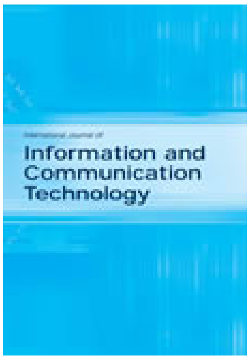


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

PointNet++ model based on directional attention for insulators segmentation

Shucai Li, Enliang Chu, Pengcheng Wang, Xi Gao, Na Zhang

Article History:

Received:	11 October 2024
Last revised:	06 November 2024
Accepted:	06 November 2024
Published online:	20 January 2025

Copyright © 2025 Inderscience Enterprises Ltd.

PointNet++ model based on directional attention for insulators segmentation

Shucai Li, Enliang Chu,
Pengcheng Wang and Xi Gao

State Grid Shanxi Lvliang Power Supply Company,
Lvliang, Shanxi, 033000, China
Email: louyimei@ldy.edu.rs
Email: lehaia@ldy.edu.rs
Email: qiuqinwei@ldy.edu.rs
Email: zhufeiyao@ldy.edu.rs

Na Zhang*

State Grid Shanxi Electric Power Research Institute,
Taiyuan, Shanxi, 030001, China
Email: xiedongnan@ldy.edu.rs
*Corresponding author

Abstract: Insulators are crucial components for the safety and reliability of power systems. However, due to their small size and complex structure, precise segmentation of insulators is challenging. To address this issue, this paper proposes a directional attention-based PointNet++ model (PDA). The core module of PDA is the directional attention (DA) module, which consists of spatial self-attention (SSA) and channel self-attention (CSA). This module is designed to establish long-range relationships in both spatial and channel directions of the feature map, enabling global modelling. Additionally, to reduce computational costs, multi-scale pyramid pooling is embedded in both the SSA and CSA modules. Notably, by integrating DA into PointNet++, the model enhances the correlation between point cloud features and the long-range dependency of positional information without significantly increasing the computational burden. Experimental results demonstrate that the PDA model significantly outperforms existing models in segmenting insulator point clouds from multiple power transmission corridors.

Keywords: PointNet++; PDA; spatial self-attention; SSA; channel self-attention; CSA; insulators.

Reference to this paper should be made as follows: Li, S., Chu, E., Wang, P., Gao, X. and Zhang, N. (2025) 'PointNet++ model based on directional attention for insulators segmentation', *Int. J. Information and Communication Technology*, Vol. 26, No. 1, pp.1–22.

Biographical notes: Shucai Li received his Master's degree from Huazhong University of Science and Technology in 2017. His main research direction is intelligent operation and inspection of substation equipment.

Copyright © The Author(s) 2024. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Enliang Chu received his Master's degree from Shaanxi University of Science and Technology in 2020. His main research direction is intelligent operation and inspection of substation equipment.

Pengcheng Wang received his Master's degree from Shanghai Maritime University in 2020. His main research direction is intelligent operation and inspection of power transformation equipment.

Xi Gao received his Bachelor's degree from Shanxi University in 2013. He worked in State Grid Shanxi Lvliang Power Supply Company. His research interests include power equipment testing technology.

Na Zhang has a Master's degree and is a Senior Engineer. She worked in State Grid Shanxi Electric Power Research Institute. Her research interests include visual intelligent diagnosis technology for power equipment.

1 Introduction

In contemporary electrical networks, guaranteeing the secure and dependable operation of the grid is vital for preserving the smooth functioning of society. As integral elements of the power network, the precise detection and efficient handling of insulators are critical for avoiding malfunctions and enhancing system stability. However, owing to the compact size and intricate design of insulator point clouds, current segmentation techniques encounter numerous difficulties.

Traditional methods typically rely on manual inspections and the use of devices such as laser rangefinders to measure the exact position of power lines. These methods are not only inefficient but also pose certain safety risks.

Insulators influence the performance of voltage drop and leakage current in distribution systems (Murthy et al., 2011) and ensuring the safety of power systems remains a core task. Traditional grid inspection methods typically rely on professionals manually inspecting the lines and using devices such as laser rangefinders and optical theodolites to measure the exact position of power lines (Yang et al., 2020; Ellis, 2013). However, this method poses certain safety risks during inspections, is inefficient, and fails to meet the requirements of modern power system management (Bayindir et al., 2016). With advancements in technology, point cloud and drone technologies are increasingly playing a key role in power inspections (Mohsan et al., 2022). By equipping drones with sensors like LiDAR, it is possible to quickly obtain 3D point cloud data of power equipment, providing detailed spatial information for inspections (Guan et al., 2021). Although 3D data is rich, it is often irregular, uneven, and unordered, making the accurate identification of insulators within point clouds a major challenge (Zhang et al., 2019b).

In the field of intelligent monitoring and management of transmission corridors, LiDAR remote sensing technology has garnered significant attention from researchers due to its high precision and broad applicability (Matikainen et al., 2016). The 3D point cloud data obtained through laser scanning not only allows for the precise identification of power transmission lines and tower structures but also enables the monitoring of vegetation growth height and distribution density, especially in complex terrains such as tropical rainforests (Matikainen et al., 2016). Vegetation grows rapidly and is subject to varying climatic conditions, posing a threat to transmission lines. In response, Taylor et al. (2022) proposed a dynamic modelling method based on forest growth simulation to predict vegetation expansion trends. By integrating real-time monitoring data, this model provides decision support for vegetation management in transmission corridors. Implementing protective measures in advance can effectively prevent trees from coming into contact with power lines, thereby ensuring the safe operation of the transmission system (Zhang et al., 2019b).

These methods can effectively identify power lines and towers based on the high similarity of their geometric shapes and successfully separate them from surrounding vegetation. The underlying principle involves utilising the local spatial topological relationships within the point cloud data to achieve more precise object recognition (Chen et al., 2018) Moreover, Li and Zhang highlighted that clustering algorithms based on local distribution features can improve the differentiation of various target objects in complex terrain, thereby enhancing the efficiency and accuracy of data processing (Zhang et al., 2019b).

As data scale and complexity grow, traditional methods encounter limitations when processing large-scale point cloud data. To address this, Poux and Billen (2019) proposed a voxel-based deep learning classification algorithm. By dividing the point cloud data into three-dimensional voxels for analysis, the model can capture nonlinear features, significantly improving classification accuracy (Zhang et al., 2020). Building on this, Fryskowska (2019) combined the random sample consensus (RANSAC) algorithm with supervised learning, proposing an improved model capable of handling noisy data. This method not only enhanced the recognition accuracy of power lines and towers but also improved the ability to accurately position power lines in cases of vegetation occlusion.

To further improve model performance, Faustino (2022) proposed a point cloud classification method that integrates global and local feature fusion, leveraging ensemble learning techniques such as JointBoost and random forest classifiers. These methods can not only handle multi-dimensional features but also effectively identify power lines, towers, and vegetation in complex forest environments (Cao et al., 2022). By employing this multi-feature fusion approach, the robustness of point cloud classification is enhanced, enabling the model to perform exceptionally well across various terrains and climatic conditions (Peng et al., 2019). Their research demonstrates that combining the macro-level perspective of global features with the detailed capture of local features significantly improves vegetation monitoring and safety management in transmission corridors.

In the field of deep learning, classification and segmentation techniques for three-dimensional point cloud data have developed rapidly (Zhang et al., 2019a). Traditional two-dimensional convolutional neural networks (CNNs) face limitations when processing 3D data, as they require voxelisation of point clouds, often leading to reduced resolution and increased memory usage. In response, researchers have developed various direct and indirect methods to optimise the processing of point cloud data (Pastucha et al., 2020).

Su et al. (2015) introduced the multi-view convolutional neural network (MVCNN), a representative indirect approach. This method projects three-dimensional point cloud data onto multiple two-dimensional views, which are then processed by a 2D CNN. The resulting features from these views are subsequently merged for classification. MVCNN captures rich features within the point cloud by combining multi-view information (Wang et al., 2024b). Although this method has shown success, challenges remain in selecting optimal viewpoints and avoiding information loss, particularly when processing high-resolution data (Zhou et al., 2019).

