*electronics*

MDPI

*Article*

# Spatial Attention-Based Kernel Point Convolution Network for Semantic Segmentation of Transmission Corridor Scenarios in Airborne Laser Scanning Point Clouds

**Fangrong Zhou, Gang Wen *, Yi Ma, Hao Pan, Guofang Wang and Yifan Wang**

Joint Laboratory of Power Remote Sensing Technology, Electric Power Research Institute, Yunnan Power Grid Company Ltd., China Southern Power Grid, Kunming 650217, China; fangrongzhou2023@163.com (F.Z.); mayi@yn.csg.cn (Y.M.); yndwpanhao@163.com (H.P.); wangguofang202311@163.com (G.W.); wangyifan@yn.csg.cn (Y.W.)

* Correspondence: wen_gang2005@163.com

**Abstract:** Accurate semantic segmentation in transmission corridor scenes is crucial for the maintenance and inspection of power infrastructure, facilitating the timely detection of potential hazards. In this study, we propose SA-KPConv, an advanced segmentation model specifically designed for transmission corridor scenarios. Traditional approaches, including Random Forest and point-based deep learning models such as PointNet++, demonstrate limitations in segmenting critical infrastructure components, particularly power lines and towers, primarily due to their inadequate capacity to capture complex spatial relationships and local geometric details. Our model effectively addresses these challenges by integrating a spatial attention module with kernel point convolution, enhancing both global context and local feature extraction. Experiments demonstrate that SA-KPConv outperforms state-of-the-art methods, achieving a mean Intersection over Union (mIoU) of 89.62%, particularly excelling in challenging terrains such as mountainous areas. Ablation studies further validate the significance of our model's components in enhancing overall performance and effectively addressing class imbalance. This study presents a robust solution for semantic segmentation, with considerable potential for monitoring and maintaining power infrastructure.

**Keywords:** semantic segmentation; transmission corridor scenarios; kernel point convolution; point clouds

check for updates

**Citation:** Zhou, F.; Wen, G.; Ma, Y.; Pan, H.; Wang, G.; Wang, Y. Spatial Attention-Based Kernel Point Convolution Network for Semantic Segmentation of Transmission Corridor Scenarios in Airborne Laser Scanning Point Clouds. *Electronics* **2024**, *13*, 4501. https://doi.org/10.3390/electronics13224501

Academic Editors: Akshya Swain and Chathura Wanigasekara

Received: 10 October 2024
Revised: 5 November 2024
Accepted: 12 November 2024
Published: 15 November 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Transmission corridors are essential components of power systems, as their operational status significantly influences the safety and stability of electricity supply. The expansion of modern power grids has led to an increase in the scale and complexity of these lines, underscoring the importance of accurate monitoring and management of their operational status. Regular inspections play a crucial role in understanding the operational conditions of these lines and identifying any changes in their surrounding environments, which can reveal potential equipment defects and safety hazards. This information enables the formulation of targeted maintenance recommendations to address these issues promptly, thereby preventing accidents or mitigating their impacts [1]. Moreover, as electricity demand grows, the distribution range of transmission lines expands, leading to a significant increase in inspection workload. The extensive coverage and complex terrain of these lines pose challenges, reducing inspection efficiency and increasing safety risks [2].

Traditional inspection methods for transmission corridor primarily rely on manual operations, encountering challenges such as long distances, high labor intensity, and inadequate inspection techniques. Moreover, unfavorable weather conditions, including thunderstorms, snowfall, earthquakes, and landslides, severely restrict the extent of inspections. In contrast, inspection technologies that employ active and passive remote

sensing technologies provide increased efficiency while decreasing time and labor expenses. Notably, Light Detection and Ranging (LiDAR) technology has become a significant player in the power sector due to its all-weather capability, proactive approach, and high accuracy, significantly boosting the efficiency of data collection in transmission corridors. However, as the efficiency of data collection improves, the industry faces an urgent need for efficient techniques to analyze and interpret the data from transmission corridor scenes swiftly and effectively.

The primary objective of transmission corridor engineering applications is to extract and segment features from noisy point cloud data, thereby achieving a three-dimensional semantic understanding of the scene. To date, various techniques for 3D semantic understanding have been developed, including unsupervised methods, such as region growing [3], principal component analysis [4], and clustering [5]. However, these unsupervised approaches predominantly rely on manually defined rules or physical models for feature segmentation, limiting their broad applicability and resulting in significant parameter dependence. Some supervised methods based on machine learning, such as random forests [6–9] and support vector machines (SVM) [10], utilize manually selected feature datasets for point cloud classification in transmission lines. Nevertheless, as the number of classification categories and data volume increase, these methods struggle to meet current production demands due to parameter sensitivity and inadequate generalization, particularly in areas with diverse terrains within transmission corridors. Recently, data-driven deep learning methods have gained significant attention, achieving notable success in fields such as natural language processing [11] and computer vision, including image classification [12] and segmentation [13]. The introduction of PointNet [14] and PointNet++ [15] has further validated the effectiveness of applying deep learning techniques for convolutions on unordered 3D point clouds. Methods such as RandLA-Net [16] and KPConv [17] have opened new avenues for the semantic segmentation of large-scale 3D point cloud data. These studies confirm that learning-based segmentation of outdoor large-scale 3D point clouds is effective, with both efficiency and accuracy surpassing those of methods based on handcrafted features.

It is crucial to note that transmission corridor encompass extensive, large-scale scenes, with point cloud data between towers often exceeding tens of millions of points. However, critical power infrastructure, such as power towers and lines, comprises only about 3–5% of the overall scene. Additionally, the varied terrain in transmission corridor environments adds complexity to understanding these landscapes. Conventional semantic segmentation methods primarily focus on indoor and street scenarios that utilize high-quality, dense point cloud data (e.g., MLS and TLS), while research on sparse, relatively lower-quality airborne laser scanning (ALS) point cloud data remains limited.

Building on this foundation, the current work aims to achieve semantic segmentation of transmission corridor scenes using airborne point cloud data. By considering various terrain types and the class imbalance present in transmission corridors, we have developed a semantic segmentation model that integrates kernel point convolution and attention mechanisms. Numerous experiments have been conducted to validate the model's effectiveness and performance.

The main contributions of this study can be summarized as follows:

1. To address the issues of sparsity in airborne transmission corridor point clouds and the presence of numerous small-category targets, we employ deformable kernel point convolution (KPConv) to learn the spatial geometric features of the power scene point clouds. By updating point weights through kernel functions, we enhance the learning of point cloud convolution, thereby improving the feature extraction capability for various land cover targets.
2. To achieve precise semantic segmentation of transmission corridor scenes at a large scale, we introduce a spatial attention module. This module models the complex interactions between points to enhance the model's perception of contextual information, thereby improving point cloud segmentation accuracy from a global perspective.

3. Experiments were conducted on field-collected transmission scene data, achieving an average intersection over union (IoU) of 89.62%, demonstrating superior performance compared to other methods of the same type. Additionally, experiments conducted across multiple terrains yielded mean intersection over union (mIoU) values exceeding 87%, confirming the robustness of our approach.

The remainder of the paper is structured as follows. Section 2 provides a review of traditional methods and recent deep learning approaches for semantic segmentation of point clouds in transmission line scenarios. In Section 3, we introduce the data preprocessing steps, the architecture of the SA-KPConv network, and detail the kernel point convolution and spatial attention modules. Section 4 presents the results on our self-collected dataset, along with a comparison and analysis of other methods. Finally, Section 5 concludes the paper.

## 2. Relate Work

Semantic segmentation of transmission corridors aims to accurately distinguish transmission corridors and their components (e.g., conductors and towers) from complex backgrounds by classifying each point. This process requires not only the precise identification of the transmission corridor itself, but also effective differentiation from other structures and related objects, such as surrounding vegetation, buildings, and roads. Based on the type of features, point cloud semantic segmentation can be categorized into two types: handcrafted methods and data-driven methods.

### *2.1. Hand-Craft Based Methods*

Point cloud semantic segmentation initially relied on geometric constraints and statistical rules, using manually defined point cloud features for data segmentation. Early methods included unsupervised clustering [18–20] and principal component analysis (PCA) [21–23]. K-means clustering, a classic unsupervised method, requires prior determination of the number of classes for segmentation, optimizing classification by minimizing intra-class distance and maximizing inter-class distance. Although this method is fast, it is sensitive to outliers and may fail to identify a globally optimal solution [18]. PCA reduces data dimensionality through linear transformations, aiming to concentrate effective information in fewer dimensions; however, this linear approach limits its ability to handle complex image data, contributing minimally to enhancing segmentation performance [21]. In supervised learning, random forests enhance model accuracy and stability by aggregating predictions from multiple decision trees, which reduces variance and improves generalization, making them particularly effective with high-dimensional data and complex patterns. However, compared to a single decision tree, they offer weaker interpretability and may struggle to distinguish geometrically similar objects, such as flat roads [9]. Support vector machines (SVM) aim to identify hyperplanes in high-dimensional space to separate different classes; however, their performance may degrade in multi-class semantic segmentation, as hyperplanes may not fulfill all requirements [24]. Overall, methods based on manually selected features are highly dependent on feature selection and parameter settings, rendering them ill-suited for complex geometric structures or variable point clouds, making them susceptible to overfitting in high-dimensional data scenarios. In contrast, data-driven methods can automatically learn features and optimize in an end-to-end manner, often demonstrating superior performance and robustness in handling complex point cloud data.

