

## Bridge inspection using image–point cloud fusion with image filtering, damage detection and 3D registration

Chao Lin <sup>a</sup>, Yu Chen <sup>a,b</sup>, Kenta Itakura <sup>c</sup>, Shreejan Maharjan <sup>a</sup>, Pang-jo Chun <sup>a,b</sup>, <sup>\*</sup>

<sup>a</sup> Department of Civil Engineering, The University of Tokyo, Tokyo, 113-8656, Japan

<sup>b</sup> Institute of Engineering Innovation, The University of Tokyo, Tokyo, 113-8656, Japan

<sup>c</sup> ImVisionLabs Inc., Tokyo, 113-8485, Japan

### ARTICLE INFO

#### Keywords:

Automated bridge inspection  
Damage detection  
Damage location registration  
Image and point cloud fusion  
2D–3D bidirectional projection

### ABSTRACT

Complex image backgrounds often compromise the reliability of damage detection. In bridge inspection, a further challenge lies in accurately recording and localizing the detected damage onto a 3D model. Based on image and point cloud data (PCD) fusion, this paper proposes a five-step methodology for detecting bridge damage and registering it on a 3D model. High-quality images and PCD files are simultaneously collected using a LiDAR 3D camera with their relationships clearly recorded. The complete bridge PCD is segmented and subsequently utilized to select images containing needed components and filter out the background via 3D-to-2D projection. Damage is detected from background-filtered images and then registered on the bridge PCD through 2D-to-3D projection. An experiment conducted on an actual bridge validated the feasibility of the proposed framework, confirming that the methodology not only produces clear and intuitive 3D visualizations of damage but also effectively supports detailed inspection and maintenance tasks.

### 1. Introduction

The significance of bridges in the modern transportation network is well known to the public [1,2], but the challenges they face, such as traffic load, environmental impact and natural disasters, etc. have always been neglected or underestimated. Regular inspections are thus needed to maintain the bridges in good condition for continuous service [3,4] where detailed information on damage should be recorded, including types and locations. Specifically, multiple types of damage need to be included for more comprehensive and reliable inspections (e.g. cracking, corrosion, spalling concrete and exposed rebar [5]) and the damage location should be precisely recorded to support subsequent maintenance tasks. To improve the efficiency of the inspection process, manual inspections have been gradually replaced by automatic methods, among which deep learning (DL)-based methods have been broadly developed [6].

Images have become the most widely used type of data for damage detection with this trend [7,8]. However, images alone have inherent limitations for inspection tasks due to the nature of the data they represent. Background elements may be misidentified as damage, leading to false positives and affecting the accuracy of damage detection. Additionally, damage locations can be non-intuitive when images are

selected as the medium for damage recording, as they often lack sufficient spatial context and orientation. Point cloud data (PCD) [9] offers opportunities to mitigate these limitations when used collaboratively with images. Images represent dense 2D data that is rich in texture and color but lacks depth information, which makes accurate spatial reasoning challenging. In contrast, PCD consists of sparse 3D data that provides depth and spatial information but lacks appearance data and tends to miss fine details or small objects. Therefore, image-PCD fusion enables a more comprehensive and accurate understanding of targets by using the complementary information from both modalities. Furthermore, PCD serves as an effective platform for recording damage due to its capability for 3D model reconstruction [10,11].

Image-PCD fusion requires high-quality data with clear and accurate relationships between the data elements. Most studies only acquire images or PCD due to equipment limitations and obtain the other through post-processing, which tends to be low-quality. Structure from motion (SfM) can be utilized to get photogrammetric PCD from images [12,13], but this process typically requires a large number of images and the resulting PCD is prone to significant errors. Moreover, a major and critical limitation is that SfM-based approaches can only generate PCD without an absolute scale [14]. This PCD is reconstructed within a relative coordinate system, making it scale-ambiguous. As a result, scale-ambiguous PCD is not applicable to advanced tasks such as 3D

\* Corresponding author at: Department of Civil Engineering, The University of Tokyo, Tokyo, 113-8656, Japan.  
E-mail address: [chun@g.ecc.u-tokyo.ac.jp](mailto:chun@g.ecc.u-tokyo.ac.jp) (P.-j. Chun).

geospatial visualization, bridge information modeling, digital twinning, and finite element analysis, etc. PCD can be projected onto a 2D plane by specifying viewpoints and projection methods to generate images [15–17]. However, the resulting images tend to be low-resolution and lack sufficient detail for accurate damage detection. A handful of studies have collected both types of data with high quality using correspondingly appropriate equipment (e.g. images obtained using a camera and PCD using a terrestrial laser scanner (TLS)). However, it is challenging to establish the precise relationship between images and PCD. The registration process is computationally intensive due to the large volume of high-resolution images and dense PCD, and it typically requires a substantial number of pre-installed ground control points. These factors not only increase operational complexity but also reduce efficiency. Moreover, distinct alignment errors still remain despite the labor-intensive and complicated procedures. Such errors can propagate through subsequent steps, negatively impacting the reliability of inspection tasks. Therefore, appropriate data organization forms are needed where high-quality images and PCD with registered reference frames are collected simultaneously and stored in a systematic arrangement to clearly reflect their relationship.

This paper proposes a methodology for damage detection and location registration in 3D models for bridge inspection based on a novel organizational form of image-PCD fusion. A LiDAR 3D camera was used to simultaneously collect high-resolution images and absolute-scale PCD. E57 was selected as the file format to store the collected data, and it explicitly records the relationship between the two modalities. Damage was detected from images that were background-filtered through 3D-to-2D projection between images and PCD after semantic segmentation (SS). Finally, the detected damage was recorded into PCD through 2D-to-3D projection, which offers not only a 3D environment for intuitive location visualization, but also a versatile platform for post-inspection tasks. A case study on a real-world bridge demonstrated the practical applicability of the proposed methodology, with corrosion selected as a representative damage type to rigorously validate the effectiveness of the framework. Background elements in the images were carefully filtered out, enabling precise damage detection with the help of the established effective detection model. Moreover, the identified damage was accurately mapped onto different components of the bridge PCD, facilitating intuitive documentation and visualization of the structural issues.

