Multi-scale Spatial Position Encoding for Enhanced Feature Learning in Bridge Point Cloud Semantic Segmentation

Yu Chen[a][0000-0002-1187-4761], Chao Lin[a][0000-0002-1517-8392], Shiori Kubo[b][0000-0002-8416-3864], Tatsuro Yamane[c][0000-0003-4027-1959], Wakana Asano[d], Ichiro Iwaki[d], Pang-jo Chun[a*][0000-0002-9755-8435]

[a] The University of Tokyo, Tokyo, Japan
[b] Kagawa University, Kagawa, Japan
[c] National Institute of Technology, Tokuyama College, Yamaguchi, Japan
[d] Nihon University, Tokyo, Japan
*chun@g.ecc.u-tokyo.ac.jp

Abstract. Aging bridges worldwide require frequent inspections to ensure public safety. Traditional visual inspection methods are slow, require significant labor, and often vary between inspectors. This research develops a new method for automatically identifying bridge components in 3D point cloud data. We present a multi-scale spatial position encoding framework designed specifically for bridge structures. Our method features a BriStruc-Encoding module that works through two main pathways: one for position information and another for extracting local structural features. The system combines a Multi-scale Bridge Structure Set Abstraction encoder with a Residual Attentive Feature Propagation decoder to capture both the shape details and spatial relationships of bridge parts. The framework uses geometric features such as linearity, planarity, and sphericity along with measurements like local radius to distinguish different structural elements. We tested our approach on a dataset of highway bridges, achieving an overall mean IoU of 96.58%. Performance was particularly strong for pier structures (98.84%) and girder components (97.28%). Analysis shows that all structural elements were classified with over 95% accuracy, with very little confusion between different bridge parts. The system performs consistently across various bridge types with an F1 score of 98.40%, recall rate of 98.39%, and mean accuracy of 98.31%. This automated approach provides reliable identification of bridge components, supporting better inspection systems and more efficient maintenance planning for infrastructure management.

Keywords: Bridge maintenance, Point cloud datas, Semantic segmentation, Deep learing,

1 Introduction

Infrastructure management, particularly bridge maintenance, has emerged as a critical global challenge in the 21st century. Following World War II, starting in the 1950s, a significant surge in infrastructure development was experienced across Europe, North America, and Asia. However, recent data indicates a concerning trend: a majority of bridges in many developed countries have surpassed their 50-year design life span [1–4]. This widespread aging of critical infrastructure necessitates more frequent and thorough inspections to ensure public safety [5].

Traditional bridge inspection methods, which heavily rely on visual assessments by certified inspectors, face significant limitations in meeting these growing demands. These conventional approaches are often time-consuming, labor-intensive, and subject to human bias [6–8], highlighting the urgent need for automated inspection technologies. Early efforts to address these limitations explored image recognition combined with structure from motion and 3D integration of multiple images for damage detection and documentation [9–11]. In response to these challenges, three-dimensional (3D) point cloud data (PCD) has emerged as a revolutionary solution in infrastructure inspection and management [12–14]. PCD, particularly when acquired through light detection and ranging (LiDAR) technology, offers superior accuracy, comprehensive spatial representation, and the capability to generate detailed digital models for various inspection and maintenance applications.

Despite the rich geometric information contained in PCD, processing infrastructure-scale point clouds presents significant challenges [15]. Raw point clouds lack semantic information, making it difficult to identify and analyze individual structural components. To address this limitation, semantic segmentation techniques have been developed [16,17], with deep learning (DL)-based methods demonstrating particularly promising results. Notable architectures such as PointNet [18], PointNet++ [19], and more sophisticated approaches incorporating graph convolutional networks [20] have shown superior performance in handling diverse bridge structures.

However, existing approaches face several critical limitations. First, current models are typically designed for either small-scale objects (e.g., indoor scenes) or large-scale environments (e.g., urban streets), failing to adequately address the unique intermediate scale of bridge structures. Second, bridges exhibit distinct structural patterns and regular component distributions in both horizontal and vertical directions, yet current semantic segmentation models do not effectively utilize this inherent spatial information. Finally, many existing bridge-specific models are limited to particular bridge types, restricting their broader applicability.

