HOKKAIDO UNIVERSITY

| | |
|---|---|
| Title | An Empirical Study of Bridge Point Cloud Semantic Segmentation Across Structural Spans |
| Author(s) | Sun, Zhongyi; Itakura, Kenta; Chen, Yu et al. |
| Relation | Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026), June 28 - July 1, 2026, Sapporo, Japan |
| Citation | Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026) |
| Issue Date | 2026-07-01 |
| Doc URL | https://hdl.handle.net/2115/100466 |
| Type | conference paper |
| File Information | C000189_1.pdf |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

*10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026)*,
28 June-01 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

# An Empirical Study of Bridge Point Cloud Semantic Segmentation Across Structural Spans

Zhongyi Sun[1], Kenta Itakura[2], Yu Chen[3], Chao Lin[1], and Pang-jo Chun[1]*

[1]*Department of Civil Engineering, School of Engineering, The University of Tokyo, Tokyo, Japan, 113-8656.*
*E-mail: zhongyis219@g.ecc.u-tokyo.ac.jp*
[2] *ImVisionLabs Inc., Tokyo, 113-0023, Japan, 113-0023.*
*E-mail: kentai@imvisionlabs.com*
[3]*Department of Global Environment and Disaster Prevention Sciences, Hirosaki University, Aomori, Japan, 036-8561.*
*E-mail: yu.chen@hirosaki-u.ac.jp*
**Corresponding Author: Pang-jo Chun (chun@g.ecc.u-tokyo.ac.jp)*

Semantic segmentation of bridge point clouds is important for automated inspection and digital asset management. Although Transformer-based methods perform well on benchmark datasets, the effect of bridge span scale on segmentation has not been systematically studied. This paper presents an empirical study of point cloud semantic segmentation across small-, medium-, and large-span bridges under a unified protocol. Using Point Transformer V3 (PTv3) with consistent preprocessing, training, and evaluation settings, we analyze segmentation performance for five structural components: background, piers/abutments, girders, decks, and superstructure accessories. Results show that performance varies across structural components and that the effect of span scale is non-uniform. Deck-related elements remain relatively stable across span scales, whereas piers and abutments show greater variability in large-span scenes. We also find that metrics computed on subsampled point clouds do not always match qualitative visual results, indicating that evaluation is influenced by sampling strategy and evaluation perspective. These findings show the importance of considering bridge span scale and structural characteristics when applying point cloud semantic segmentation to bridge inspection, and motivate further study of scale-aware modeling and evaluation for large infrastructure point clouds.

***Keywords***: Point cloud semantic segmentation, Bridge span scale, Point Transformer V3, Bridge inspection, Structural component recognition.

## 1 Introduction

The peak period of infrastructure construction in Japan occurred during the rapid economic growth era of the 1960s and 1970s. Infrastructure built during this period has now been in service for more than 50 years and is gradually entering an aging stage. As infrastructure ages, the probability of functional degradation and safety issues increases significantly. Consequently, how to effectively inspect, maintain, and manage aging infrastructure has become an urgent social challenge. This issue is particularly prominent in the field of bridges, and many economies that have experienced rapid development face similar problems. Meanwhile, traditional manual inspection methods are often time-consuming, inefficient, and costly. With the rapid advancement of artificial intelligence technologies, AI-assisted bridge inspection has gradually attracted increasing attention. In particular, the application of point cloud technologies in bridge inspection has gained significant interest in recent years. Among these, the accurate identification of structural components within bridge point cloud data is particularly important, as it serves as a prerequisite for tasks such as dimensional measurement and structural inspection based on point clouds. If component segmentation errors occur, they may lead to model inconsistencies or measurement inaccuracies. Therefore, improving the accuracy of structural component recognition in bridge point clouds is of great importance.

For example, in recent years various approaches have been proposed for structural component recognition in bridge point clouds. These include methods that generate training data using synthetic models (e.g., Jing et al. 2022; Rahman and Hoskere 2025), or digital twins (e.g., Yan and Hajjar 2021; Yan et al. 2025), approaches that combine multi-view images or multimodal information for segmentation (e.g., Song et al. 2025; Zeng et al. 2025), and techniques that incorporate geometric rules or domain knowledge from engineering structures to improve segmentation accuracy (e.g., Yang et al. 2022; Yang et al. 2024; Zhang et al. 2025). While these methods have achieved promising results for specific bridge types or application scenarios, challenges still remain when dealing with complex structural environments and large-scale point cloud data.

