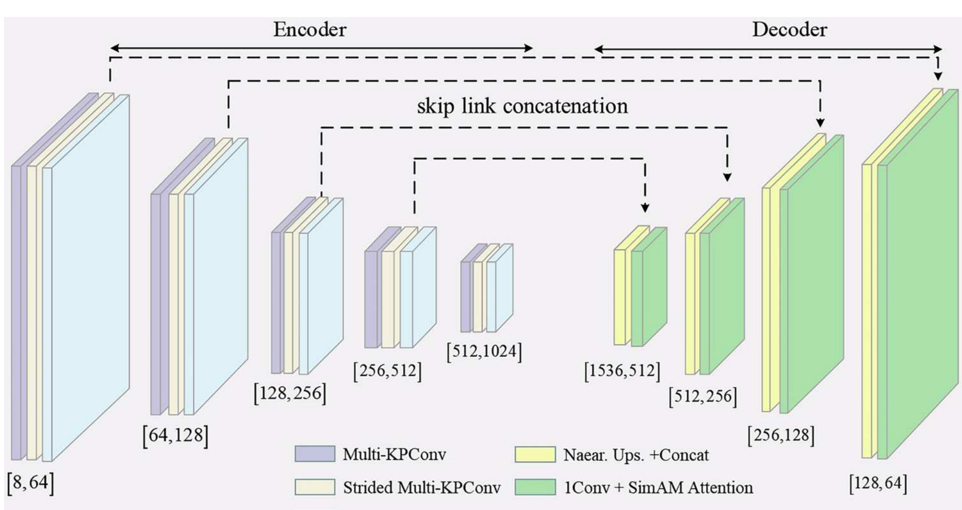

Figure 1. Multi-dimensional integrated kernel point convolution model architecture.

3.2. Multi-dimensional integrated kernel point convolution model

In Figure 1, we propose a multi-dimensional integrated kernel point convolution model based on an encoder-decoder structure for point cloud segmentation tasks. The model utilizes xyz coordinates and shallow features of the point cloud as inputs. The encoder network comprises five layers of convolutional modules, each containing two convolutional blocks, with each block including one attention block. The structure of the convolutional blocks resembles Bottleneck ResNet blocks (He et al. 2015), utilizing Multi-KPConv in place of traditional image convolution.

Multi-KPConv leverages the concept of grouped convolution in image processing, utilizing multi-branch kernel point convolution to extract multi-level features of the point cloud. The input features are initially mapped from [n, 11] to [n, d] using a linear function, aiming to enhance the feature diversity and information content of the point cloud while improving the model's expressive capacity and performance. Subsequently, the multi-dimensional grouped convolution of the point cloud [n, d] is performed to achieve the transformations and aggregations of the features. Point convolution is defined by Equation (6), where x_i represents the point in $P \in R^{N\times 3}$, f_i corresponds to the feature in $F \in R^{N\times D}$, N_x denotes the spherical radius neighborhood $N_x = \{x_i \in P| \|x_i - x\| \leq r\}$, and $g^{(w)}$ represents the kernel function. Employing spherical neighborhoods enables the network to more effectively learn shallow features with enhanced robustness in point cloud regions of varying densities (Thomas et al. 2019, 2018). The definition of the kernel function is crucial in point convolution. In this paper, we propose a multimodal kernel function $g^{(w)}$, defined by Equation (7), with the specific computational flow illustrated in Figure 2.

$$(\mathcal{F} * g)(x) = \sum_{x_i \in \mathcal{N}_{\S}} g^{(w)}(x_i - x) f_i \tag{6}$$

$$g^{(w)} = \sum_{j=1}^{m} w_j u^{(j)} \tag{7}$$

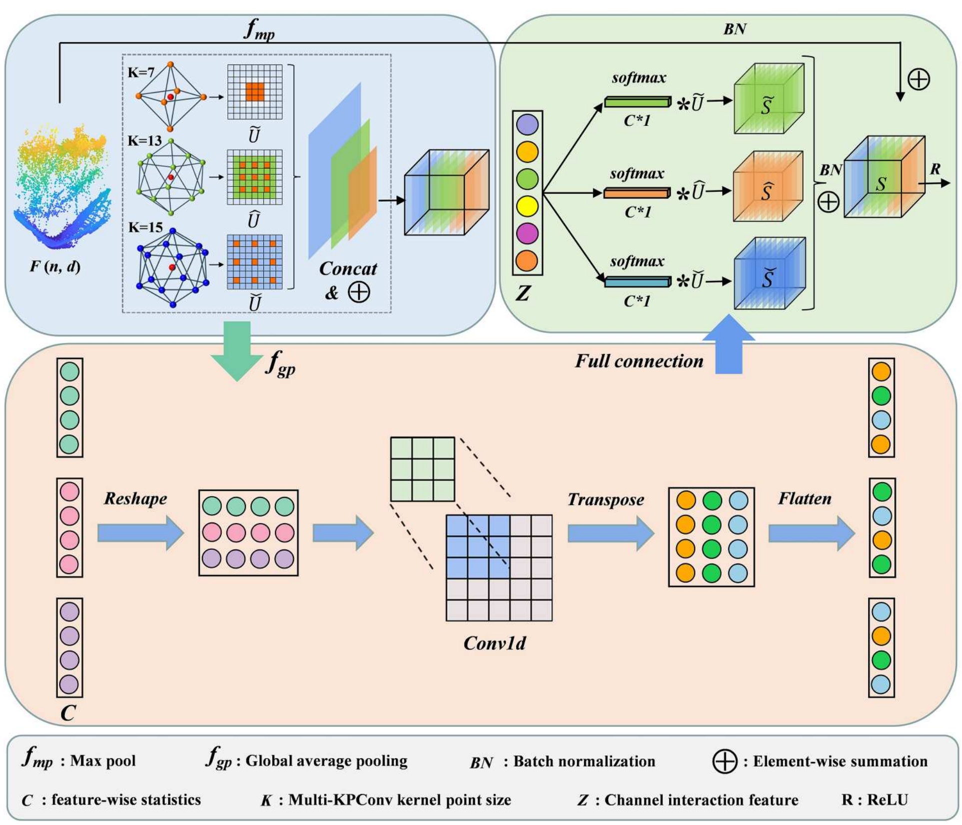

Figure 2. Flow of multi-dimensional integrated kernel function calculation.

where $w = [w_1, w_2, \ldots, w_m]$, w_j represents the weight of the j-th dimension, $\sum_{j=1}^{m} w_j = 1$ denotes a weight sum of 1, and $u^{(j)}$ represents the feature vector of the j-th dimension.

The multi-dimensional integrated kernel function $g^{(w)}$ enables the neural network to adaptively tune the receptive field based on the multiple-dimensional features of the input. Although only three dimensions are shown in Figure 2, this method is readily extendable to multiple branches. For a specific shallow feature, the feature is extracted using the kernel point g as defined in KPConv (Thomas et al. 2019), which directly utilizes the positional relationship between x and its neighboring point x_i as input. The defining equation is presented below:

$$g(y_i) = \sum_{k<K} h(y_i, \tilde{x}_k) W_k \tag{8}$$

where $y_i = x_i - x$ denotes the relative position of the sampled point to its neighboring points; $\tilde{x}_k$ represents the kernel points, and k denotes the maximum number of kernel points, defined by the operator; W_k represents the weight matrix, mapping features from the input dimensions to the specified output dimensions; h denotes the correlation between Δx and $\tilde{x}_k$.

Building on the multi-dimensional feature map, we have developed an information fusion method utilizing a gating mechanism, enabling the neural network to automatically adjust the receptive field size according to the scene. This method facilitates dynamic adjustments of the receptive field by controlling the information flow into various branches in the subsequent convolutional layer, thereby assigning varied weights to each kernel point. To ensure the gate fully integrates the information from all modalities, we employ an element-by-element summation strategy for data fusion, yielding the feature map **U**. Furthermore, we perform global average pooling on the feature map **U** to embed the global characteristics of the point cloud and generate feature-wise statistical data C, formatted as a one-dimensional array c. The detailed formulation is presented below.

$$U = \tilde{U} + \hat{U} + \check{U} \tag{9}$$

$$\mathcal{C} = \mathcal{F}_{gp}(U) = \frac{1}{n} \sum_{i=1}^{n} U_d(i) \tag{10}$$

In order to strengthen the feature connection between shallow features, such as higher-order moments and surface curvature, this paper derives more useful features by calculating the information interaction between channels in the feature map. First, the feature map of array c is reshaped to [1, c], and then the feature map is input into a one-dimensional convolutional layer to calculate the local dependency information between each channel and its k neighboring channels. Second, by transposing and reshaping, the channel order of the feature map is randomly rearranged to obtain features of dimension [1, c]. Thus, a distant channel before disruption may become a neighboring channel, and more global dependencies are obtained after inputting the features into the next convolutional layer. Finally, the features of dimension [1, c] are flattened to form array c, and the corresponding multidimensional branches are recovered by a fully connected layer to obtain the multidimensional integrated augmented features Z after channel interaction.

Guided by the enhanced feature descriptor Z, the soft attention mechanism is employed to derive dynamic parameters, enabling the model to focus on critical regions and adaptively select information from various dimensions. a_c, b_c, and d_c represent the dynamic weight matrices of $\tilde{U}$, $\hat{U}$, and Ǔ, respectively, corresponding to the weights of the fusion ratio matrices for the features of different receptive fields. These matrices are continuously adjusted and optimized during network training. The calculation formula is presented in Equation (11-12).

$$\text{softmax}(A_c z, B_c z, D_c z) = \left(\frac{e^{A_c z}}{e^{A_c z} + e^{B_c z} + e^{D_c z}}, \frac{e^{B_c z}}{e^{A_c z} + e^{D_c z} + e^{D_c z}}, \frac{e^{D_c z}}{e^{A_c z} + e^{B_c z} + e^{D_c z}} \right) \tag{11}$$

$$(a_c, b_c, d_c) = \text{softmax}(A_c z, B_c z, D_c z) \tag{12}$$

where $A, B, D \in R$, A_c denotes the cth column of A, B_c denotes the c-th column of B, and D_c denotes the c-th column of D.

During the feature fusion stage, unlike the simple pixel-level summation used to integrate multi-dimensional features, this paper employs adaptive multi-dimensional feature integration to achieve efficient fusion of diverse features. Specifically, it involves multiplying the features of different modalities by the corresponding learnable weights and adding the skip-connected initial eigenvalues **F** to obtain the final fused feature **S**, which is computed as follows.

$$S = BN\left(a_c \times \tilde{U} + b_c \times \hat{U} + d_c \times \check{U}\right) + f_{mp}(BN(F)),\ where\ a_c + b_c + d_c = 1 \tag{13}$$

The feature **S** continuously increases the effective receptive field of each point as the dynamic parameters change, retaining most of the effective information and improving the performance of ground point classification. Additionally, to train the model stably, the SimAM attention network (Yang et al. 2021) from computer vision is introduced to enhance the characterization ability of the multi-dimensional integrated kernel point convolution, enabling it to extract more distinguishable and consistent features.