To directly handle three-dimensional point cloud information, Qi et al. (2017a, 2017b) proposed the pioneering PointNet framework. PointNet was the first deep neural network capable of processing unvoxelised point cloud data by extracting global characteristics and applying symmetric operations to each point, successfully tackling the unordered structure of point clouds. While the model performs well on simple classification tasks, it struggles to capture local details in complex geometric structures. To address this limitation, Qi and colleagues later developed the PointNet++ model, which combines local feature extraction with global feature aggregation. PointNet++ enhances classification and segmentation in complex scenes by progressively extracting local geometric information through hierarchical local region grouping. However, despite its improvements, PointNet++ incurs substantial computational costs when processing high-density point clouds, and challenges remain in handling irregular and sparse regions.

To address these constraints, Maturana and Scherer (2015) introduced voxel-based architectures such as VoxNet. This method converts point cloud data into uniform voxel grids, allowing the application of three-dimensional CNNs for analysis. Although VoxNet can capture spatial structural information using 3D convolutions, voxelisation leads to reduced resolution and high memory consumption, limiting its scalability. To overcome these issues, researchers have introduced more efficient convolution methods that directly process point cloud data.

For example, Li et al. (2018) presented the PointCNN model, which enhances classification and segmentation by learning convolutional relationships between points and organising the point cloud in an ordered manner. Wang et al. (2019) introduced the dynamic graph convolutional neural network (DGCNN), which dynamically builds adjacency graphs within point clouds and applies convolution operations on the graph structure, effectively capturing local spatial dependencies among points. Compared to the PointNet series, DGCNN demonstrates significant improvements in capturing local geometric features.

More advanced architectures, such as ShellNet, proposed by Zhang et al. (2019c), process point clouds in hierarchical levels to capture multi-dimensional local spatial configurations, thereby enhancing the accuracy of point cloud classification and segmentation.

To enhance the learning of local details, Xu et al. (2021) proposed PAConv, which constructs convolution kernels using dynamically combined weight matrices, leading to improved segmentation results. To further capture fine-grained local features in point clouds, recent studies have incorporated local self-attention mechanisms into deep learning models. For example, Yan et al. (2020) proposed PointASNL, which integrates local self-attention with spatial non-local mechanisms, enabling adaptive adjustments of features for each point in the point cloud. Li et al. (2024a) utilised multi-scale feature extraction, established a local coordinate framework, and incorporated an offset attention mechanism to further boost the model's performance. Guo et al. (2023) merged a partitioning head architecture with self-attention and transformers (HST-Net), efficiently addressing the segmentation of damaged regions such as insulator degradation and electrical arcing. Xuan et al. (2022) employed a SAG-Mask with spatial attention strategies to extract insulator mask images, enhancing the precision and stability of the algorithm. Zeng et al. (2022) introduced a multi-tier attention framework structured as a feature pyramid, boosting the semantic and intricate details of feature maps at various resolutions. Su et al. (2021) proposed DLA-Net, which more precisely models fine-grained local features in point clouds by integrating local self-attention mechanisms.

Despite significant advancements in existing methods, several challenges remain. For example, the segmentation focus of PointNet++ has primarily been on visible transmission lines and towers, with less attention given to insulators, which are crucial components of transmission lines. Additionally, during the segmentation of insulator point clouds, challenges such as similarity and occlusion often arise, making it difficult to accurately classify insulators. Variations in point cloud density across different categories can also complicate the separation of points at boundaries. PointNet++ records feature values during the downsampling process in the encoding layer but does not fully utilise the positional information between different point clouds.

To address these limitations, we propose an improved PointNet++ model by embedding a directional attention module (DAM) to enhance feature correlations between point clouds, allowing for better extraction of local positional features. Similarly, Ji and Zhong (2024) proposed a bidirectional attention module that reasonably allocates weights among different features and focuses on the relationships between adjacent pixels of the same feature, thereby enhancing the segmentation capability. The proposed PDA model accurately segments point clouds of transmission lines, towers, ground, and insulators by utilising contextual information, with the DA module refining the semantic information of insulator point clouds.

Building on the previous discussion, our key contributions are as follows:

1 We introduce a PointNet++ model based on directional attention that effectively utilises the spatial relationships between point clouds and neighbouring areas by incorporating a DAM. This module suppresses irrelevant information using contextual cues, significantly improving the segmentation accuracy of insulator point clouds.

2 The suggested model is utilised for the semantic segmentation of insulators on transmission line pylons, effectively isolating the insulators from the pylons, thereby proving the efficacy of the PDA model.

3 We performed thorough experimental evaluations between the PDA model and other current models in the context of transmission line pylons, highlighting the enhanced performance of our model in this scenario 2.

2 Related work

2.1 Insulator semantic segmentation based on point cloud

In recent years, with the growing demand for inspection and maintenance in power systems, the technology for extracting insulators based on images and point clouds (3D spatial data obtained through LiDAR) has gained significant attention as a research hotspot. Insulators, being essential elements of the electrical transmission network, have a direct influence on the security and reliability of the grid. Therefore, accurately and efficiently extracting insulators in complex power scenarios has become a key topic in electrical research. This challenge has driven the development of various solutions, particularly through advancements in image processing and LiDAR point cloud technologies, leading to notable progress in this area (Zhang et al., 2019b). Several review studies have repeatedly highlighted the importance and complexity of insulator extraction (Miao et al., 2019).

Extracting insulators from sources like optical overhead images and infrared imaging has emerged as a key research focus. Optical images provide high-resolution data, while thermal imaging can reveal potential defects in insulators. However, these methods often face challenges such as data sparsity (uneven data coverage) and occlusion in complex environments. To overcome these challenges, combining multiple data sources and processing methods, especially using LiDAR point cloud technology, has proven to be an effective solution (Sohn et al., 2012; Ma et al., 2021).

Point cloud-based extraction methods for electrical components typically focus on large-scale objects with prominent features, such as power lines and towers. For instance, Arastounia and Lichti proposed a point cloud extraction method that combines prior knowledge of the main directional information, which was applied to insulator identification in substation environments (Zhang et al., 2019b). This method achieved high precision in insulator extraction by leveraging structured scene understanding. In transmission corridors. Their method improved the efficiency and accuracy of insulator extraction by identifying cylindrical segments (Sohn et al., 2012; Tao et al., 2018).

Additionally, Zhang et al. (2019a) determined the position of guy insulators by identifying points at the ends of power lines and the centres of towers (Tao et al., 2018). However, many of these methods heavily rely on the accuracy of power line extraction and often overlook the impact of different insulator types, variations in tower shapes, and functions (Ma et al., 2021). This limitation reduces the applicability of existing methods in complex power scenarios, especially when dealing with varying voltage levels, wire connection methods, and jumper distributions (Jenssen and Roverso, 2018).

These algorithmic studies also tend to neglect the differences between various components of transmission towers, such as shape, size, occlusion, and proximity to boundaries. In this research, we incorporate directional attention into the PointNet++ model, taking into full account the diverse shapes, sizes, types, and scenarios of different insulators. The introduced directional attention (PDA) not only improves the segmentation precision of insulator point clouds but also offers more dependable technical assistance for the smart management and upkeep of electrical equipment (Sun et al., 2019; Wu et al., 2018). Moreover, Tang et al. (2023) suggested a reliable extraction approach that combines multi-scale neighbourhoods and multi-feature entropy weighting to segment entire insulator chains. Lv et al. (2023) proposed a PointNet-MLS hybrid framework, which successfully performs component segmentation of insulating devices in challenging backgrounds. Sun et al. (2024) created the Insulator Segmentation Network (ISNet), which improves the segmentation efficiency of insulators.