In recent years, research that explicitly models internal relationships within point clouds using Transformer architectures or graph-based structures has attracted increasing attention (Jing et al. 2024). These methods are able to capture long-range spatial dependencies in point clouds through self-attention mechanisms, making them well suited for structured environments such as bridges. Among these approaches, Point Transformer is a neural network architecture specifically designed for point cloud data (Zhao et al. 2021). By introducing self-attention mechanisms within local neighborhoods, it effectively models geometric relationships among points. Subsequently, Point Transformer V3 further improves the network architecture and computational efficiency, enabling more effective processing of large-scale point cloud scenes (Wu et al. 2024). Therefore, this method shows strong potential for applications in structured point cloud scenarios such as bridge infrastructure.

In practical engineering applications, bridge spans exhibit considerable diversity, ranging from small-span bridges of approximately 10 m to large-span bridges extending several hundred meters. The point cloud data of bridges with different spans show significant differences in the distribution of structural components, resulting in varying proportions of each component within the overall point cloud. Meanwhile, Transformer-based point cloud semantic segmentation methods tend to rely heavily on the density of the input point cloud. Therefore, under the same segmentation model, point cloud data from bridges with different spans may lead to variations in component recognition performance, and the model's sensitivity to different structural components may also vary accordingly. Such differences may not only affect the final semantic segmentation accuracy but may also influence parameter settings during the early stages of model training.

Based on the issues discussed above, this study aims to further investigate the sensitivity of point cloud semantic segmentation models to structural components of bridges with different span scales. Fig. 1 illustrates the overall research framework of this study. Specifically, Point Transformer V3 is applied to point cloud segmentation tasks for small-, medium-, and large-span bridges. This study systematically analyzes the influence of bridge span variation on semantic segmentation results. Furthermore, it examines the sensitivity of a single model to different structural components across bridges with varying spans, thereby providing insights and references for model parameter configuration and practical applications in bridge point cloud semantic segmentation.

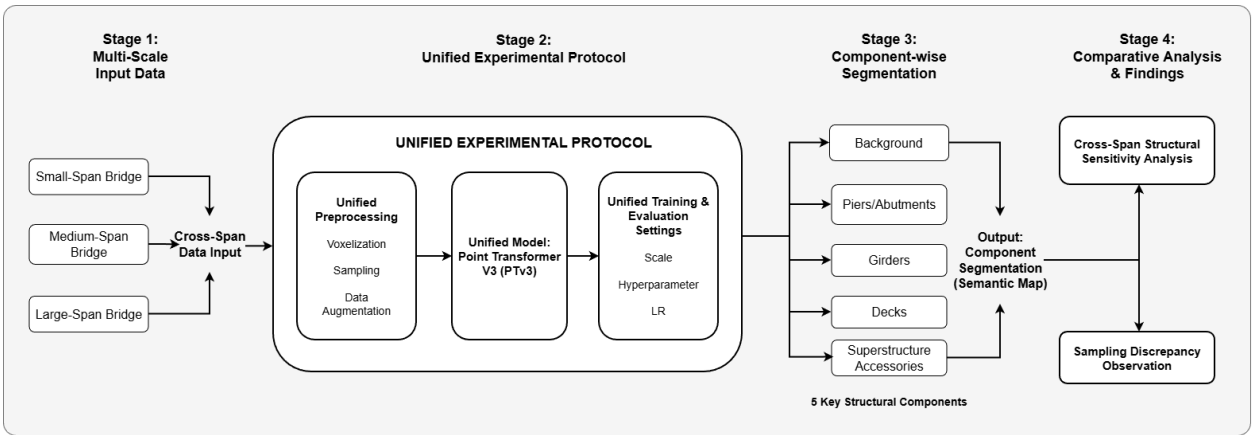


Figure 1. Research framework

## 2 Related Work

### *2.1 Point Cloud Semantic Segmentation*

Point cloud semantic segmentation is a core task in 3D scene understanding. Early deep learning methods such as PointNet and PointNet++ (Qi et al. 2017a; Qi et al. 2017b) established end-to-end learning on unordered point sets. More recent Transformer-based methods model long-range dependencies through self-attention and perform strongly on benchmark datasets. To reduce attention cost, many methods combine Transformers with hierarchical structures, local geometric encoding, or sparse convolution, improving scalability and efficiency (Zhou et al. 2022; Jing et al. 2024; Zhang et al. 2024).