To address these challenges, this study proposes a novel multi-scale spatial position encoding framework for bridge point cloud semantic segmentation. Our approach introduces two key innovations: (1) a multi-scale feature learning mechanism tailored to bridge-scale structures, (2) an explicit encoding scheme that captures both relative and absolute spatial positions of bridge components. By incorporating spatial position information into the feature learning process, our method better preserves the structural relationships and geometric patterns inherent in bridge architectures.

2 Methodology

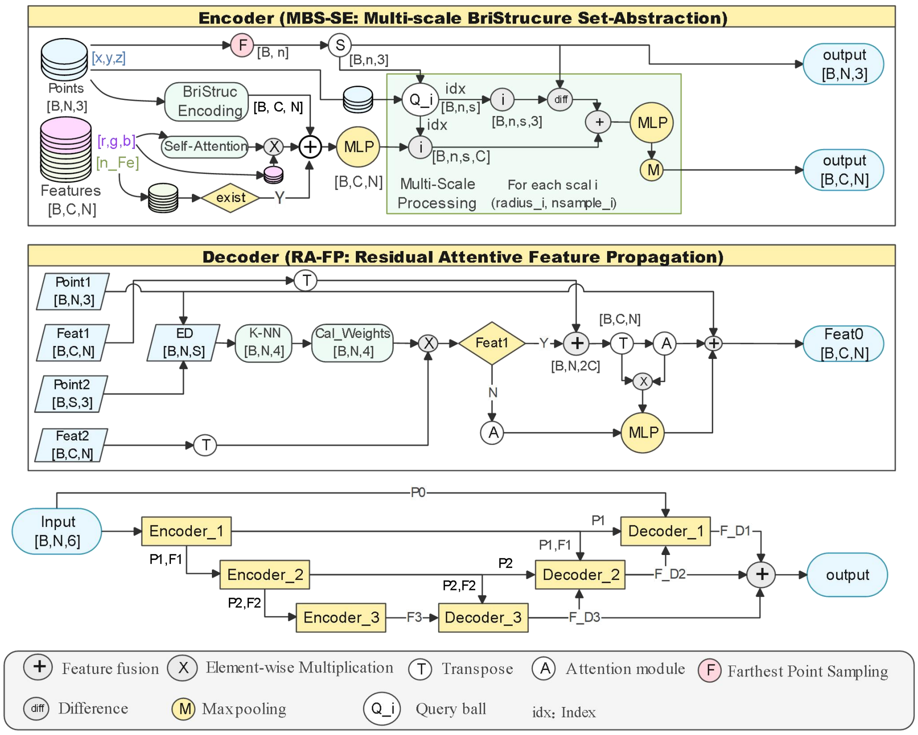


Fig. 1 Deep learning model architectures

As shown in Fig. 1, we propose a novel point cloud semantic segmentation framework specifically designed for bridge structures, consisting of a Multi-scale Bridge Structure Set Abstraction encoder (MBS-SE) and a Residual Attentive Feature Propagation decoder (RAFP).

The MBS-SE encoder effectively captures the geometric characteristics of bridge structures by first extracting structural features from the raw xyz coordinates. These structural features are then fused with RGB information through an attention-enhanced MLP module, allowing the network to leverage both geometric and visual cues. A key innovation in our encoder is the multi-scale processing approach, where features are sampled at different radius values to capture both local details and global context. To preserve spatial information, we implement residual connections between the original coordinates and processed features, along with difference operations that enhance the model's sensitivity to structural variations.

The RAFP decoder builds upon the PointNet feature propagation (FP) mechanism while incorporating attention modules to emphasize relevant features during the up-sampling process. Our framework employs a three-layer encoder-decoder architecture, where features from each decoder level are concatenated in the final layers. This hierarchical feature fusion strategy ensures that information captured at different scales is preserved throughout the network, leading to more accurate segmentation results. The residual connections within the decoder further facilitate gradient flow during training and enhance the model's ability to reconstruct fine-grained details of bridge components.