The enhanced features will be referred to as $\tilde{S}$.

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \mu)^2 + 2\sigma^2 + 2\lambda} \tag{14}$$

$$\tilde{S} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot S \tag{15}$$

.

where μ and σ^2 represent the mean and variance, respectively, of all neurons excluding t, e_t^* represents the minimum energy function. S represents the input feature, $\odot$ denotes the Hadamard product operation, E is the energy function applied to each channel, and the Sigmoid function limits any excessive values in E.

The SimAM module is integrated following a multi-dimensional integrated kernel point convolution module to facilitate a comprehensive assessment of both spatial location features and channel information, without the addition of extra parameters. By focusing specifically on key neurons, the network's feature extraction capabilities and training speed are enhanced.

4. Experiments and analyses

4.1. Dataset

4.1.1. Overview of the study area

In this paper, the Dafosi coal mine in the Loess Plateau of Xianyang City, Shaanxi Province, China, is selected as the study object. The study area is situated in the north of Xianyang City, Shaanxi Province, China, at the confluence of Binzhou City and Changwu County, with geographic coordinates ranging from longitude 107°49'40" to 108°01'00" East and latitude 35°00'00" to 35°05'00" North. As illustrated in Figure 3, the area encompasses complex terrain, including hills, gullies, terraces, collapses, and landslides. The raw point cloud data were acquired using a DJI M600 drone equipped with an SZT-R250 airborne laser measurement system, flying at an altitude of 120 meters and achieving a point density of 60 points per square meter.

4.1.2. Test Set Creation

This study evaluates the Multi-KPConv model's ability to extract ground points using six point cloud samples of varying geomorphological features, which were not included in the training set. The test sets were categorized using the methods in the Data Preprocessing section. Figure 4 shows the top view and profile of each test set. Test set (a) represents a loess gully with sparse vegetation, containing 36,837 ground and 48,233 non-ground points at 44 points per square meter. Test set (b) covers a loess gully with dense vegetation, with 58,386 ground and 63,236 non-ground points, showing non-uniform density due to vegetation. Test set (c) includes a loess gully with broken terrain and dense vegetation, holding 57,079 ground and 97,519 non-ground points with uneven density. Test set (d) is a convex slope area with dense vegetation, containing 66,542 ground and 69,811 non-ground points, with uneven density. Test set (e) covers a sparsely vegetated convex slope with 103,229 ground and 102,949 non-ground points at 86 points per

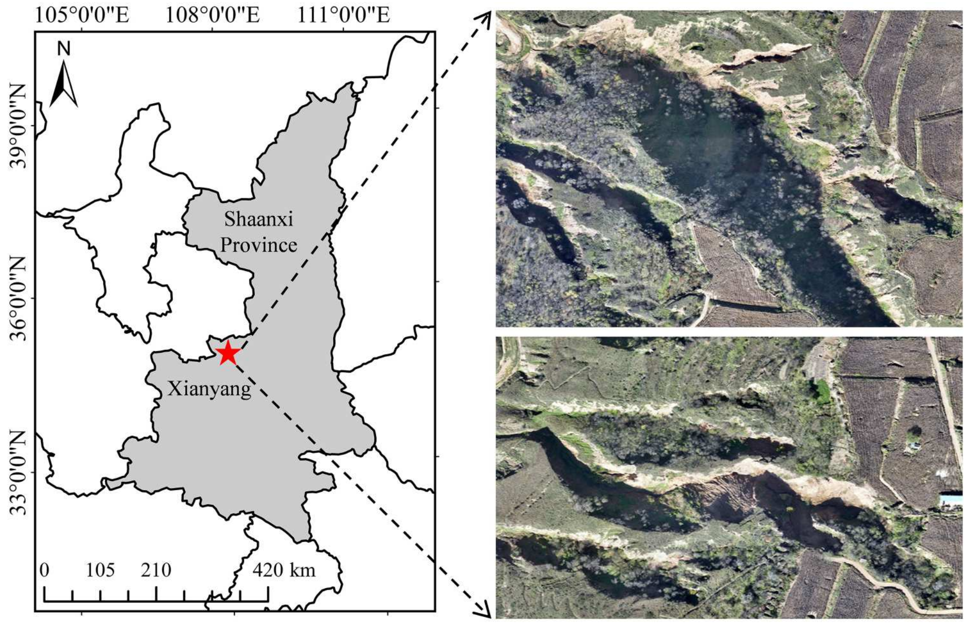

Figure 3. Overview of the study area.

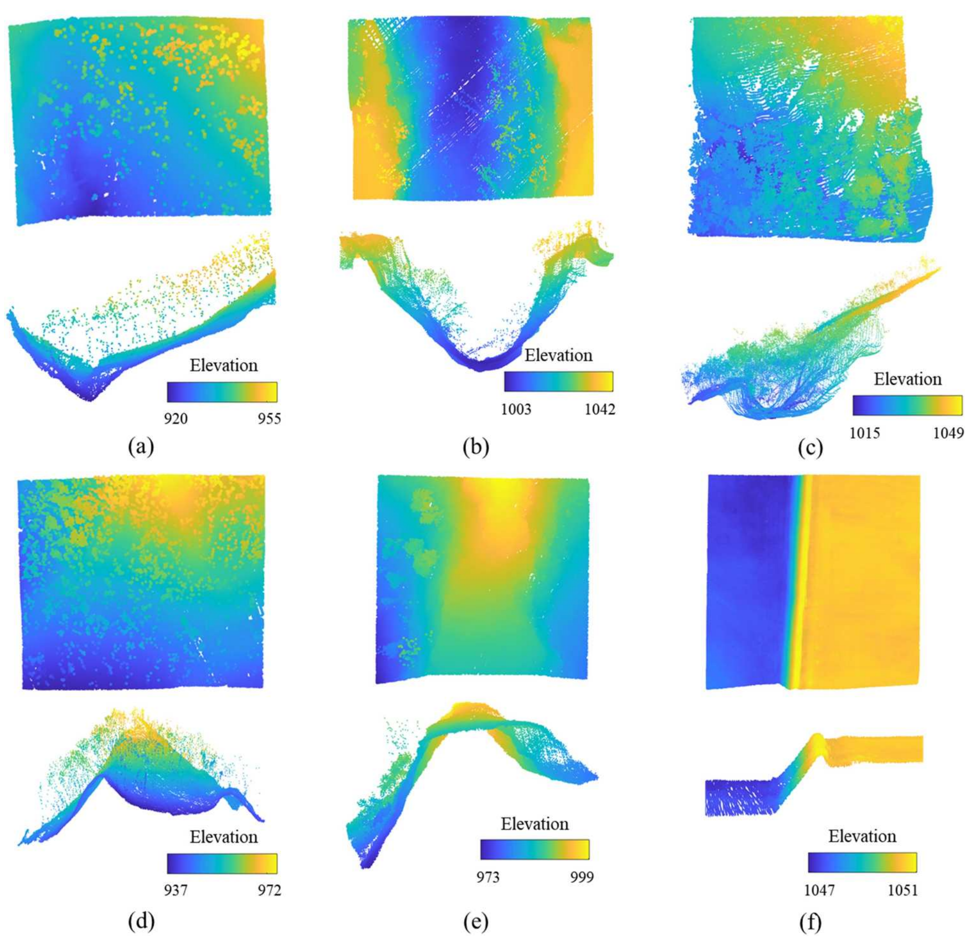

Figure 4. Point cloud topography of the test set.

square meter. Test set (f) represents an ascending slope with simple topography, containing 73,446 ground and 122,012 non-ground points at 100 points per square meter.4.2 Implementation.

4.1.3. Baseline

For comparative analysis, we selected established methods from the literature, adapted them for 3D terrain scene recognition, and employed them as a baseline. In this study, the Multi-KPConv model is compared with PointNet++ (Qi et al. 2017b), RandLA-Net (Hu et al. 2020), SCF-Net (Fan et al. 2021), Point Transformer (Engel, Belagiannis, and Dietmayer 2021), Stratified Transformer (Lai et al. 2022) and SPoTransformer (Park et al. 2023) to assess performance differences in extracting ground point cloud data. Although these models perform satisfactorily in point cloud semantic segmentation tasks, their adaptability to large-scale LiDAR datasets, particularly those with loess hilly and gully characteristics, has not been empirically validated. Additionally, this paper employs the traditional Cloth Simulation Filtering (CSF) algorithm (Zhang et al. 2016) alongside the Improved Progressive TIN Densification (IPTD) algorithm (Zhao et al. 2016) for ground point filtering, and these are analyzed relative to other methods discussed herein. All methods are implemented as per the original authors' specifications, with hyperparameters and model parameters fine-tuned according to the officially prescribed tuning schemes.

4.1.4. Evaluation Criteria

To quantitatively assess the model's accuracy, this study employs formulas (16-21) for evaluation. Here, IoU_1 and IoU_2 represent the intersection-over-union ratios for non-ground points and ground points, respectively. OA denotes the overall accuracy as the proportion of correctly classified samples relative to the total sample count. Kappa indicates the accuracy and consistency of the classification model, while MCC represents the Matthews correlation coefficient.

$$IoU_1 = \frac{TP_1}{TP_1 + FP_1 + FP_2} \tag{16}$$

$$IoU_2 = \frac{TP_2}{TP_2 + FP_2 + FP_1} \tag{17}$$

$$OA = \frac{TP_1 + TP_2}{TP_1 + FP_1 + TP_2 + FP_2} \tag{18}$$

$$Kappa = \frac{OA - P_e}{1 - P_e} \tag{19}$$

$$P_e = \frac{(TP_1 + FP_2) \times (TP_1 + FP_1) + (TP_2 + FP_1) \times (TP_2 + FP_2)}{(TP_1 + FP_1 + TP_2 + FP_2)^2} \tag{20}$$

$$MCC = \frac{TP_1 \times TP_2 - FP_1 \times FP_2}{\sqrt{(TP_1 + FP_1) \times (TP_1 + FP_2) \times (TP_2 + FP_1) \times (TP_2 + FP_2)}} \tag{21}$$

where TP_1 and TP_2 are the numbers of correctly predicted non-ground and ground points, respectively, and FP_1 and FP_2 represent the numbers of incorrectly predicted non-ground and ground points, respectively.

4.2. Experimental results

In this section, we quantitatively evaluate the ground point cloud extraction performance of the Multi-KPConv model along with nine other deep learning models across the specified test sets (test set (a) to test set (f)), according to Equations (16-21). Simultaneously, the effectiveness of ground point extraction by each model, under various landforms, is qualitatively analyzed through the visualization of error point clouds.