The principal contributions of this paper are:

- (1) Novel multimodal fusion framework: a unified approach that efficiently and seamlessly integrates 2D images with 3D PCD, achieving precise and repeatable 2D–3D bidirectional projections through fixed calibration parameters.
- (2) Enhanced damage detection: a background-filtering mechanism that improves the signal-to-noise ratio in damage recognition, resulting in higher detection accuracy compared to conventional methods.
- (3) Application-ready output: the absolute-scale PCD with registered damage to support advanced management tasks, enhancing improved decision-making and optimized long-term infrastructure management.
- (4) Practical and reproducible system design: an inspection framework designed for ease of operation and reproducibility, relying on standard, well-established algorithms to ensure straightforward deployment in real-world bridge inspection scenarios.

The remainder of this paper is organized as follows. Section 2 reviews the applications of image-PCD fusion in civil engineering, the current research state of image-based damage detection, and the multiple mediums used to record damage. Section 3 introduces the methodology framework of this study, which features an innovative workflow for damage detection and location registration based on image-PCD fusion. Section 4 presents the experimental results of the proposed methodology on a real bridge, demonstrating its feasibility for efficient bridge inspection. Section 5 provides conclusions, limitations and potential future research directions for this work.

## 2. Literature review

To investigate the workflow of damage detection and location registration using image-PCD fusion, this study reviews three key aspects. First, state-of-the-art applications of image-PCD fusion in civil engineering – particularly those related to inspection – are surveyed to explore its advantages and potential. Second, cutting-edge image-based damage detection methods are analyzed to support the development of an efficient detection model. Third, various platforms for damage location registration are compared to identify the most suitable option for bridge inspection tasks.

### 2.1. Image-PCD fusion in inspection

The combined information from both data formats demonstrated that image-PCD fusion is ideal for inspection tasks, as the complementary nature of image texture and color with PCD geometry and depth provides a more comprehensive understanding of the scene. Among studies dedicated to inspection tasks utilizing image-PCD fusion, most acquired only images and generated photogrammetric PCD afterwards through SfM. The image-based PCD segmentation is a relatively primary application in this aspect. The images are processed to generate masks of certain categories that are subsequently projected onto the photogrammetric PCD to achieve PCD segmentation. For example, Saovana et al. [18] proposed an image-based PCD instance segmentation method which was demonstrated as feasible to classify photogrammetric PCD of motorway pier and girder bridge samples. Ma et al. [19] presented a 3D segmentation and reconstruction system for pipeline PCD generation with marked defects. Two DL models were proposed to detect defects and generate depth maps from images. Then the pipeline PCD was reconstructed through a 2D-to-3D mapping method with outlier removal. Xiao et al. [20] utilized photogrammetric PCD to improve UAV-based bridge inspection. The bridge PCD was established through SfM and then segmented for 2D region of interest (ROI) extraction. After ROI extraction, more than half of the grid cells were filtered as background and the crack identification accuracy was thus improved.

Gradually, an increasing number of studies chose to collect images and PCD separately using corresponding equipment to ensure higher-quality data. Li et al. [21] presented a UAV-based bridge inspection method to detect pits and cracks on a bridge surface. A UAV was equipped with a camera and a LiDAR to acquire images and PCD simultaneously. Images and PCD were fused after complicated data processing to support damage detection with acceptable errors. Hu et al. [22] proposed a structural 3D crack detection robot system that integrates LiDAR and depth cameras. Experimental results demonstrated that the system can automatically identify and locate cracks, and ultimately obtain 3D crack geometric size information with high accuracy. Similarly, Ge and Sadhu [23] integrated a LiDAR and a camera on an unmanned ground vehicle to achieve structural mapping, damage localization, quantification, and 3D visualization. The image-PCD fusion is not restricted to equirectangular images, 360° panoramic images can also be an alternative. Chow et al. [24] developed a mobile data collection system that comprised a 360° camera and a LiDAR. The collected data were associatively processed for defect detection, scene reconstruction, defect assessment, and data integration.

Image-PCD fusion has demonstrated promising performance in infrastructure inspection. As the demand for more complex and precise inspection tasks continues to grow, acquiring both image and PCD data using specialized equipment will become the mainstream trend since high-quality data serves as the foundation for achieving accurate and reliable results. Furthermore, the precise alignment and correlation between images and PCD play a pivotal role in ensuring seamless integration, which will remain a key point in future studies and applications.

## 2.2. Image-based bridge damage detection

Prior studies predominantly focused on one particular defect type, among which the crack is the most popular. Deng et al. [25] presented a crack-boundary refinement framework (CBRF) for crack segmentation from high-resolution (HR) images, which includes three main components: a triple-scale feature extraction module, a multiscale repair fusion module comprising global and local joint operations, and an active boundary loss. According to experiments, CBRF outperformed conventional DL-based architectures by meticulously segmenting cracks from HR images, thus improving the safety and efficiency of UAV-based bridge detection. Subsequently, Chu et al. [26] and Chu and Chun [27] further enhanced the performance of HR crack image segmentation by upgrading existing modules and incorporating more DL techniques. While maintaining high detection accuracy, enhancements to the method improved inspection efficiency by allowing flexible distances between the UAV and targets. Chun and Kikuta [28] proposed a self-training framework for unsupervised domain adaptation in crack segmentation. Bayesian neural networks were incorporated to estimate the uncertainty of pseudo-labels, and spatial priors of cracks were introduced for screening noisy labels. As a result, the presented method improved the F1 score and achieved a performance comparable to supervised models.

Corrosion was also referred to in image-based bridge inspection. Katsamenis et al. [29] employed a U-Net model to simultaneously localize and classify corrosion, and recognize rust grade from RGB images on metal structures. Hattori et al. [30] developed a DL-based method to automatically measure the position and area of corroded parts on various bridge components, such as girders, bolts and cables.

