



OPEN ACCESS

EDITED BY

Zhiheng Liu,
Xidian University, China

REVIEWED BY

Mahmoud Emam,
Menoufia University, Egypt
Tingting Wu,
Anhui University, China

*CORRESPONDENCE

Su Zhang,
✉ zhangsu615@163.com

RECEIVED 18 June 2025

ACCEPTED 31 July 2025

PUBLISHED 21 August 2025

CITATION

Zhang S, Liu H, Rong J and Zhang Y (2025)
GKCAE: A graph-attention-based encoder for
fine-grained semantic segmentation of
high-voltage transmission corridors scenario
LiDAR data.
Front. Earth Sci. 13:1649203.
doi: 10.3389/feart.2025.1649203

COPYRIGHT

© 2025 Zhang, Liu, Rong and Zhang. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

GKCAE: A graph-attention-based encoder for fine-grained semantic segmentation of high-voltage transmission corridors scenario LiDAR data

Su Zhang*, Haibo Liu, Jingguo Rong and Yaping Zhang

State Grid Economic and Technological Research Institute Co., Ltd., Beijing, China

Accurate semantic segmentation of airborne LiDAR point clouds is essential for the intelligent inspection and maintenance of high-voltage transmission infrastructure. While existing methods predominantly focus on major structural components such as towers and conductors, they often fail to address the fine-grained segmentation of smaller yet critical elements, including ground wires, crossing lines, and insulators. To tackle this limitation, we propose a novel network architecture—Graph-Kernel Convolution Attention Encoder (GKCAE)—designed for multi-class, fine-grained semantic segmentation of transmission corridor point clouds. GKCAE first captures local geometric features using Kernel Point Convolution, and then models inter-class spatial relationships through Graph Edge-Conditioned Convolution to incorporate global contextual information. Additionally, a Channel-Spatial Attention Module is introduced to enhance point-level feature representations, particularly for small or geometrically similar classes. Experiments conducted on three realworld transmission corridor datasets demonstrate that our method achieves a mean Intersection over Union (mIoU) of 81.93% and an Overall Accuracy (OA) of 94.1%, outperforming existing state-of-the-art approaches.

KEYWORDS

ALS point clouds, semantic segmentation, graph edge convolution, high-voltage transmission corridors, deep learning

1 Introduction

High-voltage transmission line inspection is a vital and routine task in power systems, crucial for ensuring the safety and stability of electricity supply. Its primary objective is to assess the structural integrity of power infrastructure components—including conductors, towers, and fittings—while identifying potential hazards such as vegetation encroachment, geological disasters, and tower tilts (Yang et al., 2020). Due to its high accuracy, rapid data acquisition, and all-weather operational capability, airborne Light Detection and Ranging (LiDAR) technology has been widely adopted for transmission line inspection. Semantic segmentation of LiDAR

point clouds, as a fundamental step in intelligent inspection workflows, enables precise differentiation of key elements such as conductors, towers, and vegetation, thereby enhancing automation and supporting fine-grained management and intelligent analysis.

Previous research on the semantic segmentation of transmission line point clouds has predominantly relied on handcrafted features and physics-based constraints (Huang et al., 2021; Rejichi and Chaabane, 2015). Conventional methods include terrain classification using machine learning algorithms such as Random Forests and Support Vector Machines (SVM), as well as line- and surface-based feature extraction using Principal Component Analysis (PCA) and geometric descriptors. Although these approaches have achieved promising results, they suffer from limitations such as high parameter sensitivity, low computational efficiency, and poor generalization capability. As a result, they are typically constrained to specific datasets and narrowly defined scenarios, rendering them inadequate for the large-scale, high-performance, and fine-grained requirements of modern intelligent inspection tasks.

Recently, deep learning-based 3D point cloud semantic segmentation has been extensively explored in domains such as indoor navigation, smart cities, digital twins, plant inspection, and autonomous driving (Hu et al., 2020; Fan et al., 2021; Yin et al., 2023; Landrieu and Simonovsky, 2018; Han et al., 2023; Xiang et al., 2023; Ghahremani et al., 2021; Xin et al., 2023). In parallel, fine-grained semantic segmentation methods have attracted increasing attention due to their ability to differentiate visually similar and spatially adjacent components in complex environments. For instance, SSC-Net employs a multi-task joint learning strategy for segmentation and classification, demonstrating high performance in challenging biomedical imaging scenarios (Sha et al., 2025). Similarly, IndVisSGG introduces a vision-language model-based scene graph generation framework for industrial applications, effectively capturing small-scale structures with contextual reasoning (Wang et al., 2025). In the domain of multimodal learning, Vman integrates visual-modified attention mechanisms to enhance semantic recognition across tasks involving subtle feature variations (Song et al., 2025). Moreover, sparse Bayesian learning techniques have been successfully applied to radar data processing for superresolution and small-target enhancement, providing a principled approach to fine-grained signal discrimination (Zhang and Chen, 2014). Correspondingly, there is growing interest in leveraging deep learning techniques for semantic understanding of transmission line environments (Shen et al., 2023; Su et al., 2022; Zhou et al., 2024; Shi et al., 2023; Zhao et al., 2020). Despite notable advances, several critical challenges remain. (1) Most existing methods focus on segmenting broad object categories (e.g., ground, buildings, vegetation, towers, conductors) while neglecting the fine-grained segmentation (Lin et al., 2021) of essential metallic components such as ground wires, insulators, and drainage wires, which are indispensable for power system safety. Although certain studies have incorporated spatial topological constraints to distinguish ground and cross-span wires, the accurate segmentation of insulators and drainage wires remains challenging due to their similar geometric characteristics and close physical proximity to towers and conductors (Wang et al., 2023). Some approaches introduce additional post-processing steps to differentiate conductors from insulators; however, these techniques

are typically time-consuming and highly sensitive to parameter settings, which limits their practical applicability. (2) Transmission corridor point clouds exhibit a strip-like, continuous spatial distribution. Traditional point-based sampling strategies (e.g., k-nearest neighbors, farthest point sampling) are often inadequate due to scale disparities along the XY axes, thereby hindering full-scene sampling and reducing the model's capacity to learn discriminative features for small-class objects. Moreover, a single transmission span—comprising two towers and the connecting segment—can contain tens of millions of points, while critical components such as conductors, insulators, and drainage wires may account for less than 5% of the total data. Therefore, developing effective encoding and feature learning strategies specifically tailored to the unique structure of transmission corridor point clouds remains an open and significant research challenge.

To address these issues, we propose a fine-grained semantic understanding encoder network for transmission corridors based on airborne LiDAR point clouds. We introduce a Graph Edge-Constrained Convolution with Attention Encoder (GECCAE), which performs local-to-global feature encoding from both point-wise and class-aware perspectives, enabling high-accuracy, robust, and fine-grained semantic segmentation of transmission corridor components. The main contributions of this study are as follows:

1. A Graph Edge-Constrained Convolution (GECC) module guided by segmentation, which models inter-class geometric relationships to enhance geometric feature extraction and significantly improves inter-class distinction for precise semantic boundary delineation.
2. A novel Channel-Spatial Attention Module (CSAM) that captures cross-scene contextual dependencies and jointly models local and global features, thereby enhancing point-level semantic understanding.
3. A systematic classification schema tailored for modern power grid asset management. Extensive quantitative and qualitative evaluations on multiple real-world transmission corridor datasets demonstrate that the proposed method outperforms existing state-of-the-art approaches in various power line segmentation tasks.

In addition, recent advances in multi-scale spatiotemporal interactive fusion networks (Ma et al., 2025), deep core node information embedding (Fei et al., 2025), history-enhanced 3D scene graph reasoning (Feng et al., 2025), and pixel-level noise mining (Liu et al., 2025) have provided new insights and methodologies for addressing the challenges of fine-grained semantic segmentation in transmission corridor point clouds. These studies have significantly improved model performance in complex scene understanding and segmentation accuracy by introducing techniques such as multi-scale feature fusion, deep information embedding, and temporal context enhancement. In particular, the multi-scale spatiotemporal interactive fusion network enhances the ability to capture dynamic changes and spatial relationships of key components by integrating features across multiple spatial and temporal scales. Meanwhile, pixel-level noise mining offers an effective strategy for mitigating noise in point cloud data, further improving the accuracy and robustness of segmentation results.

The structure of this paper is organized as follows. Section 1 outlines the research objectives and summarizes the main

contributions. [Section 2](#) reviews the current state-of-the-art methods and analyzes their advantages and limitations. [Section 3](#) presents the proposed methodology, including the Kernel Point Convolution (KPCnv) module, the Graph Edge-Conditioned Convolution (GECC) module, the Channel-Spatial Attention Module (CSAM), and the overall network architecture. [Section 4](#) reports the experimental results on real-world transmission corridor datasets, compares our method with existing approaches, and discusses its effectiveness. Finally, [Section 5](#) concludes the paper and outlines potential future work.

2 Related works

In recent years, the semantic segmentation of LiDAR point clouds has emerged as a key research focus for the intelligent inspection of power transmission corridors. Existing approaches can be broadly categorized into traditional methods and deep learning-based techniques.

2.1 Traditional methods

Traditional methods primarily employ either unsupervised or supervised machine learning techniques, relying heavily on handcrafted features and predefined physical constraints.

Unsupervised methods, such as Principal Component Analysis (PCA), are widely used to identify linear structures such as power lines by computing dominant eigenvectors. These approaches offer high computational efficiency; however, they are sensitive to outliers and exhibit limited capacity in modeling complex spatial configurations, resulting in suboptimal performance for large-scale semantic segmentation tasks ([Rejichi and Chaabane, 2015](#); [Hui et al., 2021](#); [Nurunnabi et al., 2012](#); [Cao et al., 2025](#)).

Supervised classification methods, including Random Forest (RF) ([Jiang et al., 2022](#); [Liao et al., 2022](#); [Ni et al., 2017](#); [Tang et al., 2023](#)) and Support Vector Machines (SVM) ([Chen et al., 2019](#); [Zhang et al., 2013](#)), utilize manually crafted feature vectors to perform object classification at the scene level. RF enhances robustness through ensemble learning, but suffers from limited interpretability and difficulty in distinguishing geometrically similar objects. SVM seeks an optimal hyperplane in high-dimensional feature space; however, its performance often deteriorates in multi-class segmentation scenarios due to challenges in defining consistent decision boundaries across diverse object classes ([Zafar et al., 2018](#)).

Overall, these traditional approaches are inherently constrained by their reliance on domain-specific rules and limited generalization capabilities, rendering them inadequate for modern large-scale, fine-grained monitoring of power transmission systems ([Mirzaei et al., 2022](#); [Grothum et al., 2023](#); [Shen et al., 2024](#)).

2.2 Deep learning methods

With the rapid advancement of 3D deep learning, point cloud semantic segmentation has achieved significant progress across a wide range of domains, including autonomous driving, digital twins, and smart grids. Deep learning-based methods can generally be

categorized into point-based methods, convolution-based methods, and hybrid approaches.

2.2.1 Point-based methods

Point-based methods directly operate on 3D spatial coordinates, thereby preserving the raw geometric structure of point clouds. PointNet and its variants, such as PointNet++ ([Charles et al., 2017](#); [Qi et al., 2017](#)) and RandLA-Net ([Hu et al., 2020](#)), utilize multi-layer perceptrons (MLPs) and farthest point sampling (FPS) strategies to perform per-point semantic prediction ([Lawin et al., 2017](#); [Lai et al., 2022](#)). SCF-Net ([Fan et al., 2021](#)) and Stratified Transformer ([Lai et al., 2022](#)) further extend these frameworks by aggregating local contextual information through neighborhood-based encodings and attention mechanisms. While these models offer architectural simplicity and computational efficiency, they often struggle to accurately segment small object categories due to downsampling-induced feature dilution.

Recent studies have applied point-based techniques to power grid scenarios. For instance, Zhang et al. enhanced PointNet by incorporating geometric feature extraction (GFE) and neighborhood information aggregation (NIA) modules to segment power lines and towers in railway environments ([Zhang et al., 2022](#)). Su et al. (2022) and Yu et al. (2023) employed PointNet++ for high-precision segmentation of transmission structures, while Yu et al. additionally proposed a dual-stage sampling strategy based on RandLA-Net to improve segmentation accuracy for wires and towers.

2.2.2 Convolution-based and graph-based methods

Convolution-based methods aim to regularize the irregular structure of point clouds through voxelization or kernel-based strategies, thereby enabling the application of convolutional neural networks (CNNs). Representative models such as KPCnv and PointCNN employ kernel point convolution and learned transformation metrics, respectively, to effectively capture local geometric context ([Thomas et al., 2019](#); [Thomas et al., 2024](#); [Huang et al., 2024](#)).

Graph-based methods transform point clouds into graph structures that explicitly represent spatial relationships among neighboring points. Graph neural networks (GNNs), including DGCNN, SPG, and SPT, utilize these representations to facilitate both local and global feature learning ([Wang et al., 2019](#); [Simonovsky and Komodakis, 2017](#); [Robert et al., 2023](#)). These models have demonstrated superior capability in modeling topological relationships, making them particularly well-suited for corridor-like scenes with strong spatial continuity.

In the power grid domain, several studies have adopted PointCNN for UAV-based transmission line segmentation ([Zhao et al., 2023](#); [Zhou et al., 2024](#)), while others have incorporated attention mechanisms into KPCnv-based frameworks to enhance feature extraction in power corridor environments.