2.1 Data Pre-processing

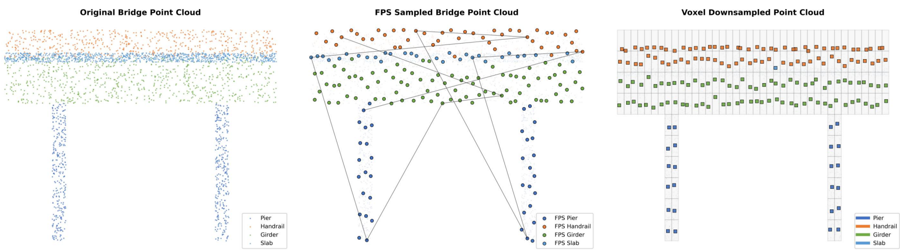


Fig. 2 Visualization of down sampling methods

In the data processing phase, the preprocessing stage initially involved applying voxel down sampling to the entire dataset to achieve uniform point cloud density. Region block sampling was implemented locally to preserve structural information. Furthermore, Farthest Point Sampling (FPS) was employed within the network architecture to effectively reduce the number of points while maintaining the structural characteristics of the bridge components (piers, parapets, girders, decks, etc.) as illustrated in the Fig. 2. This multi-level sampling strategy ensures computational efficiency without compromising the geometric integrity of critical structural elements.

2.2 BriStruc-Encoding

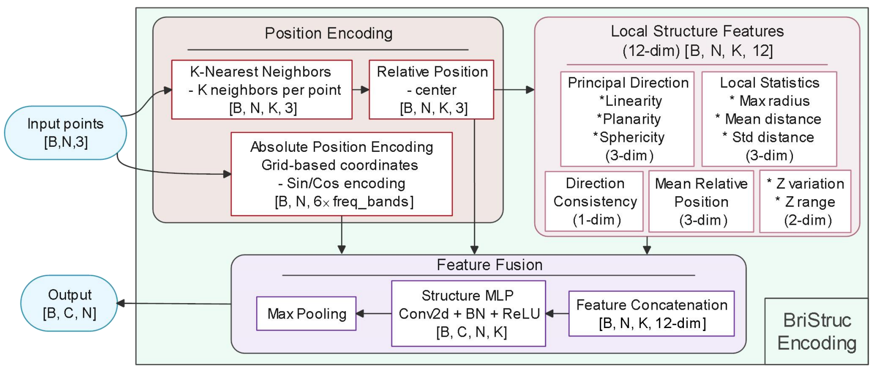
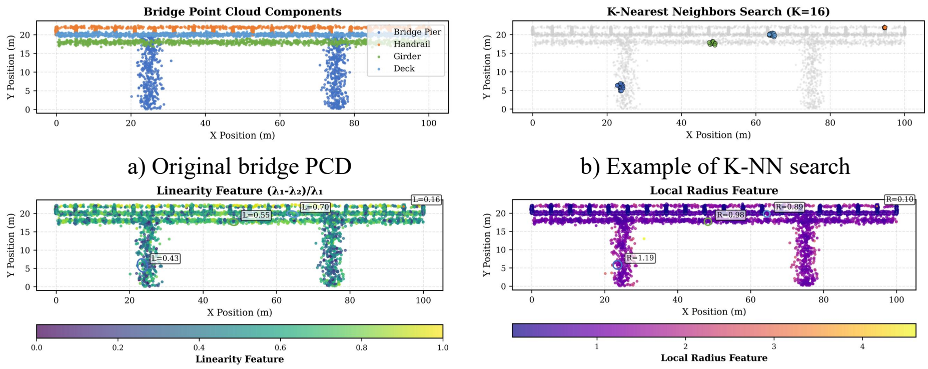