A growing body of research has focused on multi-type damage detection. Huang et al. [31] presented a DL-based approach for surface damage detection and segmentation in various bridge components covering five major categories of surface defects, namely spalling, cracks, corrosion, clogging and honeycombs. Santos and Carvalho [32] introduced an approach for object detection and instance segmentation to identify multi-class structural damage, including cracks, scaling, rust stains, exposed reinforcement, and efflorescence. You Only Look Once (YOLO) [33] was utilized for object detection and the Segment Anything Model (SAM) [34] for instance segmentation. Furthermore, three modified detection layers were fine-tuned to detect small, medium, and large damage, respectively. As a result, a wide range of both minor and major damage was detected with high accuracy.

In addition to damage detection, images can also be utilized for damage diagnosis. For instance, image-text models can be employed to analyze visual data in conjunction with textual descriptions, enabling more comprehensive assessments of damage severity, type, and possible causes. Chun et al. [5] developed a DL-based image captioning method to automatically generate explanatory texts concerning the location and type of bridge damage. The model achieved reliable and comprehensive results by considering 27 types of damage (e.g. cracking, corrosion, spalling concrete) and 38 types of members (e.g. main girders, floor slabs). Subsequently, Chun et al. [35] developed a web system to support real-time utilization of the explanatory text generation model during bridge inspections. The proposed method enables practitioners with different levels of technical expertise to understand the bridge damage in text form. Kunlamai et al. [36] explored the application of visual question answering in bridge inspection. When feeding in an input image, the fine-tuned model could provide natural language answers to three types of questions (i.e., yes/no, member-class, and damage-class questions) with reliable accuracy.

Damage detection models capable of identifying and classifying multiple types of damage offer more comprehensive and accurate assessments. However, the scarcity of training data and the significant computational burden remain major challenges. To address these limitations, the development of powerful yet efficient models that are easier to fine-tune will become increasingly valuable, enabling broader applications and improved accessibility in real-world scenarios.

## 2.3. Damage location registration

Damage information can be recorded in multiple mediums, such as text, image, PCD, the mesh and finite element method (FEM) model, the building information model (BIM), digital twin (DT).

Li et al. [21] marked detected pits and cracks with textual depth information in a planar panoramic image of the bridge deck. Chen et al. [14] reconstructed the detected defects in the form of PCD. Subsequently, the 3D defect points were successfully integrated into BIM with high accuracy. Ge and Sadhu [23] registered the detected damage in a sparse PCD file, where points corresponding to segmented damage areas were marked in red, while all other points remained white. The marked points in the PCD accurately aligned with the actual damage and were utilized for damage size measurements.

Kong et al. [37] updated the geometry of concrete components in an FEM model by removing concrete elements affected by structural damage. Similarly, Zhang et al. [38] selected the FEM model as the platform for 3D geometric updating. By performing local surface mesh reconstruction and surface cutting, the damage information was accurately synchronized. Yamane et al. [39] integrated and recorded the bridge damage detected from images into a 3D pre-built triangular mesh model. The damage-detected results are reflected in corresponding meshes, and the mesh model utilizes different colors to distinguish the probability of damage in different regions.

Chow et al. [24] integrated the inspection results into existing BIM files. Defect information and image data were linked to structural elements in BIM to enhance facility management. Gao et al. [40] updated local damage into a DT based on PCD. Moreover, FEM and BIM were also synchronized using the updated DT.

Compared to text or image, 3D models were more widely favored for their ability to provide intuitive and immersive visualizations. Among various 3D data representations, PCD has been proven to be highly versatile, serving as a foundational format for constructing functional models such as BIM, the FEM model, and DT. Given its adaptability and comprehensive spatial representation, utilizing PCD as a platform for recording and managing structural damage is an effective and practical approach.

## 3. Methodology

The framework of the proposed methodology is illustrated in Fig. 1. First, high-quality images and PCD are collected simultaneously using a LiDAR 3D camera and then extracted, and clear relationships between them are shown. Second, the collected PCD is segmented through an efficient two-stage PCD SS framework. Third, an algorithm is designed to select the images that contain needed bridge components and filter out their backgrounds through 3D-to-2D projection between the segmented PCD and images. Fourth, a lightweight yet efficient DL model is established for damage detection. Finally, a process is implemented to register the detected damage on the bridge PCD based on the principles of 2D-to-3D projection.

Notably, the entire framework runs on a PC equipped with an Intel Core i9-14900KF CPU (clocked at 3.20 GHz), 128 GB of RAM, and an NVIDIA GeForce RTX 4090 GPU.

### 3.1. Data preparation

Accurate data preparation forms the foundation for reliable bridge damage detection and registration. This process involves organizing raw sensor outputs into a unified format, establishing precise geometric relationships between 2D images and 3D PCD, and calibrating the equipment to minimize 2D–3D bidirectional projection errors. By systematically addressing file formats, projection principles, and calibration procedures, the data preparation stage ensures that subsequent segmentation, detection, and registration steps are performed on high-quality and well-aligned datasets.