Most existing methods are still evaluated on datasets with relatively consistent scene scales or controlled settings, so robustness to large structural-scale variation, especially in engineering point clouds, remains unclear.

### *2.2 Bridge and Infrastructure Point Cloud Segmentation*

For bridges and civil infrastructure, point cloud segmentation supports automated inspection, geometric measurement, and digital twin construction. Existing studies use synthetic point clouds or digital twins to reduce labeling cost (Lamas et al. 2024, Rahman and Hoskere 2025).

Other studies use multi-view projections or multimodal information to map point clouds into image or video representations and leverage mature vision models for high-accuracy segmentation (Inadomi and Chun 2025, Zeng et al. 2025). Several others design structure-aware pipelines for specific bridge types or components, including suspension, truss, and masonry arch bridges (Jing et al. 2022, Lamas et al. 2023, Park et al. 2025).

Heuristic and hybrid methods that incorporate geometric rules, topological constraints, or finite-element-derived knowledge have also been proposed to improve robustness and interpretability (Yan and Hajjar 2021, Yang et al. 2024, Chen et al. 2025). Recent work has likewise introduced local-global attention into bridge point cloud segmentation to better handle large scenes and complex spatial relations (Yang et al. 2022, Jing et al. 2024).

Although these studies report promising results, most emphasize algorithm design or application-level validation under fixed-scale conditions. Structural scale, especially bridge span variation, is usually treated as an implicit rather than explicit variable.

### *2.3 Generalization, Domain Shift, and Scale Variation in 3D Understanding*

Generalization under domain shift remains a major challenge in 3D point cloud segmentation because point density, sampling strategy, sensor modality, and scene scale can all affect performance (Xia et al. 2022, Kopiika et al. 2025). This problem is especially clear in infrastructure scenes, where components with the same label may vary substantially in size and point distribution. Existing work addresses it through multi-scale modeling, hierarchical representations, superpoint methods (Yang et al. 2022), and synthetic-data or digital-twin strategies (Jing et al. 2022). However, bridge span scale is still rarely analyzed explicitly.

Although substantial progress has been made in point cloud segmentation for bridge and infrastructure inspection (Sun et al. 2020, Di Mucci et al. 2024), the influence of bridge span scale has not been systematically examined. Existing studies typically use single bridge types, fixed-scale datasets, or mixed-scale data without isolating structural scale, limiting interpretability and comparability.

Small-, medium-, and large-span bridges differ markedly in component proportions, spatial configuration, point density, and occlusion. These differences may directly affect the learning and generalization behavior of segmentation models, yet the literature still lacks a unified protocol for comparing performance across span scales. A unified empirical study is therefore needed to clarify how bridge span scale affects segmentation performance and to identify scale-sensitive and scale-robust components. This gap motivates the present work.

## 3 Datasets

This study employs three bridge point cloud datasets representing small-, medium-, and large-span bridges. The small- and large-span datasets were collected by our team, including the Fukushima bridge dataset with 10 reinforced concrete bridges in Japan (10–20 m span) and a large-span urban viaduct dataset in Japan covering 10 bridges (100–200 m span). The medium-span dataset is derived from a public University of Cambridge dataset, containing 10 bridges with spans of 60–70 m. All point clouds were manually harmonized into five classes: background, pier/abutment, girder, deck, and superstructure accessories. The background class includes non-bridge elements (e.g., vegetation, ground, utilities, surrounding structures, and noise), improving realism for practical scenarios.

Annotations were performed manually with consideration of both geometry and structural configuration, ensuring consistent spatial and vertical relationships among components. Due to acquisition limitations, some scans are partially incomplete, particularly in lower structural regions or areas focusing on bridge-side accessories.