Fig. 3 Workflow of BriStruc-Encoding module

We propose the BriStruc-Encoding module to effectively capture bridge-specific structural characteristics in point clouds as shown in Fig. 3. This module processes input points [B,N,3] through two parallel pathways: position encoding and local structure feature extraction. The position encoding pathway utilizes K-Nearest

Neighbors to compute relative positions and grid-based coordinates for absolute position encoding. Concurrently, the local structure pathway extracts 12-dimensional features including principal direction attributes (linearity, planarity, sphericity) and statistical measurements (radius, distances, direction consistency, and height variations). These complementary features are fused through concatenation and processed by a structure-aware MLP with max pooling, generating an output representation [B,C,N] that preserves both positional and structural information essential for accurate bridge component segmentation.



a) Original bridge PCD

b) Example of K-NN search

c) Linearity feature of each point

d) Local radius feature of each point

Fig. 4 Visualization of feature extraction process

As illustrated in Fig. 4, the extracted features effectively differentiate between structural components. For instance, the linearity feature shows distinct values across different bridge elements: pier sections exhibit low linearity (L=0.43) due to their columnar structure, while deck sections demonstrate high linearity (L=0.55) reflecting their elongated geometry. Similarly, the local radius feature reveals characteristic patterns with bridge piers showing smaller radii (R=1.19) compared to the larger values observed in deck components (R=0.89). These quantified geometric attributes enable precise discrimination between structural elements, demonstrating the importance of numerical feature representation for accurate semantic segmentation of bridge point clouds.

2.3 Optimization Strategy

For model training, we employ the Adam optimizer with initial learning rate set to the configured value, momentum parameters β_1=0.9 and β_2=0.999, and weight decay of 1e-4 to prevent overfitting. A ReduceLROnPlateau scheduler monitors validation performance, reducing the learning rate by a factor of 0.1 when improvement plateaus for 5 consecutive epochs. The classification task is guided by a structure-oriented loss (SOL) function that incorporates spatial constraints defined by Structure-Oriented Concept(SOC) [21], enabling the model to leverage the inherent structural relationships between bridge components during training.

3 Experiment and results

3.1 Dataset preparation

Dataset 1 - publicly available dataset

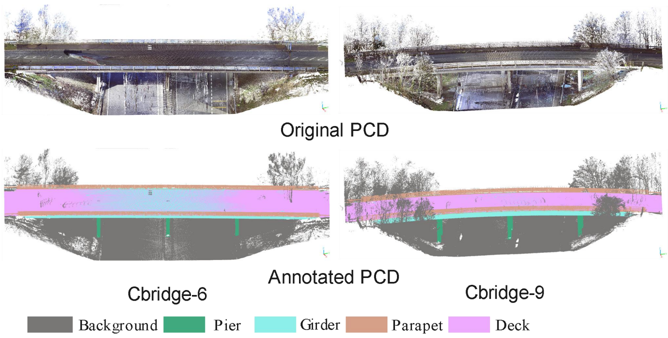


Fig. 5 Visualization of two randomly selected bridges from the benchmark dataset. Top row: original PCD; bottom row: manually annotated PCD.

The dataset-1 utilized in this study is a publicly available benchmark dataset from Lu et al. [22], accessible through Zenodo (https://zenodo.org/record/1240534). The dataset comprises PCD collected from ten highway bridges in Cambridgeshire, United Kingdom, using a FARO Focus 3D X330 terrestrial laser scanner. This dataset was divided into two parts: eight bridges were utilized for training the semantic segmentation model, whereas two bridges were randomly selected for framework validation and dimension estimation, as shown in Fig. 5. This random selection strategy was adopted to ensure unbiased evaluation and showcase the robustness and generalizability of our proposed framework. Similar to our in-house dataset, all point clouds in this benchmark dataset were manually annotated following a consistent five-class classification scheme: pier, girder, deck, parapet, and others.