2.2 Application of attention mechanism in semantic segmentation of insulator point cloud

The attention mechanism was initially proposed by Bahdanau (2014), when it was integrated into neural machine translation tasks. This enabled models to dynamically allocate importance to various segments of the input sequence, allowing the model to concentrate on the most pertinent parts of the input during translation generation. This approach effectively mitigated the issues of information compression and loss that were common in traditional encoder-decoder architectures. The introduction of this mechanism led to significant improvements in sequence-to-sequence models across various natural language processing tasks.

In 2017, Vaswani et al. revolutionised this concept by proposing the Transformer architecture, which is based entirely on self-attention mechanisms. Unlike RNNs, which process sequences sequentially, the Transformer eliminated the need for sequential dependency, processing the entire input sequence simultaneously (Vaswani et al., 2017). This not only enhanced computational efficiency but also improved the model's ability to capture global context, quickly making it a dominant technology across multiple fields. In image processing, Wang et al. (2018) introduced non-local neural networks, which leveraged a non-local self-attention mechanism to establish connections between distant pixels, significantly benefiting large-scale image tasks.

In the realm of point cloud segmentation, Qi et al. (2017a) made early advancements with the introduction of PointNet in 2017. PointNet could directly process irregular point cloud data by extracting features from individual points and performing global feature fusion for classification and segmentation. However, the model lacked the ability to capture local geometric structures and dependencies between points. To address this limitation, Qi et al. (2017b) subsequently proposed PointNet++, which improved segmentation accuracy by introducing sampling and multi-scale feature extraction, allowing the model to capture finer details in local regions.

Building on this progress, Wang et al. introduced DGCNN (dynamic graph CNN) in 2019 (Wang et al., 2019). This approach constructed dynamic adjacency graphs for point clouds, enabling the network to learn the evolving relationships between points and better reflect changes in local structures. DGCNN demonstrated exceptional performance in multiple 3D segmentation tasks and proved particularly adaptable to handling complex geometric scenes.

In the specific context of insulator segmentation, Arastounia and Lichti (2013) proposed a method based on 3D point cloud data that extracted insulators' geometric features in 3D space to address the challenges of insulator segmentation in complex environments. Wang et al. (2023) integrated the coordinate attention (CA) module with PointNet++ to form CA-PointNet++, which captures contextual features and achieves more accurate segmentation. Wang et al. (2024a) combined convolutional networks (ConvNet), Vision Transformers (ViT), and attention modules to enable the perception of insulators and defect regions. Li and Cai (2022) designed a Point Transformer layer that uses a self-attention mechanism to extract features from point clouds, thereby improving the model's performance. Later, Liu and Huang (2024) enhanced this approach by incorporating self-attention mechanisms and designing a multi-level feature fusion scheme. This enabled the model to allocate attention weights adaptively at both local and global scales, leading to improved segmentation precision (Li et al., 2024b). Building on this, Liu and Huang (2024) introduced a cross-modal point cloud segmentation method that integrated point cloud features from different perceptual dimensions, effectively enhancing insulator segmentation in complex scenarios.

3 Method

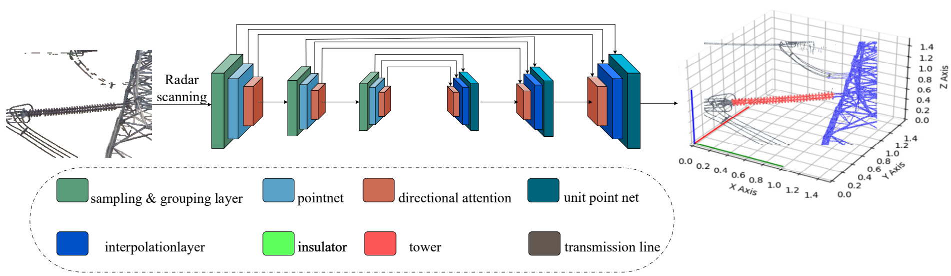
We first briefly introduce the architecture of the PDA model, and then introduce the DA module and the specific structure of embedding DA into PointNet++ in detail.

3.1 Overall Architecture of PDA

Deep learning technology is pivotal in 3D point cloud segmentation tasks, particularly when addressing complex spatial relationships and computational costs. The primary goal of 3D point cloud semantic segmentation is to classify each point within the point cloud data and assign it a corresponding semantic label. This process typically involves extracting high-dimensional features and understanding intricate spatial structures. Consequently, the design of the network architecture is essential for enhancing segmentation accuracy.

As illustrated in Figure 1, the proposed directional attention (PDA) model is an improved network architecture based on PointNet++. It integrates a directional attention (DA) module to more effectively capture the spatial structures and features of the point cloud. This architecture leverages the set abstraction and feature propagation modules of PointNet++ for downsampling and upsampling the point cloud data, respectively, allowing for the efficient extraction and transfer of spatial information throughout the encoder and decoder processes. The encoder of the PDA model comprises a set abstraction module with an embedded DA, while the decoder consists of a feature propagation module that also incorporates DA.

Figure 1 Overall architecture of the proposed PDA (see online version for colours)



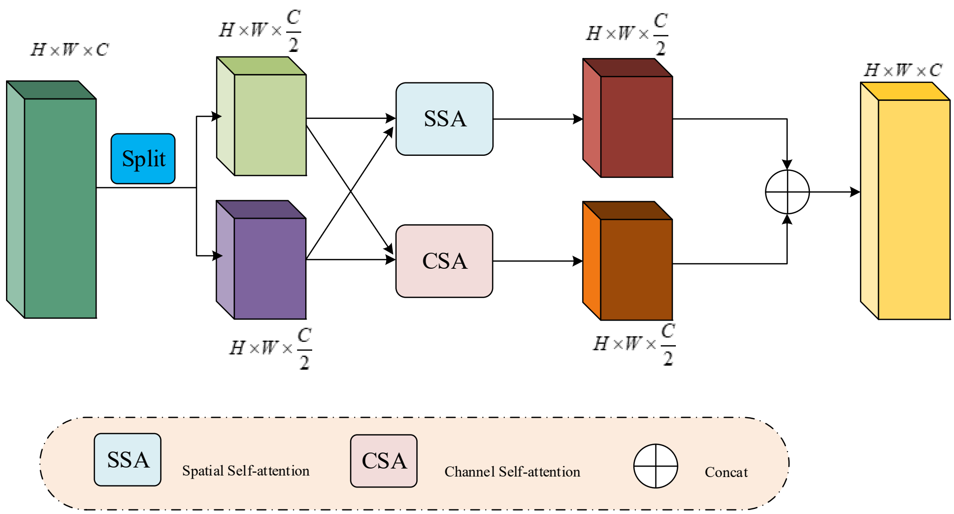
3.2 Directional attention

The squeeze-and-excitation (SE) attention module (Hu et al., 2018) effectively enhances the evaluation of channel importance; however, its design overlooks spatial information, which limits its performance in complex scenes. This shortcoming makes it challenging for the SE module to optimally capture fine-grained local features or long-range dependencies. Similarly, the convolutional block attention module (CBAM) (Woo et al., 2018) achieves a balance between channel and spatial attention but tends to focus more on local features, lacking effective integration of global context information. As a result, this limitation may prevent the model from fully leveraging global features when processing large-scale or complex images, adversely affecting overall performance.

While the self-attention mechanism (Vaswani et al., 2017) is theoretically capable of capturing global features, its computational complexity increases quadratically with input size, resulting in significant computational and memory overhead when dealing with high-resolution images. This limitation restricts its applicability in resource-constrained environments. Although the non-local attention module (Wang et al., 2018) can capture long-range dependencies, its high computational demands—especially in high-resolution images—may lead to inefficiencies in real-time applications, impacting practicality.

As illustrated in Figure 2, our DAM comprises spatial self-attention (SSA) and channel self-attention (CSA). In contrast to previous attention modules, the DAM effectively balances accuracy and parameter efficiency.