Table 1 summarizes dataset statistics and scene-level splits. The small- and medium-span datasets each include 10 scenes (7 train, 2 validation, 1 test), while the large-span dataset contains 25 scenes extracted from 10 bridges (17 train, 5 validation, 3 test). This bridge-level split enables fair evaluation of generalization across span scales.

Table 1. Dataset statistics and scene-level splits for bridges of different span scales

| Span Scale | Scenes (Train / Val / Test) | Total Scenes |
|---|---|---|
| Small-span | 7 / 2 / 1 | 10 |
| Mid-span | 7 / 2 / 1 | 10 |
| Large-span | 17 / 5 / 3 | 25 |

## 4 Model Used

### *4.1 Point Transformer V3*

Point Transformer V3 (PTv3) is adopted as the baseline model in this study. PTv3 extends Point Transformer to improve scalability and efficiency for large point clouds. Whereas earlier versions rely mainly on local attention, PTv3 uses a hierarchical architecture that integrates local and global context, allowing it to capture fine geometric detail while remaining efficient on large scenes. A key PTv3 component is efficient feature aggregation that combines sparse convolution with attention-based encoding. By organizing points hierarchically and progressively reducing spatial resolution, the model learns multi-scale geometric features across abstraction levels, substantially improving its ability to handle large scenes such as outdoor and infrastructure environments. PTv3 also includes architectural optimizations for training stability and efficiency. It therefore performs strongly on large-scale semantic segmentation benchmarks while maintaining relatively low computational cost, making it well suited to large and complex engineering point clouds.

### *4.2 Model Configuration*

To attribute performance differences to span-scale-related data characteristics rather than model design, all experiments were conducted under a unified Point Transformer V3 configuration. The task was formulated as a five-class semantic segmentation problem, including background, pier/abutment, girder, deck, and superstructure

accessories, with identical preprocessing, optimization, and evaluation protocols across all datasets. The PTv3 model was implemented using a standard configuration with 6 input channels and 64 backbone output channels. The encoder and decoder depths were set to (2, 2, 2, 6, 2) and (2, 2, 2, 2), respectively. For optimization, AdamW was used with an initial learning rate of 0.006 and a weight decay of 0.05, together with a OneCycleLR schedule. Each model was trained for 50 epochs with a batch size of 4. Mixed precision (FP16) was adopted to improve computational efficiency. The loss function combined cross-entropy loss and Lovasz loss with equal weights. In all experiments, the number of input points per sample was fixed to 80,000 to ensure consistent comparison across span scales.

### *4.3 Computational Environment*

All experiments were implemented using the Pointcept framework in a WSL2-based Linux environment. Training and evaluation were conducted with Python 3.10.19 and CUDA 12.9 on a workstation equipped with an NVIDIA GeForce RTX 5090 GPU (32 GB memory), 16 physical CPU cores (32 threads), and 33.4 GB system memory. Using a consistent software and hardware environment across all datasets minimized system-level variability and ensured a fair comparison of span-scale effects.

### *4.4 Evaluation Methods and Metrics*

This study uses mean Intersection over Union (mIoU), mean class accuracy (mAcc), and overall accuracy (allAcc) as semantic segmentation metrics. Overall accuracy is the proportion of correctly classified points, mean class accuracy averages per-class accuracy, and IoU measures overlap between predicted and ground-truth labels for each class. Mean IoU averages IoU across classes, reducing class-imbalance effects and providing a balanced view of performance across structural components.

## 5 Experiments

### *5.1 Overall Quantitative Results*

Table 2 summarizes the overall segmentation performance of PTv3 on the three span-scale bridge datasets, including mIoU, mAcc, and allAcc. For clearer comparison, the results are also visualized in Fig. 2 (left).

On the test sets, the medium-span dataset achieved the highest mIoU of 0.8779, followed by the small-span dataset (0.8196) and the large-span dataset (0.7723). Similar trends are observed for mAcc and allAcc. These results indicate that performance varies across span scales even under a unified experimental protocol, but not monotonically with scale.

