Multi-scale Transformer-based classification of bathymetric LiDAR data in shallow water environments

Lida Asgharian Pournodrati [1], Mohammad Mahdi Baba [1], Jannis Gangelhoff [2], Uwe Sörgel [1]

[1] Institute for Photogrammetry and Geoinformatics, University of Stuttgart, Germany – {lida.asgharian-pournodrati, uwe.soergel @ifp.uni-stuttgart.de, mmahdiibaba@gmail.com}
[2] Fraunhofer Institute for Physical Measurement Techniques IPM, Freiburg, Germany – Jannis.Gangelhoff@ipm.fraunhofer.de

Keywords: Multiscale Dependency, Transformers, Bathymetric LiDAR, Point Cloud Classification, Shallow Waters.

Abstract

Bathymetric LiDAR data plays a crucial role in mapping underwater topography, enabling applications in coastal monitoring, environmental assessment, and seabed classification. However, the inherent complexity and noise in 3D bathymetric point clouds pose challenges for accurate classification. To address this, we propose a voxel-based method for efficient classification of bathymetric LiDAR data, moving beyond traditional point-wise processing of unstructured point sets. In our approach, 3D points are aggregated into structured voxel grids, and their features are embedded within each voxel. To capture spatial dependencies between voxels, we employ a window-based attention mechanism that partitions voxel features into local windows where self-attention is applied. To enhance contextual learning across regions, we adopt a shifted window strategy inspired by Swin3D, allowing voxels near window boundaries to interact with adjacent regions and reducing the locality limitation of fixed windows. To improve computational efficiency, we use a voxel selection mechanism. Using HDBSCAN, we cluster voxel features within each window based on density and retain representative voxels with distinct characteristics. This reduces redundant attention operations while preserving critical structural information. Furthermore, to capture both fine-grained and large-scale spatial patterns in bathymetric data, we design transformer heads grouped by scale. Each head group processes voxels from windows of varying sizes, enabling the model to learn multi-scale representations. The fused output captures both detailed local variations and broader contextual cues. Experimental results demonstrate the effectiveness of our method, achieving an overall classification accuracy of 75.4% on bathymetric LiDAR datasets, highlighting its capability in underwater terrain analysis.

1. Introduction

Bathymetric Light Detection and Ranging (LiDAR) technology plays a crucial role in mapping underwater environments, particularly in shallow waters where accurate classification is essential for applications such as coastal monitoring, habitat analysis, and hydrographic surveying. However, the complexity of shallow water regions, influenced by surface reflections (material), depth variations, and submerged vegetation, presents significant challenges for automated classification methods (Mandlburger et al., 2015; Rhomberg-Kauert et al., 2024).

Traditional classification approaches rely on handcrafted features, segmentation techniques, or conventional machine learning models, which may struggle to capture complex spatial relationships within the data. To address this, deep learning-based methods have been explored, leveraging feature extraction from point cloud data. However, many existing approaches have limitations in effectively modeling the dependencies between extracted features across different scales.

In this study, we propose a multi-scale transformer-based model to learn feature dependencies for bathymetric LiDAR data classification. Once features are extracted from raw point clouds, our model learns relationships between pre-extracted features. Extracted features can be handcrafted based on geometric properties and statistical descriptors or they can be derived from other feature extraction techniques like KPConv (Thomas et al., 2019). By incorporating multi-scale attention mechanisms with different window sizes for query and key features, the suggested approach effectively captures both local and global dependencies within the data. We evaluate our approach on a real-world bathymetric LiDAR dataset captured by a new developed sensor from Fraunhofer Institute for Physical Measurement Techniques (IPM).

The classification method introduced in this study is tailored to address the complex variability observed in LiDAR signal returns, which can arise due to fluctuating water clarity and varying levels of submerged vegetation. By integrating both spatial structures and spectral characteristics from the collected data, the method ensures accurate identification and separation of diverse environmental features.

This refined technique offers significant advancements for ecosystem assessment, water resource modeling, and environmental planning, positioning it as a practical solution for both academic exploration and operational use in marine and freshwater analysis. Through application to actual field datasets, the study highlights the method's strength in delivering high accuracy while maintaining efficient processing performance.