4.2.1. Analysis of Test(a) and Test(b) results

Table 1 illustrates the quantitative analysis results. The Stratified Transformer achieves the highest accuracy in the evaluation metric IoU_1 (95.80) in Test(a), while the SPoTransformer demonstrates excellent

Table 1. Quantitative evaluation results of the six methods in Test(a) and Test(b) (unit: %).

	Test(a)					Test(b)				
Model	IoU_1	IoU_2	Kappa	OA	MCC	IoU_1	IoU_2	Kappa	OA	MCC
CSF	75.55	65.48	65.34	83.30	65.86	71.88	69.83	65.88	82.97	65.89
IPTD	81.09	71.98	73.45	87.27	74.48	79.32	77.42	75.74	87.90	75.78
PointNet++	87.22	85.71	84.72	92.07	85.81	85.59	84.21	84.09	90.06	84.11
RandLA-Net	92.44	90.62	91.16	95.63	91.30	88.68	88.71	88.03	94.01	88.19
SCF-Net	93.02	91.30	91.84	95.97	91.95	87.99	88.43	87.51	93.74	87.92
Point Transformer	94.57	92.48	93.31	96.75	93.46	91.53	90.68	90.70	95.35	90.77
Stratified Transformer	95.80	92.77	94.22	96.15	94.22	92.57	91.37	92.18	96.09	92.21
SPoTransformer	95.49	95.21	94.85	96.95	95.42	92.42	92.18	91.99	95.99	92.02
KPConv	92.21	89.70	90.52	95.36	90.53	89.16	88.54	88.19	94.10	88.19
Multi-KPConv	95.50	94.41	94.87	96.97	94.87	92.83	92.62	92.45	96.22	92.48

performance in the evaluation metric IoU_2 (95.21). The performance of CSF, IPTD, PointNet++, and SCF-Net degrades in the Test(b) scenario, suggesting that these models lack sufficient adaptability to complex terrain. Figures 5 and 6 present the qualitative analysis results. The misclassified point cloud points of PointNet++ are primarily distributed in the gully area. In Test(a), IPTD, RandLA-Net and SCF-Net misclassify a large number of non-ground points as ground points, while in Test(b), the misclassified points are mainly located in the terrace area. Conversely, CSF and Point Transformer misclassifies a large number of ground points as non-ground points, resulting in sparse ground points. The SPoTransformer with Multi-KPConv exhibits relatively few misclassified points in both Test(a) and Test(b) scenarios, with no obvious distribution pattern, and demonstrates high classification accuracy. Multi-KPConv demonstrates a more substantial improvement in accuracy compared to KPConv.

4.2.2. Analysis of Test(c) and Test(d) results

Table 2 presents the quantitative results. In Test(c), the Multi-KPConv model exhibits excellent performance across all evaluation indices, achieving accuracies of 90.51% in IoU_2 and 91.84% in MCC,

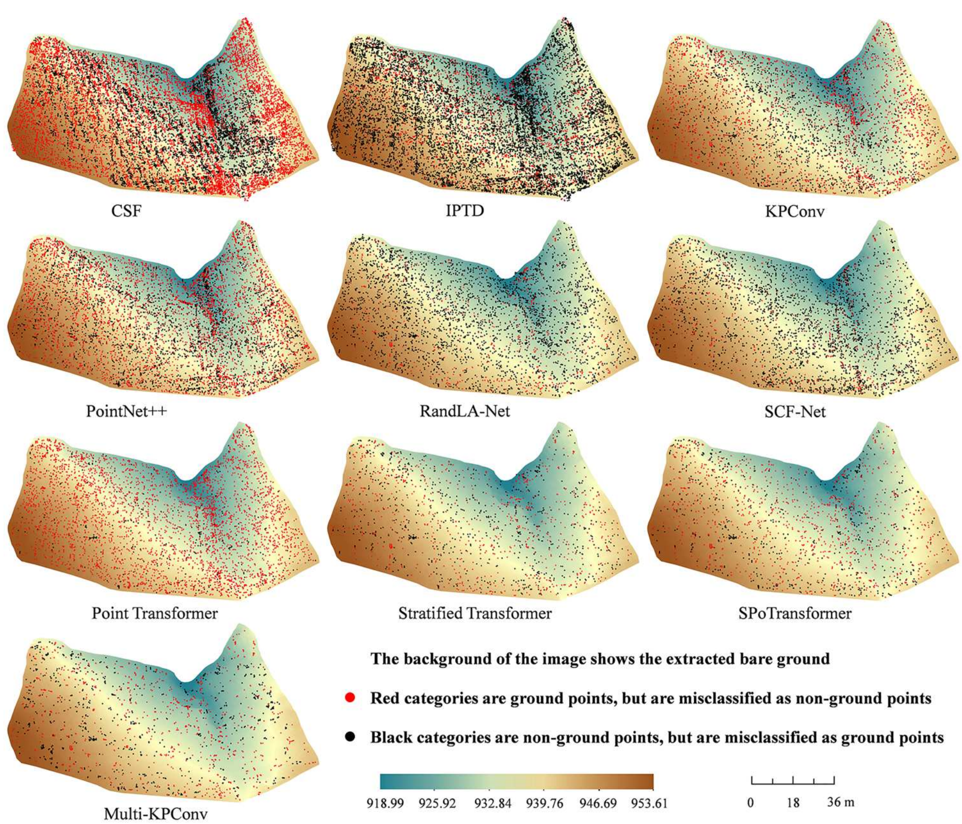

Figure 5. Qualitative analysis of the six methods in Test(a) (The extracted bare ground is shown in the background.).

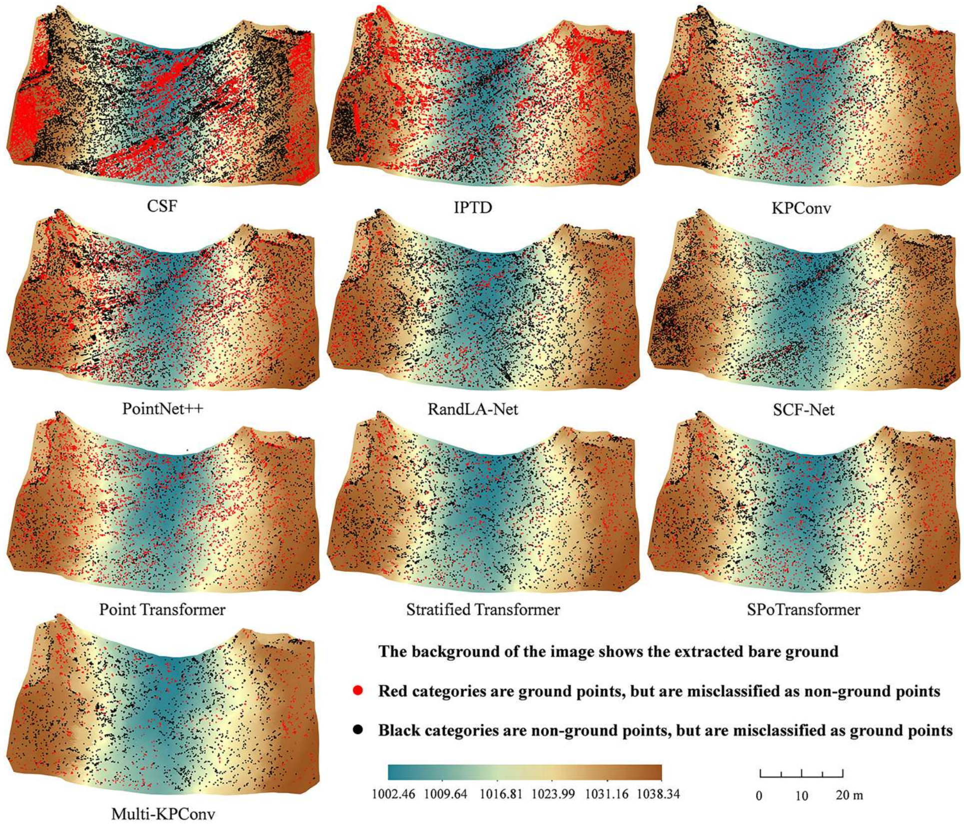

Figure 6. Qualitative analysis of the six methods in Test(b) (The extracted bare ground is shown in the background.).

demonstrating its high classification accuracy and consistency. The Stratified Transformer model achieves an overall accuracy (OA) of 95.83%, the highest correct classification rate among the models. In Test(d), the Multi-KPConv model outperforms others, particularly in the Kappa and OA metrics, with scores of 92.73% and 96.36%, highlighting the model's high reliability and accuracy. The Stratified Transformer model excels in the IoU_1 and MCC metrics, at 92.83% and 92.56%, respectively. The SPoTransformer and Stratified Transformer models excel in the IoU_1 and MCC metrics, while the remaining models underperform in several metrics.

Figures 7 and 8 display the results of the qualitative analysis; a significant number of misclassifications occur in CSF and IPTD, with their accuracy being lower than that of other deep learning models. Most of the misclassification points in PointNet++, RandLA-Net, and Point Transformer involve over-segmentation; that is, they excessively focus on certain categories, resulting in a category imbalance. SCF-Net, Stratified Transformer, and SPoTransformer misclassified a large number of non-ground points (i.e. vegetation points) in the map as ground points. In the gully area, the lush vegetation and lack of RGB information caused the high-density vegetation point cloud to appear highly similar in character to the slopes, leading

Table 2. Quantitative evaluation results of the six methods in Test(c) and Test(d) (unit: %).

