iBALR3D: imBalanced-Aware Long-Range 3D Semantic Segmentation [†]

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

1 China Southern Power Grid Digital Power Grid Group Co., Ltd., Guangzhou 510663, China
2 School of Computer Science and Technology, Xidian University, Xi'an 710126, China
* Correspondence: zhangky@csg.cn
† Presented at the 2nd AAAI Workshop on Artificial Intelligence with Biased or Scarce Data (AIBSD), Vancouver, BC, Canada, 26 February 2024.

Abstract: Three-dimensional 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.

Keywords: semantic segmentation; transmission line segmentation; point cloud segmentation

1. Introduction

Three-dimensional point cloud semantic segmentation is an important task that classifies all points into their corresponding categories [1]. 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 [2]. 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 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. [3] 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. [4] 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. [5] proposed a robust method for segmenting point cloud data of communication towers and accessory equipment based on geometrical shape context from a 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 and the framework is shown in Figure 1. 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., 500 KV) 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.
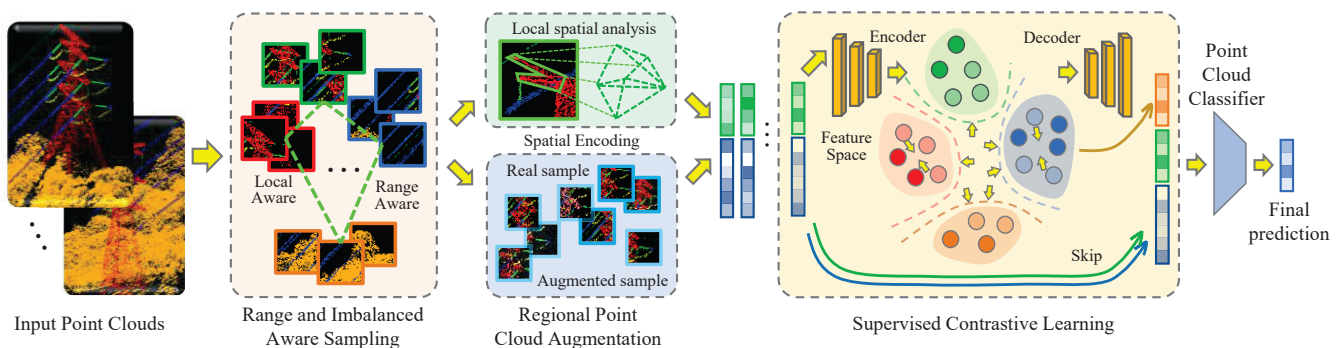

Figure 1. Framework of our iBALR3D model. A long-range and imbalanced-aware sampling strategy is deployed to balance the significant data imbalance issue and align point clouds in the long-range distance. An adaptive spatial encoder is designed to extract indistinguishable junctional regions across simple shapes. A contrastive training associated with an augmentation module is used to enhance the learning capacity of tail categories and achieve the overall highest performance.

2. 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. Ref. [6] reformulated point-based methods to operate in the projection space, which significantly improved the efficiency of processing LiDAR point clouds. Ref. [7] proposed an efficient and lightweight neural architecture to di-

rectly interpret point semantics for large-scale point clouds. Similarly, ref. [8] designed a self-positioning point-based transformer that shows promising results in point cloud understanding. Other classical research includes [9–16]. 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++ [17] jointly learns the sparsity and geometric properties of a point cloud with a voxel-as-point principle. Ref. [18] introduced a Geometry-aware Sparse Network (GASN) which leverages the sparsity and geometric properties of point clouds within a unified voxel representation. HilbertNet [19] preserves the benefits of voxel-based methods while significantly reducing computational costs through a Hilbert curve-based flattening mechanism. Ref. [20] proposed a teacher–student strategy, which eventually uses a small network to perform 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. Ref. [21] introduced an attention mechanism into the graph convolution process, thereby improving the model's capacity to concentrate on crucial points. Ref. [1] 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. Ref. [22] presented a method that utilized point and edge features in a hierarchical graph framework to label 3D scenes with semantic categories. Ref. [23] proposed PointASNL, which processes noisy point clouds robustly using adaptive sampling and local–nonlocal modules. Other research works include [21,24,25]. 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 [26] are gaining attention for their proficiency in capturing long-range data dependencies. Xin Lai et al. [27] proposed a stratified strategy for sampling keys to harvest long-range contexts, demonstrating the potential of transformers in this field. SPFormer [28] 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. Ref. [29] 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 are scarce.