2.2.3 Multi-modal and hybrid methods

Some recent studies have explored multi-modal fusion strategies to overcome the limitations of single-representation learning. RPNVNet employs a deep adaptive Range-Point-Voxel fusion framework to integrate complementary information from

different modalities (Xu et al., 2021). PVCNN Liu et al. (2019) combines point-based and voxel-based features to enhance semantic representations. Wang et al. (2023) further introduced coordinate attention mechanisms into PointNet++, resulting in an end-to-end CA-PointNet++ model that improves spatial feature awareness.

Despite considerable progress, directly applying existing deep learning models to transmission corridor scenarios remains challenging. Point cloud data in these environments often exhibit high sparsity, severe class imbalance, and a large number of small-scale target components. These characteristics significantly hinder model performance, particularly in identifying critical but underrepresented elements such as insulators, ground wires, and drainage lines. Moreover, the elongated spatial distribution of corridor scenes complicates traditional point sampling strategies (e.g., KNN, FPS), thereby weakening the network's ability to capture global and contextual features. Consequently, there is an urgent need for customized semantic segmentation models specifically designed to address the unique spatial and structural properties of high-voltage transmission corridors.

3 Methodology

Due to the characteristics of airborne LiDAR scanning and the highly unbalanced spatial distribution of scene elements in high-voltage transmission corridors, the resulting point clouds typically exhibit significant sparsity and severe class imbalance. Although some studies have attempted to enhance segmentation performance for small-sample classes using data augmentation techniques, these approaches operate primarily at the data preprocessing level and fail to address the fundamental limitations of network generalization.

The core challenge in semantic understanding of transmission corridor scenes lies in achieving accurate and fine-grained segmentation of key power infrastructure components, including towers, conductors, and associated metallic fittings (e.g., insulators and jumper wires).

To tackle this problem, we propose a novel architecture termed Graph Edge-Conditioned Convolution Attention Encoder (GKCAE). This network integrates graph convolution and point convolution to jointly encode local-global relationships at both the point-level and class-level representations. Furthermore, a context-aware attention module is incorporated to aggregate multi-scale semantic features and enhance deep feature extraction. This design enables robust, high-precision, and fine-grained semantic segmentation of complex point cloud data from power line corridors.

3.1 KPConv block

Our work focuses on three-dimensional point cloud data from high-voltage transmission corridors scenario. We employ a convolutional architecture that integrates kernel point convolutions and graph-based modeling to extract point-wise semantic features. The KPConv block is implemented following the variational convolutional framework proposed in Thomas et al. (2019), and its structure is illustrated in Figure 1.

Our work focuses on three-dimensional point cloud data in high-voltage transmission corridor scenarios. We employ a convolutional architecture that integrates kernel point convolutions with graph-based modeling to extract point-wise semantic features. The KPConv block is implemented based on the variational convolutional framework proposed in Thomas et al. (2019), and its architecture is illustrated in Figure 1.

In KPConv, each convolutional kernel is composed of a set of kernel points with predefined coordinates in 3D Euclidean space. For a given input point, convolution is performed by computing weights based on its relative position to these kernel points. The input to the KPConv layer consists of both spatial coordinates and associated feature vectors. The spatial coordinates, typically represented as (x, y, z) , encode the geometric structure and spatial relationships among points in the cloud. Formally, a KPConv kernel is defined by a set of kernel points $\{p_i\}_{i=1}^I$, where each p_i denotes a kernel point in 3D space. These kernel points serve as learnable positions analogous to filter weights in conventional image convolution, but are distributed irregularly to adapt to the unstructured nature of point clouds. During training, the positions of the kernel points are optimized to improve feature extraction capability, allowing KPConv to capture local geometric variations more effectively than grid-based methods.

Given an input point cloud $\mathcal{P} \in \mathbb{R}^{N \times 3}$ and its associated features $\mathbf{F} \in \mathbb{R}^{N \times D}$, the kernel g convolution for a point x_i can be defined as follows (Equation 1):

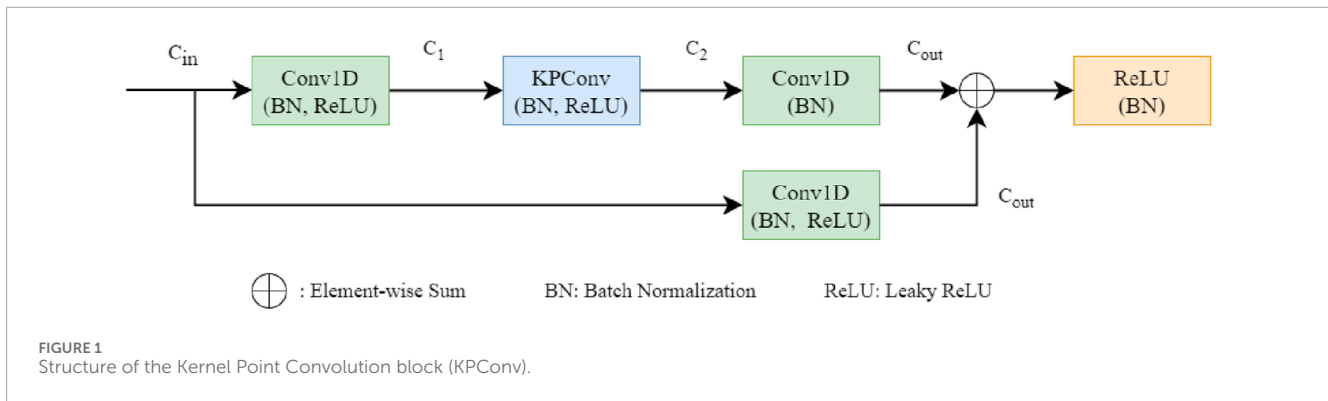
$$(F * g)(x) = \sum_{x_i \in N_x} g(x_i - x) f_i \quad (1)$$

Further, it can be explained that for each input point x_i , the network searches for a set of neighboring points $N(x_i)$. This is typically accomplished using a nearest neighbors search, where neighbors are determined by calculating the Euclidean distance $\|x_j - x_i\|$ between points and selecting the nearest n points (or within a spherical neighborhood). KPConv performs a distance-weighted convolution operation. The output feature f'_i for each input point x_i is computed as follows (Equation 2):

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

Here, $h(\cdot)$ is a weighting function, such as a Gaussian kernel, that calculates the weight 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 the neighboring points through weighted summation to produce a new feature vector. This new feature effectively captures local geometric and attribute information, representing the characteristics of the point cloud.

Additionally, to enhance the learning capability for varying object sizes, the convolution can incorporate deformable kernels. While some scholars argue that deformability has negligible effects on ALS data (Lin et al., 2021; Wen et al., 2021; Wen et al., 2020), we believe that in the context of high-voltage transmission line scenes, deformable KPConv can adapt to local geometric shapes, thus enhancing the representation of detailed features. This is particularly beneficial for capturing the significant geometric variations of objects such as power towers and buildings. Therefore, in our work, we still adopt a mixed approach as suggested by Thomas et al. (2019),



Thomas et al. (2024), incorporating both rigid and deformable kernels within the overall semantic segmentation structure. The calculations and definitions for the deformable kernel features are expressed in Equations 3, 4:

$$(F * g)(x) = \sum_{x_i \in N_x} g_{\text{deform}}(x_i - x, \Delta(x)) f_i \quad (3)$$

$$g_{\text{deform}}(y_i, \Delta(x)) = \sum_{k < K} h(y_i, (\tilde{x}_k + \Delta_k(x))) W_k \quad (4)$$

Here, $\Delta_k(x)$ represents the offsets generated by the rigid KPCnv, mapping the dimension of input features to $3K$ values. During training, the network simultaneously learns to generate shifts for the rigid kernel and output features for the deformable kernel, with the learning rate for the former set to 0.1 times that of the global network learning rate. After the convolution operation, nonlinear activation functions, specifically ReLU, along with batch normalization, are applied to enhance the model's nonlinear capacity and stability.

3.2 Graph edge-conditioned convolution block

Although KPCnv can learn local features of points through direct point convolution, its neighborhood learning limitations prevent it from obtaining a broader receptive field, particularly for capturing global information. Additionally, point-wise features fail to encode relationships between objects within a scene, resulting in challenges when exploring interactions among objects. This limitation affects our task by hindering the accurate segmentation of transmission lines and other linear targets, as well as distinguishing between poles and buildings that may be obscured by vegetation. Previous work has demonstrated that incorporating global contextual information can enhance model performance for large-scale semantic segmentation tasks (Thomas et al., 2019; Huang et al., 2024). To address this issue, we introduce a Graph Edge-Conditioned Convolution (GECC) block to encode and extract features based on dependencies between global objects. Inspired by SPG and DGCNN, we designed the GECC block, illustrated in Figure 2. We construct graphs from geometric homogeneous points along line segments to capture relationships between objects. By combining segment features with point features, the network can adaptively encode local and global features, thereby achieving improved semantic predictions on the ALS dataset.

For the definition of the graph structure and embedding encoding, our process is as follows. First, for edge-conditioned convolution, we consider it as a directed or undirected graph $G = (V, E)$, where V is a finite set of vertex geometries with $|V| = n$ and $E \in V \times V$ is a set of edges with $|E| = m$. Let $I \in \{0, \dots, I_{\max}\}$ be the layer index in the feedforward neural network. There exists a function $X^I: V \rightarrow \mathbb{R}_I^d$ that assigns labels to each vertex, and $L: E \rightarrow \mathbb{R}^s$ that assigns labels to each edge. We utilize the concept of the Superpoint graph from SPG and apply an unsupervised L_0 -cut pursuit algorithm (Landrieu and Obozinski, 2017) to partition the point cloud into individual clusters. This step involves pre-computing semantic labels and predefined geometric features before training. As a result, the graph structure remains fixed, eliminating the need to update segment labels repeatedly during training, thereby avoiding unnecessary computational overhead.

Next, we construct the graph structure based on the input. Given a point cloud \mathcal{P} with its geometric features X_p (such as linearity or normal vectors), we build a directed graph $G = (V, E)$ and set its labels X^0 and L as follows. We create a vertex $i \in V$ for every point $p \in \mathcal{P}$ and assign the respective signal to it by $X^0(i) = X_p(p)$ (or 0 if there are no features $X_p(p)$). Then, we connect each vertex i to all vertices j in its spatial neighborhood via a directed edge (j, i) . The edge-conditioned convolution can dynamically produce filtering weights $\Phi_{(j,i)}^l \in \mathbb{R}^{d_j \times d_{i-1}}$ based on the edge (j, i) and flexibly handle varying neighborhood sizes to capture contextual information between different segments. Algorithm 1 presents the pseudocode for constructing the graph structure and computing the features using the Graph Edge-Conditioned Convolution (GECC) module. The computation of ECC is as follows (Equation 5):

$$X^l(i) = \frac{1}{|N(i)|} \sum_{j \in N(i)} \Phi_{(j,i)}^l X^{(l-1)}(j) + b^l \quad (5)$$

We learn the parameters $\Phi_{(j,i)}^l$ from a multilayer perceptron. The edge features (j, i) are processed by $\Phi_{(j,i)}^l$ to generate a weight matrix, which is then used to perform matrix-vector multiplication with the features of neighboring nodes. Finally, the updated node features are directly backpropagated to each point, resulting in $F_{\text{gecc}} \in \mathbb{R}^{N \times C}$ as the output of this block. Physically, this formulation enables each point to aggregate and filter geometric features from its spatial neighbors, allowing the model to learn contextual relationships based on local geometry.

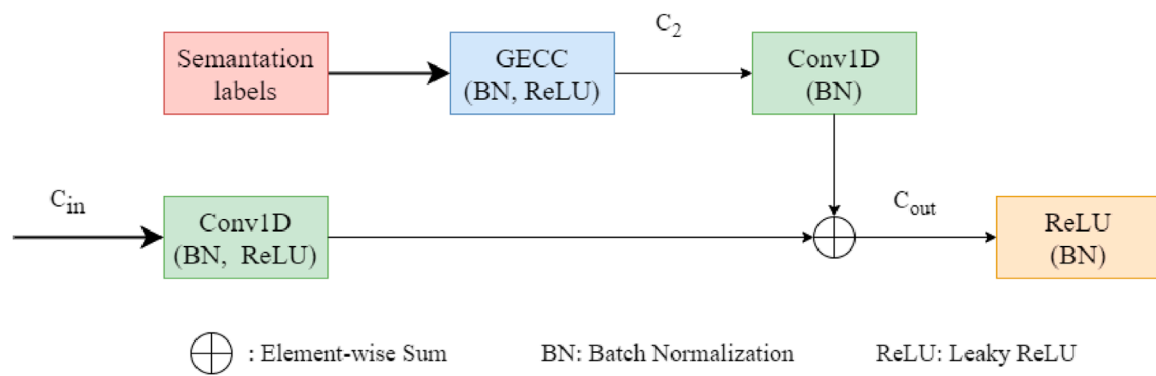


FIGURE 2
Structure of the Graph Edge-Conditioned Convolution block (GECC).