Key technical specifications of the system include a laser pulse duration of approximately 1 *ns* and a ground footprint measuring 5 *cm*. The dual-wavelength system operates with green light at 532 *nm* and infrared at 1064 *nm*, classified under laser safety category 2M. It supports a pulse emission frequency of 35,000 pulses per second, with data sampled at an ultra-high rate of 5 giga samples per second. Both laser beams green and infrared are emitted simultaneously along a shared axis and alignment, and the scanning mechanism follows an elliptical trajectory with a beam divergence angle of ±15 degrees.

The key contributions of this paper are: (1) A multi-scale transformer-based approach for learning feature dependencies in bathymetric LiDAR data classification. (2) Integration of multi-

scale attention mechanisms to capture spatial relationships at different scales. Utilization of handcrafted or pre-extracted features instead of direct point cloud processing. Finally, evaluation on real-world datasets, showcasing efficiency of the proposed method in shallow water environments.

2. Related Works

2.1 Hand Crafted Features

LiDAR systems tailored for underwater topography commonly known as bathymetric LiDAR have become increasingly valuable for mapping shallow marine ecosystems. By delivering fine-scale elevation and reflectance data, this technology provides crucial insight for identifying seafloor habitats and ecological patterns. Numerous investigations have assessed its suitability for differentiating underwater substrates and biological zones across a variety of aquatic settings.

For instance, Zavalas et al. (2014) showcased how bathymetric LiDAR can effectively delineate habitats dominated by temperate macroalgae. Utilizing decision tree classifiers and leveraging morphological indicators like slope and surface complexity, their methodology achieved classification accuracy exceeding 70%. Their findings affirm the viability of LiDAR for capturing ecological variation in dynamic coastal regions.

In another study, Kumpumäki et al. (2015) introduced a technique centered on waveform analysis, where the raw LiDAR signals were decomposed to understand interactions between the seabed and the water column. By clustering these signal features using a Self-Organizing Map (SOM), they were able to generate habitat maps that closely corresponded with known ecological distributions. This unsupervised approach underlined the promise of full-waveform LiDAR for autonomous seabed mapping.

Shifting focus to tropical environments, Su et al. (2018) conducted research in the South China Sea to distinguish coral reef habitats. They fused bathymetric terrain metrics with intensity features extracted from airborne laser bathymetry (ALB) waveforms. Their machine learning model, based on Support Vector Machines (SVM), achieved a classification accuracy greater than 93%, demonstrating how the integration of shape and reflectivity data can significantly enhance reef detection.

Tulldahl et al. (2013) examined the benefits of using multiple sensors by combining airborne bathymetric LiDAR data with high-resolution imagery from satellites. By applying maximum likelihood and random forest algorithms, they were able to differentiate six benthic categories with around 80% accuracy. The fusion of satellite imagery was particularly effective for improving depth calibration and compensating for water turbidity, thus enhancing mapping accuracy in complex nearshore environments.

Collectively, these studies provide compelling evidence for the growing role of bathymetric LiDAR in seabed classification. Techniques combining waveform signal interpretation, geomorphic analysis, and advanced classification algorithms alongside the fusion of different remote sensing platforms have significantly boosted the reliability and scalability of habitat mapping efforts.

More recently, advancements in LiDAR applications have extended to detecting specific seabed structures such as submerged boulders. Hansen et al. (2021) proposed a semi-automated framework that merges topo-bathymetric LiDAR data with Random Forest algorithms. By filtering point clouds to isolate the sea floor and deriving geometric and intensity-based features, they successfully identified boulder locations. The approach achieved a 57% recall rate and 27% precision, underscoring the promise of automated detection in coastal habitat assessments.

In a novel direction, Rhomberg-Kauert et al. (2024) applied unsupervised clustering techniques to classify aquatic vegetation using 3D LiDAR point clouds. Their method uses UMAP for dimensionality reduction, allowing for more nuanced grouping of points with similar characteristics. Employing the surface-volume-bottom separation method (Schwarz et al., 2019) to extract underwater data, they applied density-based clustering to distinguish vegetative from non-vegetative areas.

2.2 Deep Learning-based Methods

Prior to transformers, point cloud classification predominantly relied on deep learning models such as PointNet (Qi et al., 2017a) used shared multilayer perceptrons (MLPs) and max pooling to extract global features. PointNet++ (Qi et al., 2017b), introduces hierarchical feature extraction through local neighborhoods. DGCNN (Wang et al., 2019), utilizes edge convolution to capture local geometric relationships. Voxel-based networks (Graham et al., 2018) that convert point clouds into voxel grids and apply 3D convolutions. While these methods excel at feature extraction, they struggle with capturing long-range dependencies efficiently. Transformers, originally introduced for natural language processing, have been adapted to point cloud data due to their capability to model non-euclidean spatial relationships.