3. 500KV3D Dataset

We present the 500KV3D dataset, a large, high-quality, and well-structured point cloud dataset collected from 500 KV 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 2.
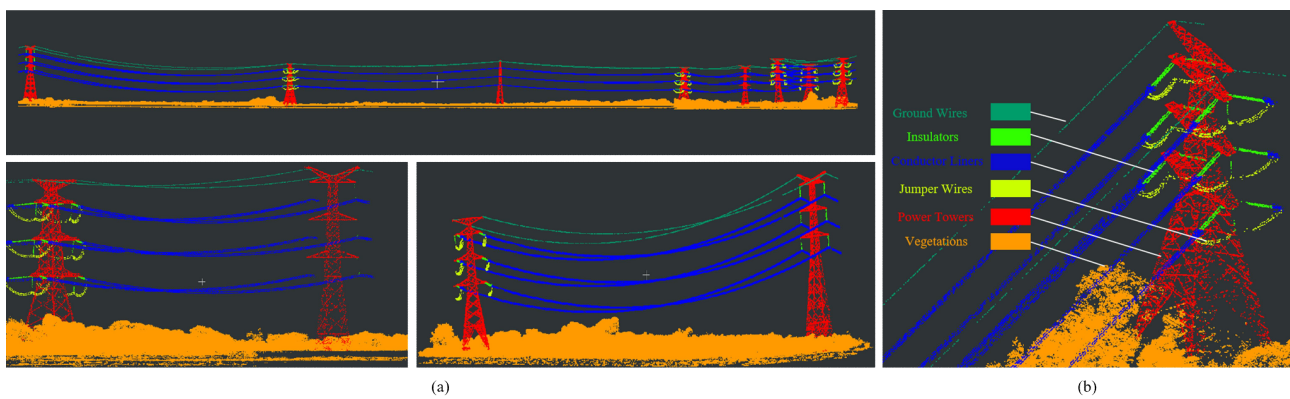
Figure 2. We introduce a novel 500KV3D dataset. 500KV3D is a large-scale long-range 3D point cloud dataset, which is collected from a high-voltage-level, 500 KV smart-grid infrastructure. (**a**) illustrates a few distant views and (**b**) is the zoomed-in view. We consider that 500KV3D could provide more insights into deploying multimedia models in electrical grid-related topics. Details and statistical analysis are provided in the 500KV3D dataset in Section 3.

3.1. Data Collection

Due to the extremely high voltage level, the power lines are usually in high and inaccessible locations, which leads to difficulties 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) (https://globalgpssystems.com/liair-220n/, accessed on 26 February 2024). It is specifically designed for mounting on drones (Unmanned Aerial Vehicles, UAVs). The system is equipped with a Hesai Pandar40P laser scanner (https://www.hesaitech.com/product/pandar40p/, accessed on 26 February 2024), making it one of the most cost-effective options in GVI's LiAir Series. More detailed configurations are listed in Table 1.

Table 1. Specifications of LiDAR sensor LiAir 220N, which is used to collect 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

3.2. Pre-Processing

Despite using professional LiDAR sensors, outliers and noise are 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.

3.3. Labeling

We consider six 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, and 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, and carry electric current from the power plants to the substations and consumers; (6) *vegetation* is considered to be any ground objects which contain trees, shrubs, hedges, bushes, etc. To streamline the labor-intensive process of 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 (https://www.danielgm.net/cc/, accessed on 26 February 2024) is used for conducting the annotation; it 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 m and point scales per segment from 10 k to almost 2 M.