Figure 2 Overall architecture of the proposed DA (see online version for colours)



Given a feature map $F \in \mathbb{R}^{H\times W\times C}$, we separate it into two feature maps of equal size, specifically $F_1 \in \mathbb{R}^{H\times W\times \frac{C}{2}}$ and $F_2 \in \mathbb{R}^{H\times W\times \frac{C}{2}}$. We then compute SSA and CSA for the two features separately. Finally, we reshape and stack the output feature maps to obtain a robust feature map $F_{12} \in \mathbb{R}^{H\times W\times C}$.

Next, we explore the advantages of separately computing SSA and CSA. First, the refinement of feature representation is achieved by focusing on the specific properties of different feature maps. SSA emphasises the spatial distribution of features, enabling the capture of subtle variations in local features, which is crucial for recognising important details in images (Vaswani et al., 2017). In contrast, CSA focuses on the relationships between different channels, helping to uncover the complementarity among channel features, thus enhancing the overall feature discrimination capability (Hu et al., 2018; Niu et al., 2021).

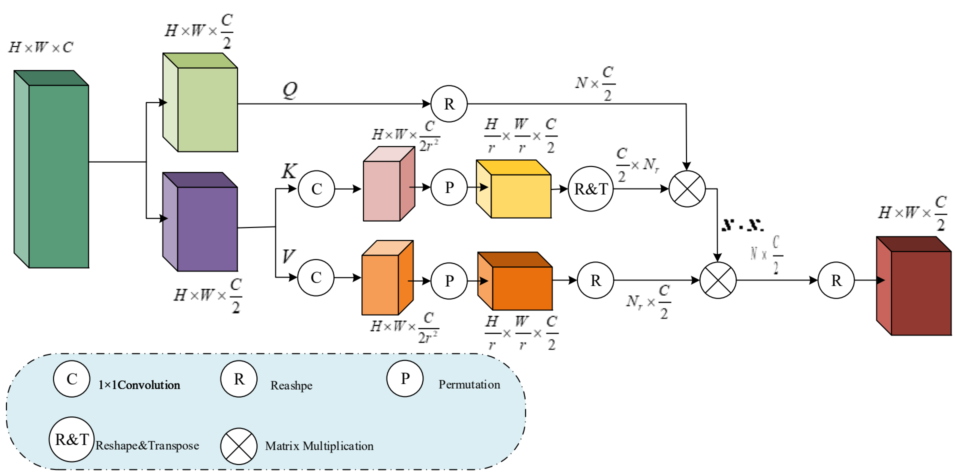
Secondly, performing self-attention calculations on smaller feature maps can significantly reduce computational complexity. The computational complexity of traditional self-attention mechanisms is proportional to the square of the input size, which results in substantial resource consumption when processing high-resolution images (Shen et al., 2021). By processing feature maps separately, the model can handle local features with linear complexity, thereby improving computational efficiency, which is particularly important in large-scale datasets (Touvron et al., 2021).

Finally, this separate processing strategy can enhance information interaction among features. The combination of SSA and CSA allows the model to leverage both local and global information simultaneously, establishing long-range dependencies. This multi-scale information fusion enhances the model's expressive capability, enabling it to capture more complex feature patterns and thereby improving task performance (Guo et al., 2022).

3.2.1 Spatial self-attention

This paper proposes a spatial attention module, as shown in Figure 3. To reduce the computational cost of matrix multiplication, we compress the channels of the input feature map by a factor of r^2. Next, the pixels in the spatial direction are rearranged along the channel direction of the feature map, which reduces the spatial resolution while retaining spatial information. This operation allows the same number of pixels to participate in the self-attention computation while decreasing the computational burden, enhancing the integration capability of spatial information and enriching the contextual information of local regions. Specifically, the channels that were compressed are restored to the input feature map size of $H \times W \times \frac{C}{2}$.

Figure 3 Overall architecture of the proposed SSA (see online version for colours)



Given the input point cloud feature map $F \in \mathbb{R}^{H \times W \times C}$, we first separate the feature map height, width, and number of channels of the feature map, respectively.

We perform a flattening operation on the Q feature map, while applying channel compression on K and V:

$$X_Q = \text{reshape}(Q), X_K = \text{Conv}_{1\times 1}(K), X_V = \text{Conv}_{1\times 1}(V) \tag{1}$$

where $X_Q \in \mathbb{R}^{H \times W \times \frac{C}{2}}$, and $X_K, X_V \in \mathbb{R}^{H \times W \times \frac{C}{2r^2}}$. After that, by rearranging the operation, we align the spatial pixels of the feature map along the channel direction. This operation enables each pixel to participate in the attention computation while reducing the computational burden.

We utilise the SSA mechanism to establish long-range dependencies between pixels. The spatial attention calculation formula is as follows:

$$\mathrm{SSA}\left(X_Q, X_K, X_V\right)=\left(\frac{X_Q X_V^T}{D}+B\right)\cdot X_V \tag{2}$$

where B represents the relative position encoding, and $\mathrm{D}=\frac{\mathrm{r}^2}{2}$ is used for normalisation. Through this operation, although the spatial resolution of our feature map is reduced, the number of pixels participating in the computation remains unchanged. This strengthens the network's ability to extract spatial and local information. Next, we reshape the feature map to match the size of the input feature map for subsequent processing. We denote the output feature map of equation (2) as $\mathrm{A} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$, with the reshaping process described in equation (3):

$$V_{SSA}=\mathrm{reshape}(A) \tag{3}$$

where $\mathrm{V_{SSA}}$ represents the feature map generated by the SSA module, and $\mathrm{V_{SSA}} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\frac{\mathrm{C}}{2}}$.

3.2.2 Channel self-attention

In Figure 4, we compressed the number of channels in the feature map and rearranged the spatial pixels into the channel direction of the feature map, achieving participation of each pixel in the computation while reducing the computational load. This operation preserves spatial information and enhances the extraction of local information. To further strengthen the long-range relationships between pixels, as shown in Figure 4, we designed the channel attention module to extract information from the feature map in the channel direction.

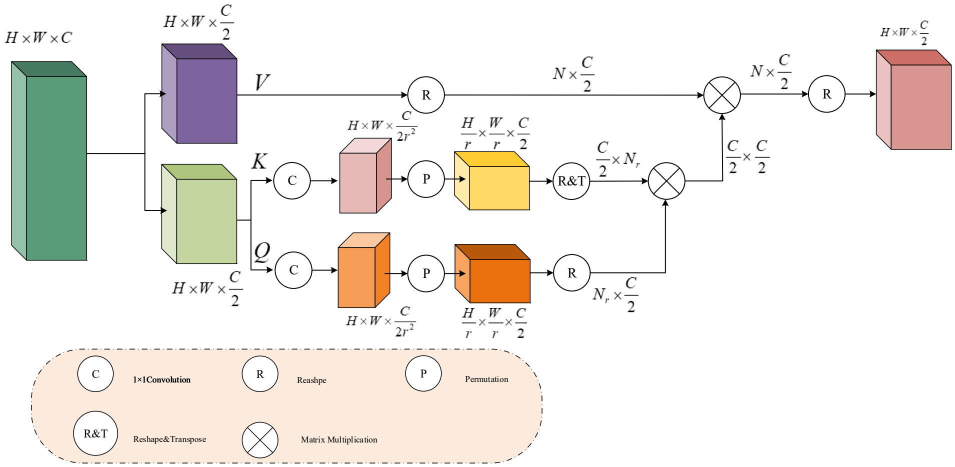
$$\mathrm{CSA}\left(X_Q, X_K, X_V\right)=\frac{X_Q^T X_K}{D}\cdot X_V \tag{4}$$

Here, set the feature map output by Formula (4) to B, and then change the shape of B to a feature map with the same size as the input feature map.

$$V_{CSA}=\mathrm{reshape}(B) \tag{5}$$

The pixel suggestions for remote dependencies along the channel direction of the feature map can aggregate information from different subspaces, demonstrating good performance in processing spatial and channel information. By calculating channel dependencies, it is possible to aggregate the receptive fields of channel information, establishing long-range interdependencies between each channel and the others. Without the effect of channel attention, the information of each channel does not establish connections with the information from other channels, which is detrimental to point cloud segmentation. On the other hand, CSA enables information interaction between channels and establishes long-range interdependencies between pixels.