Point transformer (Zhao et al., 2021), introduces self-attention mechanisms that directly operate on points, leveraging local feature aggregation and attention-based feature learning. PCT (Guo et al., 2021) employs offset-attention to enhance feature extraction while maintaining permutation invariance.

In order to enhance the ability of transformers to better determine class differences and to find out a strong dependency between different features, hybrid transformer models are introduced. Swin3D (Liu et al., 2021), extends Swin transformer to 3D point clouds, employing hierarchical feature learning with shifted windows. SparseFormer (Wang et al., 2023) exploits a sparse MLP component to effectively capture features while accounting for the distinct nature of 3D point clouds. Furthermore, a multi-scale feature aggregation module is integrated to enrich contextual understanding.

Although transformer-based models have advanced point cloud classification, several challenges still remain regarding computational complexity, since transformers often have high memory and computational requirements. On the other hand, handling irregular and sparse point distributions is an open problem. Finally, sensitivity to noise and occlusions requires further exploration. Therefore, optimizing attention mechanisms for 3D data, and exploring self-supervised learning for point cloud representations plays an important role for improving results based on transformers.

3. Methodology

In this study rather than working directly on unstructured point sets voxel-based representation is used to process 3D point

clouds. Therefore, points are grouped in discrete voxel structures. Once the raw point clouds are embedded by voxels, feature extraction step will be applied on these voxels.

3.1 Feature Extraction

Accurate data classification fundamentally relies on the extraction of discriminative features. In this study, both geometric and colour information are utilized to derive meaningful features for model training. Nevertheless, the proposed architecture is not limited to a specific type of feature representation. It is designed to be flexible and compatible with a wide range of feature extraction techniques, including but not limited to deep learning-based methods such as PointNet++, KPConv, and similar architectures. Therefore, the proposed technique can improve the performance of existing deep learning-based methods by explicitly modeling the dependencies between neighboring voxels. By more effectively analysing the relationships among adjacent voxel features, the method enhances the ability to distinguish between classes with overlapping characteristics and ambiguous boundaries.

The primary objective of the proposed method is to enhance classification performance in scenarios involving classes with closely overlapping boundaries and high inter-class ambiguity, which are common in bathymetric datasets. Such challenges are particularly pronounced in regions where aquatic vegetation intersects with the water surface or seabed. The presence of additional classes, such as coral reefs, further increases the complexity by introducing more class intersections, thereby elevating the risk of misclassification near class boundaries. The proposed framework is specifically designed to address these issues and mitigate boundary-related classification errors.

To effectively describe the spatial organization of point clouds within individual voxels, we analyze their geometric configuration through a variety of shape descriptors. These include indicators such as sphericity, linearity, planarity, surface variation, omnivariance, and anisotropy. Such metrics, derived from the eigenvalues of the covariance matrix of the local point distribution, provide insight into whether the underlying structure resembles a linear, flat, or spherical formation, following the methodologies outlined by Brodu and Lague (2012) and Weinmann et al. (2013). These geometrically-derived attributes enable robust shape recognition within localized point sets.

In addition to geometric descriptors, we compute a range of statistical metrics from the intensity and elevation values associated with points in each voxel. This includes basic statistical moments such as the mean, median, mode, standard deviation, and skewness. Both green and near-infrared (NIR) returns are incorporated to ensure a comprehensive representation of the point cloud data.

Further enhancing the feature set, height-based parameters are calculated to capture vertical structure. Treating the center of a voxel as a reference point along the Z axis, the maximum and minimum elevation values of the contained points (denoted as Z_{max} and Z_{min}, respectively) are used to describe vertical extent.

3.2 Multi-scale Attention

Each voxel comprises a collection of 3D points and their associated feature representations. Let the set of voxels be defined as $\upsilon = \{v_i \mid v_i = (P_i, F_i)\}$, $i=1, \ldots, \mid \upsilon \mid$, where v_i denotes the i^{th} voxel. Each voxel v_i contains a set of 3D points $P_i \in \mathbb{Z}^3$, representing the spatial coordinates of the points, and a corresponding set of feature vectors $F_i \in \mathbb{R}^C$, where each vector describes C dimensional features for the associated point.