3.4. 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 box-plot. 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. These data reflect the complexity of the real-world transmission line environment and reveal 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 varied.
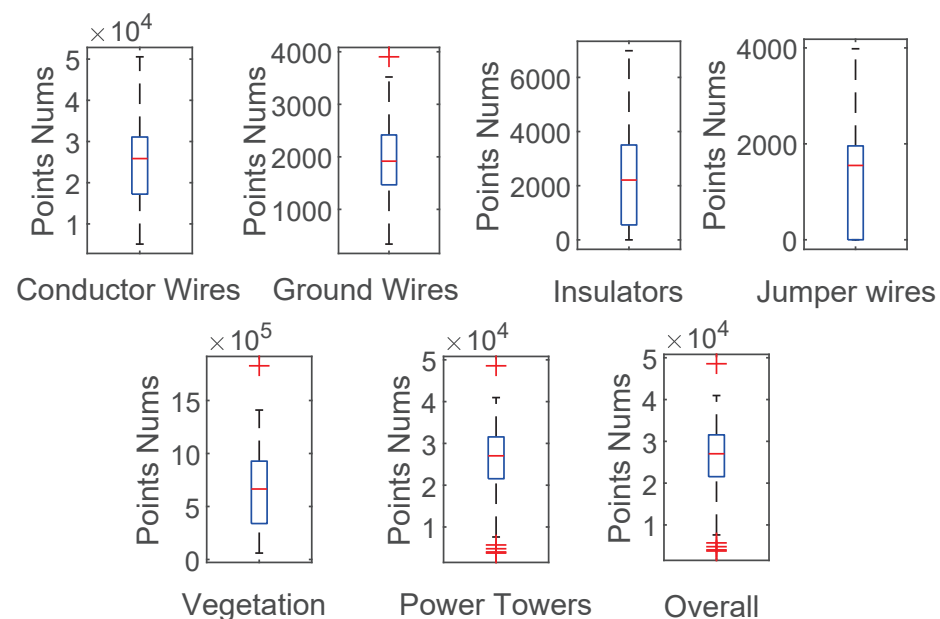

Figure 3. Point number distribution analysis of our 500KV3D dataset. All points are separated into 42 sections; the box plots illustrate the point number distributions across different semantic categories as well as different sections.

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

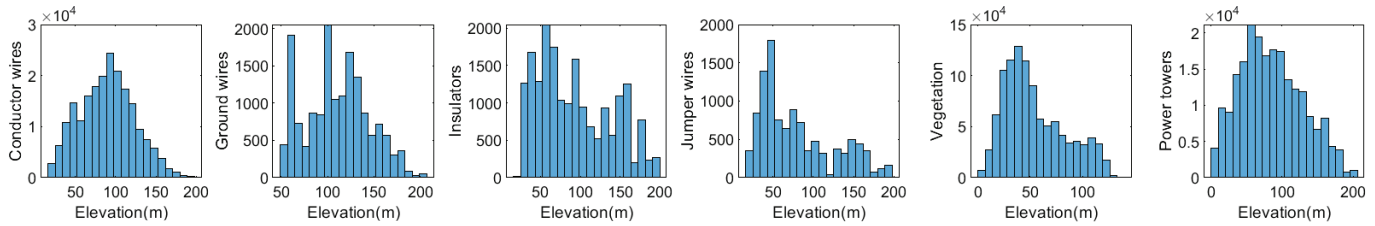| Categories	Overall Number of Points	Overall Ratio (%)	Training Set Number of Points	Training Split Ratio (%)	Testing Set Number of Points	Testing Split Ratio (%)
Conductor Lines	1,032,617	3.50	839,454	81.29	193,163	18.71
Ground Wires	80,767	0.27	64,683	80.09	16,084	19.91
Insulators	95,193	0.27	73,996	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
Vegetation	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 wire-related points are more fluctuated.

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 that it can contribute more to related research communities.

4. 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 the section below.

4.1. Enhanced Imbalanced Contrastive Learning

The significant imbalanced data distribution leads to difficulty for the model in learning 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.

We initialize a possibility for each point in a scene based on the 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 x by searching the nearest 40,960 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.

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

where $\tilde{x}$ 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 neighboring points in terms of different categories. The objective function can be represented as:

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

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 Equation (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:

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

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

4.2. 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 that the 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 4.1. In this study, the optimal value for k is 8, and the algorithm can be represented as:

$$
P(\vec{n}, d) = \arg\min_{(\vec{n}, d)} \sum_{i=1}^{k} (\vec{n} \cdot p_i - d)^2
\tag{4}
$$

We perform eigenvalue decomposition on the covariance matrix M in Equation (4) 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 neighborhood is, the larger δ is, the greater the fluctuation of the neighborhood is. We concatenate the calculated normal vectors and curvature to the original coordinates before conducting contrastive learning.