### *2.2. Data-Driven-Based Methods*

Deep learning-based semantic segmentation methods utilize deep network architectures to autonomously extract both explicit and implicit features from raw data, continuously enhancing classifier accuracy through training sets. This facilitates effective semantic segmentation of large-scale point cloud data. Currently, deep learning semantic segmentation methods can be classified into three categories: projection-based segmentation, voxel-based segmentation, and point-based segmentation.

Projection-based semantic segmentation methods involve projecting 3D point clouds onto 2D planes to obtain 2D images of the target point cloud from multiple viewpoints, which are subsequently used as inputs for deep neural networks for semantic segmentation. Xu et al. input sequential 2D projection images from known viewpoints into long short-term memory (LSTM) networks to predict 2D projections from alternative angles, thereby integrating effective features for point cloud segmentation [12]. Lawin et al. employed projection images as inputs to 2D convolutional neural networks, predicting segmentation based on depth, surface normals, and additional information [25]. Huang et al. introduced a view fusion model (VMM) to extract and reduce the dimensionality of features from projection images, subsequently utilizing the network for semantic segmentation [26]. While these methods leverage 2D information from multiple planes for comprehensive semantic analysis, the quality of the projected 2D images is influenced by the projection angles, which can result in the loss of spatial geometric information.

Voxel-based semantic segmentation methods partition the original point cloud into voxels of a specific spatial size, facilitating the pixelation of the point cloud in 3D space. This approach preserves the neighborhood structure of the original point cloud while offering good scalability for the transformed data. Tchapmi et al. voxelized point clouds and utilized 3D fully convolutional networks to obtain voxel labels corresponding to the original point cloud, followed by label prediction using fully connected conditional random fields [27]. Wu et al. developed the ShapeNets network specifically for training with voxelized point clouds to achieve semantic segmentation [28]. Meng et al. interpolated individual voxels and utilized variational autoencoders to encode the geometric structures within the voxels, employing this encoded information for semantic segmentation [29]. Although voxel-based methods address the unstructured nature of point clouds, voxelization introduces inherent limitations.

Point-based semantic segmentation preserves the geometric structure of point clouds and facilitates direct processing in 3D space. This category can be further classified into per-point multilayer perceptron (MLP) methods, point convolution methods, recurrent neural network (RNN) methods, and graph optimization segmentation methods. The per-point MLP method represents pioneering work in point-based segmentation, ensuring permutation invariance of the point cloud through transformation matrices and addressing unordered issues using max pooling functions, thus achieving semantic segmentation at the point level [15]. Point convolution methods preprocess point clouds to reweight and reorder features of each point, achieving "regularization" of the point cloud and enabling further processing with convolutional neural networks (CNNs) for semantic segmentation [30]. RNN methods leverage spatial context information from the point cloud to learn local features and perform 3D semantic segmentation using bidirectional RNNs [31]. Graph optimization methods convert point cloud data into graph data, describing relationships among neighboring points and contextual information, and utilize graph convolutional networks for learning [32]. While point-based semantic segmentation allows for the direct input of the original point cloud for end-to-end segmentation, the large volume of raw data results in prolonged training times and increased model complexity.

Some studies have utilized multimodal data, such as satellite imagery, optical photographs, and millimeter-wave radar, to enhance LiDAR point cloud data, thereby improving semantic segmentation accuracy [33–37]. However, the requirements for these multimodal data are quite stringent for transmission corridor scenes. Additionally, several new network frameworks have been proposed, including those based on Transformer architectures [38–40]. Many of these approaches have not yet been validated for application in large-scale scenarios. Among them, SPT [41,42], an optimization of its predecessor SPG [43], leverages SuperPoint to express geometric attributes while employing Transformer attention to capture relationships between multiscale superpoints. Although this approach incurs additional computational overhead, it offers a novel perspective for point cloud semantic segmentation.

Currently, several data-driven methods have achieved notable success in the semantic segmentation of power line scenes. PointNet [14] and PointNet++ [15] offer efficient and accurate solutions for processing point cloud data. Zhang et al. enhanced the PointNet model by integrating a geometric feature extraction (GFE) module and a neighborhood information aggregation (NIA) module, thereby improving the segmentation of railway power lines and towers from railway scene point clouds through the extraction of local geometric information and global contextual relationships [32]. Su employed the PointNet++ model for high-precision segmentation of power lines and towers [44]. Wang integrated a coordinate attention (CA) module with PointNet++ to develop an end-to-end CA-PointNet++ architecture, effectively capturing long-range spatial contextual features for enhanced segmentation precision [45]. Zhao utilized the PointCNN model to segment power lines and towers from drone-acquired point cloud data, demonstrating another effective approach for these tasks [46]. Yu combined a two-step sampling strategy with a local feature aggregation module based on the RandLA-Net model, resulting in improved extraction outcomes for power lines and towers [47].

It is evident that deep learning methods are highly effective for understanding point cloud scenes in large-scale environments. However, existing methods for transmission corridors [44–46] continue to rely on convolutional techniques exemplified by dot nets, which are inefficient for processing large-scale point cloud data. Furthermore, most of these methods do not account for global contextual information within transmission corridors, which is essential for effectively extracting small targets such as pylons.

## 3. Methodology

This study presents a point cloud semantic segmentation network (SA-KPConv) for transmission scene perception and applications, integrating kernel point convolution with spatial attention mechanisms. The primary workflow, as illustrated in Figure 1, comprises two key phases: (1) data preprocessing and (2) semantic segmentation.

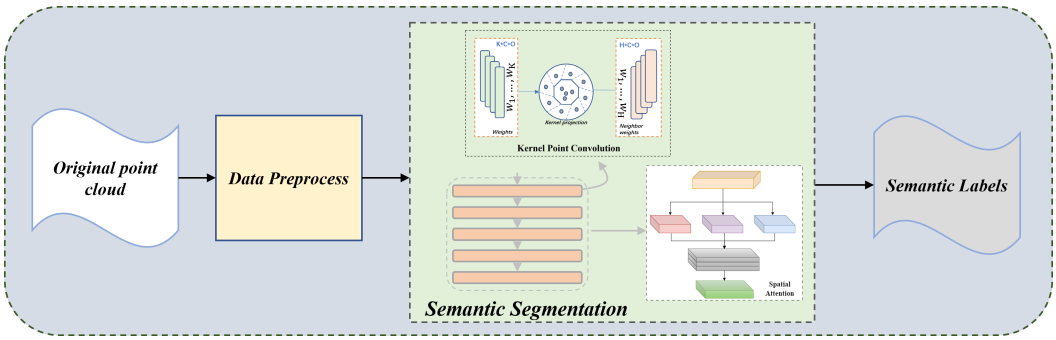



**Figure 1.** Overall processing workflow of transmission corridor scene semantic segmentation, including two stages: data preprocessing and semantic segmentation.

### *3.1. Preprocess*

A primary objective of semantic perception in power transmission scenes is to segment critical infrastructure, such as power lines and towers, from noisy raw point clouds. However, due to the scanning characteristics of airborne LiDAR, point clouds of power towers and lines typically represent a small sample size within the overall 3D point cloud. This leads to a significant class imbalance issue during the semantic segmentation of transmission corridor scenes. To enhance the network's segmentation performance for these critical facilities, we incorporated grid sampling and data augmentation strategies during the data preprocessing phase.

#### 3.1.1. Grid Sampling

Compared to indoor and urban scenes, airborne LiDAR point clouds in transmission corridor scenarios face challenges such as narrow corridors, large data volumes, and class imbalance. Additionally, due to the strip-scanning nature of airborne LiDAR, the resulting

3D point clouds often exhibit overlapping regions with redundant data, further increasing the data volume. To address this issue, we preprocess the raw point clouds by employing a grid sampling method for downsampling. This voxel-based downsampling involves converting the 3D point cloud into individual voxels and using the centroid of the point set within each voxel as the new point. As shown in Figure 2, this is a schematic diagram of grid downsampling for a high-voltage tower. The left side represents the original point cloud, while the right side displays the results after downsampling. This approach not only reduces the data volume, but also produces more evenly spaced points, facilitating subsequent operations such as convolutional learning. Furthermore, due to scanning angles and occlusions, the vertical resolution of airborne LiDAR is generally lower than the horizontal resolution. To better preserve the vertical structural features of the point cloud, we set the voxel length in the vertical direction to be smaller than that in the horizontal direction, thereby replacing traditional cubic voxels with rectangular prisms.

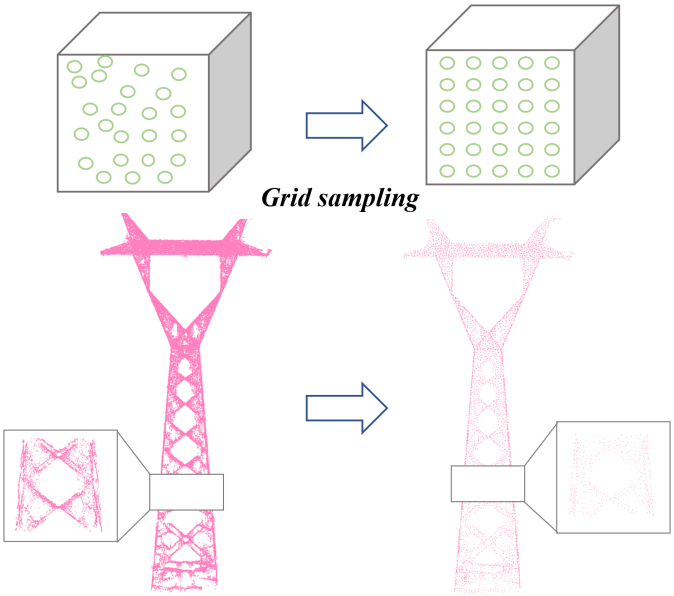



**Figure 2.** Schematic diagram of grid sampling (taking towers as an example). The left side of the diagram represents the original point cloud, while the right side illustrates the grid sampling method and its results.