In order to find out dependencies between features of voxels, window-based attention mechanism is considered. Window-based attention partitions voxelized features into local windows. For each window cantered at position C_j ($C_j \in \mathbb{Z}^3$, $j = 1, \ldots, w$), a set of voxels can be selected with determined points and corresponding features.

To train a transformer-based model for classification tasks, the input must be structured to include a set of components: query, key, and value representations. Given the adoption of a multi-scale approach for model training, it is essential to utilize windows of varying sizes. A fixed window size is employed for the query. In contrast, two distinct window sizes are utilized for the keys, referred to as key1 and key2. While additional key window sizes could be introduced, doing so would result in increased and potentially redundant computational complexity. The set of query voxels is defined as those enclosed within the query window. Similarly, the voxels associated with key1 and key2 are determined by their respective window embeddings.

Additionally, the idea of Swin3D to shift the window on the data is also employed which helps exchange information across windows, reducing the problem of local window limitations. In non-overlapping window attention, voxel features are divided into windows, and self-attention is applied inside each window. In shifted window attention, the windows shift by a certain stride (typically half the window size in each spatial dimension) which allows voxels near window boundaries to interact with voxels in adjacent windows. This shifting helps merge information between neighboring voxel regions across different parts of the model.

3.3 Voxel Sampling by HDBSCAN Clustering

To reduce the computational cost associated with attention computation at each window stage, which involves considering features from embedded voxels within a window, we propose a strategy of selecting features from voxels that exhibit distinct characteristics. Specifically, we utilize HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) (Campello et al., 2023) to automatically determine the number of clusters based on the feature density within each window. Voxels belonging to the same cluster are expected to share almost similar feature representations, thus we prioritize retaining voxels with differing feature information to minimize redundant attention computation within each window.

HDBSCAN is a hierarchical clustering algorithm module designed to discover clusters of varying densities in high-dimensional or spatial data. It generalizes the DBSCAN algorithm by replacing the fixed-density threshold with a hierarchical approach based on mutual reachability distances. The algorithm begins by estimating the core distance for each point defined as the distance to its k^{th} nearest neighbor, where k determines the minimum number of samples. Using these core distances, HDBSCAN constructs a mutual reachability graph, which is then transformed into a minimum spanning tree (MST) that captures density-connected structures in the data. The MST is progressively condensed to form a cluster hierarchy, and the most stable clusters are selected using the excess of mass (EOM) or leaf selection methods. Points not belonging to any stable cluster are labeled as noise. HDBSCAN is particularly

well-suited for clustering noisy data, non-globular clusters, and data with varying local densities. It supports arbitrary feature spaces by allowing the use of custom distance metrics. In this study, HDBSCAN was employed to cluster a matrix of high-dimensional features extracted from 3D point clouds inside of each attention window.

3.4 Transformer-based Classification

In order to have a multi-scale attention to capture fine-grained details and long-range dependencies, we prefer to have different transformer heads which learn distinct levels of self-attention. Therefore, transformer heads should be explicitly partitioned into multiple groups (Dong et al., 2022). Note that, each group processes voxels sampled from windows of varying sizes. We want to ensure that each head group specializes in capturing information at a specific scale. By integrating outputs from all head groups, this approach effectively captures mixed-scale features, encompassing both broad contextual information and intricate local details.

To capture contextual relationships across different spatial scales, we employ a transformer encoder that applies independent multi-head attention operations over two voxel groupings. Given the query voxel group $V_q = (\mathbf{X}_q, \mathbf{F}_q)$, where $\mathbf{X}_q \in \mathbb{Z}^{Nq\times 3}$ and $\mathbf{F}_q \in \mathbb{R}^{Nq\times C}$, the encoder also receives two key-value groups $(\mathbf{X}_{k1}, \mathbf{F}_{k1})$ and $(\mathbf{X}_{k2}, \mathbf{F}_{k2})$ at different scales or receptive fields.

