iBALR3D: imBalanced-Aware Long-Range 3D Semantic Segmentation

Keying Zhang[1], Ruirui Cai[2], Xinqiao Wu[1], Jiguang Zhao[1], Ping Qin[1]

[1]China Southern Power Grid Digital Power Grid Group Co., Ltd., Guangzhou, China
[2]School of Computer Science and Technology, Xidian University, Xi'an, China
{zhangky, wuxq, zhaojiguang, qinping1}@csg.cn, cairuirui3@gmail.com

Abstract

3D semantic segmentation is crucial for comprehending transmission line structure and environment. This understanding forms the basis for a variety of applications, such as automatic risk assessment of line tripping caused by wildfires, wind, and thunder. However, the performance of current 3D point cloud segmentation methods tends to degrade on imbalanced data, which negatively impacts the overall segmentation results. In this paper, we proposed an imBalanced-Aware Long-Range 3D Semantic Segmentation framework (iBALR3D) which is specifically designed for large-scale transmission line segmentation. To address the unsatisfactory performance on categories with few points, an Enhanced Imbalanced Contrastive Learning module is first proposed to improve feature discrimination between points across sampling regions by contrasting the representations with the assistance of data augmentation. A structural Adaptive Spatial Encoder is designed to capture the distinguish measures across different components. Additionally, we employ a sampling strategy to enable the model to concentrate more on regions of categories with few points. This strategy further enhances the model's robustness in handling challenges associated with long-range and significant data imbalances. Finally, we introduce a large-scale 3D point cloud dataset (500KV3D) captured from high-voltage long-range transmission lines and evaluate iBALR3D on it. Extensive experiments demonstrate the effectiveness and superiority of our approach.

Introduction

3D point cloud semantic segmentation is an important task that classifies all points into their corresponding categories (Landrieu and Simonovsky 2018). The potential of implementing associated technologies in large-scale electrical grids is substantial. However, the research progress in the power grid domain is still relatively limited, primarily due to the scarcity of well-labelled data.

More specifically, there are a few unique challenges in electrical grid applications (e.g., risk assessment and prediction under different weather conditions). The high demand for accuracy in transmission line segmentation is one primary aspect. The segmentation output will be utilized to simulate the actions of insulators or jumper wires under varying wind speeds. It can also be applied to measure the probability of wildfire-induced tripping on transmission lines. All the applications rely heavily on precise labels. Particularly, the data imbalance issue in the domain poses an inevitable challenge. General semantic segmentation algorithms usually assume a roughly balanced number of points from different categories. However, these assumptions do not hold in the context of transmission line data, leading to biased results and inaccurate representations. Furthermore, transmission lines usually contain long-range structures. To this end, the model should be capable of extracting both the

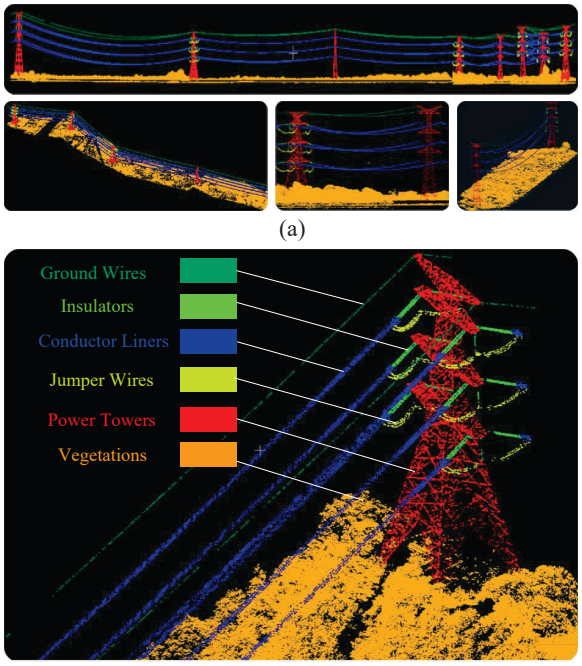

Figure 1: We introduce a novel 500KV3D dataset. 500KV3D is a large-scale long-range 3D point cloud dataset, which is collected from a high voltage-level, 500KV smart-grid infrastructure. (a) illustrates a few distant views and (b) is the zoomed in view. We consider 500KV3D could provide more insights into deploying multimedia models in electrical grids-related topics. Details and statistical analysis are provided in 500KV3D Dataset Section .

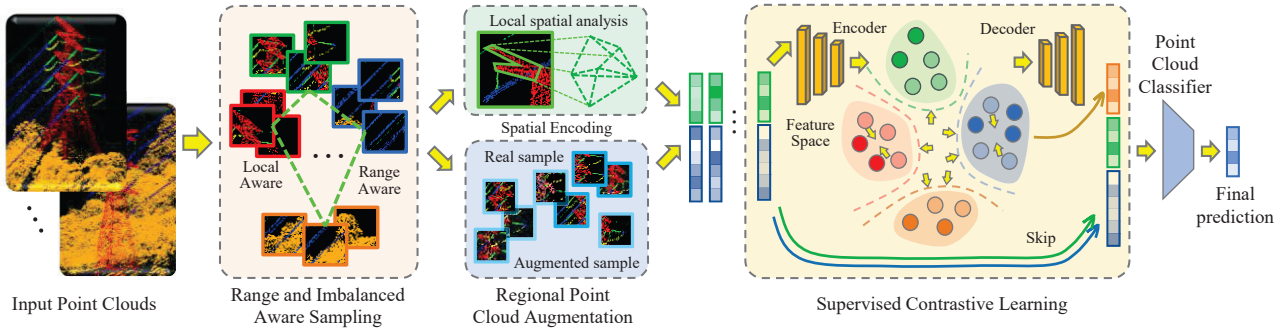

Figure 2: Framework of our iBALR3D model. A long-range and imbalanced aware sampling strategy is deployed to balance the significant data imbalanced issue and align point clouds in the long-range distance. An adaptive spatial encoder is designed to extract indistinguishable junctional regions across simple shapes. A contrastive training associated with an augmentation module is used to enhance the learning capacity of tail categories and achieve overall highest performance.