3.2 Implementation

The effectiveness of the semantic segmentation model was assessed using three standard metrics: mean class accuracy (mAcc), intersection over union (IoU), and mIoU. These metrics offer a comprehensive understanding of the model's segmentation capabilities on both a global and class-specific basis.

$$mIoU = \frac{1}{M}\sum_{i=1}^{M} IoU_i \tag{1}$$

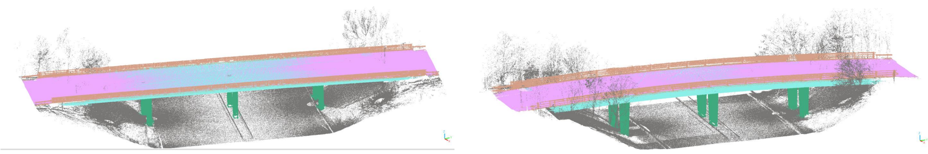
$$IoU_i = \frac{TP_i}{TP_i + FN_i + FP_i} \tag{2}$$

$$mAcc = \frac{1}{M}\sum_{i=1}^{M} \frac{TP_i}{N_i} \tag{3}$$

where M represents the total class number and N_i denotes the point count in class i. The evaluation terms include true positive (TP_i), indicating correctly classified points in class i; false negative (FN_i), representing points of class i incorrectly assigned to other classes; and false positive (FP_i), tallying points inaccurately classified as class i. The model was implemented using the PyTorch framework, incorporating PointNet++ with its multi-scale grouping strategy [18,19]. The proposed framework was implemented in Python using the scikit-learn and Open3D libraries. The training process utilized the Adam optimizer with an initial learning rate of 0.001 and a decay rate of 0.0001. Model training and evaluation were conducted on a Windows-based workstation equipped with an AMD Ryzen 9 9950X processor, 128GB RAM, and an NVIDIA GeForce RTX 4090 GPU.

3.3 Results and evaluation

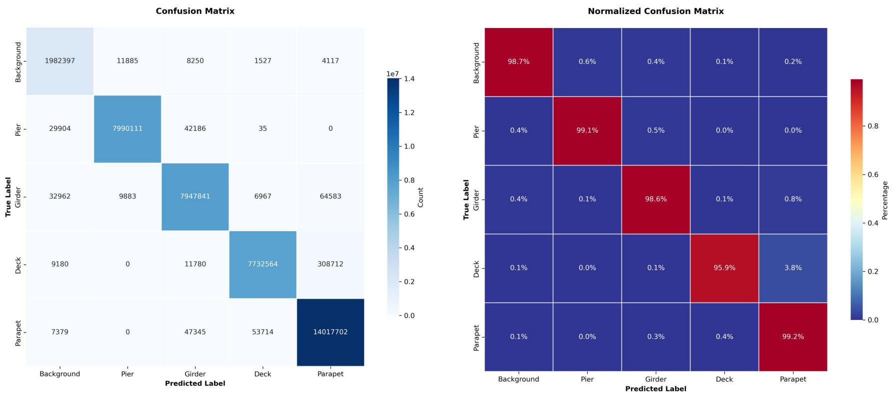
Fig. 6 shows the semantic segmentation results of our proposed method applied to two representative bridge PCDs (Bridge 6 and Bridge 9). The visualization effectively demonstrates the algorithm's capability to accurately classify different structural components, with distinct colors representing bridge decks (pink), piers (green), and other structural elements. The segmentation boundaries between components are clearly delineated, indicating the model's strong performance in distinguishing between adjacent structural elements with different geometric characteristics, even in complex bridge configurations.



a) Bridge 6 b) Bridge 9

Fig. 6 Semantic segmentation results

Fig. 7 presents both the raw and normalized confusion matrices of the segmentation results across all test datasets. The normalized confusion matrix (Fig. 7b) reveals the exceptional performance of our approach, with all structural components achieving classification accuracy exceeding 95%. Specifically, the model demonstrates remarkable precision in identifying parapets (99.2%), piers (99.1%), and background elements (98.7%), followed by girders (98.6%) and decks (95.9%). The minimal misclassification between categories is particularly noteworthy, with the highest confusion (3.8%) occurring between deck and parapet elements due to their occasional geometric similarity at intersection points. This comprehensive quantitative assessment confirms the robustness of our proposed BriStruc Encoding framework for bridge component segmentation, providing reliable structural identification necessary for subsequent engineering applications including deformation analysis and damage detection.