Each scale's attention is computed independently using multi-head attention with relative positional encoding (RPE):

$$\tilde{Y}_1 = \text{MHA}\,(\mathbf{F}_q, \mathbf{F}_{k1}; \text{RPE}\,(\mathbf{X}_q, \mathbf{X}_{k1})),$$
$$\tilde{Y}_2 = \text{MHA}\,(\mathbf{F}_q, \mathbf{F}_{k2}; \text{RPE}\,(\mathbf{X}_q, \mathbf{X}_{k2})) \quad (1)$$

where *RPE* $(\mathbf{X}_q, \mathbf{X}_{ki})$ denotes a relative positional encoding module (Shaw et al., 2018; Wu et al., 2021) that encodes spatial offsets between the query and key voxel coordinates. These encodings are added to the attention logits prior to softmax, enabling the network to incorporate geometric structure into the attention mechanism.

The outputs from both attention branches are concatenated and then passed through a layer normalization, as expressed by the following formula:

$$\mathbf{Z} = \text{LayerNorm}\,(\text{Concat}\,(\tilde{Y}_1, \tilde{Y}_2)) \quad (2)$$

Subsequently, a feed-forward network (FFN) enhances the representation:

$$\mathbf{Y} = \mathbf{Z} + \text{FFN}\,(\mathbf{Z}) \quad (3)$$

where the FFN is composed of multi-layer perceptron. This architecture effectively fuses information from multiple spatial scales while leveraging geometric cues through positional encoding, improving the network's ability to model complex 3D structures.

4. Results

The developed method was evaluated using a point cloud dataset collected from a shallow lake. This section presents both numerical analyses and visual assessments of the classification performance. The classification framework targets five distinct categories: water surface, aquatic plants, lakebed, trees, and terrestrial land.

4.1 Dataset

A shallow lake environment was chosen as the study area for data acquisition. The compiled dataset contains approximately 35 million 3D points. Each point is characterized by spatial coordinates (X, Y, Z), return intensity, and four amplitude measurements obtained from green and near-infrared (NIR) laser returns. Specifically, both high-resolution and low-resolution amplitude values were recorded for each laser wavelength.

Expert annotators manually labelled the point cloud to establish a reliable reference (ground truth). The dataset was split into subsets for training, validation, and testing. Roughly 70% of the points were allocated for training, with the remaining 30% designated for testing. A portion of the training data 10% was set aside to serve as the validation set. Figure 1 displays the distribution of train and test sets, with top-down views shown in subfigures (a) and (b), respectively, and a lateral (side) view of the lake provided in subfigure (c).

4.2 Evaluation Metrics

To maintain consistency with contemporary studies, our performance evaluation aligns with the methodology outlined by Sokolova et al. (2006). The analysis incorporates several key performance indicators: overall classification accuracy, recall (Rc), precision (Pr), and the F1 metric.

Total accuracy represents the fraction of correctly predicted samples out of the entire test dataset. Recall evaluates the model's success in retrieving all relevant examples for each category. Precision focuses on the accuracy of positive predictions made by the classifier. The F1 score serves as a harmonic mean of recall and precision, offering a comprehensive performance measure.

$$Pr_i = \frac{TP_i}{TP_i+FP_i}, Rc_i = \frac{TP_i}{TP_i+FN_i}, F1_i = \frac{2.Pr_i.Rc_i}{Pr_i+Rc_i} \quad (4)$$

In this context, TP refers to instances where the model accurately identified the correct category (True Positive), FP denotes cases where the prediction was incorrect (False Positive), and FN represents instances where the model failed to detect the correct category (False Negative).

The experimental findings reveal that the proposed approach is effective in distinguishing and assigning test samples to five separate categories, achieving an overall classification rate of 75.4%. A visual depiction of classification errors is provided in Figure 1 (d), where correctly identified points are shown in gray, while misclassified points are highlighted in red.

Furthermore, a visual representation of the classification output is provided in Figure 1. A detailed breakdown of classification performance across categories is presented in the confusion matrix, found in Table 1.

5. Conclusion

In this study, we presented a multi-scale transformer-based approach for classification of bathymetric point clouds. The method leverages attention mechanisms at different spatial scales by employing varying attention window sizes during training. This multi-scale attention design improves the model's ability to effectively capture complex dependencies among

features, leading to more accurate classification across diverse seabed classes.

Each attention window encompasses a distinct number of voxels, with varying point distributions and feature characteristics. To address the computational challenges associated with attention over large voxel sets, we integrate HDBSCAN clustering within each attention window. This unsupervised clustering groups voxels in feature space, allowing for representative sampling of voxels from distinct clusters, thereby reducing the number of voxels involved in the attention computation without sacrificing performance.