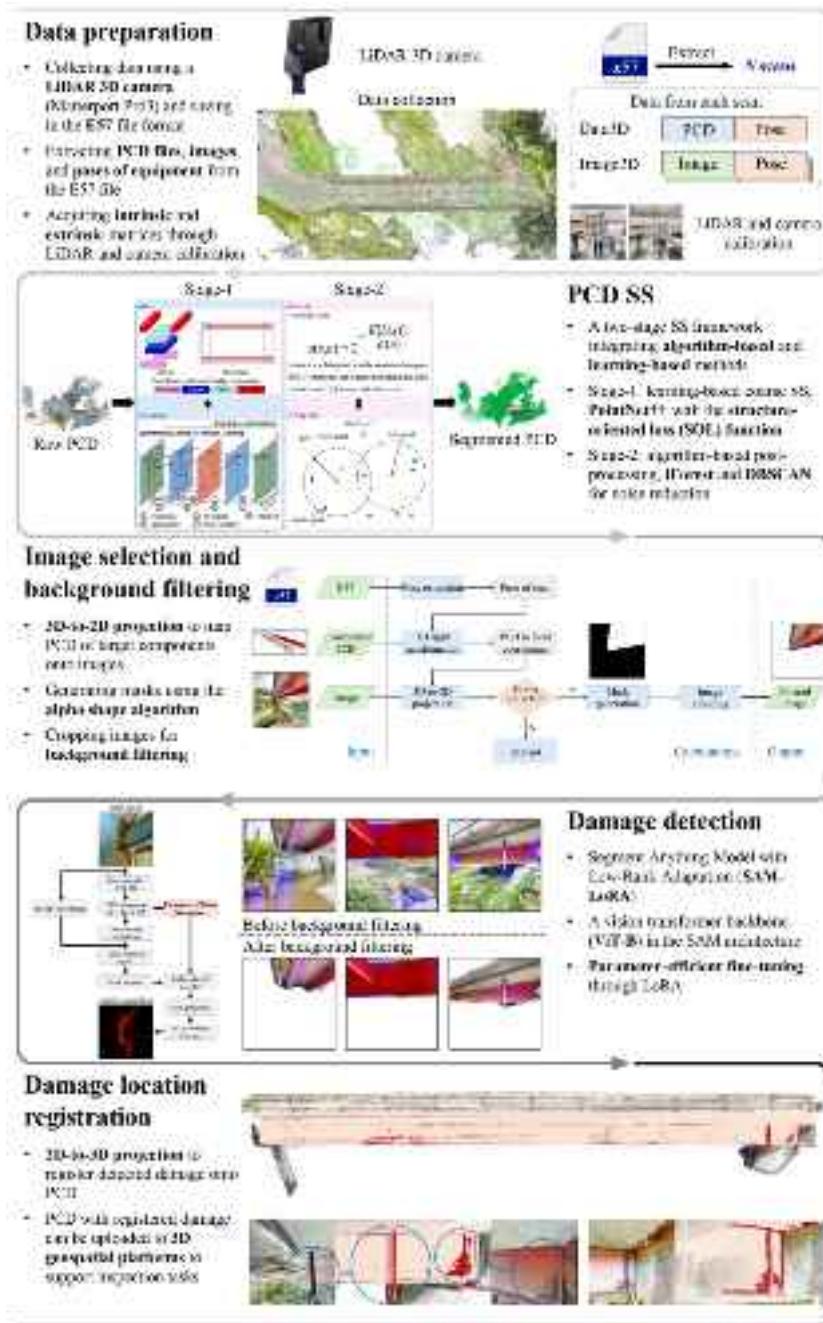


Fig. 1. Methodology framework.

### 3.1.1. File format and data collection equipment

E57 is an innovative and versatile file format that is widely used for storing PCD, images, and metadata [41]. Rather than traditional formats that store PCD and images separately, the E57 format also records the LiDAR pose for each PCD file and the camera pose for every image. This data is necessary to mate the images and PCD together [42]. A LiDAR 3D camera Matterport Pro3 that supports the E57 format is selected as the data collection equipment in this study, and its technical specifications are shown in Table 1. During each scan, the Matterport Pro3 obtains one PCD file whose points are located in its local coordinate system and takes six images of its surroundings following a fixed sequence. The directions and sequence of the six images taken in a single scan are shown in Fig. 2. Notably, there are blind spots in the center of the top and bottom images due to the restrictions of the vertical field of view. Although the Matterport Pro3

is used in this study to acquire data in the E57 format, the proposed method is applicable to any LiDAR- and camera-equipped measuring device that can output E57 files, demonstrating high generality and independence from specific hardware.

### 3.1.2. 2D-3D bidirectional projection

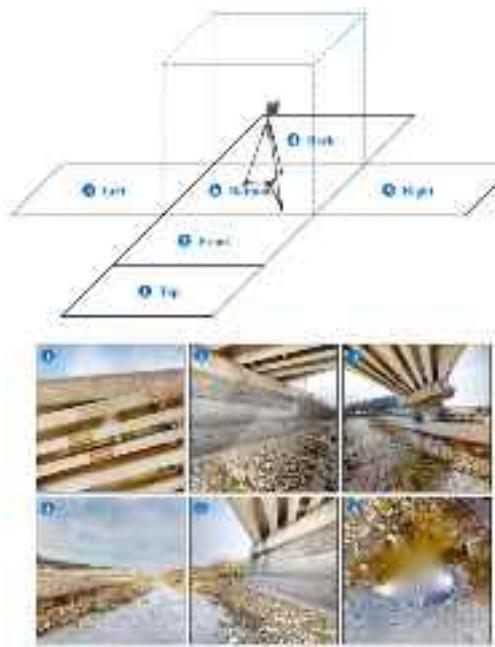
The general principle of 2D-3D bidirectional projection for the pinhole camera model is shown in Fig. 3, which is also applicable for the Matterport Pro3. Three coordinate systems are needed for the bidirectional projection between a 2D point  $p$  and a 3D point  $P$ , namely the world coordinate system (3D), the camera coordinate system (3D) whose origin is at the camera's optical center, and the image coordinate system (2D) whose origin is the top-left corner of the image.

For every image  $I$ ,  $(u, v)$  represent the coordinates of  $p$  in the image coordinate system, and the coordinates of  $P$  under the camera

**Table 1**

Technical specifications of the LiDAR 3D camera.

Category	Parameter	Value
Scanning	Wavelength	904 nm
	Field of view	360° Horizontal/295° Vertical
	Accuracy	±20 mm @ 10 m
	Depth resolution	10 <sup>5</sup> points per second
	Operating range	0.5–100 m
	Output pano pixels	134.2 MP
Imaging	Number of exposures per HDR frame	5
	Image size	4096 × 4096

**Fig. 2.** Images taken by the Matterport Pro3 during each scan.

coordinate system are  $(x, y, z)$ . The relationship between  $(u, v)$  and  $(x, y, z)$  is

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{K} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1)$$

where  $\mathbf{K}$  ( $3 \times 3$ ) is the intrinsic parameter matrix of the camera.

The directions and origins of the camera coordinate systems vary for images taken at different locations, so the world coordinate system is needed. The coordinates of the 3D point  $P$  under the world coordinate system are  $(X, Y, Z)$ . Two coordinates of the same point  $P$  under two coordinate systems can be transformed through a 3D rigid geometric transformation, which is represented as

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = \mathbf{R} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathbf{T} = [\mathbf{R} \mid \mathbf{T}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2)$$

where  $\mathbf{R}$  ( $3 \times 3$ ) and  $\mathbf{T}$  ( $3 \times 1$ ) are the rotation matrix and translation matrix that consist of extrinsic parameters, so  $[\mathbf{R} \mid \mathbf{T}]$  is the extrinsic matrix of the image  $I$ .