Require: Point features $F_{in} \in \mathbb{R}^{N \times C_{in}}$, segment labels $S \in \{1, \dots, N_s\}^N$

Ensure: Output features $F_{out} \in \mathbb{R}^{N \times C_{out}}$

```

1:  $F_1 \leftarrow \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(F_{in}, C_1)))$ 
2:  $F_{local} \leftarrow \text{KPCConv}(F_1)$   $\triangleright$  Local features in  $\mathbb{R}^{N \times C_3}$ 
3: for  $k=1$  to  $N_s$  do
4:    $M_k \leftarrow \frac{1}{|\{p: S(p)=k\}|} \sum_{S(p)=k} F_{local}[p]$ 
5: Construct graph  $G=(V, E)$  with segments  $\{s_k\}$  as
nodes and fully connected edges
6: for each node  $i \in V$  do
7:    $agg \leftarrow \emptyset$ 
8:   for each neighbor  $j \in \mathcal{N}(i)$  do
9:      $e_{ij} \leftarrow M_i - M_j$ 
10:     $\Phi_{(j,i)}^1 \leftarrow \theta(e_{ij}; W_e) \in \mathbb{R}^{C_3 \times C_3}$ 
11:     $agg \leftarrow agg + \Phi_{(j,i)}^1 \cdot M_j$ 
12:     $M'_i \leftarrow \frac{1}{|\mathcal{N}(i)|} \cdot agg + b^1$   $\triangleright$  Eq. (5)
13:   for each point  $p$  do
14:     $F_{ecc}[p] \leftarrow M'_{S(p)}$ 
15:    $F_{cat} \leftarrow \text{Concat}(F_{local}, F_{ecc})$ 
16:    $F_{mixed} \leftarrow \text{BN}(\text{Conv}_{1 \times 1}(F_{cat}, C_{out}))$ 
17:    $F_{res} \leftarrow \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(F_{in}, C_{out})))$ 
18:    $F_{out} \leftarrow \text{ReLU}(\text{BN}(F_{mixed} + F_{res}))$ 
19: return  $F_{out}$ 

```

Algorithm 1. GECC Block Feature Generation Algorithm.

3.3 Channel-spatial attention module

Contextual information is essential for capturing global relationships within a scene. In point cloud semantic segmentation, local features typically focus on the geometric properties of points within a local neighborhood, but points belonging to the same category can share similar characteristics despite being spatially distant. Identifying the correlations among points in feature space can significantly improve the model's predictive accuracy. The Channel-Spatial Attention Module (CSAM) is frequently used in remote sensing to model the global contextual information

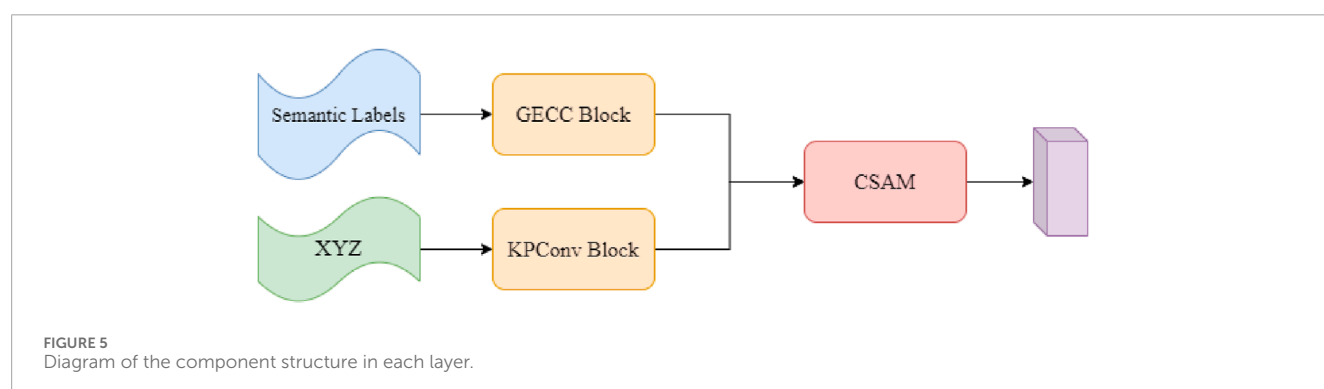
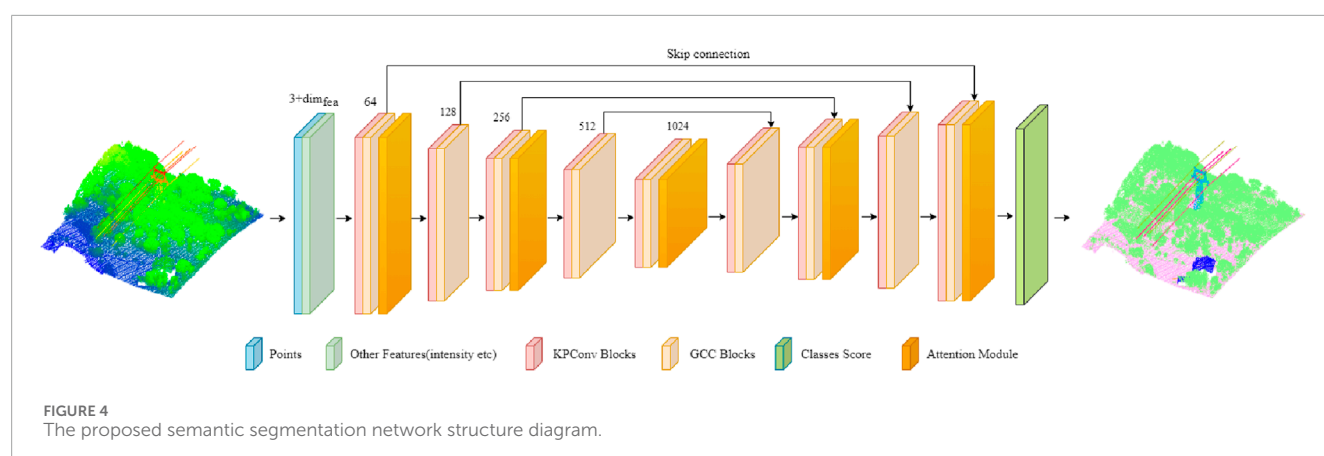
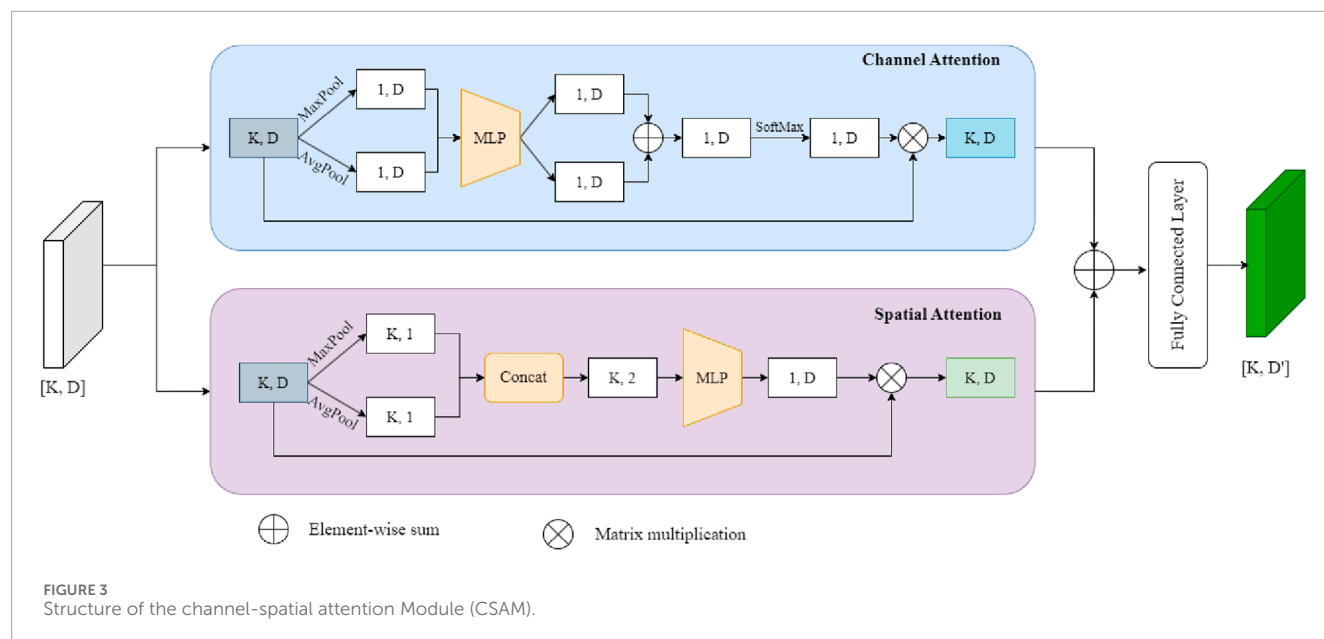
of targets in images, enhancing the global feature representation. Similarly, we employ a CSAM to compute the attention of features processed by the graph-point convolution, allowing for the modeling of similarities between distant targets and improving the global understanding of the scene.

As shown in Figure 3, the input features are projected into different feature subspaces through various learnable fully connected layers, which are then used to construct queries, keys, and values for the attention function. The output of the attention function serves to enhance the features by encoding global contextual information. Our channel attention mechanism focuses on selecting and amplifying the features that are most beneficial for the network. In Figure 3, the weights for different features are computed by analyzing the relationships between them. These computed weights are then multiplied with the features to emphasize those that are crucial for network classification. During this weight computation, we reduce the spatial dimensions of the input features to improve computational efficiency. For multi-feature aggregation, unlike PointNet, which directly applies max pooling, we use average pooling to better capture the extent of target categories. Thus, before summation, we perform feature aggregation using both max pooling and average pooling. The specific operations of the channel attention mechanism are outlined as follows.

First, we aggregate spatial dimension information between different feature channels using both average pooling and max pooling to obtain the average pooled feature $F_{x(i,j)}^{avg}$ and the max pooled feature $F_{x(i,j)}^{max}$. Then, these two distinct features are input into a multi-layer perceptron (MLP) network to generate an important feature mapping function $M_c \in \mathbb{R}^{1 \times D}$. This MLP network consists of two fully connected layers, an activation function, and a dropout mechanism applied after the second fully connected layer to enhance the network's generalization capability. Finally, the results from the MLP processing of the two different features are summed element-wise, which can be expressed as follows (Equations 6, 7):

$$M_c = \sigma \left(\varphi \left(F_{x(i,j)}^{avg} \right) \oplus \varphi \left(F_{x(i,j)}^{max} \right) \right) \quad (6)$$

$$F'_{x(i,j)} = M_c \left(F_{x(i,j)} \right) \odot F_{x(i,j)} \quad (7)$$



where \oplus and \otimes represent element-wise addition and matrix multiplication, respectively, σ denotes the activation function, and ϕ represents the MLP network.

The spatial attention module is specifically designed to select neighborhoods that are more beneficial for expressing point cloud shape information, as illustrated in the flowchart below in

Figure 3. For the input features, we first apply average pooling and max pooling operations to enhance the focus on the relationships between different object categories. Notably, the features here are the aggregated feature matrices from GECC and KPCov, representing a point-object level feature aggregator. Next, we concatenate the resulting features $M_k \in \mathbb{R}^{K \times 1}$ and perform convolution operations to

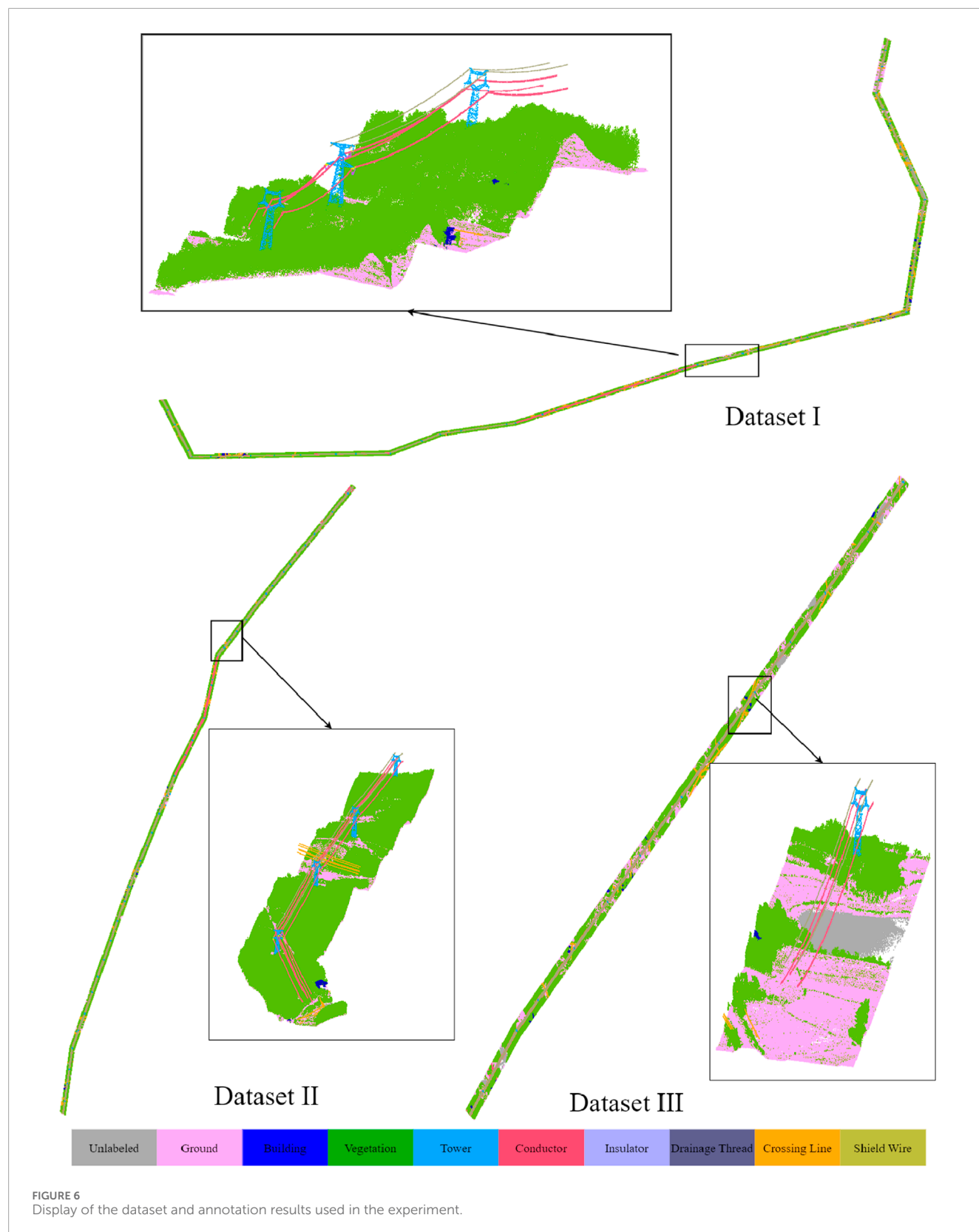


FIGURE 6
Display of the dataset and annotation results used in the experiment.

generate different attention coefficients:

$$M_k = \sigma \left(\varphi \left[\theta_{\max} \left(F'_{x_{(ij)}} \right) \parallel \theta_{\text{avg}} \left(F'_{x_{(ij)}} \right) \right] \right) \quad (8)$$

in Equation 8, where σ stands for the activation function, φ stands for the MLP network, \parallel represents the concatenation operation, and θ_{\max} and θ_{avg} represent the max pooling and average pooling operations, respectively.