Figure 4 Overall architecture of the proposed CSA (see online version for colours)



3.3 The workflow of PDA network architecture

PointNet++ builds upon PointNet by constructing a hierarchical point grouping, aggregating larger local regions within the hierarchical point cloud structure. The aggregation abstraction layer of PointNet+ + consists of a sampling layer, a grouping layer, and a PointNet layer. The PDA receiver is an N × (d + C) matrix of point cloud input, representing N points with d dimensional geometric information and C-dimensional features.

In the sampling layer, one point from the point cloud serves as a centroid, scanning surrounding points within a radius r to construct a local region. The grouping layer then collects and groups the points, where the input size is N × (d + C) and the centroid coordinates are of size N′ × d. For each centroid, K points are sampled within a radius r. Consequently, the dimensionality of the input point cloud data transforms from N × (d + C) to N′ × K × (d + C), where each group represents a local region. Here, K denotes the number of points in the neighbourhood of the centroid.

Subsequently, the PointNet layer receives point cloud data of size N′ × K × (d + C), where N′ represents the number of points in the local region. The output point cloud data from the PointNet layer is generated by aggregating the centroids and the encoded local features of the centroid neighbourhoods, resulting in point cloud data of size N′ × (d + C′). The point cloud features and spatial information are combined and grouped into a larger point cloud set, serving as input for the next DA module, generating higher-level point cloud features until all point sets are aggregated to form the complete point cloud features. The DA module establishes global long-range dependencies within the point cloud through a stacked CSA module and a SSA module.

Similarly, the decoder of the PDA consists of three feature propagation modules. It first receives high-level features from PointNet++, which are extracted through a multi-level point feature abstraction (set abstraction, SA) and feature propagation (feature propagation, FP) process, capable of representing different hierarchical information of the point cloud. During the feature propagation phase, the decoder utilises connections from low level features to high-level features, propagating features to all original points through interpolation and skip connections. This process ensures that detailed features are not lost during abstraction.

The transfer of features during the propagation process can be described by the formula $N_l \times (d + c)$. Here, N_l represents the number of input points, d is the dimension of each point feature, and c is the dimension of the additional contextual features. In this process, shallow features are computed based on the deep features of each point in the current network, transferring the $N_l \times (d + c)$ point features to N_{l-1} points, where N_{l-1} and N_l represent the number of points before and after downsampling in the SA module, respectively. This mechanism ensures that each target point can obtain rich information from the original point set, thereby enhancing the quality of feature representation.

The decoder typically employs interpolation methods such as inverse distance weighting (IDW) to transfer abstract features back to the original points. This approach ensures that features closer to the target points exert a greater influence on the interpolation results, effectively capturing local contextual information. Through skip connections, the decoder fuses features from different layers, combining the abstract capabilities of high-level features with the detailed information of low-level features. This fusion significantly enhances the model's performance in complex scenes.

To further improve feature extraction capabilities, we embed the Directional Attention (DA) module into the PointNet layer. The DA module comprises both a SSA module and a CSA module. The CSA module dynamically adjusts the weights of feature channels, enabling the decoder to automatically identify and prioritise specific features while suppressing less important ones. This mechanism enhances the accuracy and effectiveness of the outputs generated by the decoder.

The SSA module improves the understanding of both local and global context by capturing the spatial relationships between points. This allows the decoder to fully consider the dependencies among different points in the point cloud, leading to the generation of detail-rich outputs and improving the model's adaptability to complex scenes. The integration of these two modules enables the decoder to more effectively synthesise information, resulting in more accurate and detailed classification results or segmentation masks.

4 Experiment and result analysis

This study employs LiDAR laser scanning to obtain point cloud data related to power transmission lines and towers across provinces such as Jiangsu, Sichuan, and Jiangxi. Laser scanning is a precise measurement technology particularly well-suited for capturing data from complex structures. We utilised high-precision ground laser scanners, including models from Leica and Faro. Selecting appropriate scanning positions is crucial to cover all key areas; therefore, multiple scans from different angles and heights are recommended to enhance data completeness and accuracy. For post-processing, professional software such as Cyclone or Cloud Compare was used for data stitching, denoising, and generating 3D models, resulting in clear point cloud data that facilitates subsequent analysis.

In addition, drone aerial photography provides an efficient method for covering large areas. Drones equipped with high-resolution cameras or LiDAR sensors, such as those from the DJI Matrice series, were employed. Before conducting aerial photography, a detailed flight plan was developed to ensure complete coverage of all important areas.

Appropriate flying heights and image overlap were set according to specific requirements to obtain high-quality images or point clouds. After the flight, software like Pix4D or Drone Deploy was utilised to convert the images into point cloud data for detailed 3D modeling and analysis.

Once the point cloud data is obtained, processing software such as MeshLab or PDAL is employed for cleaning, filtering, and feature extraction, ensuring the accuracy and reliability of the data. These processing steps aid in identifying key features, such as the height of power lines and the condition of insulators, thereby providing valuable support for subsequent analysis and decision-making.

4.1 Experiment implementation

In this research, we conducted experiments on an Ubuntu 16.04 server system equipped with an NVIDIA GeForce RTX 4090 graphics card and an Intel Core i5 10400F processor. This configuration provided ample computational power to handle the demanding tasks involved in training and evaluating our point cloud semantic segmentation model. We utilised Python 3.10.12 and PyTorch 2.0.1, which are widely used tools in the deep learning community, for implementing our experiments.

During the training phase, we meticulously tuned our hyperparameters to ensure optimal performance. The model underwent 200 epochs of training, allowing it to learn iteratively from the dataset. The learning rate was set to 0.01, and the Adam optimiser (Kingma, 2014) was used, a widely adopted method for optimising deep neural networks. These settings were chosen to balance model convergence and computational efficiency, enabling effective training.

By leveraging this experimental setup, we thoroughly evaluated the performance of our point cloud semantic segmentation model. The combination of powerful hardware and robust software frameworks provided the necessary tools to conduct comprehensive experiments and draw meaningful conclusions from our research.

4.2 Loss function

The point cloud data of insulators exhibits significant class imbalance. To address this issue, we introduced focal loss (FL) (Quach et al., 2020) into the training process of the insulator point cloud semantic segmentation model. FL is a loss function specifically designed to tackle class imbalance problems. It effectively enhances the model's focus on rare classes, accelerates the convergence speed of the model, and improves the recognition accuracy of small classes. The computation process of the FL loss function is illustrated as shown in Formula (6),

$$L_{\mathrm{FL}}(y) = -\alpha_y \log\left(p_y\right)\left(1 - p_y\right)^y \tag{6}$$

In the above formula, α represents the class weights, similar to the class weights in weighted cross-entropy. The coefficient γ is used to adjust the model's focus on hard-to-classify samples. By reducing the weight of easy-to-classify samples, γ enables the model to focus more on difficult-to-classify samples during training.

4.3 Evaluation indicators

This study utilises evaluation metrics derived from the confusion matrix to comprehensively assess the model's performance in point cloud semantic classification tasks. The confusion matrix is shown in Table 1. TP refers to instances where the model correctly identifies positive samples as positive, FN denotes cases where the model incorrectly classifies positive samples as negative (false negatives or omissions), FP indicates situations where the model wrongly labels negative samples as positive (false positives), and TN represents instances where the model correctly categorises negative samples as negative.

Table 1 Confusion matrix

		Predicted	
		Positive	*Negative*
Actual	Positive	TP	FN
	Negative	FP	TN

To provide a comprehensive evaluation of the experimental results, accuracy (pre), recall (rec), intersection over union (IoU) and mean intersection over union (mIoU).