long-range global structural information as well as the trivial local differences to obtain accurate and consistent global performance.

Addressing these challenges is key to further process segmentation. Notably, Javier Grandio et al. (Grandio et al. 2023) developed a multi-modal method for railway infrastructure point clouds, focusing on panoptic segmentation of linear and pole-like objects. Daniela Lorena Lamas et al. (Lamas et al. 2021) introduced an innovative algorithm that leverages geometry and spatial context, enhancing segmentation in railway environments (e.g., rails, masts, wiring, droppers, traffic lights, and signals). Additionally, Jingru Wang et al. (Wang et al. 2022) proposed a robust method for segmenting point cloud data of communication towers and accessory equipment based on geometrical shape context from 3D point cloud.

In this paper, we present an imBalanced-Aware Long-Range 3D Semantic Segmentation framework (iBALR3D) which is specifically engineered to tackle the challenges inherent in transmission line applications. To validate the effectiveness of the proposed model, a large-scale, high-quality, and well-organized point cloud dataset, named 500KV3D is introduced. 500KV3D is collected from extremely high-voltage (i.e., 500KV) power transmission lines. The dataset is well labelled by technicians. Extensive experiments demonstrate the effectiveness of the proposed modules especially for categories with few points. Our method achieves leading performance across all established baselines. The main contributions are as follows:

- An Enhanced Imbalanced Contrastive Learning module is proposed, which improves the representation effectively by contrasting the features across categories in a supervised fashion.
- An Adaptive Spatial Encoding is designed, which implicitly aligns object shape knowledge as well as its context.
- A strategy called Long-Range and Imbalanced Sampling is introduced. It addresses the data imbalance issue during training and aligns points over long-range distances.
- A large-scale, high-quality, and well-organized point cloud dataset of transmission lines is introduced to validate the effectiveness of our approach.

Related work

Point cloud semantic segmentation, a key task in computer vision, classifies points in a 3D cloud into specific categories. With the advancements in deep learning and 2D vision algorithms, deep learning-based approaches have outperformed traditional methods in semantic segmentation tasks. These methods generally fall into point-based, voxel-based, graph-based, and transformer-based categories.

Point-Based Methods have emerged as a popular approach due to their ability to directly process raw point clouds. (Li, Liu, and Gall 2020) reformulated point-based methods to operate in the projection space, which significantly improved the efficiency of processing LiDAR point clouds. (Hu et al. 2020a) proposed an efficient and lightweight neural architecture to directly interpret point semantics for large-scale point clouds. Similarly, (Park et al. 2023) designed a self-positioning point-based transformer that shows promising results in point cloud understanding. Other classical research includes (Garcia-Garcia et al. 2016; Ni et al. 2020; Thomas et al. 2019; Liu et al. 2019; Chiang et al. 2019; Wu, Qi, and Fuxin 2019; Mao, Wang, and Li 2019; Hu et al. 2020b). Although Point-based methods are capable of directly processing raw point clouds, making them efficient and straightforward in their approach, most of these methods can struggle with large-scale point clouds due to high computational costs. They may also have difficulty handling the irregularity and sparsity of point clouds, which can lead to less accurate segmentation results.

Voxel-Based Methods usually convert point clouds into a voxel grid, which allows for the deployments of 3D convolutional neural networks. DRINet++ (Ye et al. 2021) jointly learns the sparsity and geometric properties of a point cloud with a voxel-as-point principle. (Ye et al. 2022) introduced a Geometry-aware Sparse Network (GASN) which leverages the sparsity and geometric properties of point clouds within a unified voxel representation. HilbertNet (Chen et al. 2022) preserves the benefits of voxel-based methods while signifi-

cantly reduced computational costs through a Hilbert curve-based flattening mechanism. (Hou et al. 2022) proposed a teacher-student strategy, which eventually using a small network to do LiDAR semantic segmentation for efficient reference. Voxel-based methods are usually effective in handling large and complex point clouds. However, the voxelization process can lead to information loss, which may decrease segmentation accuracy. Additionally, these methods are also computationally expensive.

Graph-Based Methods consider point clouds as graphs, where each point is a node, and the edges represent the relationships between the points. (Wang et al. 2019a) introduced an attention mechanism into the graph convolution process, thereby improving the model's capacity to concentrate on crucial points. (Landrieu and Simonovsky 2018) introduced a new framework for semantic segmentation of large-scale point clouds using superpoint graphs and graph convolutional networks, which captured the organization and context of 3D point clouds by partitioning them into geometrically homogeneous elements. (Jiang et al. 2019) presented a method that utilized point and edge features in a hierarchical graph framework to label 3D scenes with semantic categories. (Yan et al. 2020) proposed PointASNL, which processes noisy point clouds robustly using adaptive sampling and local-nonlocal modules. Other research works include (Wang et al. 2019c; Li et al. 2019; Wang et al. 2019b). Graph-based methods are proficient at identifying relationships in point data and work well with clear graph structures. However, they can be computationally heavy due to complex graph construction and processing, and their performance can depend on the chosen parameters.