TABLE 1 Class-wise point count and percentage in the test dataset.

| | Conductor | Tower | Vegetation | Building | Insulator | Drainage thread | Crossing line | Shield wire | Ground | Sum |
|--------------------------|-----------|-----------|----------------|---------------|-----------|-----------------|---------------|-------------|----------------|-------------|
| Number of categories | 2,932,346 | 1,045,266 | 138,751,040 | 3,500,440 | 18,430 | 18,582 | 614,916 | 1,155,238 | 80,115,096 | 228,151,354 |
| Percentage of categories | 1.285% | 0.458% | 60.815% | 1.534% | 0.008% | 0.008% | 0.270% | 0.506% | 35.115% | 100.000% |

Bold values represent the proportion of the top three most frequent sample categories.

Finally, we combine the channel attention and spatial attention, passing them through a fully connected network to obtain point-wise attention scores. Through this spatial-channel attention module, point features are updated from a global perspective, enabling comprehensive learning of interactions between complex points. This process enhances the model's ability to capture dependencies across spatially distant points, thereby improving prediction accuracy.

3.4 Overall architecture

We propose the graph kernel convolution attention-based encoder (GKCAE) for local-global feature encoding of airborne LiDAR point cloud data in high-voltage transmission corridor scenes. Inspired by previous work, we use a U-Net architecture as the overall framework, and the designed semantic segmentation model is illustrated in Figure 4. The model employs a five-layer network for encoding, with the structures of graph edge-conditioned convolution (GECC), kernel point convolution (KPCnv) and channel spatial attention module (CSAM) in each layer depicted in Figure 5.

To capture local geometric information at multiple scales, we applied downsampling to gradually expand the convolutional receptive field. In the decoder, nearest-neighbor upsampling is employed to obtain the final point-wise features. Throughout the encoding layers, we extract local-global features and incorporate CSAM in the first, third, and fifth layers to capture contextual information. In addition, skip connections are used to transfer intermediate features from the encoder to the decoder, where they are combined with the upsampled features and passed through a fully connected layer to produce the final semantic predictions.

To train the network, we use a weighted cross-entropy loss, as shown in Equation 10:

w_c = \frac{N_c}{\sum_{c=1}^C N_c} \tag{9}

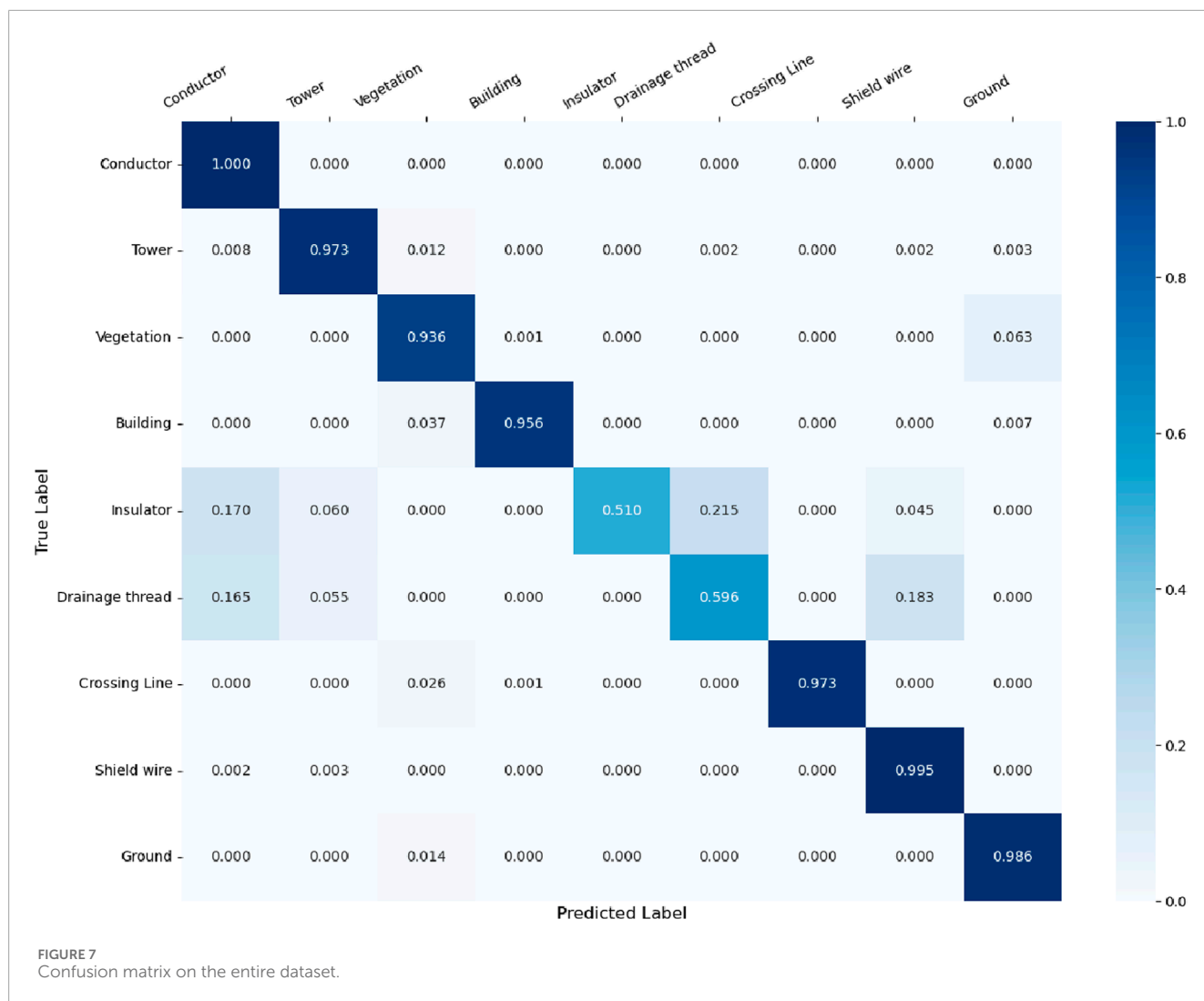
Loss = \sum_{i=1}^N \sum_{c=1}^C w_c (A'_{ic} \log \rho_{ic} + (1 - A'_{ic}) \log (1 - \rho_{ic})) \tag{10}

The weight w_c for each class is calculated based on its proportion in the total number of points, as shown in Equation 9, where N_c denotes the number of points in class c. In Equation 10, A'_{ic} indicates whether the true label of the i - th sample belongs to class c, while \rho_{ic} represents the corresponding predicted probability.

4 Experiments

4.1 Datasets and matrices

In order to verify the effectiveness of our method, we manually annotated three regions collected by airborne LiDAR. The data was collected in Anhui, China, using a DJI M600 drone equipped with a Riegl VUX-LR laser scanner. The point density of this data reaches 40–50 points per square meter. To validate the model, we employed the CloudCompare software for manual annotation. As shown in Figure 6, the three datasets used in this experiment and



their annotation results are presented. It can be observed that, in addition to the conventional categories such as ground, vegetation, buildings, poles and frames, and wires, we further classified the wires into a total of 9 categories, including crossing lines, ground wires, insulators, jumper wires, etc. For each dataset, we manually divided it into blocks, and then partitioned them into training, validation, and test sets at a ratio of 7:1.5:1.5 respectively, for the training and validation of the proposed method. Table 1 shows the number and proportion of points for each category in the test dataset. It can be observed that there is a pronounced class imbalance in high-voltage transmission line scenarios. Large-scale objects such as vegetation, buildings, and ground account for more than 97% of the points, while power facility categories represent less than 3%. In particular, the proportions of classes like insulator and drainage thread are even lower.

Specifically, the block division was carried out by cropping 50 m on each side along the direction perpendicular to the transmission line corridor. Meanwhile, along the direction of the transmission line corridor, the data was cropped into blocks every 300 m, and a 10-meter overlapping buffer zone was reserved between adjacent blocks to avoid the fragmentation of ground objects during cropping.

To quantitatively evaluate our method, we follow the protocol adopted in previous studies (Zhou et al., 2024; Wang et al., 2023; Jiang et al., 2022), use Overall Accuracy (OA) and Mean Intersection over Union (mIoU) as assessment metrics. OA measures the percentage of correctly predicted points relative to the total number of test points, while mIoU evaluates the semantic segmentation performance across various categories. The calculation formulas are as follows (Equations 11, 12):

$$OA = \frac{TP_{all}}{TP_{all} + FP_{all}} \quad (11)$$

$$mIoU = \frac{1}{C} \sum_{c=1}^C \frac{TP}{TP + FP + FN} \quad (12)$$

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

4.2 Implementation details

In terms of implementation details, we adopt the Adam optimizer to train the network, which is developed and implemented

TABLE 2 Quantitative comparison results on the entire dataset. The first column lists the different methods, the second and third columns show the overall mIoU and OA values, and the following columns represent the IoU values for each class. The best result in each column is highlighted in black.

| Methods | mIoU | OA | IoU | | | | | | | | |
|-----------------------------|--------|--------|-----------|--------|------------|----------|-----------|-----------------|---------------|-------------|--------|
| | | | Conductor | Tower | Vegetation | Building | Insulator | Drainage thread | Crossing line | Shield wire | Ground |
| PointNet++ Qi et al. (2017) | 35.96% | 73.17% | 71.56% | 24.25% | 65.44% | 68.71% | 0.00% | 0.00% | 0.00% | 23.50% | 70.21% |
| KPConv Thomas et al. (2019) | 64.01% | 91.04% | 75.07% | 92.65% | 85.14% | 82.07% | 42.80% | 19.56% | 52.07% | 48.99% | 77.74% |
| PT Wu et al. (2024a) | 71.67% | 92.48% | 87.23% | 89.34% | 87.13% | 80.61% | 32.34% | 41.23% | 54.21% | 93.10% | 79.85% |
| Ours | 81.93% | 94.10% | 88.94% | 95.34% | 86.34% | 89.22% | 68.39% | 62.41% | 69.54% | 97.91% | 79.28% |

Bold values indicate the best-performing evaluation values in their respective columns.

using the PyTorch framework. A warm-up learning rate strategy is employed to gradually adjust the learning rate during the initial training phase. Prior to training, additional grid-based downsampling is applied to the input data to ensure uniform point spacing. Specifically, the grid size is set to 0.2, and a hash mapping between the downsampled points and the original point cloud is maintained. During inference, a voting-based strategy is used to map the predicted labels back to the original point cloud based on this mapping. For the kernel point convolution layers, the spherical neighborhood radius is set to 25 m, and the initial kernel radius is set to 0.5 m. The receptive field gradually expands with increasing network depth. To improve the model’s generalization and robustness, we apply random data augmentations, including rotation, translation, and noise injection. During training, ground-truth semantic labels are utilized to construct and optimize the Graph Embedding and Coordinate Conversion (GECC) graph. However, during validation, semantic labels are excluded, and the graph is constructed solely based on the learned features and geometric coordinates. All experiments are conducted on a single NVIDIA GeForce RTX 4090 24 GB GPU.

4.3 Semantic segmentation results

In this section, to validate the effectiveness and performance of the proposed method, we conduct an analysis from both quantitative and qualitative perspectives. Additionally, we select several state-of-the-art (SOTA) methods of the same type for comparison. These include the point-based MLP method, PointNet++ (Charles et al., 2017; Qi et al., 2017), the point convolution-based method KPConv (Thomas et al., 2019), and the Transformer-based PointTransformer (PT) (Wu et al., 2024b; Wu et al., 2024a; Wu et al., 2022). In the comparative experiments, the data sample partitioning follows the strategy described in Section 4.1. For KPConv, we set the input radius to 25 m and the initial kernel radius to 0.5 m, consistent with the settings used in our model’s KPConv block. For the other methods, the hyperparameters are set to the empirical values reported in their respective original papers.