Table 2. Test performance and class-wise IoU across span-scale bridge datasets

| Dataset | Test mIoU | Test mAcc | Test allAcc | Background | Pier/abutment | Girder | Deck | Accessories |
|---|---|---|---|---|---|---|---|---|
| Small-span | 0.8196 | 0.8774 | 0.9060 | 0.8199 | 0.5798 | 0.8372 | 0.9364 | 0.9248 |
| Medium-span | 0.8779 | 0.9138 | 0.9533 | 0.9388 | 0.8778 | 0.8980 | 0.8677 | 0.8072 |
| Large-span | 0.7723 | 0.8850 | 0.8925 | 0.9257 | 0.3764 | 0.7315 | 0.9254 | 0.9028 |

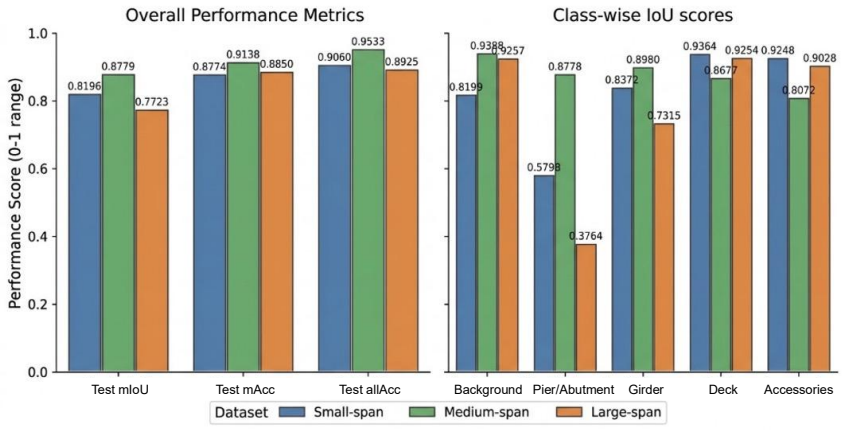


Figure 2. Performance comparison across bridge span scales

### *5.2 Component-wise Test-set Analysis*

Table 2 further reports class-wise IoU on the test sets to examine how different structural components respond to span variation, and the trends are illustrated in Fig. 2 (right).

Deck segmentation remained relatively stable across span scales, with consistently high IoU values. In contrast, pier/abutment showed the largest fluctuation, decreasing significantly from 0.8778 on the medium-span dataset to

0.3764 on the large-span dataset. Girder performance also degraded on the large-span dataset, suggesting that increased span and scene complexity negatively affect the recognition of elongated structural components.

### *5.3 Qualitative Analysis of Prediction Results*

Figure 3 shows representative prediction visualizations for the small-, medium-, and large-span test sets. PTv3 recovered the main bridge layout and separated the major structural regions in all three cases. Large continuous structures such as decks and girders were generally segmented coherently, although errors remained around structurally complex or proportionally small components, especially in the large-span case.

Another observation is that scores computed on sampled or evaluation-specific point subsets do not always fully reflect visual plausibility. In some cases, the predicted structure remains qualitatively reasonable even when the corresponding quantitative score is lower. This suggests that evaluation perspective and point sampling strategy can influence performance interpretation, especially in large infrastructure scenes.

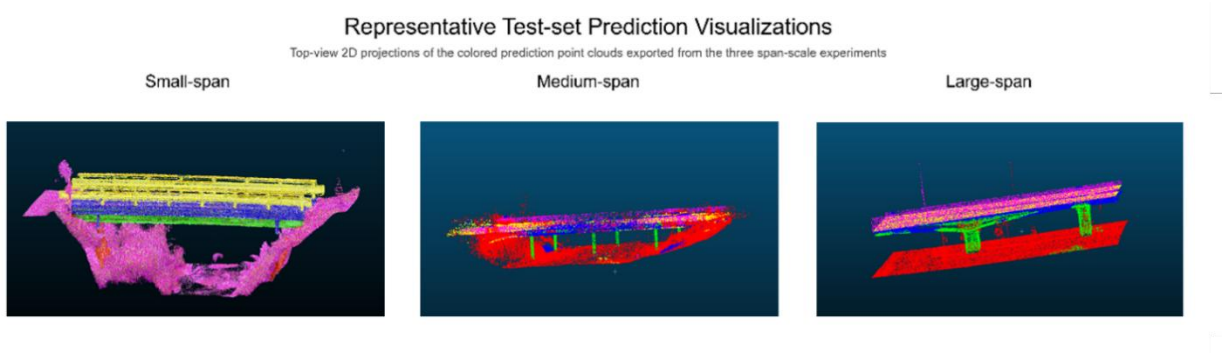


Figure 3. Cross-span qualitative visualization of bridge point cloud segmentation results