	Test (c)					Test (d)				
Model	IoU_1	IoU_2	Kappa	OA	MCC	IoU_1	IoU_2	Kappa	OA	MCC
CSF	79.98	68.59	70.29	86.07	70.56	76.21	75.83	72.76	86.38	72.79
IPTD	83.54	75.29	76.93	89.04	76.95	81.78	79.84	78.79	89.42	78.96
PointNet++	88.68	80.62	83.29	92.30	83.53	85.55	84.43	83.77	91.90	83.82
RandLA-Net	89.11	84.19	85.67	93.11	85.86	86.51	87.09	85.91	92.94	86.34
SCF-Net	91.37	87.27	88.69	94.58	88.83	88.86	89.28	88.46	94.22	88.85
Point Transformer	92.45	87.65	89.50	95.08	89.61	90.44	89.41	89.40	94.71	89.53
Stratified Transformer	93.34	88.98	91.29	95.83	91.36	92.83	90.71	92.50	96.25	92.56
SPoTransformer	92.05	88.22	89.61	95.02	89.73	92.65	92.45	92.26	96.13	92.30
KPConv	92.47	87.83	89.61	95.12	89.67	90.12	89.81	89.44	94.72	89.45
Multi-KPConv	93.72	90.51	91.78	95.37	91.84	93.04	92.92	92.73	96.36	92.78

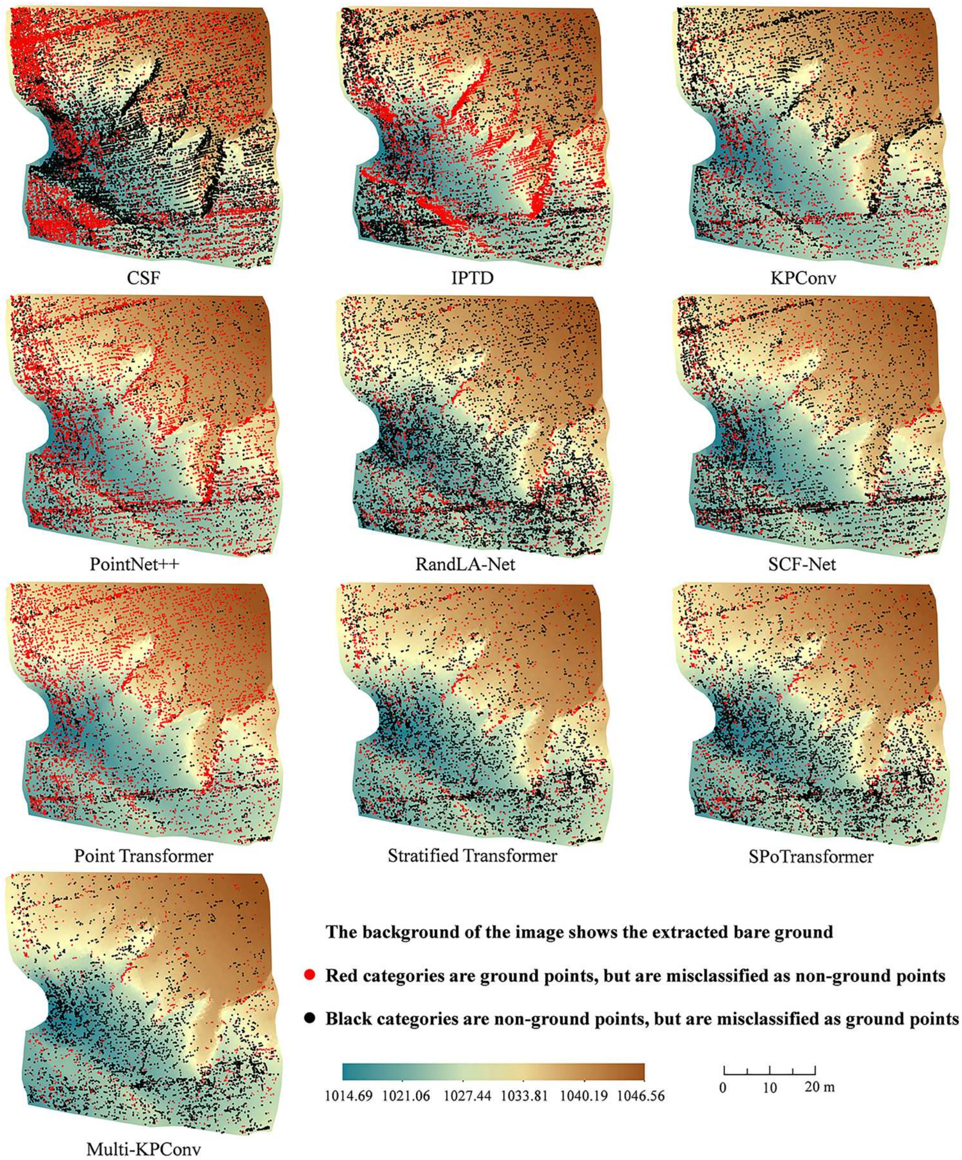

Figure 7. Qualitative analysis of the six methods in Test(c) (The extracted bare ground is shown in the background.).

to the model misclassifications. Multi-KPConv, by incorporating shallow features and utilizing multi-dimensional integrated kernel point convolution approaches to separately extract multiple receptive fields, learns global features through larger receptive fields and local features through smaller ones. This enables the model to accurately identify different types of point clouds. Consequently, Multi-KPConv can accurately discriminate between vegetation and slope points, enhancing the model's classification accuracy.

Table 3. Quantitative evaluation results of the six methods in Test(e) and Test(f) (unit: %).

	Test (e)					Test (f)				
Model	IoU_1	IoU_2	Kappa	OA	MCC	IoU_1	IoU_2	Kappa	OA	MCC
CSF	87.79	67.67	74.41	90.27	75.90	77.62	81.35	77.35	88.68	78.94
IPTD	88.96	74.09	79.28	91.61	79.34	82.57	81.17	80.09	90.05	80.41
PointNet++	85.55	84.43	83.77	91.90	83.82	84.12	85.16	83.38	91.70	83.60
RandLA-Net	86.51	87.09	85.91	92.94	86.34	88.71	87.80	87.54	93.77	87.86
SCF-Net	88.86	89.28	88.46	94.22	88.85	89.91	90.71	89.83	94.92	90.16
Point Transformer	90.44	89.41	89.40	94.71	89.53	90.32	90.41	89.87	94.94	89.87
Stratified Transformer	92.03	90.21	92.50	95.25	92.36	92.17	91.26	92.18	95.14	92.28
SPoTransformer	94.43	91.64	93.82	96.43	92.84	91.83	91.57	91.34	95.67	91.40
KPConv	94.23	86.15	89.59	95.76	89.62	84.64	86.06	84.23	92.12	84.68
Multi-KPConv	96.61	92.04	94.13	97.56	94.15	92.75	92.71	92.46	96.23	92.46

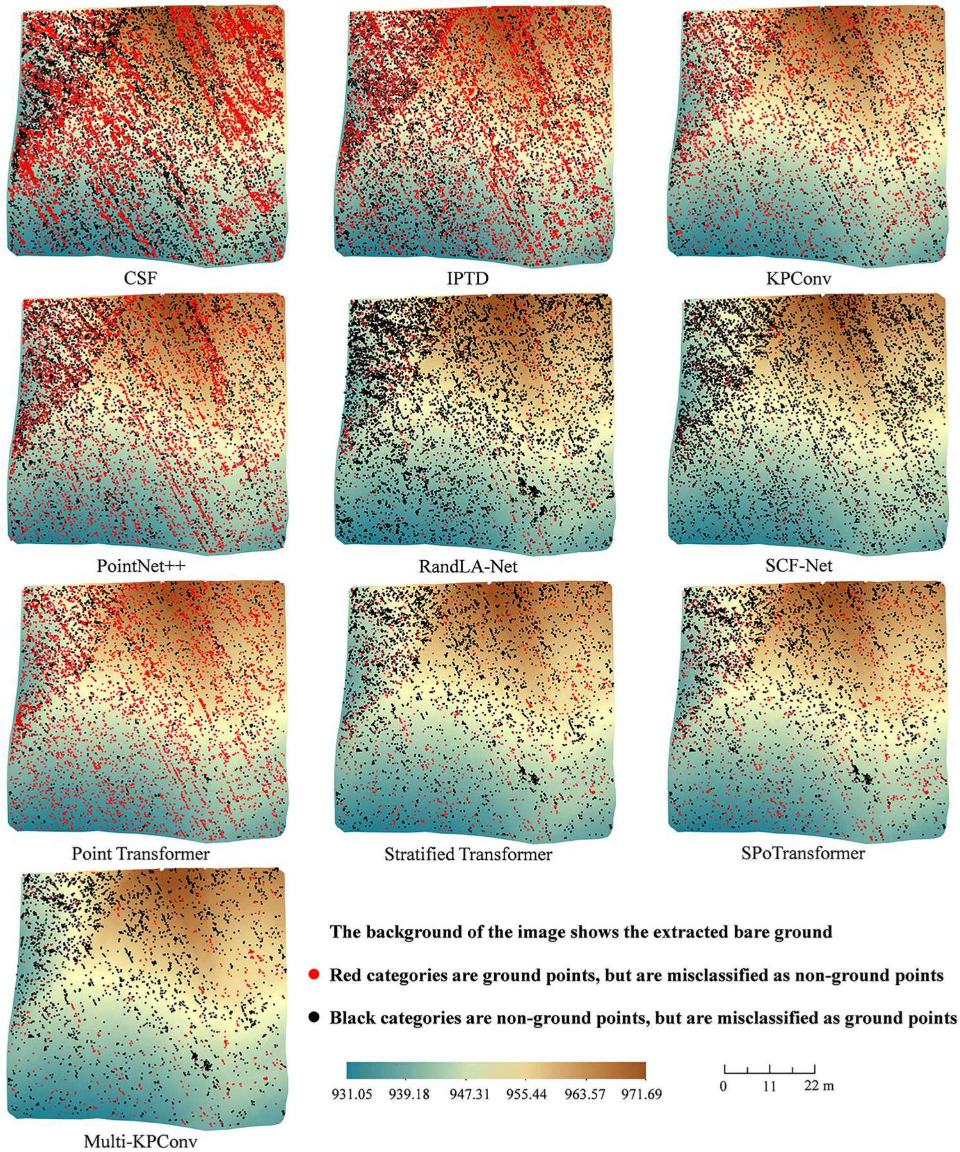

Figure 8. Qualitative analysis of the six methods in Test(d) (The extracted bare ground is shown in the background.).

4.2.3. Analysis of Test(e) and Test(f) results

Table 3 compares model performance using IoU_1, IoU_2, Kappa, OA, and MCC metrics. CSF, IPTD, PointNet++, RandLA-Net, KPConv, and SCF-Net show lower accuracy, while other models perform better. Figures 9 and 10 show that Multi-KPConv performs well in classifying difficult slope terrains, though some misclassification errors remain. In Figure 10, only the lowest elevation points are labeled as ground due to undulating terrain. For flat terrains, points within 1–3 cm above the lowest elevation are also labeled as ground. However, the model only predicts the lowest points as ground, leading to errors in labeling ground points as non-ground. In Test(f)'s flat terrain, all models show more misclassifications. Multi-KPConv maintains higher accuracy thanks to its 3D attention mechanism, which improves key feature representation and reduces the impact of complex scenes.