Transformer-Based Methods. Transformer-based methods (Zhao et al. 2021) are gaining attention for their proficiency in capturing long-range data dependencies. Xin Lai et al. (Lai et al. 2023) proposed a stratified strategy for sampling keys to harvest long-range contexts, demonstrating the potential of transformers in this field. SPFormer (Sun et al. 2022) is a method that clusters potential features from point clouds into larger units called superpoints. It then uses query vectors to directly predict instances, eliminating the need for reliance on object detection or semantic segmentation results. (Xiu et al. 2022) further extended the transformer-based methods by introducing an interpretable edge enhancement and suppression learning mechanism. Transformer-based methods are adept at handling complex point cloud segmentation by capturing long-range data dependencies. However, they are computationally intensive, require significant memory, and need ample training data, which can be problematic when labelled data is scarce.

500KV3D Dataset

We present the 500KV3D dataset, a large, high-quality, and well-structured point cloud dataset collected from 500KV power transmission lines using drones with 3D Li-DAR sensors. The dataset has been meticulously processed and checked for quality. It serves as a valuable asset for the energy industry and a practical case study for evaluating 3D applications. We discuss the data collection procedures and analysis results in this section. A sample from the 500KV3D dataset is illustrated in Figure 1.

Table 1: Specifications of LiDAR Sensor LiAir 220N, which is used to colloect our 500KV3D dataset.

Performance	Specifications
Laser Sensor	Hesai Pandar40P
Range Accuracy	±20 mm
Detection Range	200 m @ 10% reflectance
Channels	40
Power Consumption	27 W
System Accuracy	± 5 cm

Data Collection

Due to the extremely high-voltage level, the power lines are usually in high and inaccessible locations which leads difficulties to in scanning all the structural details from the ground. To this end, a powerful drone is utilized to carry a LiDAR sensor in the air, to capture even the tiny objects of the system, such as thin power lines. The LiDAR system is LiAir 220N, which is a lightweight LiDAR survey instrument manufactured by GreenValley International (GVI) [1]. It is specifically designed for mounting on drones (Unmanned Aerial Vehicle, UAVs). The system is equipped with a Hesai Pandar40P laser scanner [2], making it one of the most cost-effective options in GVI's LiAir Series. More detailed configurations are listed in Table 1.

Pre-processing

Despite using professional LiDAR sensors, outliers and noise is inevitable due to varying reflectivity properties and atmospheric interference. We use Radius and Statistical Outlier Removal techniques, followed by a manual inspection for noise reduction. The final raw dataset includes the (x, y, z) coordinates of each point.

Labeling

We consider 6 semantic categories as the critical and dominant categories for power transmission applications. More specifically, 1) ***Conductor lines*** denote quadruple split conductors that carry the electrical waves from the transmitters to the receivers; 2) ***Ground wires*** are used to protect the conductors from lightning strikes. They are usually the wires installed above conductor lines; 3) ***Insulators*** include the I-type, the II-type, and the V-type insulators which are the materials that prevent the electric current from flowing from the conductors to the ground or other objects; 4) ***Jumper wires*** are the quadruple split jumper wires that are used to connect the conductors on the poles or towers to the insulators or other equipment; 5) ***Power towers*** are three or four-circuit pole towers that support the entire transmission system overhead. It carries electric current from the power plants to the substations and consumers; 6) ***Vegetation*** are considered as any ground objects which contain trees, shrubs, hedges, bushes, etc. To streamline the labour-intensive process of

[1] https://globalgpssystems.com/liair-220n/

[2] https://www.hesaitech.com/product/pandar40p/

manual annotation for the entire point cloud data, we employ clustering algorithms to segment the data into regions. Subsequently, a manual correction procedure is implemented to refine and validate the annotation results, ensuring consistency and quality. CloudCompare [3] is used for conducting the annotation which is an open-source point cloud processing tool. The entire dataset took approximately 200 working hours for data pre-processing and labelling.

Our collection has 29M labelled points across 42 sections. We train on 34 sections and test on 8, with distances between towers ranging from 100 to 800 meters and point scales per segment from 10k to almost 2M.

Statistical Analysis

To help users better understand our dataset, more statistical details are provided in this section. Due to the nature of the transmission system, a few categories dominate the dataset,which leads to considerable imbalanced data distribution. In Figure 3, we illustrate the number of the point distribution across 42 sections of different categories via boxplot. More quantitative numbers are listed in Table 2. We can observe that some primary semantic categories (e.g., vegetation) constitute over 90 percent of the total points. In contrast, the less prevalent but crucial categories, such as jumper wires, ground wires, and insulators make up only 0.19%, 0.27%, and 0.32% respectively of the total points. This data reflects the complexity of the real-world transmission line environment and reveals a significant imbalance in the distribution of semantic classes, underscoring the difficulties in applying existing segmentation approaches universally. In addition, the elevation or height of the points across different categories are an important characteristic. In Figure 4, the histogram of the point cloud elevation is visualized. Note that most transmission system components have higher elevation, and due to the sparsity of these components, the distribution is various.