The proposed framework is versatile and can be applied to models utilizing handcrafted features, deep learning-derived representations, or a hybrid of both. Its adaptability to different feature extraction paradigms, along with its computational efficiency and classification accuracy, makes it a promising solution for large-scale bathymetric point cloud analysis and related 3D data classification tasks.

The proposed algorithm is not limited to bathymetric datasets and can be extended to non-bathymetric datasets, including terrestrial dry point clouds with severe class overlap.

Notably, the issue of overlapping classes is a common challenge across deep learning-based approaches applied to bathymetric data and is not inherent to our architecture. Rather, the proposed method is designed to mitigate this limitation. Furthermore, the algorithm is not dependent to the type of sensor footprint used for data acquisition.

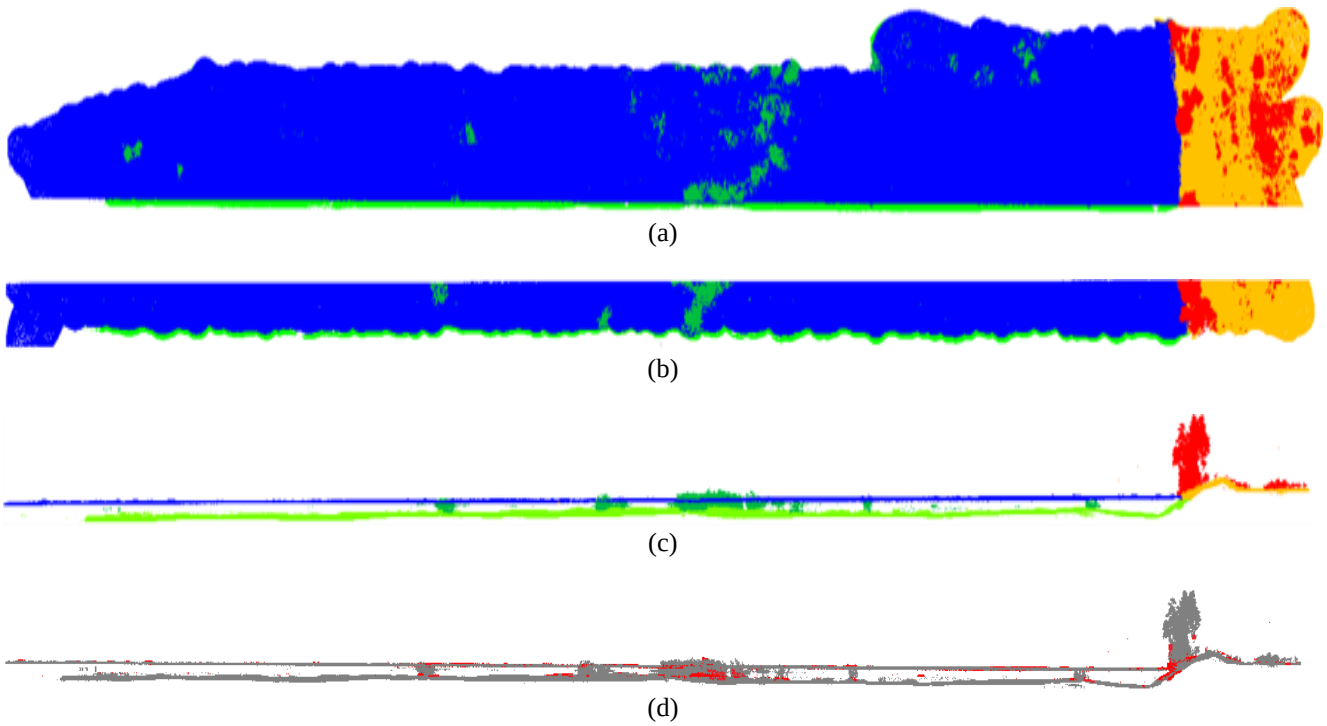

Figure 1. Train and Test sets**.** (a) Top-view of the area, annotated point clouds as train set, water surface (blue), aquatic vegetation (dark green), seabed (light green), tree (red), ground (light brown); (b) Labelled ground truth of the test set; (c) Side view of the test data to better illustrate under water region; (d) Error map of the predicted classes, correct predictions are colorized by gray and wrong predictions are colorized by red.