#### 3.1.2. Data Augmentation Strategy

As previously mentioned, the class imbalance in transmission corridor scenarios renders the effective segmentation of critical infrastructure, such as power lines and towers, particularly challenging. To address this issue, we employed a straightforward yet effective data augmentation strategy from a sampling perspective. Initially, we extracted power lines and towers from the accurately labeled samples and applied random scaling, rotation, and translation before reintegrating them into the input scene. Notably, to preserve the vertical orientation characteristic of power towers and the near-parallel alignment of power lines with the ground, we restricted rotations to the horizontal plane. Additionally, when inputting complete scene data, we applied random rotations, translations, and scaling to each input sample to enhance the model's adaptability to point clouds of varying scales. The scaling factor was randomly selected within predefined minimum and maximum ranges, simulating variations in point cloud sizes and thereby enhancing the model's generalization capability.

### *3.2. SA-KPConv*

#### 3.2.1. Overall Network Architecture

We propose a Spatial Attention Mechanism-based Kernel Point Convolution Network (SA-KPConv) for the semantic segmentation of airborne LiDAR point cloud data in transmission corridor scenes. Building on previous works, we utilize a 3D-Unet framework as the overall architecture. As illustrated in Figure 3, the model comprises various components: light green rectangles represent the kernel point convolution blocks, dark green rectangles denote the geometric features of neighborhood points at different downsampling levels, purple arrows indicate the downsampling process, blue rectangles represent unary blocks, light blue rectangles signify connection operations, orange arrows depict upsampling operations, and orange rectangles indicate spatial awareness blocks. The numbers below each structural layer indicate the dimensionality information of the data at that layer.

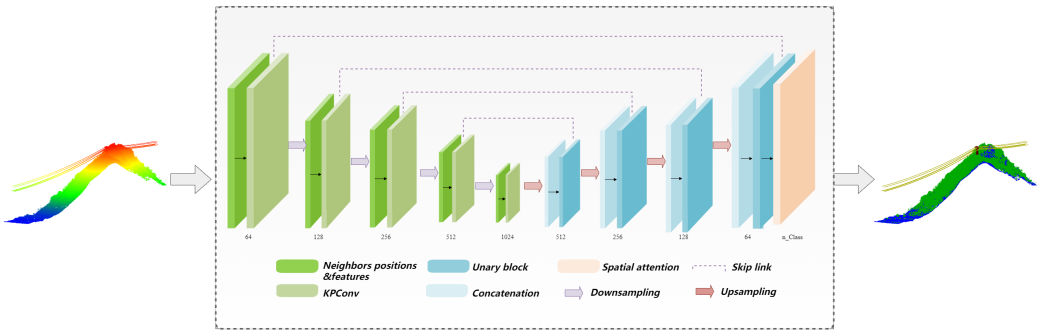



**Figure 3.** Proposed SA-KPConv network architecture. Light green rectangles represent kernel point convolution blocks, dark green rectangles indicate geometric neighborhood features, dark blue rectangles are unary blocks, and light blue rectangles show merging operations. Orange rectangles denote the spatial attention module, while purple and orange arrows indicate upsampling and downsampling processes, respectively.

The network adopts an encoder–decoder architecture to process the input point clouds. The encoder consists of five convolutional layers, each utilizing kernel point convolution to extract local geometric spatial features through Kd-tree structures. To capture multi-scale local geometric features, we implement downsampling techniques that progressively enlarge the receptive field of the convolutions. In the decoder, we employ nearest neighbor upsampling to derive the final point features. Four skip connections are established between the encoder and decoder to facilitate the transfer of intermediate features. These features are merged with the upsampled outputs and subsequently passed through a $1 \times 1$ convolutional unary block. At the end of the network, a spatial attention module is added to incorporate global context, thereby enhancing the final point-wise semantic predictions.

To further enhance the model's learning capacity in the presence of class imbalance, we employ Focal Loss [48] as the loss function during training. Its calculation is defined as follows in Equation (1):

$$L_f l = -\alpha(1 - p_t)^{\gamma} \log(p_t) \tag{1}$$

In this context, $(p_t)'$ reflects the proximity between the predicted value and the ground truth, with a larger value indicating more accurate segmentation. $\gamma > 0$ is a modulation factor. Compared to Cross Entropy Loss, Focal Loss introduces a modulating factor $(1 - (p_t)')^{\gamma}$, which effectively increases the weight of hard-to-segment samples in the loss function. This modification allows the loss function to focus more on difficult samples, thereby enhancing the accuracy of these challenging cases. $\alpha$ is used to adjust the ratio between the losses of positive and negative samples.

#### 3.2.2. Kernel Point Convolution Layer

To enhance the accuracy of point cloud semantic segmentation, we utilize kernel point convolution layers to extract local features from the point cloud. Kernel Point Convolution

(KPConv) [17] is a specialized convolution operation designed specifically for point clouds, employing kernel points to define the convolution kernel. In KPConv, each convolution kernel consists of a set of kernel points with fixed coordinates in three-dimensional space. Each point in the input point cloud is assigned a weight based on its relative position to the kernel points, resulting in the convolution output. The specific operation of kernel point convolution is illustrated in Figure 4.

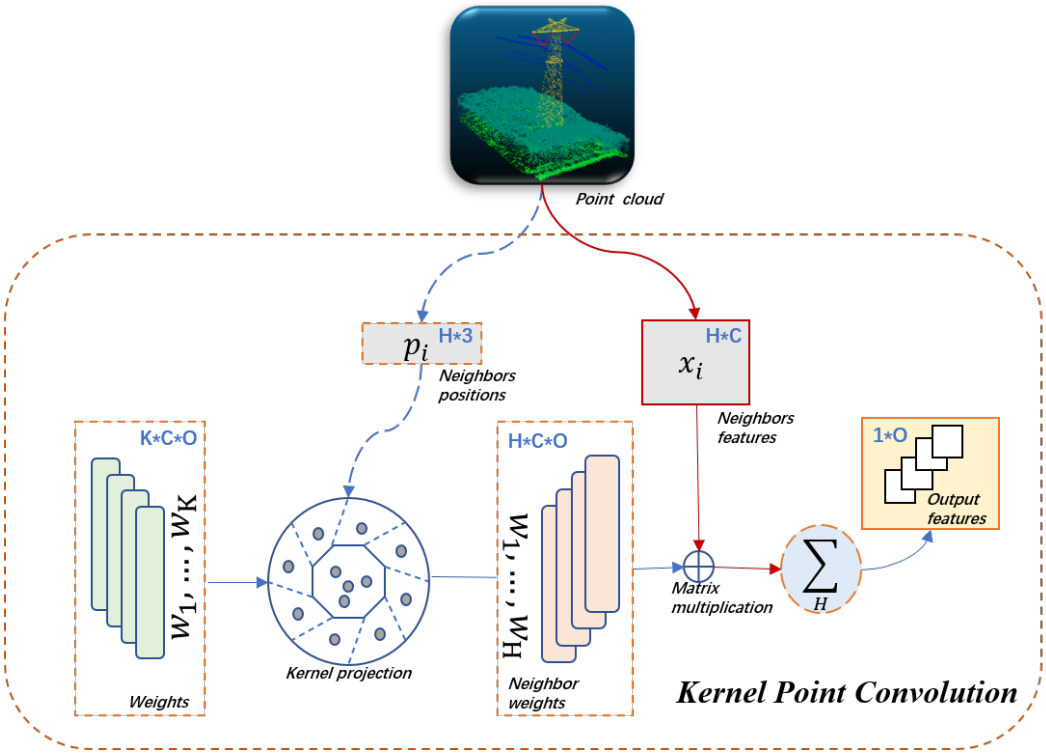



**Figure 4.** Structure of the kernel point convolution layer. This layer updates the weight of each point based on the kernel point function, facilitating the extraction of local geometric features.

The network input comprises point position information and point feature information. Point position information is conventionally represented by $x, y, z$ coordinates, which describe the specific location of each point in three-dimensional space. These coordinates are used to ascertain the spatial relationships and structural information among the points in the point cloud. Point feature information may include attributes such as color, class labels, and reflectance intensity. This feature information conveys the properties of each point, aiding the model in understanding the specific characteristics of objects based on their spatial structure.