In summary, we consider 500KV3D to be a general and practical point cloud dataset which is collected from real-world civil engineering infrastructure. We hope it could contribute more on related research communities.

Our Method

There are three main modules in our iBALR3D method, including Enhanced Imbalanced Contrastive Learning, Adaptive Spatial Encoding, and Long-range and Imbalanced Sampling. More details are introduced in this section below:

Enhanced Imbalanced Contrastive Learning

The significant imbalanced data distribution leads to the difficulty for the model to learn the distinctive structural characteristics across the tail categories. To this end, an enhanced and supervised contrastive learning strategy is proposed. Its objective is to force the model to differentiate categories. To further enhance the model learning effectiveness in the imbalanced data scenario, a data augmentation strategy is deployed, which increases the sample numbers of the tail categories.

[3]https://www.danielgm.net/cc/

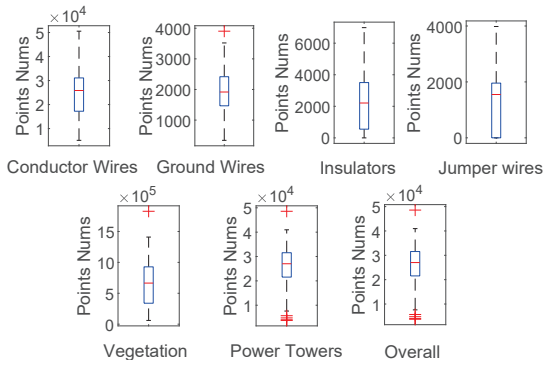

Figure 3: Point number distribution analysis of our 500KV3D dataset. All points are separated into 42 sections, the box plots illustrate and point number distributions across different semantic categories as well as different sections.

We initialize a possibility for each point in a scene based on Long-Range and Imbalanced Sampling strategy introduced in section 4.2, and we pick a point as the center point according to the generated possibility. Then, we select a sampled region $\boldsymbol{x}$ by searching the nearest 40960 points from the center point. Multiple augmentation algorithms are implemented to the region, including translation and rotation. For translation, points in the region are centered to zero by subtracting the center coordinates from the chosen point coordinates. For rotation, we randomly rotate a certain angle to the whole region.

$$\tilde{\boldsymbol{x}} = Aug(\boldsymbol{x}), \tag{1}$$

where $\tilde{\boldsymbol{x}}$ is the augment region.

For the design of the contrastive objective, we deploy the general max margin strategy, while a more sophisticated algorithm is also feasible for this module. Specifically, for a pair of sampled points, we encourage the learned representations that are more similar to their counterparts within the same category, while being as distinct as possible from neighbouring points from different categories. The objective function can be represented as:

$$\begin{aligned} &\mathcal{L}_{\text{cont}}^{m}\left(x_i, x_j; f\right) \\ &= \mathbf{1}\left\{y_i = y_j\right\} \left\|f(x_i) - f(x_j)\right\|_2^2 \\ &\quad + \mathbf{1}\left\{y_i \neq y_j\right\} \max\left(0, m - \left\|f(x_i) - f(x_j)\right\|_2^2\right), \end{aligned} \tag{2}$$

where $x_i, x_j \in \mathcal{S}_l$ is the pair of points, and the point set $\mathcal{S}_l$ contains both real and augmented samples from Eq.1. y_i,y_j denote the ground truth labels of point x_i and x_j, $f(\cdot)$ is the embedding function, and m is a hyperparameter.

For network structure design, to obtain dense and relatively low-dimension representations for downstream modules, an autoencoder network is proposed. Specifically, an encoder network projects sample points into the feature space for obtaining the representations, and a decoder recovers the representations. The equations of encoder and decoder are shown below:

$$\boldsymbol{v} = f(\boldsymbol{x}), \tilde{\boldsymbol{p}} = \tilde{f}(\boldsymbol{v}), \tag{3}$$

Table 2: Point number distributions of the training and testing sets

Categories	Overall Number of Points	Overall Ratio(%)	Training Set Number of Points	Training Split Ratio(%)	Testing Set Number of Points	Testing Split Ratio(%)
Conductor Lines	1,032,617	3.50	839,454	81.29	193,163	18.71
Ground Wires	80,767	0.27	64,683	80.09	16,084	19.91
Insulators	95,193	0.27	73,996	77.73	21,197	22.27
Jumper Wires	54,959	0.19	41,917	76.27	13,042	23.73
Power Towers	1,081,921	3.66	870,206	80.43	211,715	19.57
Vegetations	27,197,616	92.06	21,381,008	78.61	5,818,808	21.39
Overall	29,543,073	100.00	23,271,264	78.77	6,271,809	21.23