5. Discussions

5.1. Comparative analysis with the current most popular methods

As evidenced by Tables 1–3, the traditional filtering algorithm is consistently less accurate than the other deep learning models across all test sets. A reliable ground point cloud is crucial for applications such as

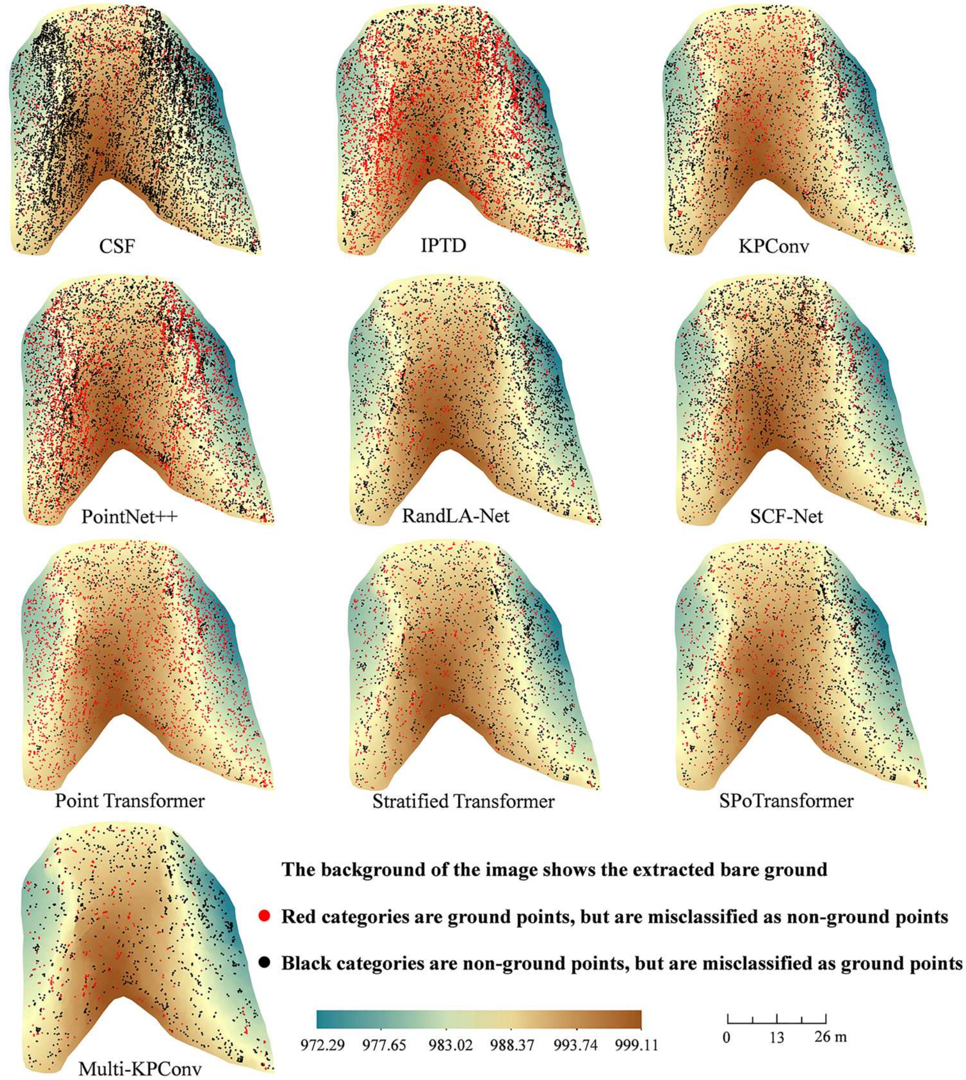

Figure 9. Qualitative analysis of the six methods in Test(e) (The extracted bare ground is shown in the background.).

geohazard monitoring and environmental protection. Consequently, this study evaluates the performance improvements of Multi-KPConv over the comparative model in terms of OA and ground point IoU. In Test(a), IoU and OA were enhanced by 0–28.93% and 0.02%–13.67%, respectively. Similar gains were observed in subsequent tests: in Test(b) improvements ranged from 0.44%–22.79% for IoU and from 0.13%–13.25% for OA; in Test(c), from 1.53%–21.92% and from 0.24%–9.3%, respectively; in Test(d), from 0.47%–17.09% and from 0.11%–9.98%; in Test(e), from 0.4%–24.37% and from 1.13%–7.29%; and in Test(f), from 1.14%–11.54% and from 0.56%–7.55%. Compared to other point cloud semantic segmentation models, Multi-KPConv consistently outperforms in nearly all evaluation metrics.

The results in Figure 5 through 10 indicate that current mainstream deep learning models tend to overemphasize a specific category, leading to numerous misclassified points in the test set, categorized either as ground or non-ground points. For example, in Figure 1, RandlaNet incorrectly classified many non-ground points as ground, whereas in Figure 10, as the scene shifted, it misclassified numerous ground points as non-ground. The comparative models struggle to capture the point cloud features of complex objects both comprehensively and accurately. This paper's method enhances the capture of subtle local spatial information by integrating shallow features, expanding the effective receptive field of each point through multi-dimensional kernel points, incorporating a three-dimensional feature attention mechanism, and preventing excessive focus on a single feature class. This approach not only surpasses other models in almost all metrics but also robustly extracts ground point clouds.

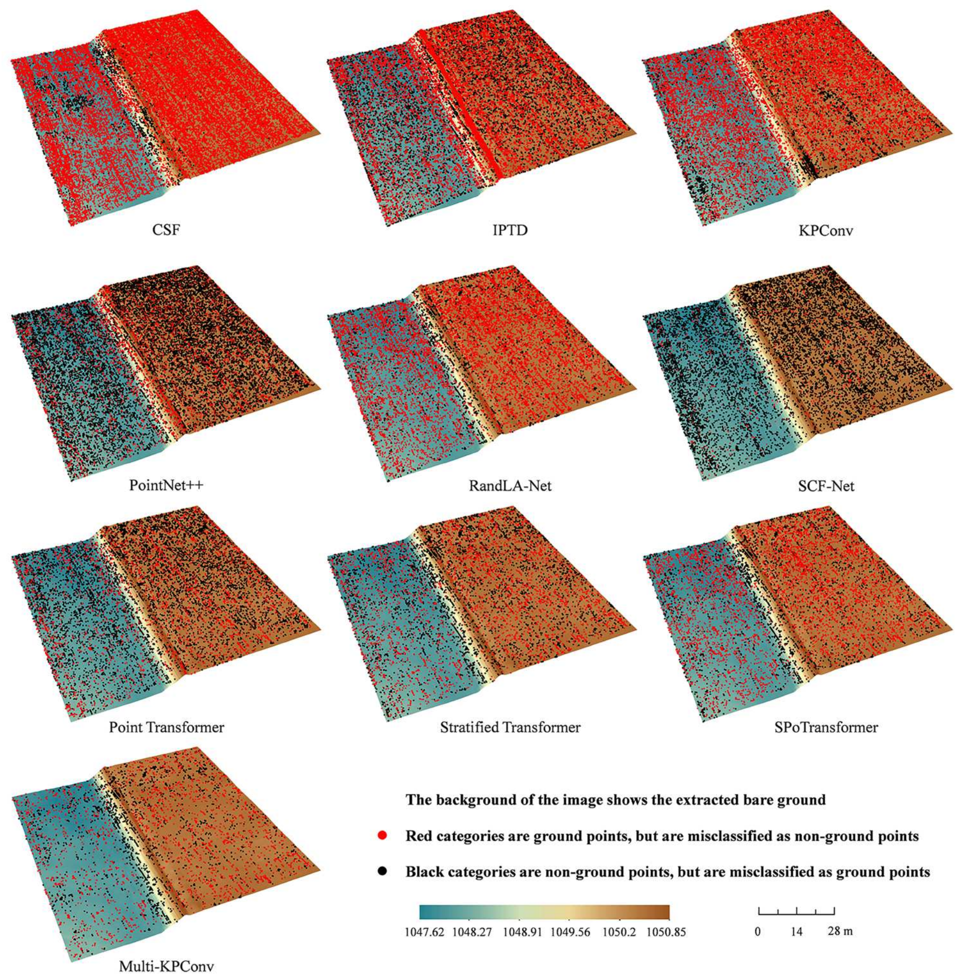

Figure 10. Qualitative analysis of the six methods in Test(f) (The extracted bare ground is shown in the background.).

5.2. Large scale complex terrain model testing

To further verify the applicability and robustness of the proposed method under multiple complex terrain conditions, a large-scale test scenario encompassing various typical landforms (including terrains such as ravines, slopes, and terraces from Test(a) to Test(f)) is constructed, and ground point extraction experiments are conducted in this scenario. Meanwhile, to assess the advantages of this paper's method in terms of accuracy improvement, Stratified Transformer and SPoTransformer, which perform excellently in small scenes, are selected as comparative models and compared with the Multi-KPConv method proposed in this paper.

As shown in Table 4, the proposed method Multi-KPConv achieves optimal performance across all evaluation metrics, with IoU_1 and IoU_2 reaching 94.48% and 91.67%, respectively, significantly outperforming Stratified Transformer (92.36%, 87.65%) and SPoTransformer (92.96%, 88.56%). In terms of the Kappa coefficient and MCC, Multi-KPConv achieves 92.27% and 92.28%, respectively, which represent significant improvements compared to other methods. Additionally, the method proposed in this paper achieves an OA of 96.41%, outperforming the comparison models with 95.05% and 95.44%, respectively. The qualitative comparison results of the various methods in the ground point extraction task are further demonstrated in

Table 4. Ground point extraction results for a wide range of complex terrain.

Model	IoU_1	IoU_2	Kappa	OA	MCC
Stratified Transformer	92.36	87.65	89.45	95.05	89.48
SPoTransformer	92.96	88.56	90.29	95.44	90.32
Multi-KPConv	94.48	91.67	92.27	96.41	92.28

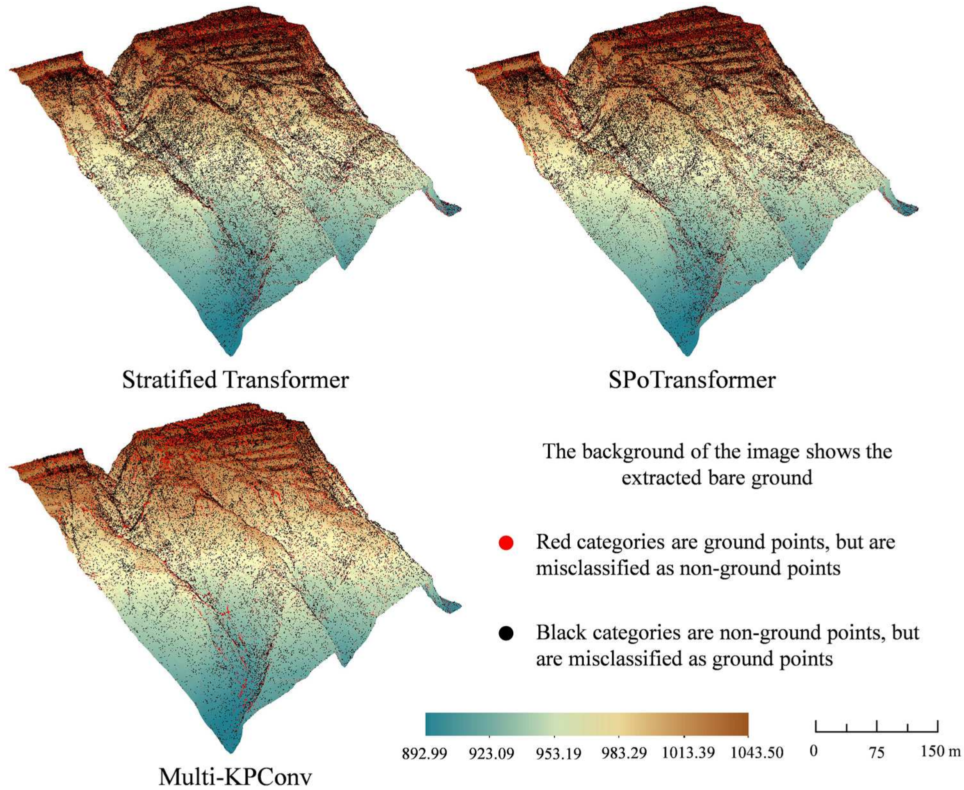

Figure 11. Results of qualitative analysis of large-scale complex terrain.