### 3.1.3. Equipment calibration and data extraction

The bidirectional projection between 2D and 3D points can be achieved using the principle illustrated in Fig. 3 when the intrinsic and extrinsic parameters in Eqs. (1) and (2) are obtained. For the LiDAR 3D camera, the intrinsic matrix (i.e.,  $\mathbf{K}$ ) is obtained by camera

calibration, and the extrinsic matrix (i.e.,  $[\mathbf{R} \mid \mathbf{T}]$ ) by calculation of the rigid transformation between the LiDAR system and the camera. As shown in Fig. 2, six images are taken during each scan while the equipment is rotating, so six groups of  $\mathbf{K}$  and  $[\mathbf{R} \mid \mathbf{T}]$  are calculated for each corresponding image. The projection errors of 2D–3D bidirectional projections remain well within the acceptable range for practical applications, as shown in Table 2. These results confirm the accuracy of the calibration process and the reliability of the 2D–3D bidirectional projection workflow. In particular, Table 2 serves as a key metric for quantitatively evaluating the precision of image background filtering and damage location registration, ensuring that the positional relationships between image and PCD are preserved with high fidelity throughout the methodology.

For the Matterport Pro3, Fig. 3 illustrates the bidirectional projections between the six images and one PCD file that are collected during the same scan. In order to register multiple PCD files with local coordinate systems, a unified reference coordinate system is required. Each PCD file must then undergo a 3D rigid transformation to align with this shared coordinate system. After the data collection, the images and registered PCD file of each scan are extracted from the stored E57 file. For every registered PCD file, one more dimension is added to all points within it to record the PCD scan index. Subsequently, all registered PCD files with scan indices are merged to form the comprehensive bridge PCD.

### 3.2. PCD SS

Semantic segmentation (SS) serves as a powerful tool for decomposing PCD into manageable components with associated geometry, features, and condition information. Algorithm-based and learning-based methods are the two main SS approaches for PCD, each with its own clear advantages and disadvantages. This study establishes a two-stage SS framework that integrates both types of methods, leveraging their complementary strengths to achieve enhanced SS performance.

The two-stage SS framework is presented in Fig. 4. Stage-1 is learning-based coarse SS where the raw bridge PCD is directly fed into a DL model. Detailed reviews of the applications of DL SS models on bridge PCD can be found in [43]. Through evaluation of state-of-the-art studies on DL-based methods for bridge PCD SS [20,43–47], PointNet++ [48] is selected as the backbone network, integrated with the structure-oriented loss (SOL) function [43], which was previously developed by our group and has demonstrated impressive performance. Stage-2 is algorithm-based post-processing, where advanced noise reduction algorithms are applied to segmented structural components for further refinement. The post-processing is respectively operated on each class of segmented structural components. First, Isolation Forest (iForest) [49] is utilized, which identifies global anomalies by recursively partitioning the data space. Second, Density Based Spatial Clustering of Applications with Noise (DBSCAN) [50] is applied to filter the noise and scattered small clusters.

To validate the proposed SS framework, the dataset collected by Lin et al. [43] is selected as the benchmark. The dataset contains annotated PCD of 12 reinforced concrete (RC) bridges that were scanned using the Matterport Pro3, and the points were labeled into five semantic classes (i.e., abutment, girder, deck, parapet, and others). Four evaluation metrics are utilized to quantitatively evaluate the SS results of bridge PCD, namely, overall accuracy (OA), mean class accuracy (mAcc), Intersection over Union (IoU), and mean IoU (mIoU).

Bridges 3, 4, and 7 in the dataset are selected for testing, and the others for training as indicated by Lin et al. [43]. The three bridges are segmented by the two-stage SS framework with hyperparameter values shown in Table 3, and the results are presented in Fig. 5 and Table 4. Overall, the framework achieved promising SS results for the testing samples.

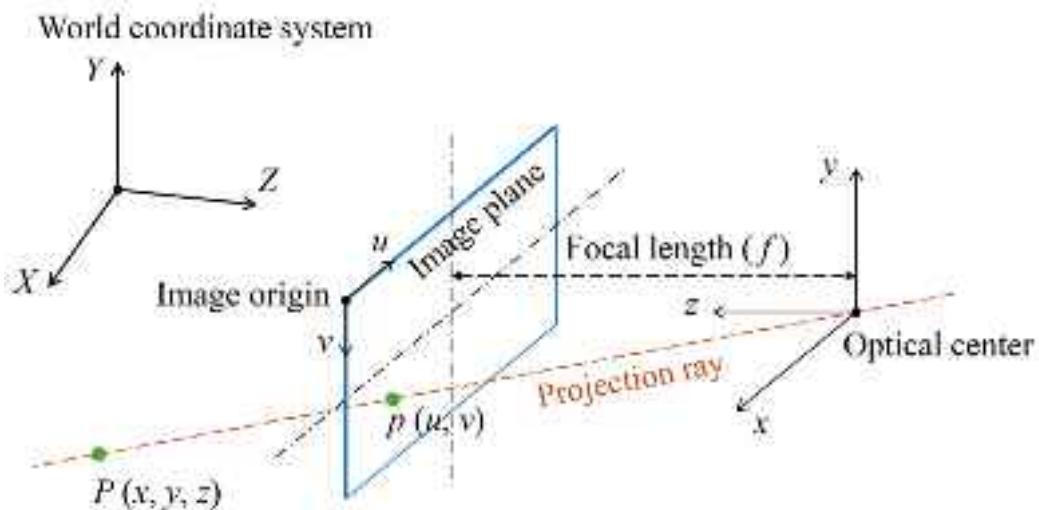


Fig. 3. Principle of 2D-3D bidirectional projection.