4.3. Long-Range and Imbalanced Sampling

Contrastive learning and spatial encoding enhance 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 a higher sampling ratio compared with their sample number ratios. Moreover, for a selected point

x_i, we measure the diversity of its neighbors. The more diverse the neighbors, the higher the learning requirements. By finding the top nearest neighbor points of x_i, our method could also reach a long range in point-sparse regions, especially for the tail categories. The sampling strategy is illustrated below:

$$\mathcal{P}(x_i) = \frac{n_{y_i}^{\alpha}}{\sum_{k=1}^{n_i} n_i^{\alpha}} + \beta \frac{n_{T_{knn}}}{\sum_{k=1}^{n_i}}, \tag{5}$$

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.

4.4. Implementation

We use a multi-layer perceptron with two 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 0.04 m 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 an RTX4090 GPU with 128 GB memory.

5. Experiments

5.1. Experimental Setup

For benchmarks, five state-of-the-art benchmarks are used for our experiments. More specifically, **PointNet** [9] 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++** [10] 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** [7] 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** [30] 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** [27] 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 [7,31–33]. It measures the average overlap between the predicted and ground truth regions for each class in a point cloud:

$$mIoU = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i}, \tag{6}$$

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 eight-dimensional embedding vectors obtained by contrastive learning.

5.2. Performance

The performance of benchmarks and our method in the 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 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.

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

Methods	Category-Level Segmentation mIoU (%)						Overall mIoU (%)
	Cro.	Gon.	Ins.	Jum.	Veg.	Pow.	
PointNet [9]	65.53	46.56	0.92	2.41	97.52	41.71	42.44
PointNet++ [10]	84.31	77.82	13.00	25.51	99.51	71.61	61.96
StratifiedTransformer [27]	89.32	85.85	19.98	40.56	99.73	85.14	70.10
RandLA-Net [7]	99.38	98.58	91.23	98.24	99.93	97.55	97.49
BAAF-Net [30]	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**

To further analyze the effectiveness of our model, t-SNE [34] is used to visualize the learned point cloud representations and the results are shown in Figure 5, where (a) and (b) denote the representations of RandLA-Net and our iBALR3D approaches, respectively, and different colors denotes different categories. Considering the significantly imbalanced point number distribution, we intentionally increase the ratios of the tail categories for better visualization. From Figure 5, we can see that our model achieves more distinguishing representations where the same categories are more clustered in the same regions.

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.
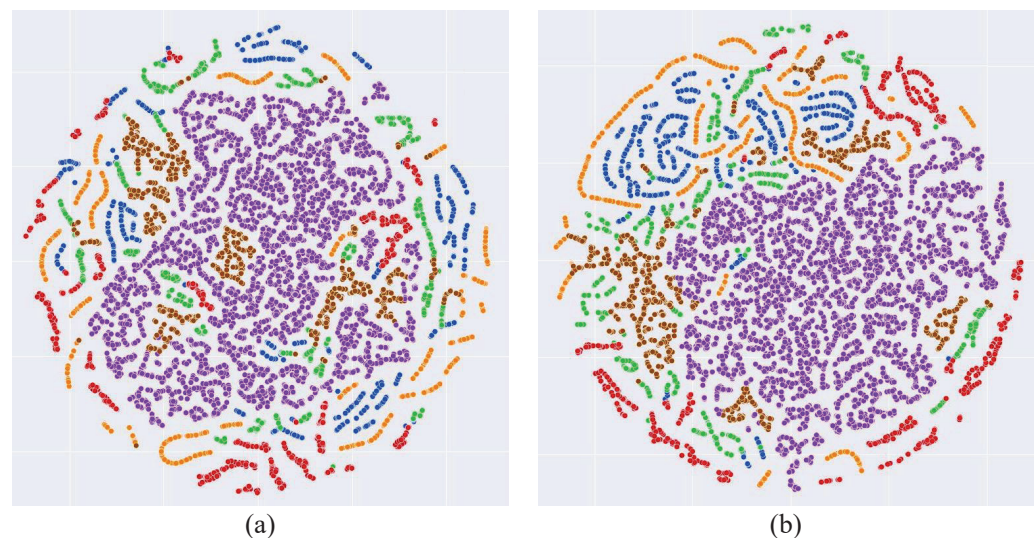