Initially, the KPConv network employs a set of kernel points $\{p_i\}_{i=1}^{I}$ to define the convolution kernel. Each kernel point $p_i$ possesses fixed coordinates in three-dimensional space. These kernel points are analogous to filter weights in traditional convolution; however, they are distributed irregularly in three-dimensional space. The positions of these kernel points are adjusted throughout the training has to optimize feature extraction. For each input point $x_i$, the network searches for a set of neighboring points, denoted as $N(x_i)$. This is typically conducted using a nearest neighbors search, where neighbors are identified by calculating the Euclidean distance $\|x_j - x_i\|$ between points and selecting the nearest $n$ points. KPConv performs a distance-weighted convolution operation. The output feature $f_i'$ corresponding to each input point $x_i$ is computed as follows:

$$f_i' = \sum_{j \in N(x_i)} \sum_{k=1}^{K} h(\|x_j - x_i - p_k\|) \cdot f_j \tag{2}$$

where $f_j$ represents the features of the neighboring point $x_j$. The function $h(\cdot)$ serves as a weighting function, such as a Gaussian kernel, which computes weights based on the distances between neighboring points $x_j$ and kernel points $p_k$. The convolution kernel defined by the kernel points aggregates features from neighboring points through weighted summation to produce a new feature vector. This new feature captures local geometric and attribute information, effectively representing the characteristics of the point cloud. Notably, both rigid and deformable versions of kernel point convolution have been defined in the literature, with the deformable version adapting to local geometric shapes and enhancing the detailed representation of features. This adaptability enables kernel point convolution to effectively capture significant geometric variations in structures, such as power towers and buildings. Following the convolution operation, nonlinear activation functions (ReLU) and batch normalization are applied to enhance the model's nonlinearity and stability. If $y_i$ denotes the output of the convolution, the activated and normalized output is given by Equation (3):

$$y_i' = \text{ReLU}(\text{BatchNorm}(y_i)) \tag{3}$$

#### 3.2.3. Spatial Attention Module

Contextual information is crucial for conveying global scene understanding. In point cloud semantic segmentation, while local features can represent the geometric attributes of the point cloud within local neighborhoods, similar feature representations can occur among points of the same category, even when they are spatially distant. Considering the correlations among points in the feature space enhances the model's prediction accuracy. To this end, inspired by previous research, we introduce a Spatial Attention Module akin to Self-Attention. This module is designed to capture attention features that reflect the correlations between points in the scene. As illustrated in Figure 5, the outputs are computed as a weighted sum of the values, with the corresponding weights derived from pairwise functions between the queries and their respective keys, representing the query-key relationships.

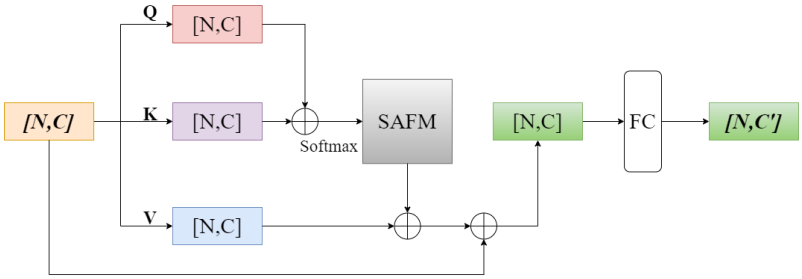



**Figure 5.** Spatial attention module. The features obtained after kernel point convolution undergo global attention updates through this module, enhancing the model's prediction accuracy.

Given a feature matrix $F \in \mathbb{R}^{(N,C)}$, three distinct sets of queries, keys, and values are generated: $Q, K, V \in \mathbb{R}^{(N,C)}$, computed as follows:

$$Q = \alpha_s(F), \quad K = \beta_s(F), \quad V = \gamma_s(F) \tag{4}$$

where $\alpha_s, \beta_s, \gamma_s$ are realized through different fully connected layers. Similarly, following the correlation calculation methods in previous studies, we use the dot product to compute point-wise correlations, leading to

$$F_{SA} = \alpha_{SA} \cdot \hat{A} \cdot V + F \tag{5}$$

It is noteworthy that $SA$ represents the spatial attention matrix, which is derived from the dot product between $K$ and the transpose of $Q$, as shown in Equation (6),

$$SA = \text{softmax}(K \cdot \hat{A} \cdot Q^T) \tag{6}$$

To fully leverage global context, we pass the updated $F_{SA}$ to a fully connected layer to obtain point-wise semantic labels. As depicted in Figure 5, $C'$ represents the final number of predicted categories. This spatial channel attention module enables the updating of point features from a global perspective, allowing for a comprehensive learning of complex interactions among points and facilitating more accurate predictions.

## 4. Experiments and Analysis

### *4.1. Datasets and Metrics*

#### 4.1.1. Datasets

We validated our method using a dataset collected from transmission corridors in 2022 in Henan, China, utilizing a Riegl VUX-LR airborne laser scanner. The collected point cloud data were manually cropped to a width of 100 m along the transmission direction, with a point density of approximately 40 points per square meter, encompassing common terrain types such as mountainous, hilly, and flat areas. We manually annotated approximately 15 km of data, comprising around 12 million points, using CloudCompare 2.13.2 Kharkiv version software. The point cloud data for the transmission corridor were categorized into six classes: ground, buildings, low vegetation, high vegetation, conductors, and tower structures. As shown in Figure 6, the distribution of samples among different categories reveals that important electrical infrastructure (wires and towers) constitutes only about 3% of the total samples, while categories such as ground and vegetation account for over 85%.

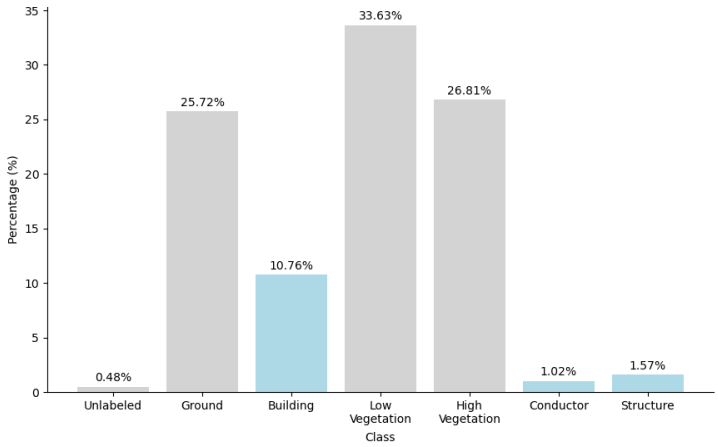



**Figure 6.** Sample proportions of each category in the datasets. Blue represents important facilities within the transmission corridor, while gray indicates other categories.

As shown in Figure 7, we present a subset of the data along with the corresponding annotation results from our experiments. The dataset was divided into training, validation, and testing sets in an 8:1:1 ratio for network training and evaluation. To validate the effectiveness of our method, we further categorized the annotated data into three terrain types—flatland, buildings, and mountainous areas—resulting in three sub-datasets. These terrain types are commonly encountered in transmission corridor scenarios.

#### 4.1.2. Metrics

To quantitatively evaluate our method, we employ Overall Accuracy (OA) and mean Intersection over Union (mIoU). OA measures the percentage of correctly predicted points relative to the total number of test points, while mIoU assesses the performance of semantic segmentation across different categories. The calculations for these metrics are detailed as follows:

$$\text{precision} = \frac{TP}{TP + FP} \tag{7}$$

$$mIoU = \frac{1}{C}\sum_{c=1}^{C}\frac{TP}{TP+FP+FN} \tag{8}$$

where $TP$, $FN$, and $FP$ represent true positives, false negatives, and false positives, respectively, in a confusion matrix.