4.3.1 Comparison results and analysis on the entire dataset

We first conducted a comparative experiment on the overall dataset, with the confusion matrix of our method presented in Figure 7. The quantitative results, comparing our method to others, are shown in Table 2. As observed, our method achieved an mIoU of 81.93% and an OA of 94.1% on the overall dataset, outperforming the other methods. Additionally, our approach achieved the best segmentation results in the categories of conductor, tower, building, insulator, drainage thread, crossing line, and shield wire. The PT method demonstrated the best segmentation results for vegetation and ground, though our method was only slightly behind. Furthermore, although other state-of-the-art (SOTA) methods, such as KPConv and PT, exhibited strong overall accuracy (with OA values exceeding 90%), their performance in terms of mIoU and class-specific segmentation accuracy was somewhat inferior. In contrast, our method demonstrates more balanced performance, particularly in the segmentation of linear features.

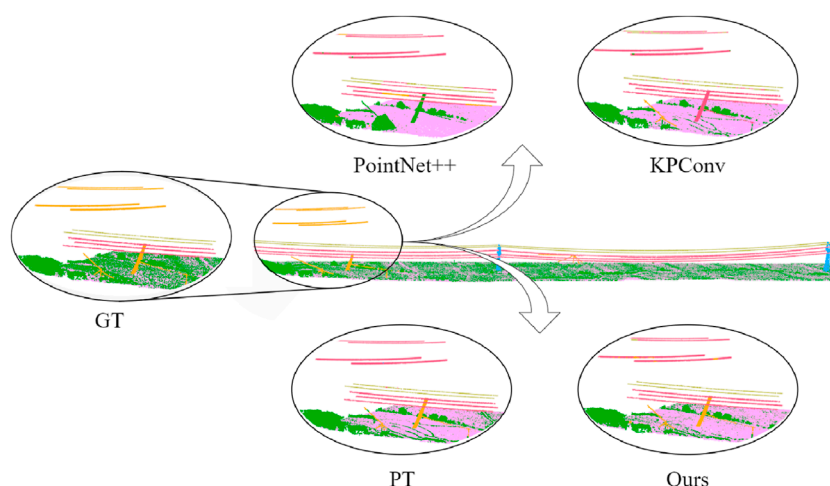


FIGURE 8

Qualitative segmentation results of different methods on the entire dataset. The center shows the ground truth point cloud and labels, while the top and bottom sections display the segmentation results of different methods within the black circle region.

Figure 8 shows the segmentation performance of different methods in the same region of a tile from the overall dataset. This region contains multiple distinct power line targets (crossing line, conductor, and shield wire) as well as common targets (ground, vegetation). It can be seen that PointNet++ struggles to effectively distinguish between different linear targets. KPConv can roughly identify the conductor and shield wire, but fails to recognize the crossing line. PT provides better recognition results for conductor and shield wire, but performs poorly on the crossing line. In contrast, our method can effectively distinguish between these three different linear targets. Although the crossing line, shield wire, and conductor have the same physical mechanisms, their applications in power transmission scenes differ. Our method achieves good segmentation performance for these targets, largely due to our GECC and CSAM modules, which can model the global relational information between different classes.

4.3.2 Comparison results and analysis on sub-datasets

Furthermore, we conducted comparative experiments and analysis of our method and other methods on the three sub-datasets. As mentioned earlier, the training samples for the three sub-datasets differ, with dataset I having the most samples and dataset III having the least. We performed comparative analysis on all three datasets, with the quantitative results shown in Table 3 and the qualitative results displayed in Figures 9–11.

From the above qualitative and quantitative comparison results, it can be seen that overall, our method performs well across all three sub-datasets, achieving the best segmentation results on datasets II and III. On dataset I, our method achieves an mIoU of 75.16% and an OA of 88.66%, slightly behind PT (mIoU 75.24%, OA 89.98%), while the results of other methods are less impressive. This is largely because both our method and PT incorporate attention mechanisms. We use spatial-channel attention, while PT uses a transformer-based self-attention mechanism. These modules help capture more information when processing a larger sequence of

samples and more diverse scenes. In contrast, methods based purely on point-MLP or convolution tend to converge earlier and reach saturation. This is mainly because high-voltage power transmission scenes are complex and the samples are highly imbalanced. Methods like KPConv and PointNet++ can easily and quickly learn large samples (such as ground and vegetation), but they struggle to learn smaller classes, especially similar ones (such as conductor, crossing line, etc.). This further validates the rationale and effectiveness of the attention mechanisms in our method's design.

Meanwhile, in data sets II and III, our method achieves the best accuracy. Although PT and other methods perform similarly in terms of overall accuracy, their segmentation results for individual classes are not as strong as ours. The inclusion of the GECC module in our method allows it to model the attributes of different categories more effectively. Graph convolutions help emphasize differences in attributes such as relative elevation, orientation, and other features of similar objects. This enables our model to better distinguish between objects that, despite sharing similar local geometric properties, differ in global attributes and topological relationships.

4.3.3 Detailed comparison and analysis

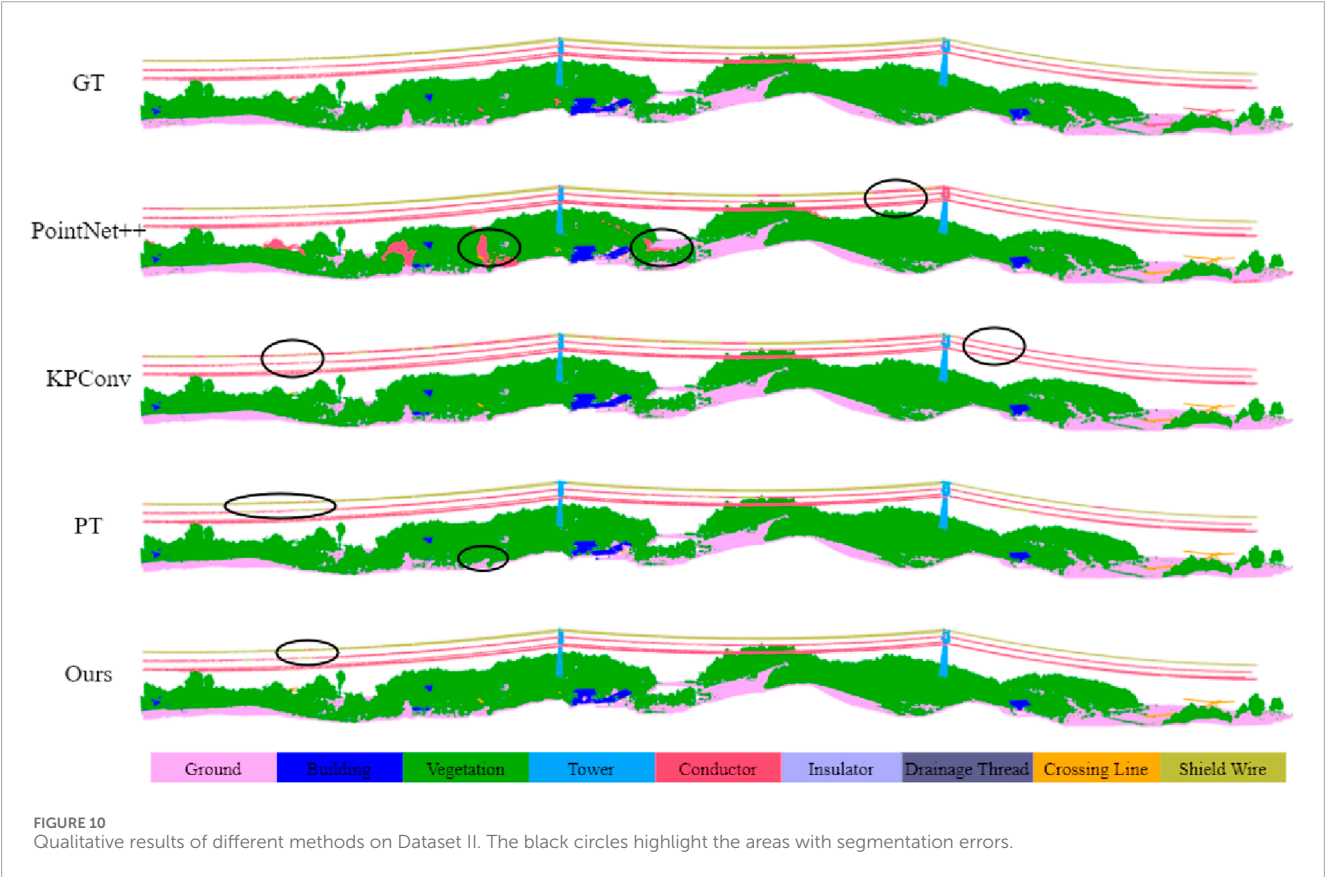
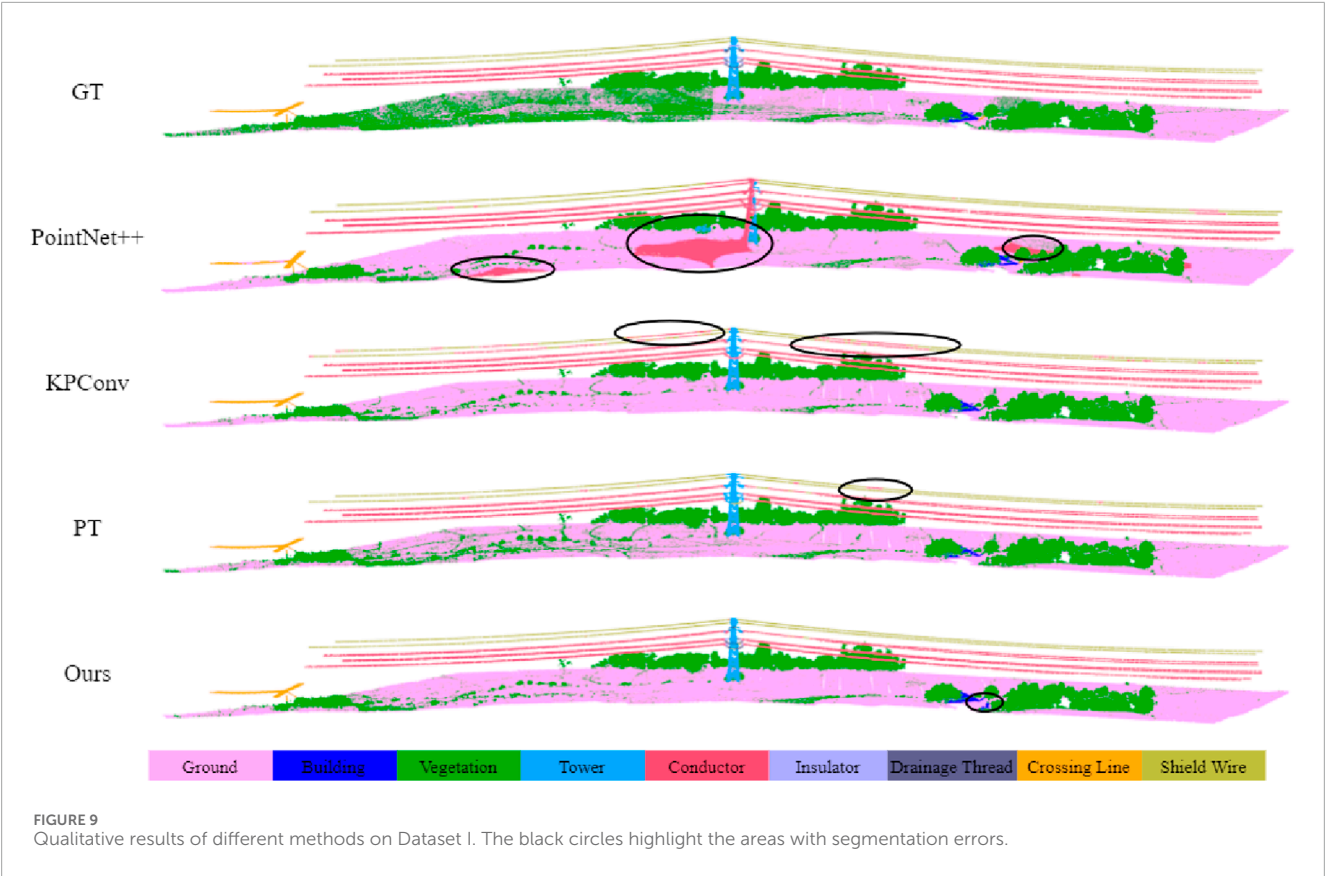
Considering that our task is the fine-grained semantic segmentation of high-voltage transmission lines, the segmentation of power facilities (especially tower areas) is a key focus. Therefore, we conduct a comparative analysis of the segmentation details of power transmission towers and their connected areas across the three subdatasets. As shown in Figure 12, the regions of power transmission towers and other power facilities are displayed in the three datasets. It can be seen that, although our method has some shortcomings in the segmentation of vegetation, it still outperforms other methods in the segmentation of various types of power lines, buildings and towers.