Categories	Water surface	Aquatic Vegetation	Seabed	Ground	Tree
Water surface	**80**	12	2	5	1
Aquatic Vegetation	10	**65**	8	10	4
Seabed	2	10	**75**	10	3
Ground	1	4	2	**80**	13
Tree	0.5	5	0.5	12	**81**
Precision	85.6	67.7	85.7	68.4	79.4
Recall	80	67	75	80	81.8
F1	82.7	67.3	79.9	73.7	80.6

Table 1. Confusion matrix of the proposed method. Precision, recall and F1 score are reported for each class.

References

Brodu, N., Lague, D., 2012: 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS journal of photogrammetry and remote sensing*, 68, 121-134.

Campello, R.J., Moulavi, D., Sander, J., 2013: Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, 160-172.

Dong, S., Ding, L., Wang, H., Xu, T., Xu, X., Wang, J., Bian, Z., Wang, Y., Li, J., 2022: Mssvt: Mixed-scale sparse voxel transformer for 3d object detection on point clouds. *Advances in Neural Information Processing Systems*, 35, 11615-11628.

Graham, B., Engelcke, M., Van Der Maaten, L., 2018: 3d semantic segmentation with submanifold sparse convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 9224-9232.

Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M., 2021: Pct: Point cloud transformer. *Computational visual media* 7, 187-199.

Hansen, S.S., Ernstsen, V.B., Andersen, M.S., Al-Hamdani, Z., Baran, R., Niederwieser, M., Steinbacher, F., Kroon, A., 2021: Classification of boulders in coastal environments using random forest machine learning on topo-bathymetric LiDAR data. *Remote Sensing*, 13(20), 4101.

Kumpumäki, T., Ruusuvuori, P., Kangasniemi, V., Lipping, T., 2015: Data-driven approach to benthic Cover type classification using bathymetric LiDAR waveform analysis. *Remote Sensing*, 7(10), 13390-13409.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021: Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012-10022.

Mandlburger, G., Hauer, C., Wieser, M., Pfeifer, N., 2015: Topo-bathymetric LiDAR for monitoring river morphodynamics and instream habitats—A case study at the pielach river. *Remote Sens (basel)* 7, 6160–6195. https://doi.org/10.3390/rs70506160

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a: Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652-660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Rhomberg-Kauert, J., Dammert, L., Groemer, M., Pfennigbauer, M., Mandlburger, G. 2024: Macrophyte detection with bathymetric LiDAR – Applications of high-dimensional data analysis for submerged ecosystems. *The International Hydrographic Review*, 30(2). https://doi.org/10.58440/ihr-30-2-a16

Schwarz, R., Mandlburger, G., Pfennigbauer, M., Pfeifer, N. 2019: Design and evaluation of a full-wave surface and bottom-detection algorithm for lidar bathymetry of very shallow waters. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 1–10.

Shaw, P., Uszkoreit, J., Vaswani, A., 2018: Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*.

Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006: Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *In Australasian joint conference on artificial intelligence*, 1015-1021.

Su, D., Yang, F., Ma, Y., Zhang, K., Huang, J., Wang, M., 2018: Classification of coral reefs in the South China Sea by combining airborne LiDAR bathymetry bottom waveforms and bathymetric features. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 815-828.

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 *Proceedings of the IEEE/CVF international conference on computer vision*, 6411-6420.

Tulldahl, H.M., Philipson, P., Kautsky, H., Wikström, S.A., 2013: Sea floor classification with satellite data and airborne lidar bathymetry. *In Ocean Sensing and Monitoring V*, 8724, 100-115.

Wang, Y., Liu, Y., Zhou, P., Geng, G., Zhang, Q., 2023: SparseFormer: Sparse transformer network for point cloud classification. *Computers & Graphics* 116, 24-32.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019: Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5), 1-12.

Weinmann, M., Jutzi, B., Mallet, C., 2013: Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 313-318.

Wu, K., Peng, H., Chen, M., Fu, J., Chao, H., 2021: Rethinking and improving relative position encoding for vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*,10033-10041.

Zavalas, R., Ierodiaconou, D., Ryan, D., Rattray, A., Monk, J., 2014: Habitat classification of temperate marine macroalgal communities using bathymetric LiDAR. *Remote sensing*, 6(3), 2154-2175.

Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V., 2021: Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 16259-16268.