The calculation process of precision is shown in Formula (7),

$$\mathrm{I}_{pre} = \frac{TP}{TP + FP} \tag{7}$$

The calculation process of recall rate is shown in Formula (8),

$$\mathrm{I}_{rec} = \frac{TP}{TP + FN} \tag{8}$$

The calculation process of IoU is shown in Formula (9),

$$IoU = \frac{TP}{TP + FP + FN} \tag{9}$$

The calculation process of mIoU is shown in Formula (10),

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{TP + FP + FN} \tag{10}$$

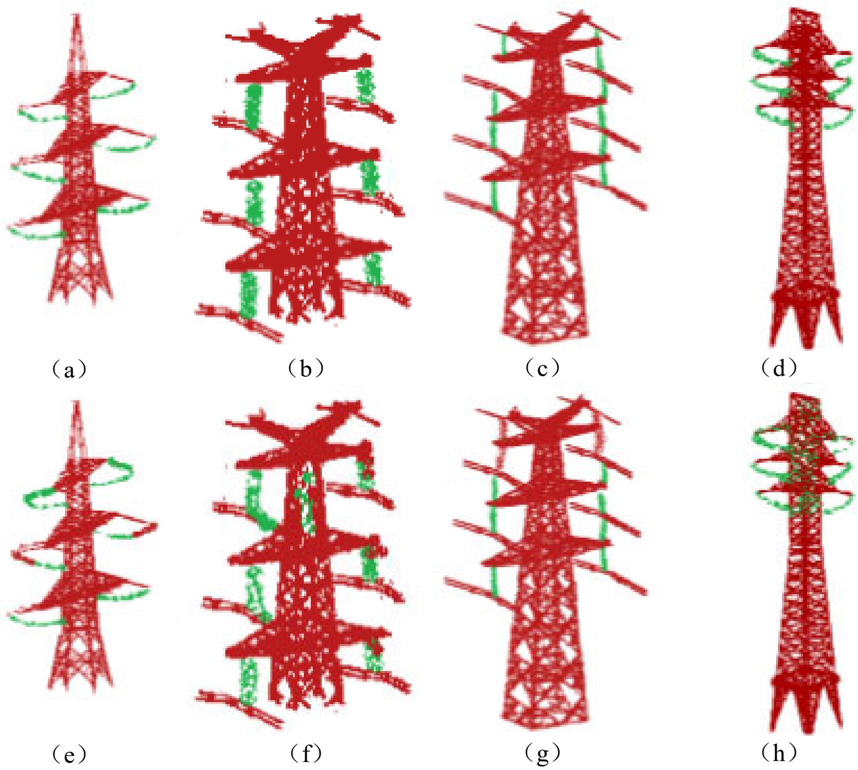
4.4 Ablation study

As shown in Table 2, we conducted a series of experiments that sequentially removed or replaced these components, demonstrating a decline in overall performance when each component was absent. The introduction of the CSA component resulted in an increase of 1.02 in I_{pre} and an increase of 1.07 in I_{rec} compared to the baseline model PointNet+ +. The component achieves feature correction by eliminating irrelevant and redundant features. Additionally, with the sequential addition of CSA and SS, the PDA model, compared to PointNet++, showed increases of 2.44 and 2.4 in I_{pre} and I_{rec}, respectively.

Table 2 Perform ablation experiments on different components of the model

Module	$I_{pre}\%$	$I_{rec}\%$
PointNet++	92.12	93.14
CSA	93.14	94.21
SSA	93.89	94.68
DA	94.12	95.23
PDA	94.56	95.54

Figure 5 Ablation experiment for visual comparison using PDANet (see online version for colours)



Note: The first row (a-d) respectively represents the real labels of four different shapes, and the second row (e-h) respectively represents the predicted segmentation result map corresponding to the first row (a-d).

To further illustrate the role of the modules, we visualised the segmentation results for different shapes. As shown in Figure 5, predictive segmentation results of the PDA model under four different shapes. We randomly selected four different types of transmission towers and visualised them. In the visualisation, green represents insulators, and e, f, g, h represents the backbone, embedded SSA, embedded CSA, and DA, respectively. From image f, it can be seen that, except for the slight imperfections in image f, all other images align with the label prediction results. This indicates that combining CSA and

SSA can establish long-range relationships from a global perspective, thereby improving the performance of predictions.

4.5 Comparative experiments with advanced methods

Based on the experimental results shown in Table 3, the PDA model shows considerable enhancements over PointNet and PointNet++ in several categories, such as lion, tower, and insulator. Notably, in the tower category, the IoU for PDA reached 83.1, highlighting its clear advantage in overall performance and achieving a mean IoU (mIoU) of 69, which far exceeds that of other models.

The effectiveness of the PDA model can be credited to its integrated channel attention and spatial attention components, which allow the model to concentrate more efficiently on key features, thus improving classification precision. In contrast, PointNet has a relatively low IoU of 33.2 for transmission lines, while both PointNet and PointNet++ have IoUs below 50 for insulators, indicating their struggles to effectively segment small point cloud objects. In contrast, our model exhibits strong performance across the lion, tower, and insulator categories.

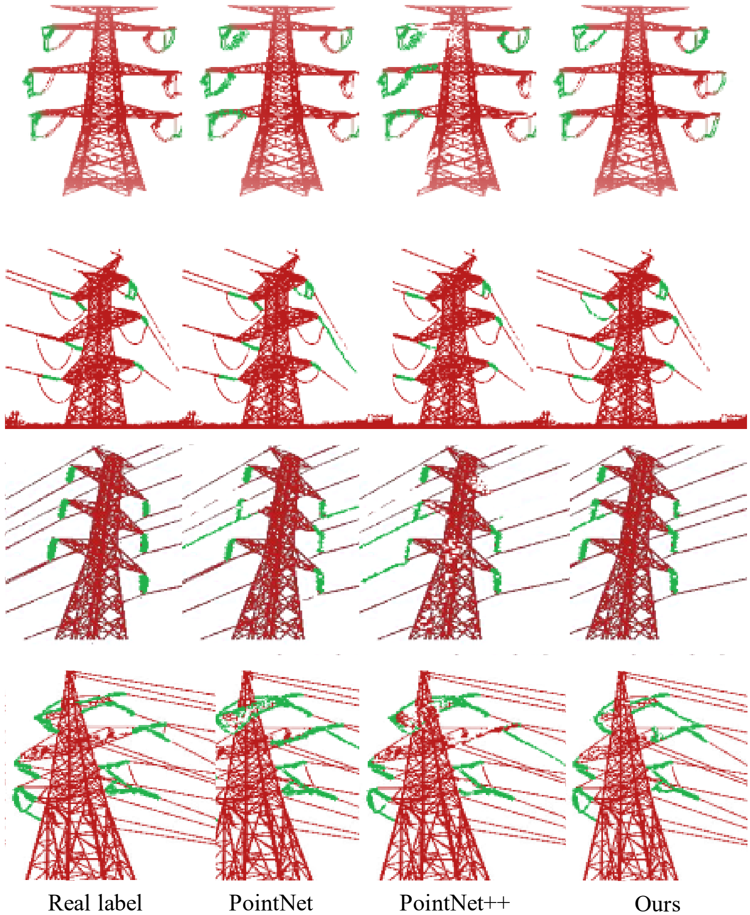
Table 3 Comparative experiments with advanced methods

Model	*IoU*			*mIoU*
	Lion	*Tower*	*Insulator*	
PointNet	33.2	55.2	34.2	40.9
PointNet++	47.2	52.9	48.1	49.4
PDA	61.3	83.1	62.6	69

To further confirm the benefits of our approach over alternative methods, we performed a comparative evaluation of the enhanced models, such as PointNet and PointNet++, as shown in Figure 6. We selected four different labeled shapes for this comparison. In the first and fourth rows of the second column, the segmentation results for insulators are clearly inconsistent with the ground truth labels. The primary reasons for this discrepancy are changes in lighting and the proximity relationship between pixels, which hinder the differentiation of insulators and lead to segmentation errors. PointNet++ addresses this issue by considering local region features, resulting in improved segmentation metrics for lines and insulators. As demonstrated in the third column, the segmentation results from PointNet++ show significant improvement over those of PointNet.