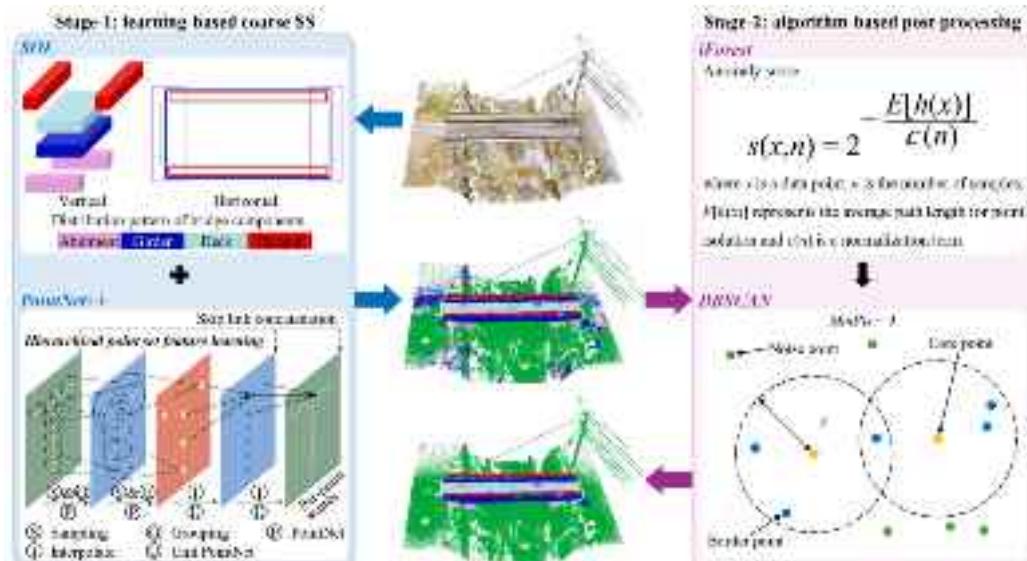


Fig. 4. Two-stage SS framework.

**Table 2**  
Projection errors of 2D-3D bidirectional projections.

Projection error	Image taken during each scan					
	Top	Front	Left	Back	Right	Bottom
3D-to-2D (pixel)	10.58	0.37	4.77	0.24	$4.55 \times 10^{-13}$	10.18
2D-to-3D (cm)	1.52	0.28	1.06	$2.60 \times 10^{-2}$	$1.38 \times 10^{-13}$	0.39

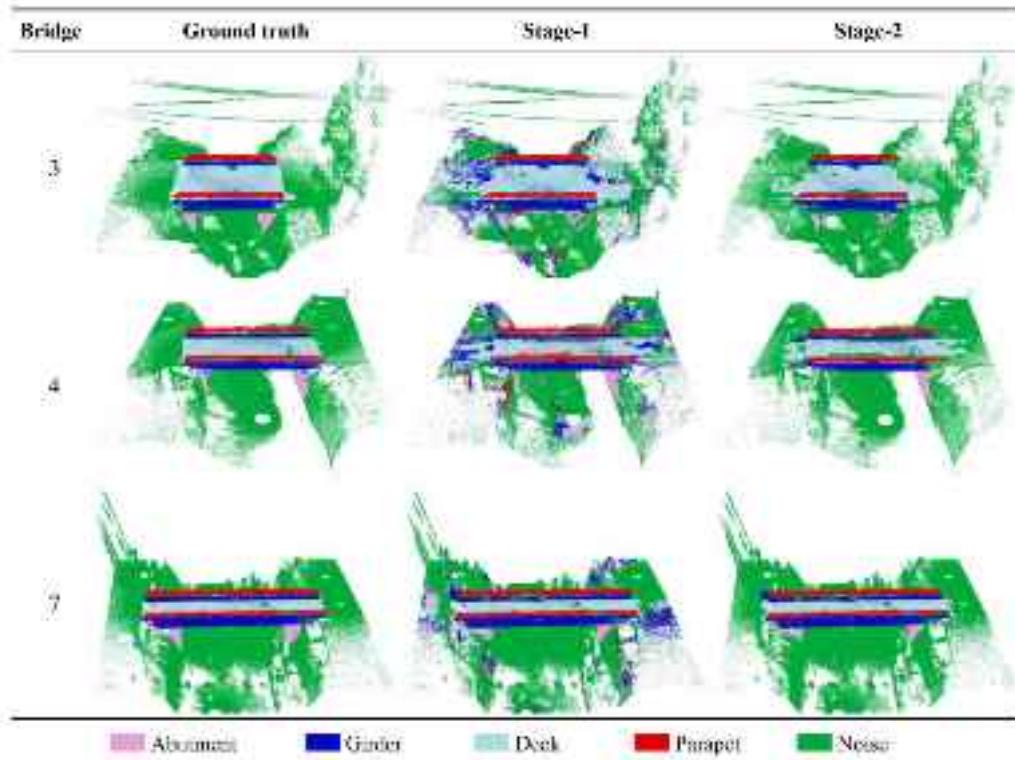
**Table 3**  
Hyperparameter values of the two-stage SS framework.

Stage-1	Stage-2		
PointNet++ with SOL function	iForest	DBSCAN	
Vertical/Horizontal weighting coefficient ( $\alpha/\beta$ )	Contamination	Epsilon Radius ( $\epsilon$ )	Minimum Samples (MinPts)
200/300	0.1	0.5	5

### 3.3. Image selection and background filtering

As shown in Fig. 2, the Matterport Pro3 takes six images from different directions at each scan spot. Therefore, there must be images that do not contain the specified bridge components, and these need to

be excluded for damage detection. Furthermore, backgrounds (i.e., non-component areas) in the selected images should also be filtered out to improve detection accuracy. The PCD of target bridge components is projected onto each image to select images with components and filter the background.