(a)   (b)

Figure 5. 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.
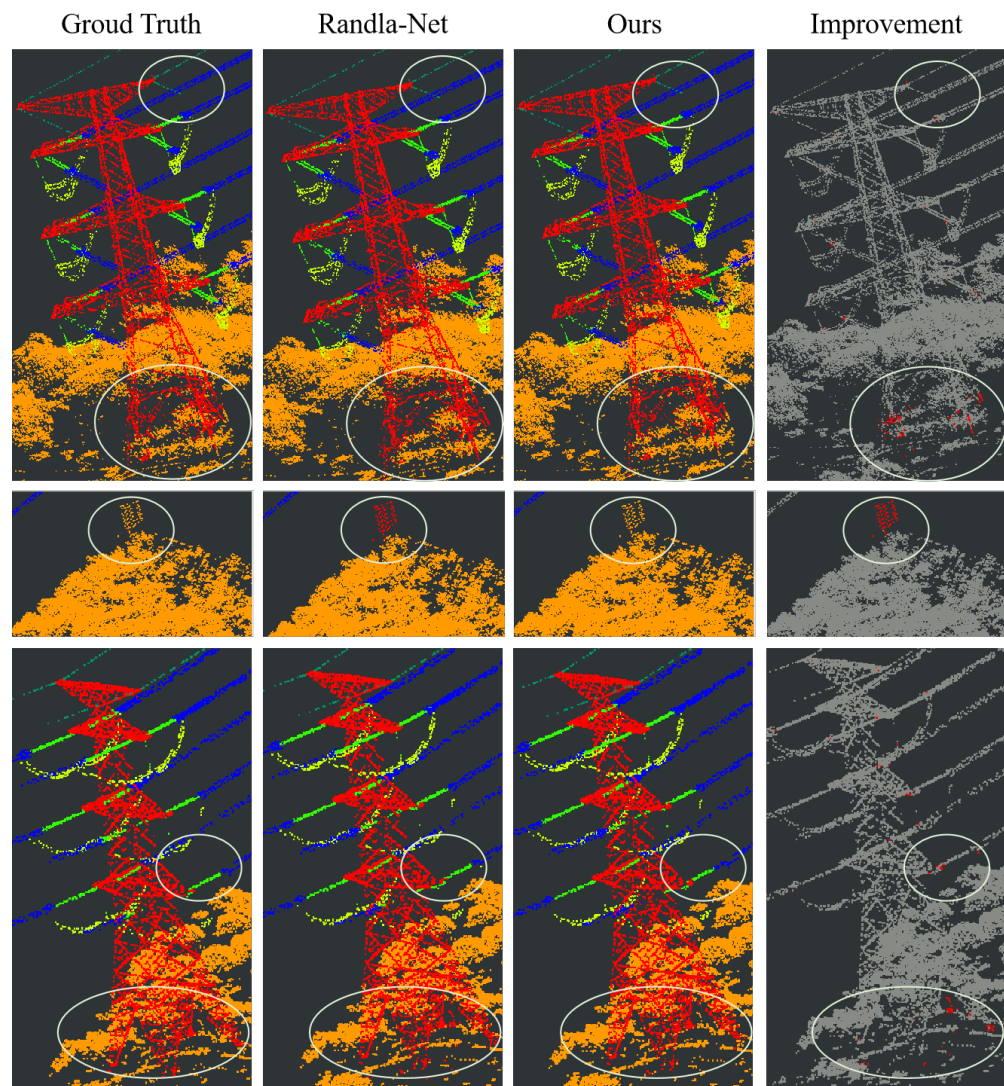
Figure 6. We visualize the results of RandLA-Net baseline and our iBALR3D and the improvements on several different scenes. We can see that iBALR3D can effectively reduce errors on the junctional regions (e.g., *Power Tower*).

5.3. 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 a random sampling strategy. The results are shown in Table 4 and comparison in the training stage can be observed in Figure 7. This ablation study demonstrates how the proposed modules synergistically improve performance.

Table 4. Ablation study of our iBALR3D model.