## 6 Discussion

These results have several implications for bridge point cloud semantic segmentation. First, even with the same Transformer-based architecture and training protocol, performance still varies markedly across bridge span scales. Span scale therefore acts not only as a geometric descriptor but also as a proxy for differences in scene composition, structural proportion, and spatial organization. Second, the influence of span scale is not uniform across semantic categories. Deck regions remain comparatively stable, whereas pier/abutment segmentation varies much more. This matters in engineering practice. Lower robustness on such components may accumulate errors in downstream tasks such as structural measurement, model reconstruction, and defect localization. Third, quantitative values from evaluation subsets do not always match qualitative structural consistency. This is especially relevant in bridge-scale scenes, where sampling strategy, fragment-based inference, and scene-size variation may alter the relation between point-wise metrics and perceived segmentation quality. Future studies should therefore complement conventional point-wise metrics with more structure-aware criteria.

Several limitations should also be noted. The datasets for different span scales were not perfectly matched in scene count and structural diversity, and the test sets were relatively small, especially for the small- and medium-span datasets. The observed differences should therefore be interpreted as empirical evidence under the current benchmark rather than strict proof that span scale alone determines segmentation difficulty. In addition, this study considered only one strong baseline, PTv3, so it remains unclear whether the same pattern holds for other architectures.

Several future directions follow from these observations: scale-aware sampling or fragment design for large bridge scenes, category-balanced optimization for important but underrepresented components such as piers/abutments, and evaluation methods that incorporate structural consistency and engineering semantics in addition to point-wise overlap. These directions may improve the reliability of AI-assisted bridge inspection.

## 7 Conclusion

This paper presented an empirical study of bridge point cloud semantic segmentation across structural spans using Point Transformer V3 under a unified protocol. Controlled experiments on small-, medium-, and large-span bridge datasets were used to examine how span-scale-related data variation affects segmentation performance for five key bridge components. Segmentation performance differed across span scales, but the effect was not uniform. The medium-span dataset achieved the best overall performance, whereas the large-span dataset showed the largest validation-to-test drop. At the component level, deck regions were relatively stable, while pier/abutment segmentation showed much greater variability, especially in large-span scenes. The study also found that quantitative scores on sampled point sets do not always align with qualitative prediction appearance, indicating that interpretation depends partly on evaluation perspective and sampling strategy.

Overall, the results indicate that bridge span scale should be considered explicitly when designing, training, and evaluating point cloud semantic segmentation systems for infrastructure inspection. The study provides an empirical reference for future work on scale-aware modeling, balanced optimization, and structure-aware evaluation in bridge point cloud understanding.