**Fig. 5.** Results visualization of the two-stage SS framework.

**Table 4**  
Results (%) of the two-stage SS framework.

Stage	OA	mAcc	mIoU	IoU for each class				
				Abutment	Girder	Deck	Parapet	Others
1	87.67	86.65	74.43	50.40	85.10	75.73	82.84	78.06
2 (only iForest)	86.87	80.22	73.87	56.96	77.37	77.48	78.83	78.69
2 (only DBSCAN)	91.56	90.85	80.83	58.22	90.74	79.47	91.06	84.68
2 (complete)	92.27	87.44	83.08	58.01	92.04	81.51	96.72	87.13

The image selection and background filtering are integrated into an automatic workflow, as shown in Fig. 6. First, each registered PCD file is transformed into its local coordinate system in preparation for 3D-to-2D projection. The 3D rigid transformation is performed, which applies the pose read from the E57 file. Second, the points of the target bridge components are mapped onto images through 3D-to-2D projection utilizing the obtained intrinsic and extrinsic parameters. Images that do not contain projected points are discarded, while the others remain for post-processing. Then the alpha shape algorithm [51] is employed to generate a polygon mask for each image that encloses the projected points on it. Finally, each image is cropped by the generated mask and saved for damage detection. The algorithm of this workflow is presented in Algorithm 1.

#### 3.4. Damage detection

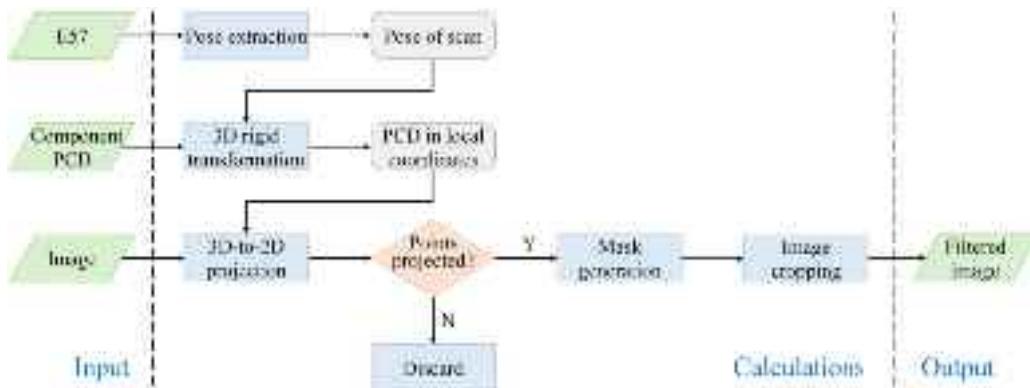
In damage detection tasks, established baseline models—such as UNet [52], DeepLabv3+ [53], and SegFormer [54]—are prone to severe domain shift when discrepancies arise between training and testing datasets, particularly in color and texture. Such shifts result in marked performance deterioration, thereby undermining the reliability of these models for practical deployment. Therefore, Segment Anything Model (SAM) [34] is selected which leverages a vision transformer backbone (ViT-B) that effectively captures hierarchical visual features across diverse scenes and is renowned for its versatility in general segmentation tasks. However, its performance tends to degrade in domain-specific

applications like infrastructure damage detection, primarily due to domain gaps. To adapt this powerful foundation model for the specific challenge of infrastructure damage detection, we fine-tuned it using Low-Rank Adaptation (LoRA) [55], which strategically modified selected network parameters while maintaining computational efficiency. The architecture of SAM-LoRA is shown in Fig. 7.

The implementation integrates rank-512 adaptation matrices within the self-attention mechanisms of the transformer architecture, enabling the model to develop increased sensitivity to damage-specific visual indicators such as discoloration, texture anomalies, and material degradation patterns. Table 5 shows the key parameters of the SAM-LoRA. This targeted adaptation approach preserves the robust general segmentation capabilities of the original SAM model while enhancing its performance in the domain-specific task. The LoRA fine-tuning strategy proved particularly valuable for the application, as it substantially reduced the computational resources required during the training phase compared to full-parameter fine-tuning, while delivering comparable or superior performance in damage detection tasks.

#### 3.5. Damage location registration

The damage detected from images is registered onto the PCD through 2D-to-3D projection utilizing the intrinsic and extrinsic parameters. The algorithm of this workflow is presented in Algorithm 2. The process of 2D-to-3D projection has a significant similarity to that of 3D-to-2D projection, but the extrinsic matrix (i.e.,  $[R | T]$ ) needs to be inverted.



**Fig. 6.** Workflow of image selection and background filtering.

**Algorithm 1:** Image selection and background filtering.

```

Input: Number of scans  $N$ , PCD files of components  $\{P_j\}_{j=1}^N$ ,
        Images  $\{I_j\}_{j=1}^{6 \times N}$ , E57 file  $E57$ , Parameter matrices
         $\{K_j, [R | T]_j\}_{j=1}^6$ 
Output: Images with background filtered  $\{I_{ij}^{filtered}\}$ 
for  $scan_i$  ( $i=1:N$ ) do
    select images  $\{I_{ij}\}_{j=1}^6$  of  $scan_i$ ;
     $pose_i$ =ReadScanPose( $E57$ ,  $scan_i$ );
    rotate  $P_i$  to its local coordinate system,
     $P_i^{local}$ =ApplyTransform( $P_i$ , invert( $pose_i$ ));
    for  $I_{ij}$  ( $j=1:6$ ) do
        match  $K_j, [R | T]_j$  to  $I_{ij}$ ;
        project  $P_i^{local}$  onto  $I_{ij}$ ,  $points_{ij}$ =Project3DTo2D( $P_i^{local}$ ,  $K_j$ ,
             $[R | T]_j$ );
        if  $points_{ij} > 100$  then
             $mask_{ij}$ =GenerateAlphaShape( $points_{ij}$ );
            crop  $I_{ij}$  using  $mask_{ij}$  to get  $I_{ij}^{filtered}$ ;
            save  $I_{ij}^{filtered}$ ;
        end
    end
end
return  $\{I_{ij}^{filtered}\}$ 

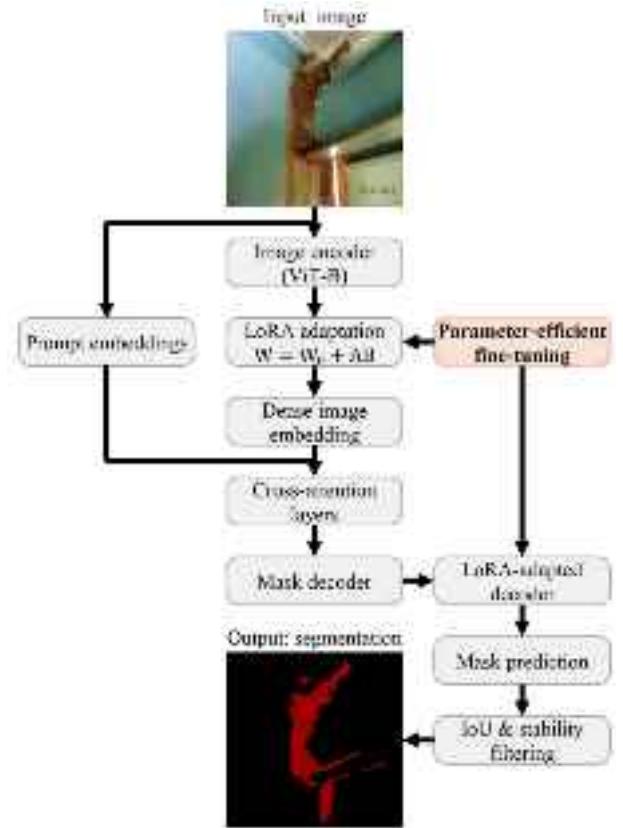
```