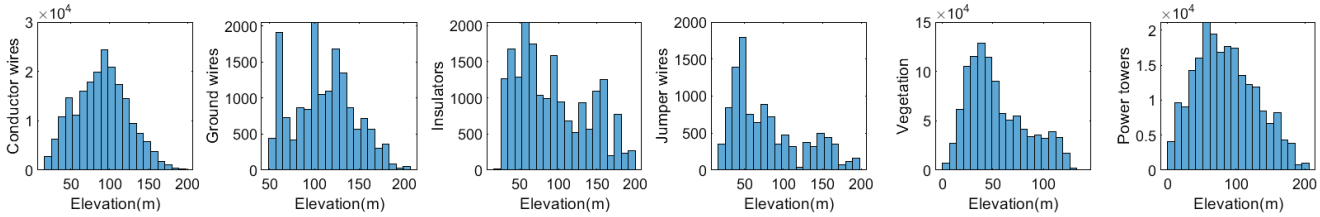

Figure 4: The elevation histogram of the point cloud in the 500KV3D dataset, where the points are separated into 6 different categories. We can see that there are considerable distribution differences across different categories. For instance, the point number of *Vegetation* considerably dominates the data while the height is relatively low. And the height distributions of wires-related points are more fluctuated.

where $f(\cdot)$ and $\tilde{f}(\cdot)$ are the embedding and decoding network. $\boldsymbol{v} \in \mathbb{R}^{d_E}$ and $\tilde{\boldsymbol{p}} \in \mathbb{R}^{d_D}$ are the encoded representation and decoded results, and $\mathbb{R}^{d_E}$ and $\mathbb{R}^{d_D}$ are the corresponding dimensions. By this way, supervised contrastive learning enhances the discrimination of features across categories, and weaken the negative influence the data imbalance challenge.

Adaptive Spatial Encoding

In transmission line related applications, we observe that the shapes of most categories are elegant with enough distance for general models to accurately recognize most regions. However, the errors usually exist in the junctional area (e.g., between *Vegetation* and *Power Tower*) due to the undistinguished transition between the simple shapes.

To this end, we proposed an adaptive spatial encoding strategy. Specifically, the normal vector and curvature are jointly deployed. We consider normal vector is able to reveal the slight surface variations. For instance, the smooth change in the normal vector suggests a relatively flat region, while a significant change in the normal vector indicates a fluctuating region. For a given point p_i, we choose its k nearest neighbors and calculate the local plane P of these points based on the least squares algorithm. Grid search is utilized to find the best value for k, based on the minimal test loss as outlined in section . In this study, the optimal value for k is 8, and the algorithm can be represented as:

$$P(\overrightarrow{n}, d) = \underset{(\overrightarrow{n},d)}{\arg\min} \sum_{i=1}^{k} \left(\overrightarrow{n} \cdot p_i - d\right)^2, \qquad (4)$$

where $\overrightarrow{\mathrm{n}}$ is the normal vector of the plane P, d is the distance from P to the origin point. The eigenvector corresponding to the minimum eigenvalue of M is the normal vector of P:

$$M = \frac{1}{k} \sum_{i=1}^{k} \|p_i - p_0\|_2^2, \qquad (5)$$

We perform eigenvalue decomposition on the covariance matrix M in Eq.5, and obtain the eigenvalues of M. If the eigenvalues satisfy $\lambda_0 \leq \lambda_1 \leq \lambda_2$, then the surface curvature δ of point p_i is $\delta = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}$. The smaller δ is, the flatter the neighbourhood is, the larger δ is, the greater the fluctuation of the neighbourhood is. We concatenate the calculated normal vectors and curvature to the original coordinates before conducting contrastive learning.

Long-Range and Imbalanced Sampling

Contrastive learning and spatial encoding enhances the model learning effectiveness. However, considering the long-range point cloud distribution as well as the significant imbalanced label, a long-range and imbalanced sampling strategy is further proposed.

In the sampling phrase, the tail categories (e.g., *Jumper Wires*) will have higher sampling ratio compared with their sample number ratios. Moreover, for a selected point x_i, measure the diversity of its neighbors. The more diverse of the neighbors indicates the higher learning requirements. By finding the top nearest neighbor points of x_i, our method could also reach long-range in point sparse regions, especially for the tail categories. The sampling strategy is illustrated below:

$$\mathcal{P}(x_i) = \frac{n_{y_i}^{\alpha}}{\sum_{k=1}^{n_i} n_i^{\alpha}} + \beta \frac{n_{T_{knn}}}{\sum_{k=1}^{n_i}}, \qquad (6)$$

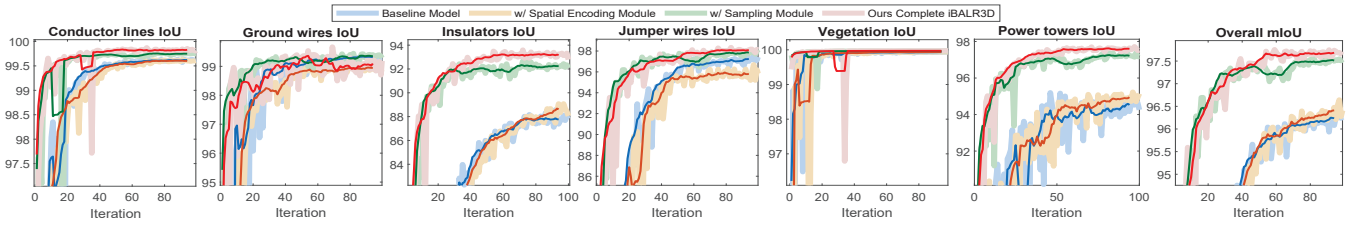

Figure 5: Ablation study of our model. We illustrate the category-wise and overall segmentation performance when different modules are included in the training stage. The thick light colour curve is the exact performance, and the darker colour denotes the smoothed result for clear comparison. Red indicates our complete iBALR3D framework, green ablated the spatial encoding, the brown curve ablated the sampling and spatial encoding modules, and the blue curve is the baseline framework. We can observe that our complete framework outperforms others, which demonstrates the effectiveness of the proposed modules.