However, it is important to note that the transmission line in the third row of the third column was not clearly segmented by PointNet++, as it was misclassified as insulator strings. A similar issue occurs in the first row of the third column, where the same phenomenon is observed. This challenge primarily arises from the lack of differentiation between classes and insufficient contextual information. To address this issue, we embedded SSA and CSA into PointNet++ to capture pixel dependencies in both channel and spatial dimensions, while distinguishing the differences among classes. Notably, in the implementation of CSA and SSA, we employed compression and rearrangement of pixels. This approach allows for achieving strong segmentation performance without increasing computational overhead.

Figure 6 Visualised experimental results compared to state-of-the-art methods (see online version for colours)



Note: From left to right are the Real labels, segmentation results of PointNet, PointNet++ and PDA models.

5 Conclusions

Segmenting insulators is essential for ensuring the reliable transmission of electricity. This paper utilises data obtained from LiDAR scanning for precise segmentation, serving as an effective tool for risk management. We propose a PointNet++ model based on directional attention, referred to as PDA, which provides a robust method for segmenting point cloud transmission corridors. PDA adopts the end-to-end structure of PointNet++ and incorporates CSA and SSA into both the encoding and decoding sections. CSA handles pixel interactions across channels in the feature maps, differentiating between pixels of the same class and those from different classes, while SSA creates spatial dependencies between pixels to capture detailed contextual information. By leveraging CSA and SSA, we obtain global pixel dependencies and effectively capture the positional and spatial relationships of various features, leading to significant improvements in point cloud segmentation. In future research, we will place greater emphasis on the refined segmentation of insulators within transmission corridors. Additionally, while prioritising accuracy, we will also consider real-time requirements to enhance the practical applicability of our model.

Acknowledgements

This work was supported by State Grid Shanxi Electric Power Company Technology Project :Research and application of precise diagnosis technology for infrared three-dimensional imaging and thermal anomalies (No.52053023000Y).

References

Arastounia, M. and Lichti, D.D. (2013) 'Automatic extraction of insulators from 3D LiDAR data of an electrical substation', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 2, pp.19–24.

Bahdanau, D. (2014) 'Neural machine translation by jointly learning to align and translate', arXiv preprint arXiv:1409.0473.

Bayindir, R., Colak, I., Fulli, G. and Demirtas, K. (2016) 'Smart grid technologies and applications', *Renewable and Sustainable Energy Reviews*, Vol. 66, pp.499–516, https://doi.org/10.1016/j.rser.2016.08.002.

Cao, S., Fan, Q., Yu, W. J., Wang, L.T., Ni, S. and Chen, J. (2022) 'Multi-sensor fusion and data analysis for operating conditions of low power transmission lines', *Measurement*, Vol. 190, p.110586, https://doi.org/10.1016/j.measurement.2021.110586.

Chen, C., Yang, B., Song, S., Peng, X. and Huang, R. (2018) 'Automatic clearance anomaly detection for transmission line corridors utilizing UAV-Borne LIDAR data', *Remote Sensing*, Vol. 10, No. 4, p 613.

Ellis, N. (2013) 'Inspection of power transmission lines using UAVs'.

Faustino, P.M.F. (2022) *Vegetation Detection and Classification for Power Line Monitoring*, Doctoral dissertation,NOVA University of Lisbon.

Fryskowska, A. (2019) 'Improvement of 3d power line extraction from multiple low-cost UAV imagery using wavelet analysis', *Sensors*, Vol. 19, No. 3, p.700.

Guan, H., Sun, X., Su, Y., Hu, T., Wang, H., Wang, H., Peng, C. and Guo, Q. (2021) 'UAV-lidar aids automatic intelligent powerline inspection', *International Journal of Electrical Power & Energy Systems*, Vol. 130, p.106987, https://doi.org/10.1016/j.ijepes.2021.106987.

Guo, J., Li, T. and Du, B. (2023) 'Segmentation head networks with harnessing self-attention and transformer for insulator surface defect detection', *Applied Sciences*, Vol. 13, No. 16, p.9109.

Guo, M.H., Xu, T.X., Liu, J.J., Liu, Z.N., Jiang, P.T., Mu, T.J., Zhang, S.H., Martin, R.R., Cheng, M.M. and Hu, S.M. (2022) 'Attention mechanisms in computer vision: a survey', *Computational Visual Media*, Vol. 8, No. 3, pp.331–368.

Hu, J., Shen, L. and Sun, G. (2018) 'Squeeze-and-excitation networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132–7141.

Jenssen, R. and Roverso, D. (2018) 'Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning', *International Journal of Electrical Power & Energy Systems*, Vol. 99, pp.107–120, https://doi.org/10.1016/j.ijepes.2017.12.016.

Ji, Z. and Zhong, X. (2024) 'Bidirectional attention network for real-time segmentation of forest fires based on UAV images', *International Journal of Information and Communication Technology*, Vol. 25, No. 6, pp.38–51.

Kingma, D.P. (2014) 'Adam: a method for stochastic optimization', arXiv preprint arXiv:1412.6980.

Li, X., Li, Y., Chen, Y., Zhang, G. and Liu, Z. (2024a) 'Deep learning-based target point localization for UAV inspection of point cloud transmission towers', *Remote Sensing*, Vol. 16, No. 5, p.817.

Li, Y., Zhu, X., Li, B., Zeng, J. and Wang, S. (2024b) 'Discriminative features based comprehensive detector for defective insulators', *Energy and AI*, p.100387.

Li, Y. and Cai, J. (2022) 'Point cloud classification network based on self-attention mechanism', *Computers and Electrical Engineering*, Vol. 104, p.108451, https://doi.org/10.1016/j.egyai.2024.100387.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X. and Chen, B. (2018) 'Pointcnn: convolution on x-transformed points', *Advances in Neural Information Processing Systems*, Vol. 31, https://doi.org/10.1016/j.compeleceng.2022.108451.

Liu, Y. and Huang, X. (2024) 'Efficient cross-modality insulator augmentation for multi-domain insulator defect detection in UAV images', *Sensors*, Vol. 24, No. 2, p.428.

Lv, C., Guan, X., Liu, J. and Liao, J. (2023) '3-D segmentation and surface reconstruction of gas insulated switchgear via PointNet-MLS architecture', in *International Symposium on Insulation and Discharge Computation for Power Equipment*, pp.187–193.

Ma, Y., Li, Q., Chu, L., Zhou, Y. and Xu, C. (2021) 'Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision', *Remote Sensing*, Vol. 13, No. 2, p.230.

Matikainen, L., Lehtomäki, M., Ahokas, E., Hyyppä, J., Karjalainen, M., Jaakkola, A., Kukko, A. and Heinonen, T. (2016) 'Remote sensing methods for power line corridor surveys', *ISPRS Journal of Photogrammetry and Remote sensing*, Vol. 119, pp.10–31, https://doi.org/10.1016/j.isprsjprs.2016.04.011.

Maturana, D. and Scherer, S. (2015) 'Voxnet: A 3d convolutional neural network for real-time object recognition', *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE*, pp.922–928.

Miao, X., Liu, X., Chen, J., Zhuang, S., Fan, J. and Jiang, H. (2019) 'Insulator detection in aerial images for transmission line inspection using single shot multibox detector', *IEEE Access*, Vol. 7, pp.9945–9956.

Mohsan, S.A.H., Khan, M.A., Noor, F., Ullah, I. and Alsharif, M.H. (2022) 'Towards the unmanned aerial vehicles (UAVs): a comprehensive review', *Drones*, Vol. 6, No. 6, p.147.