a) Confusion matrix b) Normalized confusion matrix

Fig. 7 Confusion matrix of total result

Table 1 presents a comprehensive quantitative evaluation. The overall performance metrics demonstrate the exceptional effectiveness of our approach across all bridge components. In total, our method achieves an impressive mean IoU of 96.58%, with particularly outstanding performance on pier structures (98.84%) and girder components (97.28%). The parapet and background elements show slightly lower but still excellent IoU values at 95.16% and 94.99%, respectively. Furthermore, the model exhibits remarkable consistency across other evaluation metrics, with an F1 score of 98.40%, and mean accuracy of 98.31%. These results validate the robustness of our proposed feature extraction and segmentation framework, providing reliable structural identification across different bridge configurations that can serve as a foundation for subsequent engineering analyses and applications.

Table 1 Quantitative evaluation of SS in test set (unit: %)

	IoU						F1 Score	Mean Acc.
	Back.	Pier	Girder	Parapet	Deck	Mean		
Bri-6	96.20	98.92	97.75	95.64	97.19	97.14	98.63	98.45
Bri-9	93.71	98.76	96.75	94.70	96.09	96.00	98.15	98.16
Total	94.99	98.84	97.28	95.16	96.65	96.58	98.40	98.31

4 Conclusions

This study presents a multi-scale spatial position encoding framework for semantic segmentation of bridge point clouds that effectively addresses the unique challenges associated with infrastructure-scale data processing. By incorporating the BriStruc-Encoding module, our approach successfully using both positional information and local structural features to distinguish between different bridge components with high accuracy. The integration of relative position encoding through K-nearest neighbors and absolute position encoding via grid-based coordinates, combined with eigenvalue-based geometric features, has proven highly effective for capturing the inherent structural patterns of bridge architectures.

Experimental results demonstrate the exceptional performance, achieving an overall mean IoU of 96.58% across all bridge components. The confusion matrix analysis further confirms the method's robustness, revealing classification accuracies exceeding 95% for all component with minimal misclassification between categories. This comprehensive quantitative assessment validates that our approach effectively overcomes the limitations of existing models by specifically addressing the intermediate scale and unique spatial characteristics of bridge structures.

Acknowledgements

This work was partially supported by the Council for Science, Technology and Innovation (CSTI), Cross-Ministerial Strategic Innovation Promotion Program (SIP), the 3rd period of SIP "Smart Infrastructure Management System" Grant Number JPJ012187 (Funding agency: Public Works Research Institute) and JSPS Grantin-Aid for Scientific Research Grant Numbers 21H01417, 22H01561, and 23H00198, and China Scholarship Council Grant Number 202306210048.