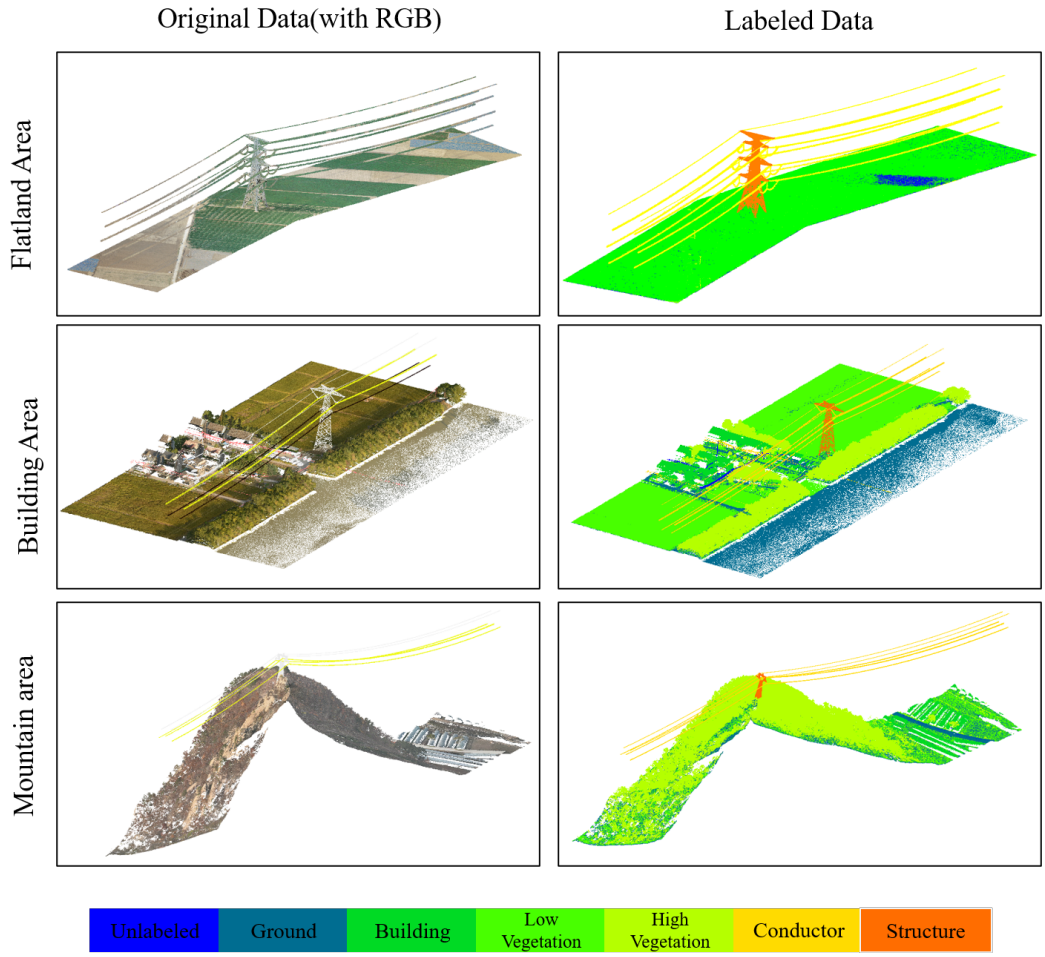



**Figure 7.** Data and annotations used in experiments, including flatland, buildings, and mountainous areas.

### *4.2. Implementation Details*

In the implementation phase, we employ the Adam optimizer with default parameters using the PyTorch framework. The initial learning rate is set at 0.01, and decreases by 5% after each epoch. For the kernel point convolution radius, based on empirical evidence from [17], we define it as 2.5 times the grid radius. The number of kernel points in the convolution is set to 15, ensuring even distribution within a specified sphere. All experiments were conducted on an NVIDIA RTX 3090 GPU. Additionally, during training, unlabeled categories are ignored, and their loss is not computed.

For testing, we randomly select spheres within the test area, ensuring that each point is input into the network a minimum of 20 times to obtain average predictive probabilities. This repetition is essential for mitigating misclassification of points near the sphere boundary, where geometry may be incomplete.

### *4.3. Semantic Segmentation Results and Analysis*

To validate the effectiveness and performance of our proposed SA-KPConv model for semantic segmentation in transmission corridor scenarios, we conduct experiments

across three distinct terrain sub-datasets and analyze the results from both quantitative and qualitative perspectives.

Figures 8 and 9 illustrate the segmentation visualization results of our method across building, mountainous, and flat terrain areas. Our model demonstrates strong performance in segmenting towers and conductors across various terrain types, which is highly valuable for transmission corridor applications. Furthermore, the SA-KPConv model yields satisfactory segmentation results for buildings, vegetation, and the ground, underscoring its effectiveness for semantic segmentation in transmission corridor scenarios. However, certain shortcomings remain, particularly at the connections between conductors and towers, as well as at the boundary contours of buildings, as indicated by the red circles in Figure 8. This may be due to our method not thoroughly considering the edge detail information of the targets, leading to semantic prediction ambiguity in the segmentation results. The combination of kernel point convolution and spatial attention enables our model to effectively model and learn local and global information. However, it lacks the capability to learn edge information between categories. While this ambiguity does not significantly impact the overall accuracy of semantic segmentation, there remains room for improvement in the completeness of segmentation for certain components. Future work could focus on incorporating edge information to address this issue.

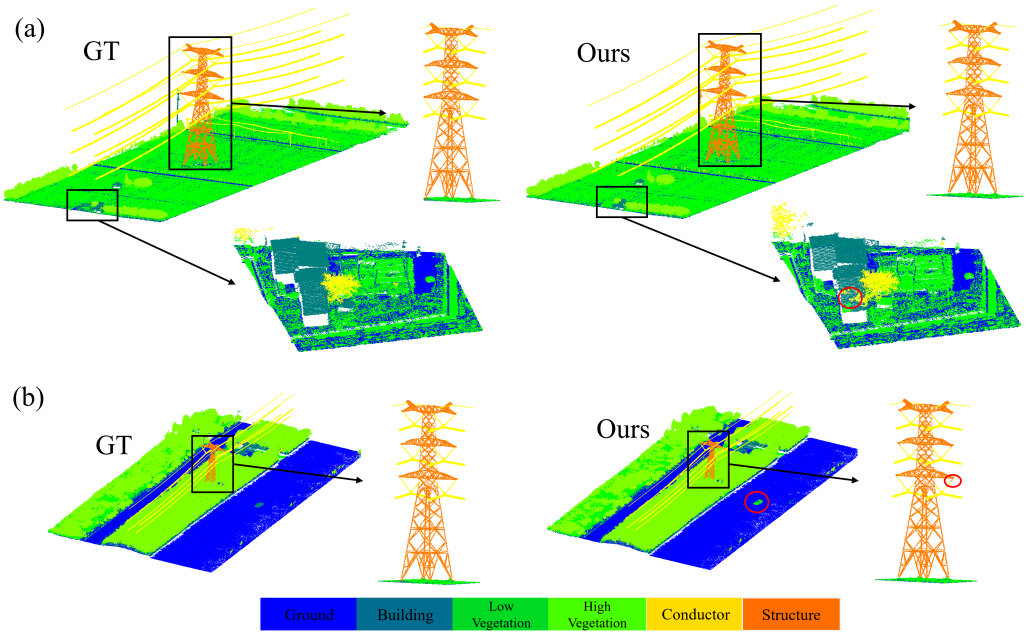



**Figure 8.** Semantic segmentation results of our methods in flat and built-up areas, where (**a**) represents built-up areas and (**b**) represents flat areas. The red circles mark the parts that were incorrectly predicted.

We conducted a quantitative evaluation of the proposed method's performance. Table 1 displays the results for various terrain sub-datasets. The findings indicate that our model performs effectively across different terrains, achieving a mean Intersection over Union (mIoU) exceeding 87% in all instances. Notably, for key components such as towers and conductors, the IoU scores surpass 84%, which is critical for downstream tasks in transmission corridor perception. These quantitative and qualitative results underscore the robustness and generalizability of our approach.

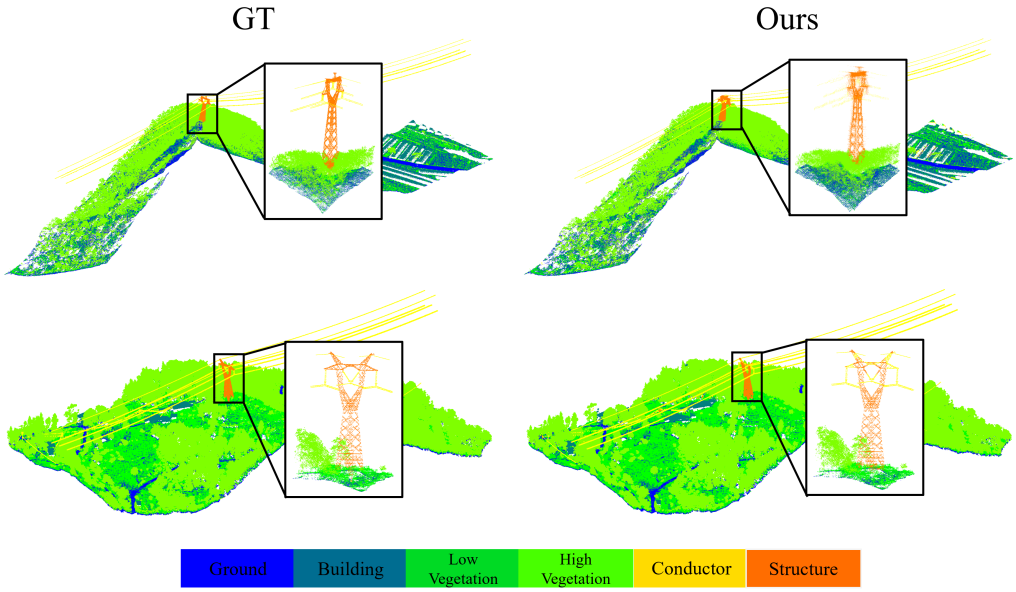



**Figure 9.** Semantic segmentation results in mountainous areas. The black box highlights the details of the power tower section.

**Table 1.** Quantitative description of the proposed SA-KPConv in different terrains.

| Sub-Datasets | IoU | | | | | | mIoU | OA |
|---|---|---|---|---|---|---|---|---|
| | Ground | Building | Low Vegetation | High Vegetation | Conductor | Structure | | |
| Building Area | 90.17% | 80.22% | 88.24% | 93.66% | 87.21% | 84.48% | 87.33% | 97.34% |
| Mountain Area | 91.40% | 85.16% | 85.65% | 95.03% | 90.95% | 84.14% | 88.72% | 95.74% |
| Flatland Area | 92.43% | 72.97% | 98.59% | 95.59% | 98.70% | 96.48% | 92.46% | 99.06% |

Further analysis reveals that the model attains the highest performance in flatland areas, achieving a mean Intersection over Union (mIoU) of 92.46% and IoU scores for both conductors and towers exceeding 96%. This superior performance can be attributed to the relatively complete and unobstructed nature of power infrastructure in flat terrain. In more complex terrains, the presence of human-made structures, particularly linear objects, adversely affects the model's accuracy for conductors and towers, resulting in slightly lower performance compared to flatland areas.

### *4.4. Comparison to State-of-the-Art Methods*

To fairly validate the effectiveness of our method, we conducted comparisons and analyses with other approaches. We selected Random Forest (RF) [4] as a representative machine learning method, PointNet++ [15] and RandLA-Net [16] as typical point-based deep learning methods, and KPFCNN [49] from the semantic segmentation frameworks referenced in our method based on KPConv. Additionally, we included SPT [41,42] as representatives of graph-based and Transformer-based methods. For RF, we utilized the feature set construction method proposed in [49] for training and testing. For the other methods, we adhered to the data partitioning strategy described in Section 4.1.1 for training and testing.