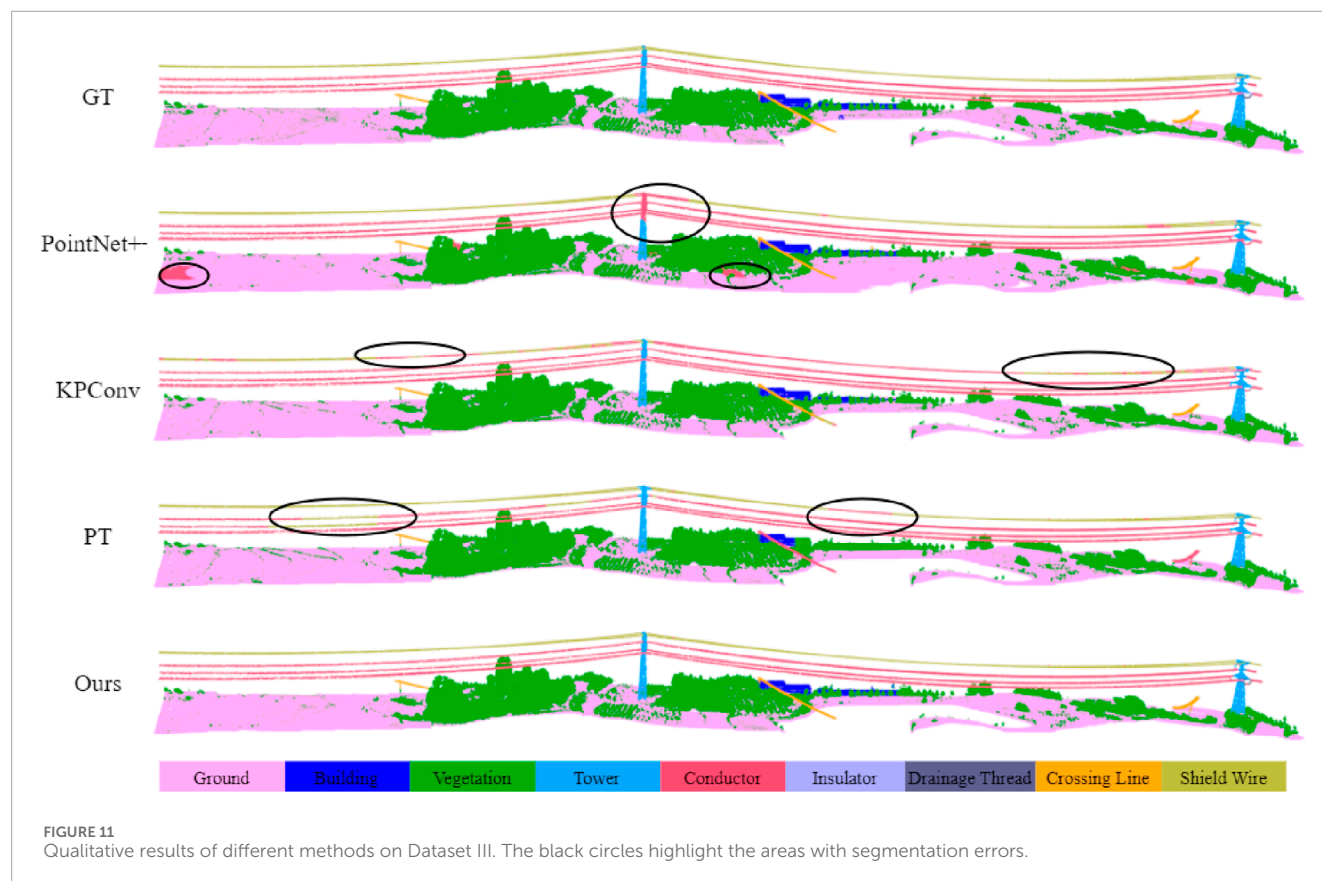
Specifically, as consistent with the previous comparative analysis, PointNet++ lacks effective segmentation of current objects, while KPConv cannot distinguish between similar linear targets of different categories. Although PT can differentiate them,

TABLE 3 Quantitative comparison results on the sub-datasets. The first four columns represent the different datasets, methods, and the mIoU and OA metrics. The remaining nine columns show the IoU values for each class for the different methods.

| Sub-datasets | Methods | mIoU | OA | IoU | | | | | | | | |
|--------------|--------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | | | Conductor | Tower | Vegetation | Building | Insulator | Drainage line | Crossing line | Shield wire | Ground |
| Dataset I | PointNet++ Qi et al. (2017) | 31.27% | 60.17% | 61.04% | 34.03% | 54.06% | 59.58% | 0.00% | 0.00% | 13.22% | 0.00% | 59.53% |
| | KPConv Thomas et al. (2019) | 59.71% | 76.53% | 80.04% | 91.39% | 66.69% | 89.93% | 17.45% | 0.00% | 64.43% | 60.70% | 66.73% |
| | PT Wu et al. (2024a) | 75.24% | 89.98% | 90.63% | 95.67% | 76.15% | 84.26% | 52.20% | 46.61% | 67.29% | 93.04% | 71.25% |
| | Ours | 75.16% | 88.66% | 90.45% | 94.47% | 75.53% | 85.77% | 50.30% | 45.61% | 67.39% | 97.86% | 69.04% |
| | PointNet++ Qi et al. (2017) | 29.18% | 70.18% | 43.72% | 30.04% | 45.70% | 62.43% | 0.00% | 0.00% | 0.00% | 18.40% | 62.32% |
| Dataset II | KPConv Thomas et al. (2019) | 61.34% | 92.35% | 67.48% | 95.80% | 96.73% | 87.59% | 18.96% | 0.00% | 54.69% | 38.85% | 91.99% |
| | PT Wu et al. (2024a) | 71.74% | 95.19% | 79.89% | 92.04% | 96.34% | 79.57% | 30.40% | 19.40% | 65.05% | 91.97% | 91.02% |
| | Ours | 78.71% | 97.65% | 81.96% | 96.62% | 96.93% | 87.87% | 34.80% | 44.21% | 75.87% | 97.70% | 92.45% |
| Dataset III | PointNet++ Qi et al. (2017) | 40.46% | 74.09% | 70.06% | 33.21% | 83.77% | 67.32% | 0.00% | 0.00% | 24.42% | 16.29% | 69.06% |
| | KPConv Thomas et al. (2019) | 64.91% | 95.11% | 79.02% | 91.90% | 90.89% | 66.24% | 20.89% | 18.90% | 87.98% | 36.34% | 92.02% |
| | PT Wu et al. (2024a) | 80.46% | 95.27% | 98.59% | 90.63% | 90.16% | 84.41% | 34.67% | 42.30% | 95.67% | 96.15% | 91.56% |
| | Ours | 88.47% | 96.30% | 99.05% | 96.13% | 92.12% | 93.32% | 69.16% | 58.01% | 96.53% | 98.62% | 93.27% |
| | | | | | | | | | | | | |

Bold values indicate the best-performing evaluation values in their respective columns.





its segmentation accuracy is insufficient, and its segmentation of buildings is less complete compared to our method and KPConv. Our method effectively segments power infrastructure and buildings, but its performance in the segmentation of cluttered vegetation is inadequate. This is primarily because, during feature extraction, the GECC module aggregates local geometric properties of the point cloud (such as linearity and relative height), and when combined with the neighborhood features obtained by the KPConv module, it strengthens the model's ability to learn artificial objects. However, its segmentation capability for natural objects is relatively weaker. Nonetheless, it is worth noting that our primary goal is to address the fine-grained segmentation of power facilities in high-voltage transmission corridor scenes, so this limitation can be temporarily overlooked.

4.4 Effectiveness analysis

4.4.1 Effectiveness of KPConv

In this section, we evaluate the effectiveness of the proposed Graph-Kernel Convolution Attention Encoder (GKCAE) through a series of ablation studies. To quantify the contribution of each key component, we systematically remove or replace specific modules and assess the resulting performance on the entire dataset. A central component of GKCAE is the Kernel Point Convolution (KPConv) module (Thomas et al., 2019), which serves as the backbone for point-wise convolution and local feature extraction. To investigate its role, we compare two commonly used variants: rigid

KPConv, which uses fixed kernel point positions, and deformable KPConv, where kernel positions are learned during training. The comparison reveals how kernel flexibility influences the network's ability to capture fine-grained geometric structures.

Table 4 presents the quantitative results comparing rigid and deformable variants of the KPConv module. As shown, the model employing deformable KPConv achieves superior performance, with an improvement of approximately 1.85% in mean Intersection over Union (mIoU) over its rigid counterpart. This performance gain can be attributed to the inherent complexity of transmission line scenes, which contain diverse object categories with significant scale variation. The rigid KPConv, constrained by fixed kernel point positions and a uniform receptive radius, lacks adaptability to local geometric structures. This limitation reduces its effectiveness in capturing small-scale components, particularly fine-grained power infrastructure distributed around transmission towers. In contrast, the deformable KPConv dynamically adjusts kernel point locations during training, allowing the convolutional operation to better align with local surface geometry. This flexibility enhances the network's capacity to extract discriminative features from irregular regions, thereby improving segmentation accuracy for critical components such as towers, conductors, and insulators.

4.4.2 Effectiveness of CSAM and GECC

To analyze the effectiveness of the CSAM and GECC modules, we progressively add these components and test the modified versions. Table 5 presents the results of our ablation experiments.

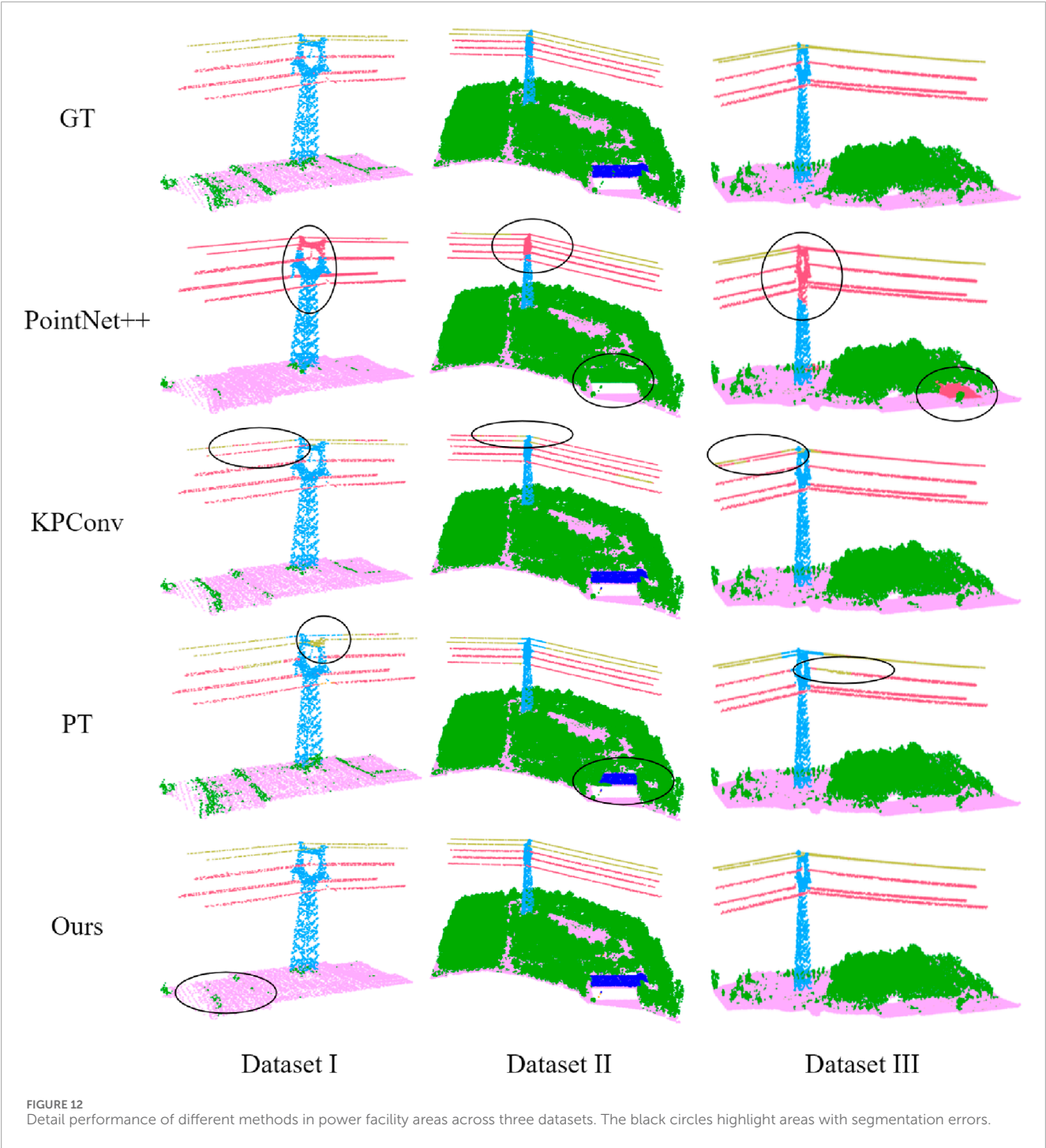


TABLE 4 Effectiveness Analysis of Rigid and Deformable Kernel Point Convolutions. "KP" represents the Kernel point convolution network.

| Methods | mIoU (%) | OA (%) |
|--------------------------|----------|--------|
| With rigid KP | 80.08 | 93.81 |
| With deformable KP(Ours) | 81.93 | 94.10 |

It can be observed that when key components are removed, both overall accuracy and performance metrics decline to varying degrees. In particular, when the GECC module is removed and only

the KPConv + CSAM combination is used, the mIoU decreases significantly. This indicates that the absence of the graph network negatively impacts the model's ability to recognize certain categories, especially for similar linear power line targets. This highlights the crucial role of the GECC module in capturing global relationships and enhancing segmentation accuracy.