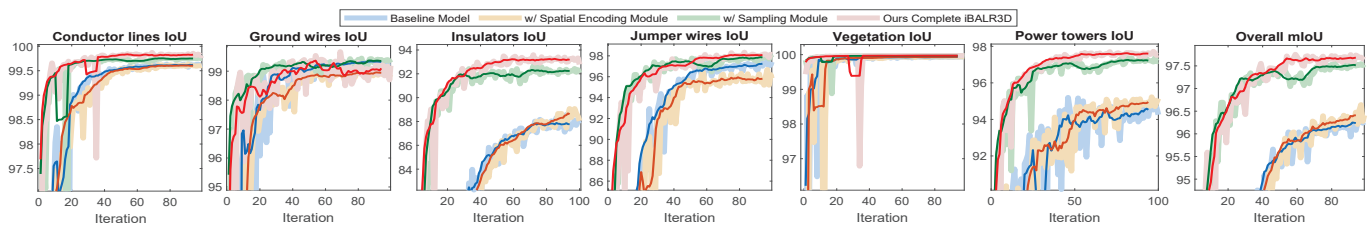| Methods	Category-Level Segmentation mIoU (%)						Overall mIoU (%)
	Cro.	Gon.	Ins.	Jum.	Veg.	Pow.	
Baseline RandLA-Net	99.38	98.58	91.23	98.24	99.93	97.55	97.49
Ours w/o spatial encoding module	99.49	99.06	92.77	97.86	99.95	98.30	97.90
Ours w/o sampling module	99.64	98.83	94.42	98.52	99.95	98.36	98.29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

Figure 7. 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 color curve is the exact performance, and the darker color 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.

6. Conclusions

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.

Author Contributions: Conceptualization, K.Z. and R.C.; methodology, K.Z. and R.C.; software, K.Z.; validation, K.Z.; formal analysis, K.Z.; investigation, K.Z.; resources, X.W.; data curation, P.Q.; writing—original draft preparation, K.Z.; writing—review and editing, K.Z.; visualization, K.Z.; supervision, X.W.; project administration, P.Q.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: Keying Zhang, Xinqiao Wu, Jiguang Zhao and Ping Qin have received research grants from China Southern Power Grid Digital Power Grid Group Co., Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Landrieu, L.; Simonovsky, M. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In Proceedings of the CVPR, Salt Lake City, UT, USA, 18–23 June 2018.
2. Zhang, K. MADA: Mask Aware Domain Adaptation for Open-set Semantic Segmentation. In Proceedings of the 2nd Workshop on Sustainable AI (SAI-AAAI2024), Vancouver, BC, Canada, 20 February 2024.
3. Grandio, J.; Riveiro, B.; Lamas, D.; Arias, P. Multimodal deep learning for point cloud panoptic segmentation of railway environments. *Autom. Constr.* **2023**, *150*, 104854. [CrossRef]
4. Lamas, D.L.; Soilán, M.; Grandio, J.; Riveiro, B. Automatic Point Cloud Semantic Segmentation of Complex Railway Environments. *Remote Sens.* **2021**, *13*, 2332. [CrossRef]
5. Wang, J.; Wang, C.; Xi, X.; Du, M.; Wang, P.; Nie, S. Segmentation of the communication tower and its accessory equipment based on geometrical shape context from 3D point cloud. *Int. J. Digit. Earth* **2022**, *15*, 1547–1566. [CrossRef]
6. Li, S.; Liu, Y.; Gall, J. Rethinking 3D LiDAR Point Cloud Segmentation. *arXiv* **2020**, arXiv:2008.03928.
7. Hu, Q.; Yang, B.; Xie, L.; Rosa, S.; Guo, Y.; Wang, Z.; Trigoni, N.; Markham, A. Randla-net: Efficient semantic segmentation of large-scale point clouds. In Proceedings of the CVPR, Seattle, WA, USA, 13–19 June 2020.

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

9.   Garcia-Garcia, A.; Gomez-Donoso, F.; Garcia-Rodriguez, J.; Orts-Escolano, S.; Cazorla, M.; Azorin-Lopez, J. PointNet: A 3D Convolutional Neural Network for real-time object class recognition. In Proceedings of the IJCNN, Vancouver, BC, Canada, 24–29 July 2016.