the large-scale, complex terrain scene shown in Figure 11. It can be observed that Stratified Transformer and SPoTransformer have a significant number of red misclassified points in regions with dense gullies and considerable slope undulations, indicating that these methods are prone to ground point omission in areas with high terrain complexity and irregular point cloud density distribution. Furthermore, there are more black misclassified points in localized areas, indicating that the model misidentifies non-ground points when dealing with areas featuring severe feature occlusion or high contrast vegetation. In contrast, the Multi-KPConv method proposed in this paper demonstrates superior classification performance in this scenario. The number of red and black misclassified points in the figure is significantly reduced, with a sparser spatial distribution and clearer boundaries, reflecting the method's stronger ability in geometric feature modeling and classification robustness under irregular geomorphic structures and dense vegetation interference. These results fully demonstrate that Multi-KPConv exhibits stronger ground point recognition capability and higher accuracy robustness when handling large-scale and complex terrain conditions. It effectively adapts to diverse terrain and non-uniform point cloud distribution scenarios, showcasing strong generalization performance.

5.3. Ablation experiment

To investigate the effectiveness of Multi-KPConv, we conducted a series of ablation experiments in Test(c) and Test(d), where the point cloud is relatively sparse, and the terrain is the most complex: Multi-

Table 5. Results of ablation experiments

	Test(c)					Test(d)				
Model	IoU_1	IoU_2	Kappa	OA	MCC	IoU_1	IoU_2	Kappa	OA	MCC
Multi-$KPConv_{PC}$	92.60	89.66	89.58	95.15	89.63	91.55	90.52	90.36	94.52	89.39
Multi-$KPConv_{AH}$	92.63	89.71	89.60	95.17	89.68	91.55	90.54	90.37	94.53	89.41
Multi-$KPConv_{SC}$	92.65	89.73	89.63	95.20	89.71	91.59	90.57	90.41	94.61	89.46
Multi-$KPConv_{HM}$	92.65	89.79	89.66	95.31	89.68	91.60	90.56	90.38	94.58	89.45
Multi-$KPConv_{RGB}$	92.40	89.52	89.57	95.12	89.55	91.75	90.67	90.54	94.64	89.51
Multi-$KPConv_{SA}$	92.96	89.81	90.76	95.38	90.83	92.24	92.10	91.85	95.32	91.91
Multi-$KPConv_{KP}$	91.29	86.18	88.02	94.35	88.04	89.90	89.81	89.31	94.35	89.38
Our	93.72	90.51	91.78	96.07	91.84	93.04	92.92	92.73	96.36	92.78

$KPConv_{PC}$ is a model that only inputs point coordinates to validate the effectiveness of the overall shallow features; Multi-$KPConv_{AH}$ is a model that only inputs point coordinates with mean elevation to verify the effectiveness of the mean elevation feature; Multi-$KPConv_{SC}$ is a model that only inputs point coordinates and surface curvature to verify the effectiveness of the surface curvature feature; Multi-$KPConv_{HM}$ is a model that only inputs point coordinates and higher-order moments to verify the effectiveness of the higher-order moments feature; Multi-$KPConv_{RGB}$ is a model that only inputs point coordinates and RGB features to verify the effectiveness of the RGB feature in complex terrain; Multi-$KPConv_{SA}$ is a model that omits the SimAM 3D attention module to verify the effectiveness of the attention mechanism, while Multi-$KPConv_{KP}$ uses the original KPConv to extract high-level semantic features instead of Multi-KPConv, verifying the effectiveness of Multi-KPConv.

Table 5 presents the results of the ablation experiments calculated according to Eqs. (20-25). In different geomorphic regions, introducing geometric relationship-based features in the shallow layer of the network effectively improves ground point extraction accuracy. Specifically, when the proposed enhanced features, based on different geometric relations constructed in the neighborhood, are combined and input, the OA improves by approximately 0.92%–1.84%, and the ground point extraction accuracy improves by about 0.85%–2.4%. This quantitative analysis demonstrates that neighborhood-based geometric relation embedding enhances the geometric semantic representation of sparse point clouds, allowing the model to mine and capture effective geometric information more effectively. Notably, the mean elevation, surface curvature, and higher-order moment features consistently enhance model performance across different terrain conditions. However, due to differences in their geometric construction relationships, the model's enhancement may vary from feature to feature.

In the ablation with RGB features, we find that OA is ineffective in Test(c) but superior in Test(d). This phenomenon is attributed to various factors, including the complexity of the terrain structure, vegetation cover features, point cloud density distribution, and the stability of RGB feature expression. Test set (c) represents the Loess Plateau ravine geomorphology area, characterized by dramatic topographic variations and deep, intertwined ravines, resulting in an uneven distribution of laser point clouds. The geometric structure of local neighborhoods is unstable, which severely interferes with the model's ability to perceive topographic features and learn classification boundaries. Moreover, the region is densely vegetated, heavily shaded, and experiences significant lighting variations, resulting in numerous shadows and reflective areas in the RGB image. This weakens the expressiveness of optical features and may introduce pseudo-differences. This unstructured and unstable optical information easily misleads the depth model, reducing the effectiveness of RGB and geometric feature fusion. In contrast, although test set (d) is also densely vegetated, its geomorphology is a continuous, regular convex slope with gentle terrain changes. While the point cloud density remains inhomogeneous, the overall spatial consistency is stronger, and the model is more likely to extract stable geometric structural features. Additionally, the RGB features in this region are less affected by occlusion and illumination, with relatively clear target boundaries, making the RGB and point cloud geometric features more complementary and improving the model's overall recognition ability. In the Loess Plateau region, where ravine terrain predominates, the use of RGB features introduces errors. However, the shallow features proposed in this paper effectively address this challenge.

The addition of the SimAM 3D attention module improves ground point extraction accuracy (IoU_2) by 0.7%–0.82% and OA by 0.69%–1.04%. SimAM calculates dynamic feature weights by learning the correlation information between channels and space, highlighting important features while suppressing the interference of irrelevant ones. After adding Multi-KPConv, ground point extraction accuracy (IoU_2) improves by 3.11%–4.33%, OA improves by 1.72%–2.1%, and Kappa improves by 3.42%–3.76%. Compared to the original KPConv, Multi-KPConv captures more contextual information through different sensory fields, enabling multilevel feature learning and improving the model's accuracy and consistency.

5.4. The analysis of generalization

To validate the generalization capability of Multi-KPConv, we performed a generalization assessment on the STPLS3D dataset (Chen et al. 2022). The dataset comprises a diverse and high-density urban point cloud, covering an area of approximately 1.27 square kilometers and is intended to serve as benchmark

Table 6. Classification results based on STPLS3D dataset.

Model	mIoU(%)	OA(%)	Per Class IoU(%)					
			Ground	Building	Tree	Car	Light pole	Fence
RandLA-Net	68.81	90.68	89.46	88.26	79.23	71.81	58.52	25.58
SCF-Net	69.03	90.86	89.06	88.3	79.91	75.39	44.91	36.61
Point Transformer	58.92	87.26	90.03	85.92	81.36	38.28	48.36	9.54
Stratified Transformer	61.12	88.37	90.76	87.52	81.53	44.39	49.86	12.65
SPoTransformer	63.60	90.38	90.94	88.64	87.40	46.71	49.57	18.34
Our	87.19	96.12	94.41	95.01	90.19	88.15	79.92	75.48

data for deep learning semantic and instance segmentation. Consequently, it is utilized in this study for testing model generalization. In this paper, the real-world USC, WMSC, and OCCC sections of the STPLS3D dataset are used as the training set, and RA is used as the test set. RandLA-Net, SCF-Net, Point Transformer, Stratified Transformer, and SPoTransformer are selected as the models for the comparative analysis. The point cloud is divided into six categories for model training: 'Ground' (32.15%), 'Building' (40.35%), 'Tree' (25.38%), 'Car' (1.08%), 'Light pole' (0.18%), and 'Fence' (0.85%). The corresponding point cloud classification results are shown in Table 6. Our method outperforms the comparative models in mIoU and OA across all categories, with the IoU exceeding 90% for Ground, Building, and Tree. From Table 6 and Figure 12, it is evident that the Transformer-based model exhibits a classification imbalance in the category 'Fence'. Compared to other methods, Multi-KPConv achieves higher accuracy with a smaller dataset and demonstrates strong generalization capabilities.

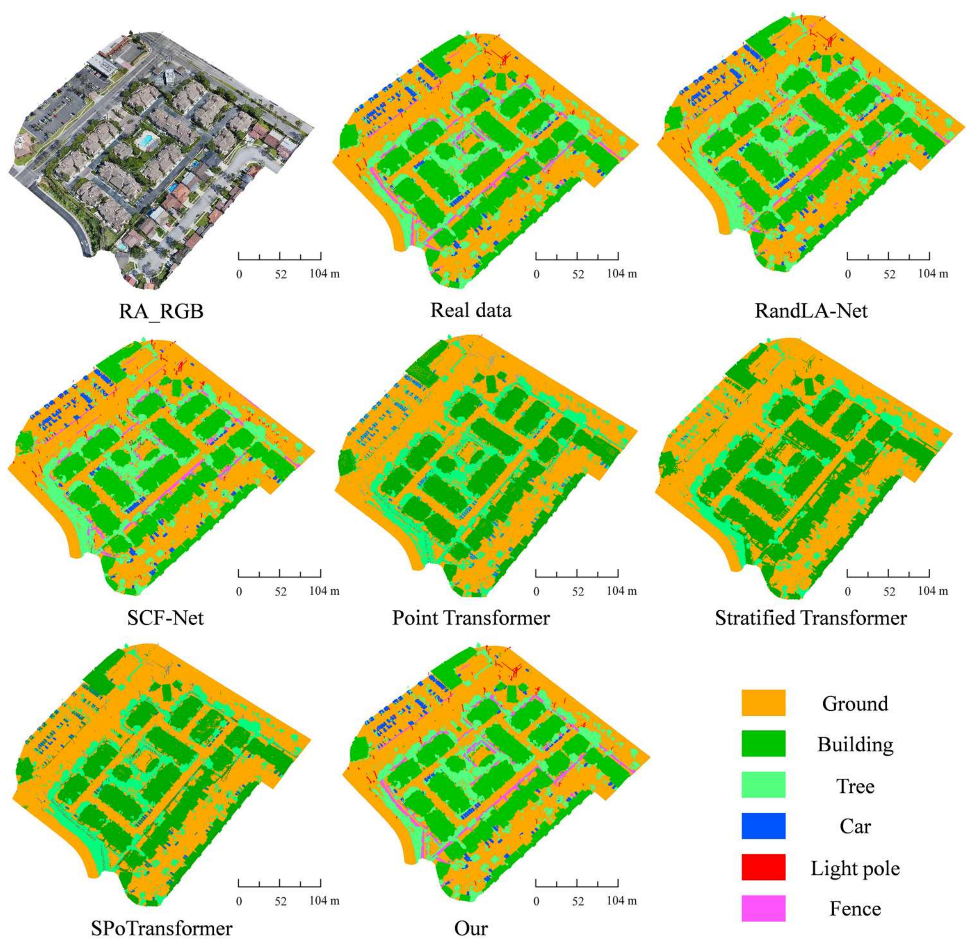

Figure 12. Classification results based on the STPLS3D dataset.