## References

Chen, Y., Lin, C., Li, X., Kubo, S., Yamane, T., Chun, P.-J., Automated dimension estimation of bridge components using semantic segmentation and geometric fitting of point cloud data, Eng. Struct., 342, 120837, 2025.
Di Mucci, V. M., Cardellicchio, A., Ruggieri, S., Nettis, A., Renò, V., Uva, G., Artificial intelligence in structural health management of existing bridges, Autom. Constr., 167, 105719, 2024.
Inadomi, S., Chun, P.-J., A comparative study of projection-based and point-based point cloud segmentation for 3D bridge modeling, Struct. Infrastruct. Eng., 1–13, 2025.
Jing, S., Cha, G., Maru, M. B., Yu, B., Park, S., Improved building MEP systems semantic segmentation in point clouds using a novel multi-class dataset and local–global vector transformer network, J. Build. Eng., 96, 110311, 2024.
Jing, Y., Sheil, B., Acikgoz, S., Segmentation of large-scale masonry arch bridge point clouds with a synthetic simulator and the BridgeNet neural network, Autom. Constr., 142, 104503, 2022.
Jing, Y., Sheil, B., Acikgoz, S., A lightweight Transformer-based neural network for large-scale masonry arch bridge point cloud segmentation, Comput.-Aided Civ. Infrastruct. Eng., 39(16), 2427–2438, 2024.
Kopiika, N., Karavias, A., Krassakis, P., Ye, Z., Ninic, J., Shakhovska, N., Argyroudis, S., Mitoulis, S.-A., Rapid post-disaster infrastructure damage characterisation using remote sensing and deep learning technologies: A tiered approach, Autom. Constr., 170, 105955, 2025.
Lamas, D., Justo, A., Soilan, M., Cabaleiro, M., Riveiro, B., Instance and semantic segmentation of point clouds of large metallic truss bridges, Autom. Constr., 151, 104865, 2023.
Lamas, D., Justo, A., Soilan, M., Riveiro, B., Automated production of synthetic point clouds of truss bridges for semantic and instance segmentation using deep learning models, Autom. Constr., 158, 105176, 2024.
Park, Y., Lee, J., Han, K., Yoon, H., Automated digital transformation for pedestrian suspension bridges using hybrid semantic structure from motion, Autom. Constr., 176, 106232, 2025.
Qi, C. R., Su, H., Mo, K., Guibas, L. J., PointNet: Deep learning on point sets for 3D classification and segmentation, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 652–660, 2017a.
Qi, C. R., Yi, L., Su, H., Guibas, L. J., PointNet++: Deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst., 30, 2017b.
Rahman, A. U., Hoskere, V., Instance segmentation of reinforced concrete bridge point clouds with transformers trained exclusively on synthetic data, Autom. Constr., 173, 106067, 2025.
Song, J., Zhou, X., Xu, C., Xu, S., Shu, J., Training-free automatic instance segmentation of girder bridge point cloud via large model fusion with reverse entity modelling verification, Autom. Constr., 179, 106484, 2025.
Sun, L., Shang, Z., Xia, Y., Bhowmick, S., Nagarajaiah, S., Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection, J. Struct. Eng., 146(5), 04020073, 2020.
Wu, X., Jiang, L., Wang, P. S., Liu, Z., Liu, X., Qiao, Y., Zhao, H., Point Transformer V3: Simpler, faster, stronger, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 4840–4851, 2024.
Xia, T., Yang, J., Chen, L., Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning, Autom. Constr., 133, 103992, 2022.
Yan, J., Zhang, H., Ai, Q., Xu, Y., Yang, J., Meng, W., Bao, T., Digital twin-based bridge geometric quality inspection using knowledge mapping and data-driven method, Autom. Constr., 178, 106383, 2025.
Yan, Y., Hajjar, J. F., Automated extraction of structural elements in steel girder bridges from laser point clouds, Autom. Constr., 125, 103582, 2021.
Yang, T., Zou, Y., Yang, X., Castillo, E. D. R., Domain knowledge-enhanced region growing framework for semantic segmentation of bridge point clouds, Autom. Constr., 165, 105572, 2024.
Yang, X., Castillo, E. D. R., Zou, Y., Wotherspoon, L., Tan, Y., Automated semantic segmentation of bridge components from large-scale point clouds using a weighted superpoint graph, Autom. Constr., 142, 104519, 2022.
Zeng, Y., Huang, F., Xiong, G., Ma, X., Peng, Y., Yang, W., Liu, J., Bridge point cloud semantic segmentation based on view consensus and cross-view self-prompt fusion, Autom. Constr., 171, 106003, 2025.
Zhang, M., Wang, L., Han, S., Wang, S., Li, H., Deep learning framework with local sparse Transformer for construction worker detection in 3D with LiDAR, Comput.-Aided Civ. Infrastruct. Eng., 39(19), 2990–3007, 2024.
Zhang, T., Chen, H., Li, P., Li, H., Two-stage bridge point cloud segmentation by fusing deep learning and heuristic methods, Measurement, 250, 117125, 2025.
Zhao, H., Jiang, L., Jia, J., Torr, P. H., Koltun, V., Point Transformer, Proc. IEEE/CVF Int. Conf. Comput. Vis., 16259–16268, 2021.
Zhou, Y., Ji, A., Zhang, L., Sewer defect detection from 3D point clouds using a Transformer-based deep learning model, Autom. Constr., 136, 104163, 2022.