10. Ni, P.; Zhang, W.; Zhu, X.; Cao, Q. PointNet++ Grasping: Learning An End-to-end Spatial Grasp Generation Algorithm from Sparse Point Clouds. In Proceedings of the ICRA, Paris, France, 31 May–31 August 2020.

11. Thomas, H.; Qi, C.R.; Deschaud, J.E.; Marcotegui, B.; Goulette, F.; Guibas, L. KPConv: Flexible and Deformable Convolution for Point Clouds. In Proceedings of the CVPR, Long Beach, CA, USA, 15–20 June 2019.

12. Liu, X.; Han, Z.; Liu, Y.S.; Zwicker, M. Point2Sequence: Learning the Shape Representation of 3D Point Clouds with an Attention-based Sequence to Sequence Network. In Proceedings of the AAAI, Honolulu, HI, USA, 27 January–1 February 2019.

13. Chiang, H.Y.; Lin, Y.L.; Liu, Y.C.; Hsu, W. A Unified Point-Based Framework for 3D Segmentation. In Proceedings of the CVPR, Quebec City, QC, Canada, 16–19 September 2019.

14. Wu, W.; Qi, Z.; Fuxin, L. PointConv: Deep Convolutional Networks on 3D Point Clouds. In Proceedings of the CVPR, Long Beach, CA, USA, 15–20 June 2019.

15. Mao, J.; Wang, X.; Li, H. Interpolated Convolutional Networks for 3D Point Cloud Understanding. In Proceedings of the ICCV, Seoul, Republic of Korea, 27 October–2 November 2019.

16. Hu, Z.; Zhen, M.; Bai, X.; Fu, H.; Tai, C.L. JSENet: Joint Semantic Segmentation and Edge Detection Network for 3D Point Clouds. In Proceedings of the ECCV, Glasgow, UK, 23–28 August 2020.

17. Ye, M.; Wan, R.; Xu, S.; Cao, T.; Chen, Q. DRINet++: Efficient Voxel-as-point Point Cloud Segmentation. In Proceedings of the CVPR, Nashville, TN, USA, 20–25 June 2021.

18. Ye, M.; Wan, R.; Xu, S.; Cao, T.; Chen, Q. Efficient Point Cloud Segmentation with Geometry-aware Sparse Networks. In Proceedings of the ECCV, Tel Aviv, Israel, 23–27 October 2022.

19. Chen, W.; Zhu, X.; Chen, G.; Yu, B. Efficient Point Cloud Analysis Using Hilbert Curve. In Proceedings of the ECCV, Tel Aviv, Israel, 23–27 October 2022.

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

21. Wang, L.; Huang, Y.; Hou, Y.; Zhang, S.; Shan, J. Graph Attention Convolution for Point Cloud Semantic Segmentation. In Proceedings of the CVPR, Long Beach, CA, USA, 15–20 June 2019.

22. Jiang, L.; Zhao, H.; Liu, S.; Shen, X.; Fu, C.W.; Jia, J. Hierarchical Point-Edge Interaction Network for Point Cloud Semantic Segmentation. In Proceedings of the ICCV, Seoul, Republic of Korea, 27 October–2 November 2019.

23. Yan, X.; Zheng, C.; Li, Z.; Wang, S.; Cui, S. PointASNL: Robust Point Clouds Processing using Nonlocal Neural Networks with Adaptive Sampling. In Proceedings of the CVPR, Seattle, WA, USA, 14–19 June 2020.

24. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.* **2019**, *38*, 146. [CrossRef]

25. Li, G.; Muller, M.; Thabet, A.; Ghanem, B. DeepGCNs: Can GCNs Go As Deep As CNNs? In Proceedings of the ICCV, Seoul, Republic of Korea, 27 October–2 November 2019.

26. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.; Koltun, V. Point Transformer. In Proceedings of the ICCV, Virtual, 11–17 October 2021. [CrossRef]

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

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

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

30. Qiu, S.; Anwar, S.; Barnes, N. Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion. In Proceedings of the CVPR, Nashville, TN, USA, 20–25 June 2021.

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

32. Hu, Q.; Yang, B.; Khalid, S.; Xiao, W.; Trigoni, N.; Markham, A. SensatUrban: Learning Semantics from Urban-Scale Photogrammetric Point Clouds. *Int. J. Comput. Vis.* **2022**, *130*, 316–343. [CrossRef]

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

34. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.