6. Conclusion

To accurately extract ground points from complex terrain point clouds, this study introduces Multi-KPConv, a novel point cloud semantic segmentation network. Given the complex topography of the Loess Plateau, directly assigning RGB information to the point clouds introduces significant noise. Consequently, this study integrates domain knowledge and designs shallow features to enrich the initial data input to the convolutional layers, thereby enhancing the network's capacity to learn more comprehensive and detailed contextual information and improve performance. Subsequently, multi-dimensional integrated kernel point convolution is employed to extract high-level semantic features from various receptive fields, enhancing the algorithm's understanding of essential data features, reducing overfitting risks, and improving generalization capabilities. Ultimately, the SimAM 3D attention network is incorporated to emphasize important features within the 3D scene, mitigate environmental interference, and enhance the model's object localization and feature expression capabilities, thereby increasing the accuracy of ground point cloud segmentation. Quantitative and qualitative analyses from Test(a) to Test(f) demonstrate that the method described in this paper surpasses prevalent point cloud semantic segmentation methods, positioning Multi-KPConv as a novel approach for intelligent point cloud filtering in complex terrains like the Loess Plateau.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (No.52274168). We acknowledge all researchers who contributed to the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Natural Science Foundation of China: [Grant Number 52274168].

Data availability

Data is available on request from the authors. The data in this study include topographic, which is not publicly released because it involves national energy-related data. The data that support the findings of this study are available from the corresponding author, [Fuquan T], upon reasonable request.

References

Aditya, A., B. Lohani, J. Aryal, and S. Winter. 2024. "Benchmarking Deep Learning Architectures for Urban Vegetation Point Cloud Semantic Segmentation from MLS." *IEEE Transactions on Geoscience and Remote Sensing* 62:1–14. https://doi.org/10.1109/TGRS.2024.3381976.

Atzmon, M., H. Maron, and Y. Lipman. 2018. "Point Convolutional Neural Networks by Extension Operators." *ACM Transactions on Graphics* 37 (4): 71:1–71-12. https://doi.org/10.1145/3197517.3201301.

Balado, J., P. van Oosterom, L. Díaz-Vilariño, and M. Meijers. 2020. "Mathematical Morphology Directly Applied to Point Cloud Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 168:208–220. https://doi.org/10.1016/j.isprsjprs.2020.08.011.

Benouini, R., I. Batioua, K. Zenkouar, A. Zahi, H. E. Fadili, and H. Qjidaa. 2019. "Fast and Accurate Computation of Racah Moment Invariants for Image Classification." *Pattern Recognition* 91:100–110. https://doi.org/10.1016/j.patcog.2019.02.014.

Ben-Shabat, Y., M. Lindenbaum, and A. Fischer. 2018. "3DmFV: Three-Dimensional Point Cloud Classification in Real-Time Using Convolutional Neural Networks." *IEEE Robotics and Automation Letters* 3:3145–3152. https://doi.org/10.1109/LRA.2018.2850061.

Boulch, A. 2020. "ConvPoint: Continuous Convolutions for Point Cloud Processing." *Computers & Graphics* 88:24–34. https://doi.org/10.1016/j.cag.2020.02.005.

Boulch, A., B. L. Saux, and N. Audebert. 2017. "Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks." In *Proceedings of the Workshop on 3D Object Retrieval (3Dor '17, 2017)*, 17–24. Goslar, DEU: Eurographics Association. https://doi.org/10.2312/3dor.20171047.

Cao, Zhen, MiX. Qiu, B. Cao, C. ZhipengLong, X. Yan, C. Zheng, Z. Dong, and B. Yang. 2025. "Cross-modal Semantic Transfer for Point Cloud Semantic Segmentation." *ISPRS Journal of Photogrammetry and Remote Sensing* 221:265–279. https://doi.org/10.1016/j.isprsjprs.2025.01.024.

Chen, M. D., Q. Y. Hu, Z. F. Yu, H. Thomas, A. Feng, Y. Hou, K. McCullough, F. Ren, and L. Soibelman. 2022. STPLS3D: A Large-Scale Synthetic and Real Aerial Photogrammetry 3D Point Cloud Dataset. https://doi.org/10.48550/arXiv.2203.09065.

Chen, N., N. Wang, Y. He, X. Ding, and J. Kong. 2023. "An Improved Progressive Triangular Irregular Network Densification Filtering Algorithm for Airborne LiDAR Data." *Frontiers in Earth Science* 10: 1015153. https://doi.org/10.3389/feart.2022.1015153.

Chen, Z. Y., B. Xu, and B. B. Gao. 2016. "An Image-Segmentation-Based Urban DTM Generation Method Using Airborne Lidar Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9:496–506. https://doi.org/10.1109/JSTARS.2015.2512498.

Chiang, H. Y., Y. L. Lin, Y. C. Liu, and W. H. Hsu. 2019. A Unified Point-Based Framework for 3D Segmentation. https://doi.org/10.48550/arXiv.1908.00478.

Demantké, J., C. Mallet, N. David, and B. Vallet. 2012. "Dimensionality Based Scale Selection in 3d Lidar Point Clouds." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-5/W12:97–102.

Du, A., T. Zhou, S. Pang, Q. Wu, and J. Zhang. 2024. "PCL: Point Contrast and Labeling for Weakly Supervised Point Cloud Semantic Segmentation." *IEEE Transactions on Multimedia* 26:8902–8914. https://doi.org/10.1109/TMM.2024.3383674.

Engel, N., V. Belagiannis, and K. Dietmayer. 2021. "Point Transformer." *IEEE Access* 9:134826–134840. https://doi.org/10.1109/ACCESS.2021.3116304.

Engelmann, F., T. Kontogianni, A. Hermans, and B. Leibe. 2019. Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds. https://doi.org/10.1109/ICCVW.2017.90.

Engelmann, F., T. Kontogianni, and B. Leibe. 2020. Dilated Point Convolutions: On the Receptive Field Size of Point Convolutions on 3D Point Clouds. https://doi.org/10.48550/arXiv.1907.12046.

Fan, S. Q., Q. L. Dong, F. H. Zhu, Y. Lv, P. Ye, and F.-Y. Wang. 2021. "SCF-Net: Learning Spatial Contextual Features for Large- Scale Point Cloud Segmentation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, 14504–14513. IEEE. https://openaccess.thecvf.com/content/CVPR2021/html/Fan_SCF- Net_Learning_Spatial_Contextual_Features_for_Large- Scale_Point_Cloud_Segmentation_CVPR_2021_paper.html.

Feng, Y. F., Z Z. Zhang, X. B. Zhao, R. Ji, and Y. Gao. 2018. "GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, 264–272. IEEE. https://openaccess.thecvf.com/content_cvpr_2018/html/Feng_GVCNN_Group-View_Convolutional_CVPR_2018_paper.html.

Guo, B., X. F. Huang, F. Zhang, and G. Sohn. 2015. "Classification of Airborne Laser Scanning Data Using JointBoost." *ISPRS Journal of Photogrammetry and Remote Sensing, High-Resolution Earth Imaging for Geospatial Information* 100:71–83. https://doi.org/10.1016/j.isprsjprs.2014.04.015.

Hao, S. J., Y. Zhou, and Y. R. Guo. 2020. "A Brief Survey on Semantic Segmentation with Deep Learning." *Neurocomputing* 406:302–321. https://doi.org/10.1016/j.neucom.2019.11.118.

He, K. M., X. Y. Zhang, S. Q. Ren, and J. Sun. 2015. Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385.

Hermosilla, P., T. Ritschel, P.-P. Vázquez, À Vinacua, and T. Ropinski. 2018. "Monte Carlo Convolution for Learning on non-uniformly Sampled Point Clouds." *ACM Transactions on Graphics* 37 (6): 235:1–235:12. https://doi.org/10.1145/3272127.3275110.

Hsu, P. H., and Z. Y. Zhuang. 2020. "Incorporating Handcrafted Features into Deep Learning for Point Cloud Classification." *Remote Sensing* 12:3713. https://doi.org/10.3390/rs12223713.

Hu, Q. Y., B. Yang, L. H. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham. 2020. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. https://doi.org/10.48550/arXiv.1911.11236.

Hua, B.-S., M.-K. Tran, and S.-K. Yeung. 2018. Pointwise Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1712.05245.

Hui, Z. Y., Y. J. Hu, Y. Z. Yevenyo, and X. Yu. 2016. "An Improved Morphological Algorithm for Filtering Airborne LiDAR Point Cloud Based on Multi-level Kriging Interpolation." *Remote Sensing* 8:35. https://doi.org/10.3390/rs8010035.

Jing, Z., H. Guan, P. Zhao, D. Li, Y. Yu, Y. Zang, H. Wang, and J. Li. 2021. "Multispectral LiDAR Point Cloud Classification Using SE-PointNet++." *Remote Sensing* 13:2516. https://doi.org/10.3390/rs13132516.

Joseph-Rivlin, M., A. Zvirin, and R. Kimmel. 2019. Momen(e)t: Flavor the Moments in Learning to Classify Shapes. https://doi.org/10.48550/arXiv.1812.07431.

Kuprowski, M., and P. Drozda. 2023. "Feature Selection for Airbone LiDAR Point Cloud Classification." *Remote Sensing* 15:561. https://doi.org/10.3390/rs15030561.

Kushwaha, S. K. P., A. Dali, C. H. Singh, A. Rai, and K. Jain. 2023. Analysis of Shadow Interference on Chromatic Information in Unmanned Aerial Vehicle Optical Imagery, in: 2023 IEEE India Geoscience and Remote Sensing Symposium (InGARSS). IEEE, pp. 1–4.