Moreover, when the CSAM module is removed and only the KPConv and GECC modules are retained, we observe a slight increase in mIoU but a decrease in overall accuracy (OA). This phenomenon can be attributed to the following: in the absence of the attention mechanism, the network still retains its ability

TABLE 5 Effectiveness analysis of CSAM and GECC. “KP” represents the use of deformable Kernel point convolution network.

| Methods | mIoU | OA | IoU | | | | | | | | |
|-------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|-----------------|---------------|---------------|---------------|
| | | | Conductor | Tower | Vegetation | Building | Insulator | Drainage thread | Crossing line | Shield wire | Ground |
| KP + CSAM | 74.50% | 92.58% | 87.09% | 93.60% | 85.02% | 68.68% | 61.94% | 50.07% | 59.43% | 86.57% | 78.10% |
| KP + GECC | 77.47% | 90.62% | 88.12% | 91.13% | 83.92% | 71.46% | 68.17% | 61.96% | 67.34% | 94.68% | 70.45% |
| KP + CSAM + GECC (Ours) | 81.93% | 94.10% | 88.94% | 95.34% | 86.34% | 89.22% | 68.39% | 62.41% | 69.54% | 97.91% | 79.28% |

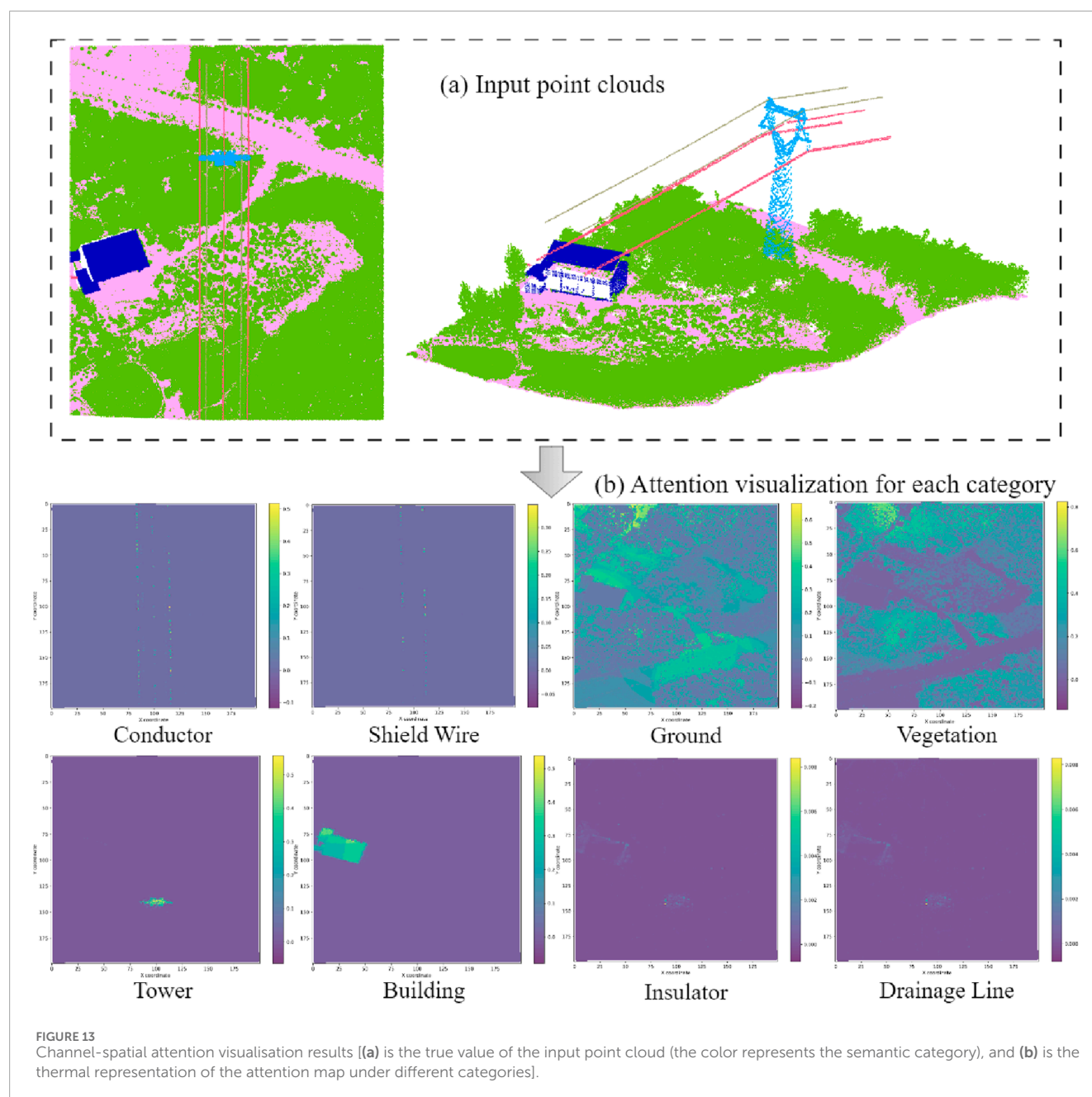
Bold values indicate the best-performing evaluation values in their respective columns.

to capture fine-grained geometric features through KPConv and GECC, which enhances per-class segmentation performance and leads to higher mIoU. However, the lack of global contextual modeling—particularly in the absence of spatial and channel-wise attention—reduces the model’s ability to distinguish dominant classes, such as vegetation, thereby lowering the OA. These results highlight the critical role of the CSAM module in enhancing global feature representation and improving class-level precision, especially for large-area categories. Although the model maintains a certain level of object recognition capability without CSAM, the segmentation accuracy is compromised due to the loss of global semantic context. This further validates the importance of integrating both GECC and CSAM modules: their combination yields the best performance in fine-grained semantic segmentation of complex transmission corridor scenes.

Furthermore, we output and visually display the attention matrix output by the CASM. As shown in [Figures 13a](#), it is the original 3D point cloud ground truth. The following figures respectively show the visualization results of the attention features of categories such as conductors, ground wires, the ground, transmission towers, and buildings. It can be seen that after the input point cloud undergoes feature learning through kernel point convolution and graph convolution, and then global point-level feature extraction is carried out by the channel-spatial attention module, it is able to effectively learn large targets (such as transmission towers, buildings, the ground, vegetation, etc.), and can also effectively learn and capture linear targets (such as conductors and ground wires). At the same time, for the learning of small targets (such as jumper wires and insulators), although the capture ability of learning is not as strong as that for large targets, the learning of these ground objects can still be achieved.

As shown in [Figure 13b](#), the ablation experiments further confirm the effectiveness and rationale behind the semantic segmentation network model we designed for high-voltage transmission line point clouds. In particular, the use of graph convolutional attention significantly enhances the model’s ability to segment and recognize similar categories of power line targets. This capability is crucial for achieving large-scale, rapid semantic understanding of transmission line scenes, enabling the extraction of detailed information about key components. Ultimately, this approach lays a solid foundation for downstream tasks, such as automated monitoring and maintenance, where precise identification of power line infrastructure is essential.

While the proposed GKCAE network demonstrates strong performance in fine-grained semantic segmentation of transmission line point clouds, several limitations remain. First, the model currently focuses on unimodal geometric and spatial features extracted from point clouds, and lacks the integration of other informative modalities such as RGB or thermal data. Future research can explore ([Cai et al., 2024](#)), such as token division strategies, to enhance feature diversity and robustness. Second, the graph structure in our method is constructed on relatively shallow geometric relationships. Advanced graph representations, such as [Jing et al. \(2025\)](#), offer an opportunity to model more complex and high-order semantic dependencies between points, potentially improving the segmentation of highly entangled or ambiguous objects like insulators and jumper wires. Moreover, our method primarily focuses on discriminative learning. Incorporating ([Xiang et al., 2025](#)) or topological regularization can further improve feature consistency among local regions and enhance structural



awareness. Additionally, for complex shape variations in small components, biomimetic representations like (Zhang and Chen, 2018) may offer novel geometric priors to better handle target deformation or noise. In future work, we plan to integrate these ideas to further enhance the generalization and interpretability of our model, particularly in large-scale, multi-source transmission line applications.

5 Conclusion

In this paper, we proposed a novel Graph-Attention-based Encoder Network (GKCAE) for fine-grained semantic segmentation of airborne LiDAR point cloud data in high-voltage transmission

corridor scenarios. The proposed model integrates Graph Edge-Conditioned Convolution (GECC) and a Channel-Spatial Attention Module (CSAM) to effectively capture both local geometric features and global contextual information, thereby addressing key challenges such as point cloud sparsity, severe class imbalance, and the high-precision requirements for segmenting power line components. Through comprehensive experiments on multiple real-world datasets, our method demonstrated superior performance compared to state-of-the-art (SOTA) techniques, particularly in the segmentation of critical transmission line elements such as conductors, towers, and various wire types. The proposed approach achieved a mean Intersection over Union (mIoU) of 81.93% and an Overall Accuracy (OA) of 94.1%, outperforming existing methods in both overall segmentation accuracy and class-wise

performance, especially for smaller and visually similar categories. Ablation studies further validated the effectiveness of each model component, showing that the integration of graph convolution and attention mechanisms significantly enhances the network's ability to distinguish between structurally similar power line elements. These findings confirm the model's suitability for fine-grained semantic segmentation in complex and cluttered corridor environments.

Despite these promising results, our method has several limitations. First, the current framework relies solely on geometric and spatial features from point clouds, without incorporating additional modalities such as RGB or infrared imagery, which could offer richer semantic cues. Second, the graph representation used is relatively shallow, limiting the model's capacity to capture higher-order semantic relationships among complex infrastructure components. Third, the absence of explicit topological or structural constraints may reduce feature consistency in highly cluttered scenes. Future work will focus on integrating multi-modal interaction mechanisms, hypergraph representations, and homology-based regularization to improve the robustness, generalization, and interpretability of the proposed framework across diverse transmission corridor environments.

In summary, the proposed GKCAE method provides a reliable and effective solution for intelligent perception and analysis in high-voltage transmission line scenarios. It demonstrates strong potential for supporting the automated monitoring and maintenance of power infrastructure. Future research will extend this work by exploring hazard detection based on fine-grained segmentation results and by evaluating model performance under varying sensor types and point cloud densities, with the aim of improving scalability and practical applicability in real-world inspection tasks.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SZ: Methodology, Writing – review and editing, Writing – original draft. HL: Writing – review and editing, Investigation,

Supervision, Data curation, Conceptualization. JR: Validation, Writing – review and editing, Funding acquisition, Visualization, Resources. YZ: Supervision, Writing – review and editing, Project administration, Methodology.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the project of State Grid Economic and Technological Research Institute (Research on Carbon Storage Estimation Technology for Power Transmission Corridors Based on Multi-Source LiDAR, 52440024000J).

Acknowledgments

The authors thank all the reviewers for their insightful comments.

Conflict of interest

Authors SZ, HL, JR, and YZ were employed by State Grid Economic and Technological Research Institute Co., LTD.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Cai, Y., Sui, X., Gu, G., and Chen, Q. (2024). Multi-modal interaction with token division strategy for rgb-t tracking. *Pattern Recognit.* 155, 110626. doi:10.1016/j.patcog.2024.110626
- Cao, D., Wang, C., Liu, H., Zhang, S., Du, M., Wang, P., et al. (2025). An automatic method for powerline extraction from als point cloud of powerline corridors. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 18, 10803–10829. doi:10.1109/JSTARS.2025.3555534
- Charles, R. Q., Su, H., Kaichun, M., and Guibas, L. J. (2017). "Pointnet: deep learning on point sets for 3d classification and segmentation," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017 (IEEE), 77–85. doi:10.1109/cvpr.2017.16
- Chen, C., Li, X., Belkacem, A. N., Qiao, Z., Dong, E., Tan, W., et al. (2019). The mixed kernel function svm-based point cloud classification. *Int. J. Precis. Eng. Manuf.* 20, 737–747. doi:10.1007/s12541-019-00102-3
- Fan, S., Dong, Q., Zhu, F., Lv, Y., Ye, P., and Wang, F.-Y. (2021). "Scf-net: learning spatial contextual features for large-scale point cloud segmentation," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021 (IEEE). doi:10.1109/cvpr46437.2021.01427
- Fei, R., Wan, Y., Hu, B., Li, A., Cui, Y., and Peng, H. (2025). Deep core node information embedding on networks with missing edges for community detection. *Inf. Sci.* 707, 122039. doi:10.1016/j.ins.2025.122039
- Feng, M., Yan, C., Wu, Z., Dong, W., Wang, Y., and Mian, A. (2025). "History-enhanced 3d scene graph reasoning from rgb-d sequences," in *IEEE transactions on circuits and systems for video technology* (IEEE Press).
- Gahremani, M., Williams, K., Corke, F. M., Tiddeman, B., Liu, Y., and Doonan, J. H. (2021). Deep segmentation of point clouds of wheat. *Front. Plant Sci.* 12, 608732. doi:10.3389/fpls.2021.608732