Figure 10 illustrates the qualitative results of various methods applied to different terrain regions within the transmission corridor scene point cloud data, with the top row depicting the ground truth and the subsequent rows presenting the segmentation results for each method across diverse terrain areas. Correspondingly, Table 2 presents the quantitative comparison results for each method. Both qualitative and quantitative analyses demonstrate that our method achieves superior performance in semantic segmentation of

power transmission corridors, attaining a mean Intersection over Union (mIoU) of 89.62% and an overall accuracy (OA) of 95.74%. This marks an improvement of approximately 8% in mIoU and 3% in OA compared to KPFCNN and SPT, respectively. Our method performs well across all categories, particularly excelling in the classification of ground, low vegetation, and power towers. While SPT yields the best segmentation results for buildings and wires, our method closely approaches this performance. Additionally, KPFCNN excels in high vegetation segmentation, with our method maintaining a difference of less than 1%, which is within the acceptable range of systemic error.

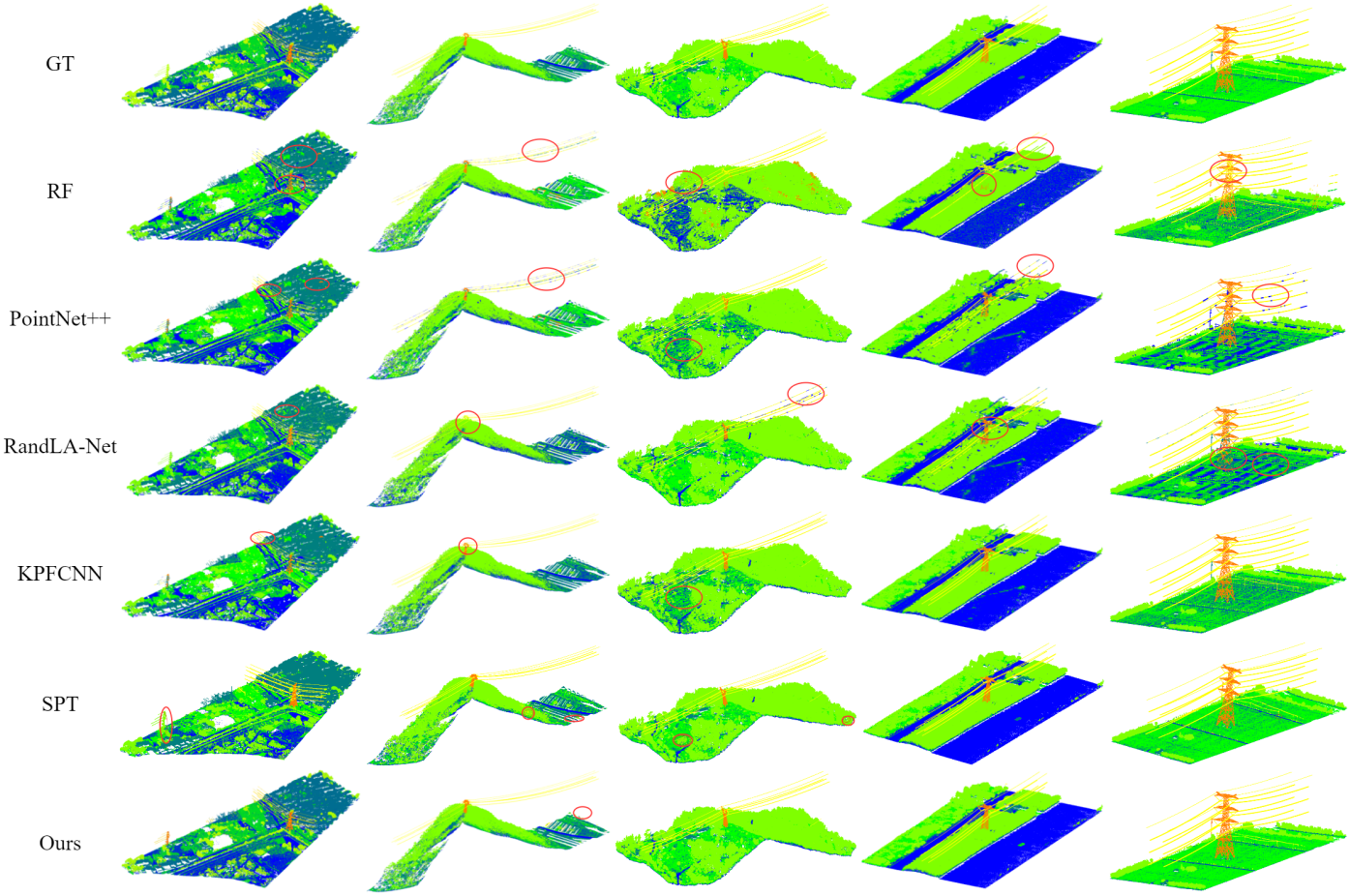



**Figure 10.** Qualitative results of semantic segmentation compared to different methods. The red circles mark the parts that were incorrectly predicted.

**Table 2.** Quantitative comparison of semantic segmentation results of different methods. The first six columns show the IoU values of the different methods in the dataset for each category, and the last two columns show the mIoU and OA values. Bold text indicates the highest value in the column.

| Methods | Ground | Building | Low Vegetation | High Vegetation | Conductor | Structure | mIoU | OA |
|---|---|---|---|---|---|---|---|---|
| RF | 43.94% | 1.67% | 35.00% | 84.11% | 68.92% | 67.18% | 50.14% | 72.75% |
| PointNet++ | 65.32% | 31.63% | 55.76% | 88.10% | 77.63% | 56.74% | 62.53% | 78.12% |
| RandLA-Net | 84.87% | 62.14% | 78.89% | 92.89% | 76.32% | 62.88% | 76.33% | 87.62% |
| KPFCNN | 80.14% | 75.71% | 59.12% | **94.15%** | 93.99% | 85.27% | 81.40% | 90.78% |
| SPT | 85.56% | **84.05%** | 79.32% | 86.50% | **95.27%** | 88.56% | 86.54% | 93.83% |
| **Ours** | **86.70%** | 83.01% | **89.74%** | 93.83% | 94.75% | **89.67%** | **89.62%** | **95.74%** |

The results clearly demonstrate that the integration of kernel point convolution and spatial attention within the proposed SA-KPConv model is more effective in addressing

the challenges posed by large-scale and imbalanced point cloud data compared to current state-of-the-art techniques.

Furthermore, we conduct a comparative analysis of various methods, specifically focusing on the critical conductor and tower infrastructure within the transmission corridor environment. Figure 11 illustrates the segmentation results of the objects surrounding towers in both mountainous and flat regions. Our method demonstrates superior segmentation results for both conductors and towers compared to other approaches.

Based on the comparative results above, we can analyze that while the Random Forest (RF) method benefits from incorporating linear features into its feature set, leading to satisfactory segmentation outcomes for conductors in specific datasets, its performance in tower segmentation remains relatively poor. This limitation stems from its reliance on manually designed features, which restricts its generalization capabilities compared to learning-based methods.

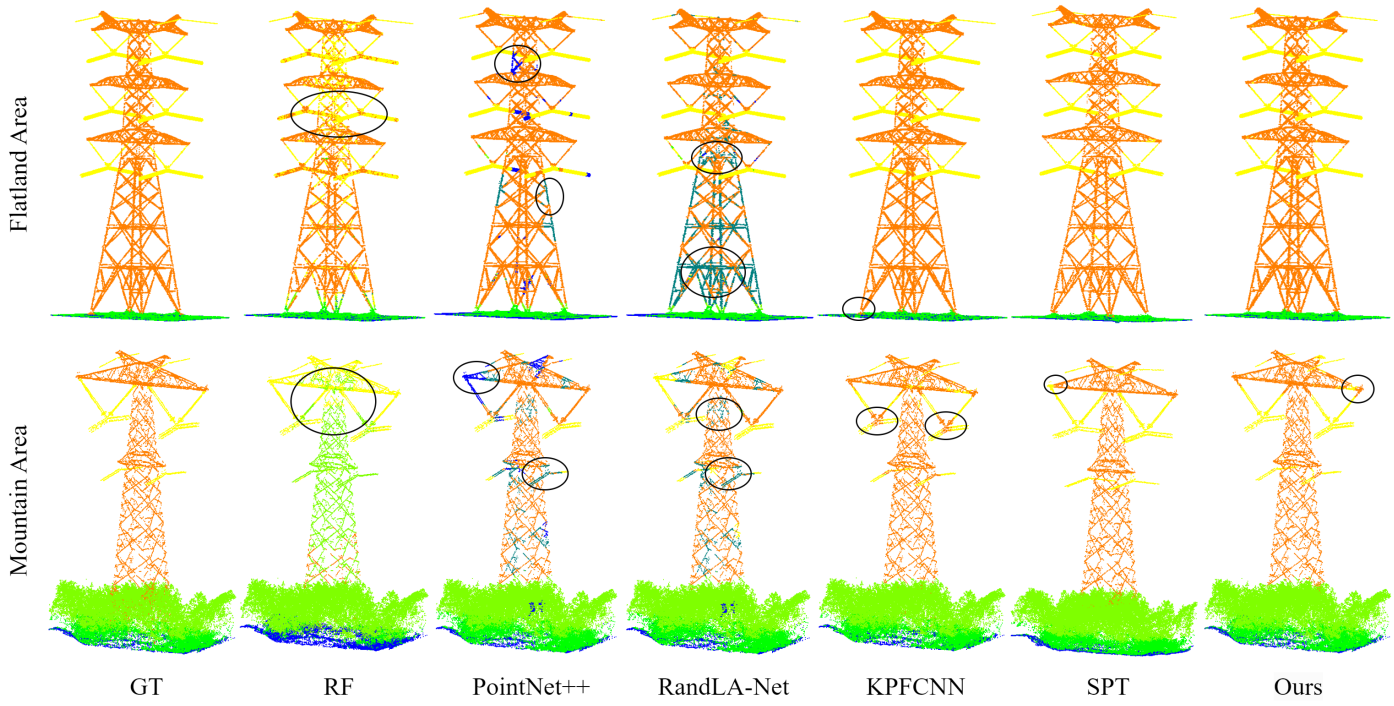



**Figure 11.** Qualitative results of semantic segmentation compared to different methods. The black circles mark the parts that were incorrectly predicted.