Lai, X., J. H. Liu, L. Jiang, L. Wang, H. Zhao, S. Liu, X. Qi, and J. Jia. 2022. Stratified Transformer for 3D Point Cloud Segmentation. https://doi.org/10.48550/arXiv.2203.14508.

Lalonde, J.-F., N. Vandapel, D. F. Huber, and M. Hebert. 2006. "Natural Terrain Classification Using Three-Dimensional Ladar Data for Ground Robot Mobility." *Journal of Field Robotics* 23 (10): 839–861. https://doi.org/10.1002/rob.20134.

Lawin, F. J., M. Danelljan, P. Tosteberg, G. Bhat, F. S. Khan, and M. Felsberg. 2017. Deep Projective 3D Semantic Segmentation. https://doi.org/10.48550/arXiv.1705.03428.

Lei, H., N. Akhtar, and A. Mian. 2019. Octree guided CNN with Spherical Kernels for 3D Point Clouds. https://doi.org/10.48550/arXiv.1903.00343.

Li, B. C., and J. Shen. 1991. "Fast Computation of Moment Invariants." *Pattern Recognition* 24 (8): 807–813. https://doi.org/10.1016/0031-3203(91)90048-A.

Li, S. T., W. W. Song, L. Y. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson. 2019. "Deep Learning for Hyperspectral Image Classification: An Overview." *IEEE Transactions on Geoscience and Remote Sensing* 57:6690–6709. https://doi.org/10.1109/TGRS.2019.2907932.

Li, Y. Y., R. Bu, M. C. Sun, W. Wu, X. Di, and B. Chen. 2018. "PointCNN: Convolution on X-Transformed Points." *Advances in Neural Information Processing Systems* 31.

Liu, B. Y., Y. Xie, Z. G. Li, Y. Liang, W. Zhang, S. Fu, S. Yin, et al. 2020. The assessment of soil loss by water erosion in China. International Soil and Water Conservation Research, Soil erosion assessment tools and data; creation, consolidation, and harmonization – A special issue from the Global Symposium on Soil Erosion 2019 (Rome, FAO HQ). 8, 430–439. https://doi.org/10.1016/j.iswcr.2020.07.002.

Luo, Z., L. Gao, H. Xiang, and J. Li. 2023. "Road Object Detection for HD map: Full-Element Survey, Analysis and Perspectives." *ISPRS Journal of Photogrammetry and Remote Sensing* 197:122–144. https://doi.org/10.1016/j.isprsjprs.2023.01.009.

Maturana, D., and S. Scherer. 2015. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* 922–928. Hamburg, Germany: IEEE. https://doi.org/10.1109/IROS.2015.7353481.

Park, J., S. Lee, S. Kim, Y. Xiong, and H. J. Kim. 2023. "Self-positioning Point-Based Transformer for Point Cloud Understanding." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21814–21823. https://openaccess.thecvf.com/content/CVPR2023/html/Park_Self-Positioning_Point-Based_Transformer_for_Point_Cloud_Understanding_CVPR_2023_paper.html.

Phan, A. V., M. L. Nguyen, Y. L. H. Nguyen, and L. T. Bui. 2018. "DGCNN: A Convolutional Neural Network over Large-Scale Labeled Graphs." *Neural Networks* 108:533–543. https://doi.org/10.1016/j.neunet.2018.09.001.

Qi, C. R., H. Su, K. Mo, and L. J. Guibas. 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. https://doi.org/10.48550/arXiv.1612.00593.

Qi, C. R., H. Su, M. Niessner, A. Dai, M. Yan, and L. J. Guibas. 2016. Volumetric and Multi-view CNNs for Object Classification on 3D Data. https://doi.org/10.48550/arXiv.1604.03265.

Qi, C. R., L. Yi, H. Su, and L. J. Guibas. 2017b. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. https://doi.org/10.48550/arXiv.1706.02413.

Qin, N. N., X. Y. Hu, and H. M. Dai. 2018. "Deep Fusion of Multi-view and Multimodal Representation of ALS Point Cloud for 3D Terrain Scene Recognition." *ISPRS Journal of Photogrammetry and Remote Sensing, ISPRS Journal of Photogrammetry and Remote Sensing Theme Issue "Point Cloud Processing"* 143:205–212. https://doi.org/10.1016/j.isprsjprs.2018.03.011.

Quan, Y. N., J. F. Song, X. Guo, Q. Miao, and Y. Yang. 2017. "Filtering LiDAR Data Based on Adjacent Triangle of Triangulated Irregular Network." *Multimedia Tools and Applications* 76 (8): 11051–11063. https://doi.org/10.1007/s11042-016-3465-4.

Roynard, X., J.-E. Deschaud, and F. Goulette. 2018. Classification of Point Cloud Scenes with Multiscale Voxel Deep Network. https://doi.org/10.48550/arXiv.1804.03583.

Su, H., S. Maji, E. Kalogerakis, and E. Learned-Miller. 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. https://doi.org/10.48550/arXiv.1505.00880.

Susaki, J. 2012. "Adaptive Slope Filtering of Airborne LiDAR Data in Urban Areas for Digital Terrain Model (DTM) Generation." *Remote Sensing* 4:1804–1819. https://doi.org/10.3390/rs4061804.

Tang, F. Q., and Q. Yang. 2023. "Progress and Prospects of Multi-source Remote Sensing Monitoring Technology for Coal Mining Subsidence in Mining Areas of the Western Loess Plateau." *Coal Science and Technology* 51:9–26. https://doi.org/10.12438/cst.2023-1113.

Tchapmi, L. P., C. B. Choy, I. Armeni, J. Gwak, and S. Savarese. 2017. SEGCloud: Semantic Segmentation of 3D Point Clouds. arXiv preprint arXiv:1710.07563v1 (October 20, 2017). https://arxiv.org/abs/1710.07563v1.

Thomas, H., J.-E. Deschaud, B. Marcotegui, F. Goulette, and Y. L. Gall. 2018. Semantic Classification of 3D Point Clouds with Multiscale Spherical Neighborhoods. https://doi.org/10.48550/arXiv.1808.00495.

Thomas, H., C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas. 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. https://doi.org/10.48550/arXiv.1904.08889.

Wang, J. Y., R. Chakraborty, and S. X. Yu. 2022. "Transformer for 3D Point Clouds." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44:4419–4431. https://doi.org/10.1109/TPAMI.2021.3070341.

Wang, S. L., S. Suo, W. C. Ma, A. Pokrovsky, and R. Urtasun. 2018. "Deep Parametric Continuous Convolutional Neural Networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2589–2597. https://openaccess.thecvf.com/content_cvpr_2018/html/Wang_Deep_Parametric_Continuous_CVPR_2018_paper.html.

Wang, W. J., T. Wang, and Y. Cai. 2022. "Multi-view Attention-Convolution Pooling Network for 3D Point Cloud Classification." *Applied Intelligence* 52 (13): 14787–14798. https://doi.org/10.1007/s10489-021-02840-2.

Wang, Y., Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. 2019. "Dynamic Graph CNN for Learning on Point Clouds." *ACM Transactions on Graphics* 38:146:1–146:12. https://doi.org/10.1145/3326362.

Wang, Y. J. 2017. "Research Progress and Prospect on Ecological Disturbance Monitoring in Mining Area." *Acta Geodaetica et Cartographica Sinica* 46:1705–1716.

Weinmann, M., B. Jutzi, S. Hinz, and C. Mallet. 2015. "Semantic Point Cloud Interpretation Based on Optimal Neighborhoods, Relevant Features and Efficient Classifiers." *ISPRS Journal of Photogrammetry and Remote Sensing* 105:286–304. https://doi.org/10.1016/j.isprsjprs.2015.01.016.

Wu, X. T., Y. P. Wei, B. J. Fu, S. Wang, Y. Zhao, and E. F. Moran. 2020. "Evolution and Effects of the Social-Ecological System over a Millennium in China's Loess Plateau." *Science Advances* 6:eabc0276. https://doi.org/10.1126/sciadv.abc0276.

Xiao, A., J. X. Huang, D. Y. Guan, X. Zhang, S. Lu, and L. Shao. 2023. "Unsupervised Point Cloud Representation Learning with Deep Neural Networks: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45:11321–11339. https://doi.org/10.1109/TPAMI.2023.3262786.

Yang, B. S., and Z. Dong. 2013. "A Shape-Based Segmentation Method for Mobile Laser Scanning Point Clouds." *ISPRS Journal of Photogrammetry and Remote Sensing* 81:19–30. https://doi.org/10.1016/j.isprsjprs.2013.04.002.

Yang, L. X., R. Y. Zhang, L. D. Li, and X. Xie. 2021 "SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks." In *Proceedings of the International Conference on Machine Learning, PMLR*, 11863–11874. https://proceedings.mlr.press/v139/yang21o.html.

Yang, Q., F. Q. Tang, F. Wang, J. Y. Tang, Z. G. Fan, T. Ma, Y. Su, and J. Xue. 2023. "A new Technical Pathway for Extracting High Accuracy Surface Deformation Information in Coal Mining Areas Using UAV LiDAR Data: An Example from the Yushen Mining Area in Western China." *Measurement* 218:113220. https://doi.org/10.1016/j.measurement.2023.113220.

Yang, Y. B., N. N. Zhang, and X. L. Li. 2017. "Adaptive Slope Filtering for Airborne Light Detection and Ranging Data in Urban Areas Based on Region Growing Rule." *Survey Review* 49:139–146. https://doi.org/10.1080/00396265.2015.1135860.

Ye, M. S., S. J. Xu, and T. Y. Cao. 2020. HVNet: Hybrid Voxel Network for LiDAR Based 3D Object Detection. https://doi.org/10.48550/arXiv.2003.00186.

Ye, X. Q., J. M. Li, H. X. Huang, L. Du, and X. Zhang. 2018. "3D Recurrent Neural Networks with Context Fusion for Point Cloud Semantic Segmentation." In *Presented at the Proceedings of the European Conference on Computer Vision (ECCV)*, 403–417. https://openaccess.thecvf.com/content_ECCV_2018/html/Xiaoqing_Ye_3D_Recurrent_Neural_ECCV_2018_paper.html.

Zhang, L., X. T. Li, A. Arnab, K. Y. Yang, Y. Tong, and P. H. S. Torr. 2020. Dual Graph Convolutional Network for Semantic Segmentation. https://doi.org/10.48550/arXiv.1909.06121.

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

Zhao, X., Q. Guo, Y. Su, and B. Xue. 2016. "Improved Progressive TIN Densification Filtering Algorithm for Airborne LiDAR Data in Forested Areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 117:79–91. https://doi.org/10.1016/j.isprsjprs.2016.03.016.

Zhao, Z. Y., M. Liu, and K. Ramani. 2019. DAR-Net: Dynamic Aggregation Network for Semantic Scene Segmentation. https://doi.org/10.48550/arXiv.1907.12022.