References

1. Collette Q, Sire S, Vermes WJ, Mesler VJ, Wouters I. Experimental investigations on hot-driven structural rivets in historical French and Belgian wrought-iron structures (1880s–1890s). Construction and Building Materials. 2014 Mar;54:258–69.
2. ASCE. Report card for America's infrastructure, American Society of Civil Engineers [Internet]. 2021 [cited 2024 Nov 13]. Available from: https://infrastructurereportcard.org/
3. MLIT. Road Maintenance in Japan: Problems and Solutions. Ministry of Land, Infrastructure, Transport and Tourism. [Internet]. Ministry of Land, Infrastructure, Transport and Tourism.; 2024. Available from: https://www.mlit.go.jp/road/road_e/pdf/RoadMaintenance.pdf
4. EC. Transport in the European Union: current trends and issues. European Commission. Directorate General for Mobility and Transport. [Internet]. LU: Publications Office; 2024 [cited 2024 Nov 13]. Available from: https://data.europa.eu/doi/10.2832/131741
5. Yamane T, Chen Y, Kubo S, Asano W, Katayama N, Iwaki I, Chun PJ. Development of an Inspection Information Management System for Infrastructures using Spherical Images. Artificial Intelligence and Data Science (in Japanese). 2025 Feb 27;(in press).
6. Lamas D, Justo A, Soilán M, Cabaleiro M, Riveiro B. Instance and semantic segmentation of point clouds of large metallic truss bridges. Automation in Construction. 2023 Jul;151:104865.
7. Dorafshan S, Maguire M. Bridge inspection: human performance, unmanned aerial systems and automation. J Civil Struct Health Monit. 2018 Jul;8(3):443–76.
8. Agnisarman S, Lopes S, Chalil Madathil K, Piratla K, Gramopadhye A. A survey of automation-enabled human-in-the-loop systems for infrastructure visual inspection. Automation in Construction. 2019 Jan;97:52–76.
9. Yamane T, Chun P jo, Dang J, Honda R. Recording of bridge damage areas by 3D integration of multiple images and reduction of the variability in detected results. Computer-Aided Civil and Infrastructure Engineering. 2023;38(17):2391–407.

10. Yamane T, Chun P jo, Honda R. Detecting and localising damage based on image recognition and structure from motion, and reflecting it in a 3D bridge model. Structure and Infrastructure Engineering. 2024 Apr 2;20(4):594–606.
11. Chun PJ, Yamane T, Maemura Y. A deep learning-based image captioning method to automatically generate comprehensive explanations of bridge damage. Computer-Aided Civil and Infrastructure Engineering. 2022;37(11):1387–401.
12. TSUJII J, GODA T, NAKANO M. Application of deep learning methods treating convolutional features of local geometry to point cloud analysis of civil infrastructures. Artificial Intelligence and Data Science. 2023;4(3):442–50.
13. Lu Y, Wang S, Fan S, Lu J, Li P, Tang P. Image-based 3D reconstruction for Multi-Scale civil and infrastructure Projects: A review from 2012 to 2022 with new perspective from deep learning methods. Advanced Engineering Informatics. 2024 Jan;59:102268.
14. Inadomi S, Chun P jo. A comparative study of projection-based vs. point-based point clouds segmentation for 3D bridge modelling. Structure and Infrastructure Engineering. 2025;(Accepted).
15. Yang T, Zou Y, Yang X, Del Rey Castillo E. Domain knowledge-enhanced region growing framework for semantic segmentation of bridge point clouds. Automation in Construction. 2024 Sep;165:105572.
16. Mansour M, Martens J, Blankenbach J. Hierarchical SVM for Semantic Segmentation of 3D Point Clouds for Infrastructure Scenes. Infrastructures. 2024 May 6;9(5):83.
17. Yang X, del Rey Castillo E, Zou Y, Wotherspoon L. Semantic segmentation of bridge point clouds with a synthetic data augmentation strategy and graph-structured deep metric learning. Automation in Construction. 2023 Jun 1;150:104838.
18. Qi CR, Su H, Kaichun M, Guibas LJ. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Honolulu, HI: IEEE; 2017 [cited 2024 Nov 14]. p. 77–85. Available from: http://ieeexplore.ieee.org/document/8099499/
19. Qi CR, Yi L, Su H, Guibas LJ. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space [Internet]. arXiv; 2017 [cited 2024 Nov 14]. Available from: http://arxiv.org/abs/1706.02413
20. Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM. Dynamic Graph CNN for Learning on Point Clouds. ACM Trans Graph. 2019;38(5):146:1-146:12.
21. Lin C, Abe S, Zheng S, Li X, Chun P jo. A structure-oriented loss function for automated semantic segmentation of bridge point clouds. Computer-Aided Civil and Infrastructure Engineering. 2025;40(6):801–16.
22. Lu R, Brilakis I, Middleton CR. Detection of Structural Components in Point Clouds of Existing RC Bridges. Computer aided Civil Eng. 2019 Mar;34(3):191–212.