Meanwhile, PointNet++ and RandLA-Net exhibit subpar performance in segmenting conductors and towers, primarily due to their reliance on sampling-based approaches for local feature learning and aggregation. As described in Section 4.1.1, these classes of samples are underrepresented, resulting in weaker feature learning for conductors and towers compared to other categories.

For KPFCNN, both our method and theirs utilize kernel point convolution layers to capture local features. Although KPFCNN can identify small targets such as wires, our method demonstrates superior completeness and accuracy in recognition. This improvement is largely attributed to our structural optimization and the incorporation of a spatial attention mechanism that emphasizes global information modeling.

Another widely used state-of-the-art method, RandLA-Net, also employs an attention mechanism; however, their attention pooling differs from ours. RandLA-Net computes attention based on the geometrical aggregation features of points within the local KNN neighborhood, feeding this information to the next network layer. In contrast, our attention mechanism calculates scores for features obtained after applying kernel point convolution to the entire input point cloud. Consequently, their acquisition of global information

largely relies on gradually expanding the receptive field through downsampling, which differs significantly from our approach. Our spatial attention mechanism shares similarities with self-attention mechanisms. Experimental results further validate our findings, indicating that our method improves mIoU by 13% compared to RandLA-Net, particularly demonstrating better performance in small target categories such as wires and towers.

For SPT, both quantitative and qualitative results are strong, especially in the segmentation of artificial structures such as buildings and wires. This success can be largely attributed to SPT being a SuperPoint-based semantic segmentation model, which calculates geometric partitions based on point neighborhoods and adapts to capture local geometric attributes. These features are beneficial for recognizing artificial targets, particularly those with linear and planar characteristics. However, in the context of transmission corridor scenes, its accuracy in vegetation segmentation is somewhat lacking.

Overall, the comparative analysis with various state-of-the-art methods clearly shows that our approach achieves superior segmentation performance, thereby establishing a novel theoretical framework for the semantic understanding of transmission corridor scenes in real-world applications.

### *4.5. Effectiveness of Each Proposed Module*

To evaluate the contributions of various components in our proposed SA-KPConv model, we conducted a series of ablation experiments. These experiments aimed to isolate the effects of individual components, including the Spatial Attention module (SA), Kernel Point Convolution (KP), and Focal Loss (FL) function, on the overall performance of the semantic segmentation task. The results presented in the first row indicate a significant decrease in mean Intersection over Union (mIoU) (approximately 5% reduction) upon removing the Spatial Attention module, underscoring the critical importance of capturing global contextual information for accurate segmentation of different objects in transmission corridor scenes. The second row of Table 3 shows the results of replacing the Focal Loss function with traditional Cross Entropy Loss. The findings demonstrate that Focal Loss improved the model's capacity to address class imbalance, thereby enhancing overall accuracy, with the mIoU increasing by 1.64% when using Focal Loss compared to Cross Entropy Loss. Additionally, we assessed the impact of the Kernel Point Convolution layer by substituting it with a point-based convolution method, specifically PointNet [14], as the backbone. Performance metrics reveal that kernel point convolution significantly outperforms standard convolution methods, achieving an mIoU improvement of 12% over the PointNet-based model. This highlights the effectiveness of our approach in capturing local geometric features and addressing the irregularities inherent in point cloud data.

**Table 3.** Effectiveness of each module in the proposed method. The first three rows represent the results of the model after removing specific components, and the last row shows the results of our full model.

| | **mIoU (%)** | **OA (%)** |
|---|---|---|
| WO SA | 83.78 | 93.78 |
| WO FL | 87.31 | 94.53 |
| WO KP | 76.05 | 82.72 |
| SA-KPConv (Ours) | 88.95 | 95.74 |

The results demonstrating the effectiveness of each module in the proposed method, confirming the significance of both the Spatial Attention module and Kernel Point Convolution in enhancing model performance. Moreover, the selection of the loss function is critical in addressing class imbalance, further validating the efficacy of our semantic segmentation approach for transmission corridor scenes.

## 5. Conclusions

In this paper, we propose the SA-KPConv model, specifically designed for the semantic segmentation of transmission corridor scenes. By integrating a Spatial Attention module and Kernel Point Convolution, the model effectively captures both local geometric features and global contextual information, thereby addressing the unique challenges posed by complex transmission corridor environments. Our experimental results, validated across various sub-datasets encompassing flatlands, mountainous areas, and urban regions, demonstrate that SA-KPConv achieves state-of-the-art performance, particularly in segmenting critical components such as power lines and towers. Furthermore, a series of ablation studies further confirm the contributions of the Spatial Attention module and KPConv to improving segmentation accuracy, with the use of focal loss demonstrating advantages in addressing class imbalance. The model's robustness and generalization capabilities across various terrains underscore its practical value for real-world transmission corridor monitoring applications.

In conclusion, our SA-KPConv model provides an effective and reliable solution for precise semantic segmentation in transmission corridor scenes, offering substantial potential for downstream tasks, including infrastructure inspection and maintenance. Future research could investigate more refined segmentation of additional categories within transmission corridor scenes, as well as the potential for achieving end-to-end 3D reconstruction of power infrastructure.

**Author Contributions:** Conceptualization, F.Z. and G.W. (Gang Wen); methodology, F.Z. and Y.W.; validation, F.Z. and G.W. (Gang Wen); formal analysis, H.P.; investigation, Y.M.; resources, G.W. (Gang Wen); data curation, G.W. (Guofang Wang); writing—original draft preparation, F.Z.; writing—review and editing, F.Z.; visualization, F.Z. and Y.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Major Science and Technology Special Project of Yunnan Province OF funder grant number 202202AD080010 and South Grid Key Science and Technology Program of funder grant number 056200KK52220011.

**Data Availability Statement:** The data used in this study are available upon request from the corresponding author. Due to privacy restrictions, these data are not publicly available.

**Acknowledgments:** We sincerely thank the anonymous reviewers for the critical comments and suggestions for improving the manuscript.

**Conflicts of Interest:** Authors F.Z., G.W., Y.M., H.P., G.W. and Y.W. were employed by the company Yunnan Power Grid Company Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflicts of interest. The authors declare that this study received funding from Major Science and Technology Special Project of Yunnan Province (202202AD080010) and Key Science and Technology Project of South China Network (056200KK52220011). The funder was not involved in the study design, collection, analysis interpretation of data, the writing of this article or the decision to submit it for publication.

## References

1. Nguyen, V.N.; Jenssen, R.; Roverso, D. Automatic Autonomous Vision-Based Power Line Inspection: A Review of Current Status and the Potential Role of Deep Learning. *Int. J. Electr. Power Energy Syst.* **2018,** *99*, 107–120. [CrossRef]
2. Yang, L.; Fan, J.; Liu, Y.; Li, E.; Peng, J.; Liang, Z. A Review on State-of-the-Art Power Line Inspection Techniques. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9350–9365. [CrossRef]
3. Vo, A.V.; Truong-Hong, L.; Laefer, D.F.; Bertolotto, M. Octree-Based Region Growing for Point Cloud Segmentation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 88–100. [CrossRef]
4. Lin, C.H.; Chen, J.Y.; Su, P.L.; Chen, C.H. Eigen-Feature Analysis of Weighted Covariance Matrices for LiDAR Point Cloud Classification. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 70–79. [CrossRef]
5. Huang, Y.; Du, Y.; Shi, W. Fast and Accurate Power Line Corridor Survey Using Spatial Line Clustering of Point Cloud. *Remote Sens.* **2021**, *13*, 1571. [CrossRef]
6. Liao, L.; Tang, S.; Liao, J.; Li, X.; Wang, W.; Li, Y.; Guo, R. A Supervoxel-Based Random Forest Method for Robust and Effective Airborne LiDAR Point Cloud Classification. *Remote Sens.* **2022**, *14*, 1516. [CrossRef]