Murthy, V.S., Mohanta, D.K. and Gupta, S. (2011) 'Video surveillance-based insulator condition monitoring analysis for substation monitoring system (SMS)', *International Journal of Information and Communication Technology*, Vol. 3, No. 1, pp.11–31.

Niu, Z., Zhong, G. and Yu, H. (2021) 'A review on the attention mechanism of deep learning', *Neurocomputing*, Vol. 452, pp.48–62, https://doi.org/10.1016/j.neucom.2021.03.091.

Pastucha, E., Puniach, E., Ścisłowicz, A., Ćwiąkała, P., Niewiem, W. and Wiącek, P. (2020) '3D reconstruction of power lines using uav images to monitor corridor clearance', *Remote Sensing*, Vol. 12, No. 22, p.3698.

Peng, S., Xi, X., Wang, C., Dong, P., Wang, P. and Nie, S. (2019) 'Systematic comparison of power corridor classification methods from ALS point clouds', *Remote Sensing*, Vol. 11, No. 17, p.1961.

Poux, F. and Billen, R. (2019) 'Voxel-based 3D point cloud semantic segmentation: Unsupervised geometric and relationship featuring vs deep learning methods', *ISPRS International Journal of Geo-Information*, Vol. 8, No. 5, p.213.

Qi, C.R., Su, H., Mo, K. and Guibas, L.J. (2017a) 'Pointnet: deep learning on point sets for 3d classification and segmentation', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.652–660.

Qi, C.R., Yi, L., Su, H. and Guibas, L.J. (2017b) *PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space*, ArXiv, abs/1706.02413.

Quach, M., Valenzise, G. and Dufaux, F. (2020) 'Improved deep point cloud geometry compression', *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP). IEEE*, pp.1–6.

Shen, Z., Zhang, M., Zhao, H., Yi, S. and Li, H. (2021) 'Efficient attention: attention with linear complexities', *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp.3531–3539.

Sohn, G., Jwa, Y. and Kim, H.B. (2012) 'Automatic powerline scene classification and reconstruction using airborne lidar data', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 1, pp.167–172.

Su, H., Maji, S., Kalogerakis, E. and Learned-Miller, E. (2015) 'Multi-view convolutional neural networks for 3d shape recognition', *Proceedings of the IEEE International Conference on Computer Vision*, pp.945–953.

Su, Y., Liu, W., Yuan, Z., Cheng, M., Zhang, Z., Shen, X. and Wang, C. (2021) *DLA-Net: Learning Dual Local Attention Features for Semantic Segmentation of Large-Scale Building Facade Point Clouds*, ArXiv, abs/2106.00376.

Sun, S., Chen, C., Yang, B., Yan, Z., Wang, Z., He, Y., ... and Fu, J. (2024) 'ID-Det: insulator burst defect detection from UAV inspection imagery of power transmission facilities', *Drones (2504-446X)*, Vol. 8, No. 7, https://doi.org/10.3390/drones8070299.

Sun, Y., Chen, X., Jian, X. and Xiao, Z. (2019) 'Identification and localization method of the insulator based on three-dimensional point cloud modeling', *2019 Chinese Control Conference (CCC), IEEE*, pp.7051–7056.

Tang, J., Tan, J., Du, Y., Zhao, H., Li, S., Yang, R., ... and Li, Q. (2023) 'Quantifying multi-scale performance of geometric features for efficient extraction of insulators from point clouds', *Remote Sensing*, Vol. 15, No. 13, p.3339.

Tao, X., Zhang, D., Wang, Z., Liu, X., Zhang, H. and Xu, D. (2018) 'Detection of power line insulator defects using aerial images analyzed with convolutional neural networks', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 50, No. 4, pp.1486–1498.

Taylor, W.O., Watson, P.L., Cerrai, D. and Anagnostou, E.N. (2022) 'Dynamic modeling of the effects of vegetation management on weather-related power outages', *Electric Power Systems Research*, No. 207, p.107840, https://doi.org/10.1016/j.epsr.2022.107840.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A. and Jégou, H. (2021) 'Training data-efficient image transformers & distillation through attention', *International Conference on Machine Learning, PMLR*, pp.10347–10357.

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) 'Attention is all you need', *Neural Information Processing Systems*.

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*, Vol. 12, No. 13, p.2829.

Wang, Q., Li, Y., Cui, S., Li, N., Zhang, X., Jiang, W., ... and Sun, J. (2024a) 'Enhanced recognition of insulator defects on power transmission lines via proposal-based detection model with integrated improvement methods', *Engineering Applications of Artificial Intelligence*, Vol. 136, p.109078, https://doi.org/10.1016/j.engappai.2024.109078

Wang, Q.Y., Lv, X.L. and Tang, S.K. (2024b) 'FVCNet: detection obstacle method based on feature visual clustering network in power line inspection', *Computational Intelligence*, Vol. 40, No. 2, p.e12634.

Wang, X., Girshick, R., Gupta, A. and He, K. (2018) 'Non-local neural networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7794–7803.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M. (2019) 'Dynamic graph CNN for learning on point clouds', *ACM Transactions on Graphics (tog)*, Vol. 38, No. 5, pp.1–12.

Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018) 'Cbam: convolutional block attention module', *Proceedings of the European Conference on Computer Vision (ECCV)*, pp.3–19.

Wu, Q., Yang, H., Wei, M., Remil, O., Wang, B. and Wang, J. (2018) 'Automatic 3D reconstruction of electrical substation scene from LiDAR point cloud', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 143, pp.57–71, https://doi.org/10.1016/J.ISPRSJPRS.2018.04.024.

Xu, M., Ding, R., Zhao, H. and Qi, X. (2021) 'Paconv: position adaptive convolution with dynamic kernel assembling on point clouds', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp.3173–3182.

Xuan, Z., Ding, J. and Mao, J. (2022) 'Intelligent identification method of insulator defects based on CenterMask', *IEEE Access*, Vol. 10, pp.59772–59781, https://doi.org/10.1109/ACCESS.2022.3179975.

Yan, X., Zheng, C., Li, Z., Wang, S. and Cui, S. (2020) 'Pointasnl: robust point clouds processing using nonlocal neural networks with adaptive sampling', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.5589–5598.

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 Transactions on Instrumentation and Measurement*, Vol. 69, No. 12, pp.9350–9365.

Zeng, J., Zhu, X., Li, B., Guo, Z., Tian, Y. and Yuan, S. (2022) 'Attention-based multiscale context awareness network for insulator defect detection', *2022 4th International Conference on Smart Power & Internet Energy Systems (SPIES)*, December, pp.2350–2355.

Zhang, J., Zhao, X., Chen, Z. and Lu, Z. (2019a) 'A review of deep learning-based semantic segmentation for point cloud', *IEEE Access*, Vol. 7, pp.179118–179133, https://doi.org/10.1109/ACCESS.2019.2958671.

Zhang, R., Yang, B., Xiao, W., Liang, F., Liu, Y. and Wang, Z. (2019b) 'Automatic extraction of high-voltage power transmission objects from UAV lidar point clouds', *Remote Sensing*, Vol. 11, No. 22, p.2600.

Zhang, Z., Hua, B.S. and Yeung, S.K. (2019c) 'Shellnet: efficient point cloud convolutional neural networks using concentric shells statistics', *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.1607–1616.

Zhang, S.C., Liu, J.Z., Niu, Z., Gao, S., Xu, H.Z. and Pei, J. (2020) 'Power line simulation for safety distance detection using point clouds', *IEEE Access*, Vol. 8, pp.165409–165418, https://doi.org/10.1109/ACCESS.2020.3022670.

Zhou, W., Jiang, X. and Liu, Y.H. (2019) 'MVPointNet: multi-view network for 3D object based on point cloud', *IEEE Sensors Journal*, Vol. 19, No. 24, pp.12145–12152.