Table 3: Semantic segmentation performance of benchmarks and our method.

Methods	Category-level segmentation mIoU(%)						Overall mIoU(%)
	Cro.	Gon.	Ins.	Jum.	Veg.	Pow.	
PointNet (Garcia-Garcia et al. 2016)	65.53	46.56	0.92	2.41	97.52	41.71	42.44
PointNet++ (Ni et al. 2020)	84.31	77.82	13.00	25.51	99.51	71.61	61.96
StratifiedTransformer (Lai et al. 2023)	89.32	85.85	19.98	40.56	99.73	85.14	70.10
RandLA-Net (Hu et al. 2020a)	99.38	98.58	91.23	98.24	99.93	97.55	97.49
BAAF-Net (Qiu, Anwar, and Barnes 2021)	99.42	98.54	91.78	98.02	99.92	97.52	97.53
iBALR3D (Ours)	**99.66**	**99.10**	**95.06**	**98.64**	**99.97**	**99.00**	**98.57**

where $\mathcal{P}(x_i)$ is the probability of sampling x_i, n_{y_i} is the point number of a given category y_i, T_{knn} is the nearest point numbers. Both α and β are trade-off parameters.

Implementation

We use a multi-layer perceptron with 2 hidden layers for $f(\cdot)$, and normalize its output, enabling distance measurement in feature space via inner product. For training, we use a batch size of 6, sample raw input points at 0.04m grid size, and fix the total input points at 40,960. The KNN parameter is set to 16, and all other configurations follow the RandLA-Net for the S3DIS Dataset. Our iBALR3D trains for 100 epochs on a RTX4090 GPU with 128GB memory.

Experiments

Experimental setup

For benchmarks, five state-of-the-art benchmarks are used for our experiments. More specifically: **PointNet** (Garcia-Garcia et al. 2016) is an innovative deep learning model. It uses raw data to create a comprehensive global feature vector, employs a symmetric function for unordered data, and incorporates a transformation network to handle rotational and translational variances. **PointNet++** (Ni et al. 2020) is an extension of PointNet. It solves the limitations of PointNet in capturing local structures by recursively applying PointNet on the nested partitions of the input point cloud. **RandLA-Net** (Hu et al. 2020a) efficiently processes large-scale 3D point clouds, eliminating pre/post-processing. It uses random point sampling and a local feature aggregation module to preserve geometric details by increasing the receptive field for each 3D point. **BAAF-Net** (Qiu, Anwar, and Barnes 2021) is designed for analyzing and segmenting real point cloud scenes. It improves the local context and fuses multi-resolution features for each point, resulting in a comprehensive and accurate analysis. **Stratified Transformer** (Lai et al. 2023) uses sparse sampling of distant points to expand its receptive field and create long-range dependencies. It also includes a first-layer point embedding and contextual position encoding to manage irregular point arrangements.

For evaluation, Overall Mean Intersection-over-Union (mIoU) is deployed, which is a common evaluation metric for semantic segmentation tasks (Tang et al. 2022; Hu et al. 2020a, 2022; Zhang et al. 2023). It measures the average overlap between the predicted and ground truth regions for each class in a point cloud:

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i}, \tag{7}$$

where N represents the number of classes, TP_i represents the number of true positives for class i, FP_i represents the number of false positives for class i, and FN_i represents the number of false negatives for class i.

For the training and testing split, since our 500KV3D dataset consists of 42 scenes with 84 towers, we randomly selected 34 scenes for training and 8 scenes for testing. Detailed statistical numbers of the training set and testing set can be found in Table 2. Our iBALR3D model is trained and tested using 3D coordinates together with the 8-dimensional embedding vectors obtained by contrastive learning.

Performance

The performance of benchmarks and our method in mIoU evaluation metric is shown in Table 3, where both the category level and overall performances are provided. The categories in the dataset include Conductor Liners, Ground

Table 4: Ablation study of our iBALR3D model

Methods	Category-level segmentation mIoU(%)						Overall mIoU(%)
	Cro.	Gon.	Ins.	Jum.	Veg.	Pow.	
Baseline RandLA-Net	99.38	98.58	91.23	98.24	99.93	97.55	97.49
Ours w/o spatial encoding module	99.49	99.06	92.77	97.86	99.95	98.30	97.90
Ours w/o sampling module	99.64	98.83	94.42	98.52	99.95	98.36	98.29
Ours complete iBALR3D model	**99.66**	**99.10**	**95.06**	**98.64**	**99.97**	**99.00**	**98.57**