7. Tang, Q.; Zhang, L.; Lan, G.; Shi, X.; Duanmu, X.; Chen, K. A Classification Method of Point Clouds of Transmission Line Corridor Based on Improved Random Forest and Multi-Scale Features. *Sensors* **2023**, *23*, 1320. [CrossRef]
8. Wang, Y.; Chen, Q.; Liu, L.; Li, X.; Sangaiah, A.K.; Li, K. Systematic Comparison of Power Line Classification Methods from ALS and MLS Point Cloud Data. *Remote Sens.* **2018**, *10*, 1222. [CrossRef]
9. Jiang, S.; Guo, W.; Fan, Y.; Fu, H. Fast Semantic Segmentation of 3D Lidar Point Cloud Based on Random Forest Method. In *Proceedings of the China Satellite Navigation Conference (CSNC 2022), Beijing, China, 25–27 May 2022*; Yang, C., Xie, J., Eds.; Springer Nature: Berlin/Heidelberg, Germany, 2022; pp. 415–424. [CrossRef]
10. Shokri, D.; Rastiveis, H.; Sheikholeslami, S.M.; Shahhoseini, R.; Li, J. Fast Extraction of Power Lines from Mobile LiDAR Point Clouds Based on SVM Classification in Non-Urban Area. *Earth Obs. Geomat. Eng.* **2021**, *5*, 63–73. [CrossRef]
11. Otter, D.W.; Medina, J.R.; Kalita, J.K. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 604–624. [CrossRef]
12. Xu, C.; Leng, B.; Chen, B.; Zhang, C.; Zhou, X. Learning Discriminative and Generative Shape Embeddings for Three-Dimensional Shape Retrieval. *IEEE Trans. Multimed.* **2019**, *22*, 2234–2245. [CrossRef]
13. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef] [PubMed]
14. Charles, R.Q.; Su, H.; Kaichun, M.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85. [CrossRef]
15. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Long Beach, CA, USA, 4–9 December 2017; pp. 5105–5114.
16. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11108–11117.
17. 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 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Seoul, Republic of Korea, 27–28 October 2019; pp. 6410–6419. [CrossRef]
18. Hui, C.; Tingting, W.; Zuoxiao, D.; Weibin, L.; Menhas, M.I. Power Equipment Segmentation of 3D Point Clouds Based on Geodesic Distance with K-means Clustering. In Proceedings of the 2021 6th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 17–20 September 2021; pp. 317–321. [CrossRef]
19. Ying, S.; Xu, G.; Li, C.; Mao, Z. Point Cluster Analysis Using a 3D Voronoi Diagram with Applications in Point Cloud Segmentation. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1480–1499. [CrossRef]
20. Cao, Y.; Wang, Y.; Xue, Y.; Zhang, H.; Lao, Y. FEC: Fast Euclidean Clustering for Point Cloud Segmentation. *Drones* **2022**, *6*, 325. [CrossRef]
21. Réjichi, S.; Chaabane, F. Feature Extraction Using PCA for VHR Satellite Image Time Series Spatio-Temporal Classification. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Athens, Greece, 7–12 July 2024; pp. 485–488. [CrossRef]
22. Nurunnabi, A.; Belton, D.; West, G. Robust Segmentation in Laser Scanning 3D Point Cloud Data. In Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), Fremantle, Australia, 3–5 December 2012; pp. 1–8. [CrossRef]
23. Duan, Y.; Yang, C.; Chen, H.; Yan, W.; Li, H. Low-Complexity Point Cloud Denoising for LiDAR by PCA-based Dimension Reduction. *Opt. Commun.* **2021**, *482*, 126567. [CrossRef]
24. Zafar, B.; Ashraf, R.; Ali, N.; Ahmed, M.; Jabbar, S.; Naseer, K.; Ahmad, A.; Jeon, G. Intelligent Image Classification-Based on Spatial Weighted Histograms of Concentric Circles. *Comput. Sci. Inf. Syst.* **2018**, *15*, 615–633. [CrossRef]
25. Lawin, F.J.; Danelljan, M.; Tosteberg, P.; Bhat, G.; Khan, F.S.; Felsberg, M. Deep Projective 3D Semantic Segmentation. In *Proceedings of the Computer Analysis of Images and Patterns, Ystad, Sweden, 22–24 August 2017*; Felsberg, M., Heyden, A., Krüger, N., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 95–107. [CrossRef]
26. Huang, J.; Yan, W.; Li, T.; Liu, S.; Li, G. Learning the Global Descriptor for 3-D Object Recognition Based on Multiple Views Decomposition. *IEEE Trans. Multimed.* **2020**, *24*, 188–201. [CrossRef]
27. Tchapmi, L.; Choy, C.; Armeni, I.; Gwak, J.; Savarese, S. SEGCloud: Semantic Segmentation of 3D Point Clouds. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 537–547. [CrossRef]
28. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D ShapeNets: A Deep Representation for Volumetric Shapes. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, 7–12 June 2015; pp. 1912–1920. [CrossRef]
29. Meng, H.Y.; Gao, L.; Lai, Y.K.; Manocha, D. VV-net: Voxel VAE Net with Group Convolutions for Point Cloud Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8499–8507. [CrossRef]
30. Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; Chen, B. PointCNN: Convolution on X-transformed Points. In Proceedings of the Advances in Neural Information Processing Systems, Curran Associates, Inc., Montréal, QC, Canada, 3–8 December 2018; Volume 31.

31. Ye, X.; Li, J.; Huang, H.; Du, L.; Zhang, X. 3D Recurrent Neural Networks with Context Fusion for Point Cloud Semantic Segmentation. In *Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; Volume 11211, pp. 415–430. [CrossRef]
32. Zhang, L.; Wang, J.; Shen, Y.; Liang, J.; Chen, Y.; Chen, L.; Zhou, M. A Deep Learning Based Method for Railway Overhead Wire Reconstruction from Airborne LiDAR Data. *Remote Sens.* **2022**, *14*, 5272. [CrossRef]
33. Ni, P.; Li, X.; Xu, W.; Zhou, X.; Jiang, T.; Hu, W. Robust 3D Semantic Segmentation Method Based on Multi-Modal Collaborative Learning. *Remote Sens.* **2024**, *16*, 453. [CrossRef]
34. Sun, T.; Zhang, Z.; Tan, X.; Qu, Y.; Xie, Y. Image Understands Point Cloud: Weakly Supervised 3D Semantic Segmentation via Association Learning. *IEEE Trans. Image Process.* **2024**, *33*, 1838–1852. [CrossRef]
35. Kang, X.; Chu, L.; Li, J.; Chen, X.; Lu, Y. Hierarchical Intra-modal Correlation Learning for Label-free 3D Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 28244–28253.
36. Xu, J.; Yang, W.; Kong, L.; Liu, Y.; Zhang, R.; Zhou, Q.; Fei, B. Visual Foundation Models Boost Cross-Modal Unsupervised Domain Adaptation for 3D Semantic Segmentation. *arXiv* **2024**, arXiv:2403.10001.
37. Xu, R.; Wang, C.; Zhang, D.; Zhang, M.; Xu, S.; Meng, W.; Zhang, X. DefFusion: Deformable Multimodal Representation Fusion for 3D Semantic Segmentation. In Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024; pp. 7732–7739. [CrossRef]
38. Park, J.; Lee, S.; Kim, S.; Xiong, Y.; Kim, H.J. Self-Positioning Point-Based Transformer for Point Cloud Understanding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 21814–21823.
39. Wu, X.; Jiang, L.; Wang, P.S.; Liu, Z.; Liu, X.; Qiao, Y.; Ouyang, W.; He, T.; Zhao, H. Point Transformer V3: Simpler Faster Stronger. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 4840–4851.
40. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.H.; Koltun, V. Point Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 16259–16268.
41. Robert, D.; Raguet, H.; Landrieu, L. Scalable 3D Panoptic Segmentation As Superpoint Graph Clustering. In Proceedings of the 2024 International Conference on 3D Vision (3DV), Davos, Switzerland, 18–21 May 2024; pp. 179–189. [CrossRef]
42. Robert, D.; Raguet, H.; Landrieu, L. Efficient 3D Semantic Segmentation with Superpoint Transformer. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–6 October 2023; pp. 17149–17158. [CrossRef]
43. Landrieu, L.; Simonovsky, M. Large-Scale Point Cloud Semantic Segmentation With Superpoint Graphs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
44. Su, C.; Wu, X.; Guo, Y.; Lai, C.S.; Xu, L.; Zhao, X. Automatic Multi-Source Data Fusion Technique of Powerline Corridor Using UAV Lidar. In Proceedings of the 2022 IEEE International Smart Cities Conference (ISC2), Paphos, Cyprus, 26–29 September 2022; pp. 1–5. [CrossRef]
45. Wang, G.; Wang, L.; Wu, S.; Zu, S.; Song, B. Semantic Segmentation of Transmission Corridor 3D Point Clouds Based on CA-PointNet++. *Electronics* **2023**, *12*, 2829. [CrossRef]
46. Zhao, W.; Dong, Q.; Zuo, Z. A Point Cloud Segmentation Method for Power Lines and Towers Based on a Combination of Multiscale Density Features and Point-Based Deep Learning. *Int. J. Digit. Earth* **2023**, *16*, 620–644. [CrossRef]
47. Yu, H.; Wang, Z.; Zhou, Q.; Ma, Y.; Wang, Z.; Liu, H.; Ran, C.; Wang, S.; Zhou, X.; Zhang, X. Deep-Learning-Based Semantic Segmentation Approach for Point Clouds of Extra-High-Voltage Transmission Lines. *Remote Sens.* **2023**, *15*, 2371. [CrossRef]
48. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *arXiv* **2017**, arXiv:1708.02002.
49. Ni, H.; Lin, X.; Zhang, J. Classification of ALS Point Cloud with Improved Point Cloud Segmentation and Random Forests. *Remote Sens.* **2017**, *9*, 288. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.