- Grothum, O., Bienert, A., Bluemlein, M., and Eltner, A. (2023). "Using machine learning techniques to filter vegetation in colorized sfm point clouds of soil surfaces" in *The international archives of the photogrammetry, remote sensing and spatial information sciences XLVIII-1/W2-2023*, 163–170. doi:10.5194/isprs-archives-xlviii-1-w2-2023-163-2023
- Han, Y., Zhang, G., and Zhang, R. (2023). Grid graph-based large-scale point clouds registration. *Int. J. Digital Earth* 16, 2448–2466. doi:10.1080/17538947.2023.2228298
- Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., et al. (2020). "Randla-net: efficient semantic segmentation of large-scale point clouds," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020 (IEEE). doi:10.1109/cvpr42600.2020.01112
- Huang, Y., Du, Y., and Shi, W. (2021). Fast and accurate power line corridor survey using spatial line clustering of point cloud. *Remote Sens.* 13, 1571. doi:10.3390/rs13081571
- Huang, S., Hu, Q., Zhao, P., Li, J., Ai, M., and Wang, S. (2024). Als point cloud semantic segmentation based on graph convolution and transformer with elevation attention. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 17, 2877–2889. doi:10.1109/jstars.2023.3347224
- Hui, C., Tingting, W., Zuoxiao, D., Weibin, L., and Menhas, M. I. (2021). "Power equipment segmentation of 3d point clouds based on geodesic distance with k-means clustering," in 2021 6th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 17–20 September 2021 (IEEE), 317–321. doi:10.1109/icpre52634.2021.9635211
- Jiang, S., Guo, W., Fan, Y., and Fu, H. (2022). *Fast semantic segmentation of 3D lidar point cloud based on random forest method*. Singapore: Springer Nature, 415–424. doi:10.1007/978-981-19-2580-1_35
- Jing, W., Zhang, W., Di, D., Li, C., Emam, M., and Mian, A. (2025). Hypergraph biformer for semantic segmentation of high-resolution remote sensing images. *IEEE Trans. Geoscience Remote Sens.* 63, 1–15. doi:10.1109/tgrs.2025.3543556
- Lai, X., Liu, J., Jiang, L., Wang, L., Zhao, H., Liu, S., et al. (2022). "Stratified transformer for 3d point cloud segmentation," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022 (IEEE). doi:10.1109/cvpr52688.2022.00831
- Landrieu, L., and Obozinski, G. (2017). Cut pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs. *SIAM J. Imaging Sci.* 10, 1724–1766. doi:10.1137/17m1113436
- Landrieu, L., and Simonovsky, M. (2018). "Large-scale point cloud semantic segmentation with superpoint graphs," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018 (IEEE), 4558–4567. doi:10.1109/cvpr.2018.00479
- Lawin, F. J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F. S., and Felsberg, M. (2017). *Deep projective 3D semantic segmentation*. Springer International Publishing, 95–107. doi:10.1007/978-3-319-64689-3_8
- Liao, L., Tang, S., Liao, J., Li, X., Wang, W., Li, Y., et al. (2022). A supervoxel-based random forest method for robust and effective airborne lidar point cloud classification. *Remote Sens.* 14, 1516. doi:10.3390/rs14061516
- Lin, Y., Vosselman, G., Cao, Y., and Yang, M. Y. (2021). Local and global encoder network for semantic segmentation of airborne laser scanning point clouds. *ISPRS J. Photogrammetry Remote Sens.* 176, 151–168. doi:10.1016/j.isprsjprs.2021.04.016
- Liu, Z., Tang, H., Lin, Y., and Han, S. (2019). "Point-voxel cnn for efficient 3d deep learning," in Conference on Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, December 8–14, 2019.
- Liu, K., Feng, M., Zhao, W., Sun, J., Dong, W., Wang, Y., et al. (2025). Pixel-level noise mining for weakly supervised salient object detection. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–15. doi:10.1109/tnnls.2025.3575255
- Ma, C., Mu, R., Li, M., He, J., Hua, C., Wang, L., et al. (2025). A multi-scale spatial-temporal interaction fusion network for digital twin-based thermal error compensation in precision machine tools. *Expert Syst. Appl.* 286, 127812. doi:10.1016/j.eswa.2025.127812
- Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y., and Behnood, A. (2022). 3d point cloud data processing with machine learning for construction and infrastructure applications: a comprehensive review. *Adv. Eng. Inf.* 51, 101501. doi:10.1016/j.aei.2021.101501
- Ni, H., Lin, X., and Zhang, J. (2017). Classification of als point cloud with improved point cloud segmentation and random forests. *Remote Sens.* 9, 288. doi:10.3390/rs9030288
- Nurunnabi, A., Belton, D., and West, G. (2012). "Robust segmentation in laser scanning 3d point cloud data," in 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), Fremantle, WA, Australia, 03–05 December 2012 (IEEE), 1–8. doi:10.1109/dicta.2012.6411672
- Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). "Pointnet++: deep hierarchical feature learning on point sets in a metric space," in *Proceedings of the 31st international conference on neural information processing systems* (Red Hook, NY, USA: Curran Associates Inc), 5105–5114.
- Rejichi, S., and Chaabane, F. (2015). "Feature extraction using pca for vhr satellite image time series spatio-temporal classification," in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015 (IEEE), 485–488. doi:10.1109/igarss.2015.7325806
- Robert, D., Raguet, H., and Landrieu, L. (2023). "Efficient 3d semantic segmentation with superpoint transformer," in 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 01–06 October 2023 (IEEE), 17149–17158. doi:10.1109/iccv51070.2023.01577
- Sha, X., Guan, Z., Wang, Y., Han, J., Wang, Y., and Chen, Z. (2025). Ssc-net: a multi-task joint learning network for tongue image segmentation and multi-label classification. *Digit. Health* 11, 20552076251343696. doi:10.1177/20552076251343696
- Shen, Y., Huang, J., Chen, D., Wang, J., Li, J., and Ferreira, V. (2023). An automatic framework for pylon detection by a hierarchical coarse-to-fine segmentation of powerline corridors from uav lidar point clouds. *Int. J. Appl. Earth Observation Geoinformation* 118, 103263. doi:10.1016/j.jag.2023.103263
- Shen, Y., Huang, J., Wang, J., Jiang, J., Li, J., and Ferreira, V. (2024). A review and future directions of techniques for extracting powerlines and pylons from lidar point clouds. *Int. J. Appl. Earth Observation Geoinformation* 132, 104056. doi:10.1016/j.jag.2024.104056
- Shi, B., Yang, M., Liu, J., Han, B., and Zhao, K. (2023). Rail transit shield tunnel deformation detection method based on cloth simulation filtering with point cloud cylindrical projection. *Tunn. Undergr. Space Technol.* 135, 105031. doi:10.1016/j.tust.2023.105031
- Simonovsky, M., and Komodakis, N. (2017). "Dynamic edge-conditioned filters in convolutional neural networks on graphs," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017 (IEEE), 29–38. doi:10.1109/cvpr.2017.11
- Song, X., Han, D., Chen, C., Shen, X., and Wu, H. (2025). Vman: visual-modified attention network for multimodal paradigms. *Vis. Comput.* 41, 2737–2754. doi:10.1007/s00371-024-03563-4
- Su, C., Wu, X., Guo, Y., Lai, C. S., Xu, L., and Zhao, X. (2022). "Automatic multi-source data fusion technique of powerline corridor using uav lidar," in 2022 IEEE International Smart Cities Conference (ISC2), Pafos, Cyprus, 26–29 September 2022 (IEEE). doi:10.1109/isc255366.2022.9922293
- Tang, Q., Zhang, L., Lan, G., Shi, X., Duanmu, X., and Chen, K. (2023). A classification method of point clouds of transmission line corridor based on improved random forest and multi-scale features. *Sensors* 23, 1320. doi:10.3390/s23031320
- Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., and Guibas, L. (2019). "Kpconv: flexible and deformable convolution for point clouds," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 27 October 2019 – 02 November 2019 (IEEE), 6410–6419. doi:10.1109/iccv.2019.00651
- Thomas, H., Tsai, Y.-H. H., Barfoot, T. D., and Zhang, J. (2024). "Kpconvx: modernizing kernel point convolution with kernel attention," in 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024 (IEEE), 5525–5535. doi:10.1109/cvpr52733.2024.00528
- Wang, L., Huang, Y., Hou, Y., Zhang, S., and Shan, J. (2019). "Graph attention convolution for point cloud semantic segmentation," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019 (IEEE), 10288–10297. doi:10.1109/cvpr.2019.01054
- Wang, G., Wang, L., Wu, S., Zu, S., and Song, B. (2023). Semantic segmentation of transmission corridor 3d point clouds based on ca-pointnet++. *Electronics* 12, 2829. doi:10.3390/electronics12132829
- Wang, Z., Yan, Z., Li, S., and Liu, J. (2025). Indvisg: vlm-based scene graph generation for industrial spatial intelligence. *Adv. Eng. Inf.* 65, 103107. doi:10.1016/j.aei.2024.103107
- Wen, C., Yang, L., Li, X., Peng, L., and Chi, T. (2020). Directionally constrained fully convolutional neural network for airborne LiDAR point cloud classification. *ISPRS J. Photogrammetry Remote Sens.* 162, 50–62. doi:10.1016/j.isprsjprs.2020.02.004
- Wen, C., Li, X., Yao, X., Peng, L., and Chi, T. (2021). Airborne LiDAR point cloud classification with global-local graph attention convolution neural network. *ISPRS J. Photogrammetry Remote Sens.* 173, 181–194. doi:10.1016/j.isprsjprs.2021.01.007
- Wu, X., Lao, Y., Jiang, L., Liu, X., and Zhao, H. (2022). *Point transformer v2: grouped vector attention and partition-based pooling*. Red Hook, NY: NeurIPS.35, 33330–33342
- Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., et al. (2024a). Point transformer v3: Simpler faster stronger. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Los Alamitos, CA: CVPR, 4840–4851.
- Wu, X., Tian, Z., Wen, X., Peng, B., Liu, X., Yu, K., et al. (2024b). Towards large-scale 3d representation learning with multi-dataset point prompt training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: CVPR, 19551–19562.
- Xiang, B., Peters, T., Kontogianni, T., Vetterli, F., Puliti, S., Astrup, R., et al. (2023). Towards accurate instance segmentation in large-scale lidar point clouds. *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.* X-1/W1-2023, 605–612. doi:10.5194/isprs-annals-x-1-w1-2023-605-2023
- Xiang, D., He, D., Sun, H., Gao, P., Zhang, J., and Ling, J. (2025). Hcmpe-net: an unsupervised network for underwater image restoration with multi-parameter

estimation based on homology constraint. *Opt. and Laser Technol.* 186, 112616. doi:10.1016/j.optlastec.2025.112616

Xin, B., Sun, J., Bartholomeus, H., and Kootstra, G. (2023). 3d data-augmentation methods for semantic segmentation of tomato plant parts. *Front. Plant Sci.* 14, 1045545. doi:10.3389/fpls.2023.1045545

Xu, J., Zhang, R., Dou, J., Zhu, Y., Sun, J., and Pu, S. (2021). "Rpvnet: a deep and efficient range-point-voxel fusion network for lidar point cloud segmentation," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021 (IEEE), 16004–16013. doi:10.1109/iccv48922.2021.01572

Yang, L., Fan, J., Liu, Y., Li, E., Peng, J., and Liang, Z. (2020). A review on state-of-the-art power line inspection techniques. *IEEE Trans. Instrum. Meas.* 69, 9350–9365. doi:10.1109/tim.2020.3031194

Yin, F., Huang, Z., Chen, T., Luo, G., Yu, G., and Fu, B. (2023). Dcnet: Large-scale point cloud semantic segmentation with discriminative and efficient feature aggregation. *IEEE Trans. Circuits Syst. Video Technol.* 33, 4083–4095. doi:10.1109/tcsvt.2023.3239541

Yu, H., Wang, Z., Zhou, Q., Ma, Y., Wang, Z., Liu, H., et al. (2023). Deep-learning-based semantic segmentation approach for point clouds of extra-high-voltage transmission lines. *Remote Sens.* 15, 2371. doi:10.3390/rs15092371

Zafar, B., Ashraf, R., Ali, N., Ahmed, M., Jabbar, S., Naseer, K., et al. (2018). Intelligent image classification-based on spatial weighted histograms of concentric circles. *Comput. Sci. Inf. Syst.* 15, 615–633. doi:10.2298/CSIS180105025Z

Zhang, H. H., and Chen, R. S. (2014). Coherent processing and superresolution technique of multi-band radar data based on fast sparse bayesian learning algorithm. *IEEE Trans. Antennas Propag.* 62, 6217–6227. doi:10.1109/tap.2014.2361158

Zhang, H.-H., and Chen, P.-Y. (2018). Biomimetic radar target recognition based on hypersausage chains. *Appl. Comput. Electromagn. Soc. J. (ACES)* 33, 1429–1438.

Zhang, J., Lin, X., and Ning, X. (2013). Svm-based classification of segmented airborne lidar point clouds in urban areas. *Remote Sens.* 5, 3749–3775. doi:10.3390/rs5083749

Zhang, L., Wang, J., Shen, Y., Liang, J., Chen, Y., Chen, L., et al. (2022). A deep learning based method for railway overhead wire reconstruction from airborne lidar data. *Remote Sens.* 14, 5272. doi:10.3390/rs14205272

Zhao, H., Jiang, L., Jia, J., Torr, P., and Koltun, V. (2020). *Point transformer*. doi:10.48550/ARXIV.2012.09164

Zhao, W., Dong, Q., and Zuo, Z. (2023). 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. Digital Earth* 16, 620–644. doi:10.1080/17538947.2023.2168770

Zhou, F., Wen, G., Ma, Y., Pan, H., Wang, G., and Wang, Y. (2024). Spatial attention-based kernel point convolution network for semantic segmentation of transmission corridor scenarios in airborne laser scanning point clouds. *Electronics* 13, 4501. doi:10.3390/electronics13224501