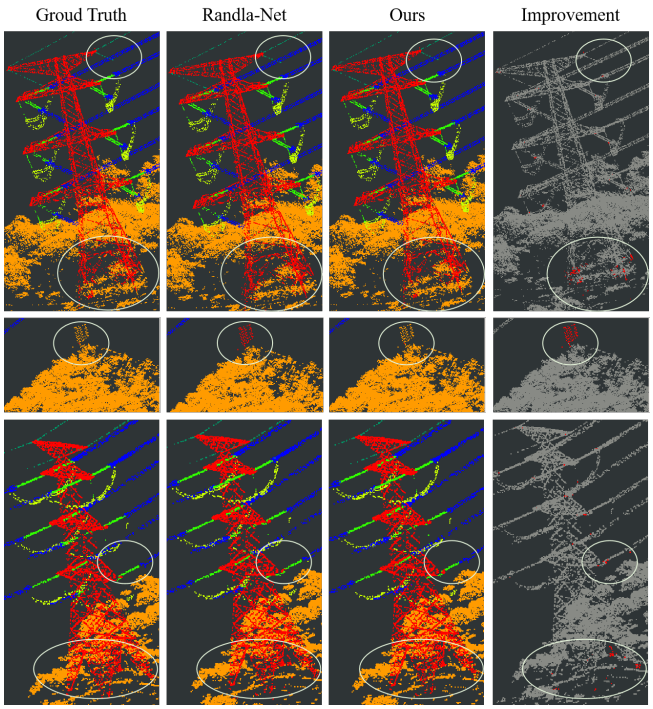

Figure 6: We visualize the results of RandLA-Net baseline and our iBALR3D and the improvements on several different scenes. We can see that iBALR3D can effectively reduce errors on the junctional regions (e.g., *Power Tower*)

Wires, Insulators, Jumper Wires, Vegetation, and Power Towers. Our approach achieved the best performance across all categories. Notably, our approach outperformed existing methods in both categories with fewer points and categories with numerous points. In particular, the performance improvement for the Insulators category was nearly 4 per cent, which is significant for applications such as insulator wind deviation checking.

To further analyse the effectiveness of our model, t-SNE (Van der Maaten and Hinton 2008) is used to visualize the learned point cloud representations and the results are shown in Figure 7, where (a) and (b) denote the representations of RandLA-Net and our iBALR3D approaches respectively, and different colours denotes different categories. Considering the significantly imbalanced point number distribution, we intentionally increase the ratios of the tail categories for better visualization. From Figure 7, we can see that our model achieves more distinguishing representations where the same categories are more clustered in the same regions.

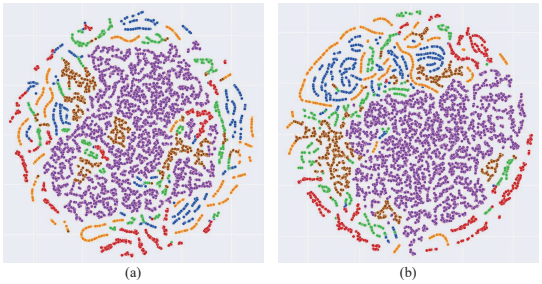

Figure 7: t-SNE visualization of the learned point cloud features. (a) denotes RandLA-Net features and (b) denotes our iBALR3D features. Different colors denotes different semantic categories. From the results, we observed that our model achieves more distinguishing features compared with other SOTA benchmarks.

A case study is shown in Figure 6 where we visualize the ground truth and the prediction results from RandLA-Net and our iBALR3D model. More importantly, we further visualize the prediction improvement compared with RandLA-Net. We can see that our approach considerably reduces the errors in the junctional region, which further demonstrates the effectiveness of our modules.

Ablation Studies

We conduct ablation studies to showcase the effectiveness of each module. Each module is individually removed, and the model is retrained and evaluated. The adaptive spatial encoding module is removed by directly inputting the original point coordinates. The long-range and imbalanced sampling module is replaced with random sampling strategy. The results are shown in Table 4. This ablation study shows how the proposed modules synergistically improve performance.

Conclusion

We proposed iBALR3D, a novel method for semantic segmentation of point clouds. It addresses the challenges of imbalanced data and long-range distribution in real-world transmission line scenarios. iBALR3D incorporates a contrastive learning algorithm, adaptive spatial encoding module, and sampling strategy to prioritize junctional regions in long-range space and learn distinctive representations for different classifications. We also introduce a new dataset, 500KV3D, for evaluation purposes. Through extensive experiments, ablation studies, and case studies, we demonstrate the effectiveness of iBALR3D.

Acknowledgments

The work was supported by China Southern Power Grid Digital Power Grid Research Institute Co., Ltd., and the project NO. is 210005KK52220019.

References

Chen, W.; Zhu, X.; Chen, G.; and Yu, B. 2022. Efficient Point Cloud Analysis Using Hilbert Curve. In *ECCV*.

Chiang, H.-Y.; Lin, Y.-L.; Liu, Y.-C.; and Hsu, W. 2019. A Unified Point-Based Framework for 3D Segmentation. In *CVPR*.

Garcia-Garcia, A.; Gomez-Donoso, F.; Garcia-Rodriguez, J.; Orts-Escolano, S.; Cazorla, M.; and Azorin-Lopez, J. 2016. PointNet: A 3D Convolutional Neural Network for real-time object class recognition. In *IJCNN*.

Grandio, J.; Riveiro, B.; Lamas, D.; and Arias, P. 2023. Multimodal deep learning for point cloud panoptic segmentation of railway environments. In *Automation in Construction*.

Hou, Y.; Zhu, X.; Ma, Y.; Loy, C.; and Li, Y. 2022. Point-to-Voxel Knowledge Distillation for LiDAR Semantic Segmentation. *arXiv preprint arXiv:2206.02099*.

Hu, Q.; Yang, B.; Khalid, S.; Xiao, W.; Trigoni, N.; and Markham, A. 2022. SensatUrban: Learning Semantics from Urban-Scale Photogrammetric Point Clouds. In *International Journal of Computer Vision*.

Hu, Q.; Yang, B.; Xie, L.; Rosa, S.; Guo, Y.; Wang, Z.; Trigoni, N.; and Markham, A. 2020a. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *CVPR*.

Hu, Z.; Zhen, M.; Bai, X.; Fu, H.; and Tai, C.-l. 2020b. JSENet: Joint Semantic Segmentation and Edge Detection Network for 3D Point Clouds. In *ECCV*.

Jiang, L.; Zhao, H.; Liu, S.; Shen, X.; Fu, C.-W.; and Jia, J. 2019. Hierarchical Point-Edge Interaction Network for Point Cloud Semantic Segmentation. In *ICCV*.

Lai, X.; Liu, J.; Jiang, L.; Wang, L.; Zhao, H.; Liu, S.; Qi, X.; and Jia, J. 2023. Stratified Transformer for 3D Point Cloud Segmentation. *arXiv preprint arXiv:2203.14508*.

Lamas, D. L.; Soilán, M.; Grandio, J.; and Riveiro, B. 2021. Automatic Point Cloud Semantic Segmentation of Complex Railway Environments. In *Remote Sensing*.

Landrieu, L.; and Simonovsky, M. 2018. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In *CVPR*.

Li, G.; Muller, M.; Thabet, A.; and Ghanem, B. 2019. DeepGCNs: Can GCNs Go As Deep As CNNs? In *ICCV*.

Li, S.; Liu, Y.; and Gall, J. 2020. Rethinking 3D LiDAR Point Cloud Segmentation. *arXiv preprint arXiv:2008.03928*.

Liu, X.; Han, Z.; Liu, Y.-S.; and Zwicker, M. 2019. Point2Sequence: Learning the Shape Representation of 3D Point Clouds with an Attention-based Sequence to Sequence Network. In *AAAI*.

Mao, J.; Wang, X.; and Li, H. 2019. Interpolated Convolutional Networks for 3D Point Cloud Understanding. In *ICCV*.

Ni, P.; Zhang, W.; Zhu, X.; and Cao, Q. 2020. PointNet++ Grasping: Learning An End-to-end Spatial Grasp Generation Algorithm from Sparse Point Clouds. In *ICRA*.

Park, J.; Lee, S.; Kim, S.; Xiong, Y.; and Kim, H. 2023. Self-positioning Point-based Transformer for Point Cloud Understanding. *arXiv preprint arXiv:2303.16450*.

Qiu, S.; Anwar, S.; and Barnes, N. 2021. Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion. In *CVPR*.

Sun, J.; Qing, C.; Tan, J.; and Xu, X. 2022. Superpoint Transformer for 3D Scene Instance Segmentation. *arXiv preprint arXiv:2211.15766*.

Tang, L.; Zhan, Y.; Chen, Z.; Yu, B.; and Tao, D. 2022. Contrastive Boundary Learning for Point Cloud Segmentation. *arXiv preprint arXiv:2203.05272*.

Thomas, H.; Qi, C. R.; Deschaud, J.-E.; Marcotegui, B.; Goulette, F.; and Guibas, L. 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. In *CVPR*.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. In *Journal of Machine Learning Research*.

Wang, J.; Wang, C.; Xi, X.; Du, M.; Wang, P.; and Nie, S. 2022. Segmentation of the communication tower and its accessory equipment based on geometrical shape context from 3D point cloud. In *International Journal of Digital Earth*.

Wang, L.; Huang, Y.; Hou, Y.; Zhang, S.; and Shan, J. 2019a. Graph Attention Convolution for Point Cloud Semantic Segmentation. In *CVPR*.

Wang, L.; Huang, Y.; Hou, Y.; Zhang, S.; and Shan, J. 2019b. Graph Attention Convolution for Point Cloud Semantic Segmentation. In *CVPR*.

Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019c. Dynamic Graph CNN for Learning on Point Clouds. In *ACM Transactions on Graphics*.

Wu, W.; Qi, Z.; and Fuxin, L. 2019. PointConv: Deep Convolutional Networks on 3D Point Clouds. In *CVPR*.

Xiu, H.; Liu, X.; Wang, W.; Kim, K.-S.; Shinohara, T.; Chang, Q.; and Matsuoka, M. 2022. Interpretable Edge Enhancement and Suppression Learning for 3D Point Cloud Segmentation. *arXiv preprint arXiv:2209.09483*.

Yan, X.; Zheng, C.; Li, Z.; Wang, S.; and Cui, S. 2020. PointASNL: Robust Point Clouds Processing using Nonlocal Neural Networks with Adaptive Sampling. In *CVPR*.

Ye, M.; Wan, R.; Xu, S.; Cao, T.; and Chen, Q. 2021. DRINet++: Efficient Voxel-as-point Point Cloud Segmentation. In *CVPR*.

Ye, M.; Wan, R.; Xu, S.; Cao, T.; and Chen, Q. 2022. Efficient Point Cloud Segmentation with Geometry-aware Sparse Networks. In *ECCV*.

Zhang, Z.; Yang, B.; Wang, B.; and Li, B. 2023. GrowSP: Unsupervised Semantic Segmentation of 3D Point Clouds. *arXiv preprint arXiv:2305.16404*.

Zhao, H.; Jiang, L.; Jia, J.; Torr, P.; and Koltun, V. 2021. Point Transformer. In *ICCV*.