**Algorithm 2:** Location registration of detected damage.

```

Input: Number of scans  $N$ , Denoised PCD files  $\{P_j\}_{j=1}^N$ , Image
        masks of detected damage  $\{I\}$ , E57 file  $E57$ , Parameter
        matrices  $\{K_j, [R | T]_j\}_{j=1}^6$ 
Output: Complete PCD with damage  $P^{damage}$ 
for  $scan_i$  ( $i=1:N$ ) do
    select images  $\{I_{ij}\}$  of  $scan_i$ ;
     $pose_i$ =ReadScanPose( $E57$ ,  $scan_i$ );
    rotate  $P_i$  to its local coordinate system,
     $P_i^{local}$ =ApplyTransform( $P_i$ , invert( $pose_i$ ));
    for  $I_{ij}$  do
        match  $K_j, [R | T]_j$  to  $I_{ij}$ ;
        project  $I_{ij}$  onto  $P_i^{local}$ ,  $points_{ij}$ =Project2DTo3D( $I_{ij}$ ,  $P_i^{local}$ ,
             $K_j$ , invert( $[R | T]_j$ ));
    end
    add  $\{points_{ij}\}$  to  $P_i^{local}$  to get  $P_i^{local\_damage}$ ;
    rotate  $P_i^{local\_damage}$  to the unified reference coordinate system,
     $P_i^{damage}$ =ApplyTransform( $P_i^{local\_damage}$ ,  $pose_i$ );
end
 $P^{damage}$ =Merge( $\{P_i^{damage}\}$ );
return  $P^{damage}$ 

```



**Fig. 7.** Parameter-efficient SAM-LoRA architecture for domain-specific segmentation in damage detection.

#### 4. Experiment and results

To validate the proposed methodology, an RC bridge named Kawadura Bridge was selected as the experimental subject. Kawadura Bridge was built in 1971 and is located in Hirata Village, Fukushima Prefecture, Japan. It has a length of 23.5 m and a width of 4.8 m, and contains a steel girder beneath its RC girder. Considering its size, Kawadura Bridge was scanned from 33 spots using the Matterport Pro3, as shown in Fig. 8. Corrosion was chosen as the target damage type for this case study, as it presents distinct challenges in detection and characterization. Moreover, while extensive studies have focused on crack detection, corrosion has received comparatively less attention and remains a critical concern in structural health monitoring.



**Fig. 8.** Scan spots of Kawadura Bridge.

**Table 5**  
Key parameters of the SAM-LoRA model for corrosion detection.

Category	Parameter	Value
Model	Base architecture	SAM (ViT-B backbone)
	Fine-tuning approach	LoRA
	Adaptation rank	512
	Optimizer/Learning rate	Adam/ $1 \times 10^{-4}$
Training	Weight decay	0.001
	Loss function	Binary cross entropy with Logits
	Class weighting	Positive sample weight = 5.0
	Feature processing	Normalized representation
Inference	Confidence threshold	0.9
	Points per side	64
	Prediction IoU threshold	0.90
	Stability score threshold	0.85

The steel girder exhibits extensive corrosion distributed across a wide range of structural components. This widespread and varied distribution of corrosion introduces significant variability in terms of geometry, texture, and surface orientation, thereby presenting a complex and realistic scenario for evaluation. Owing to its diverse spatial characteristics and occurrence patterns, corrosion represents an ideal damage type for rigorously assessing the effectiveness of the proposed methodology. It enables a thorough examination of the methodology's robustness, adaptability to complex geometries, and generalizability across different damage manifestations, ultimately validating the accuracy and reliability of the damage detection and location registration in practical infrastructure scenarios. The experiment followed the framework outlined in Fig. 1. Inputs, outputs, and time costs of major steps are shown in Table 6. All major steps were completed within a short time frame, indicating that the pipeline is not only well-optimized but also practically feasible for real-world applications. Detailed results of each step and discussions regarding the methodology are provided in the following subsections.

#### 4.1. Data preparation

After the data collection, 33 PCD files and 198 images were extracted from the E57 file. All PCD files were sequentially numbered according to their corresponding scan indices and then merged. Fig. 9 presents the merged PCD and images. For quantitative evaluation, the complete PCD was annotated into five categories (i.e., abutment, girder, deck, parapet, and others) and each image was checked to identify the types of components within. For computational efficiency,

the annotated PCD was downsampled using voxel grid filtering with a resolution of 3D voxels set to 0.02 m. The number of images that contain each type of component and the point number of each class are presented in Table 7.

#### 4.2. Bridge PCD SS

The prepared PCD was fed into the two-stage SS framework, and the results are presented in Fig. 10 and Table 8. Fig. 10 clearly demonstrates that the components are well distinguished from the background, with only a small fraction of adjacent noisy points misclassified as part of the components. Additionally, Table 8 confirms that the SS performance is satisfactory and aligns closely with the tested results presented in Section 3.2. Overall, the PCD SS of Kawadura Bridge is successful and can serve as a solid foundation for subsequent tasks.

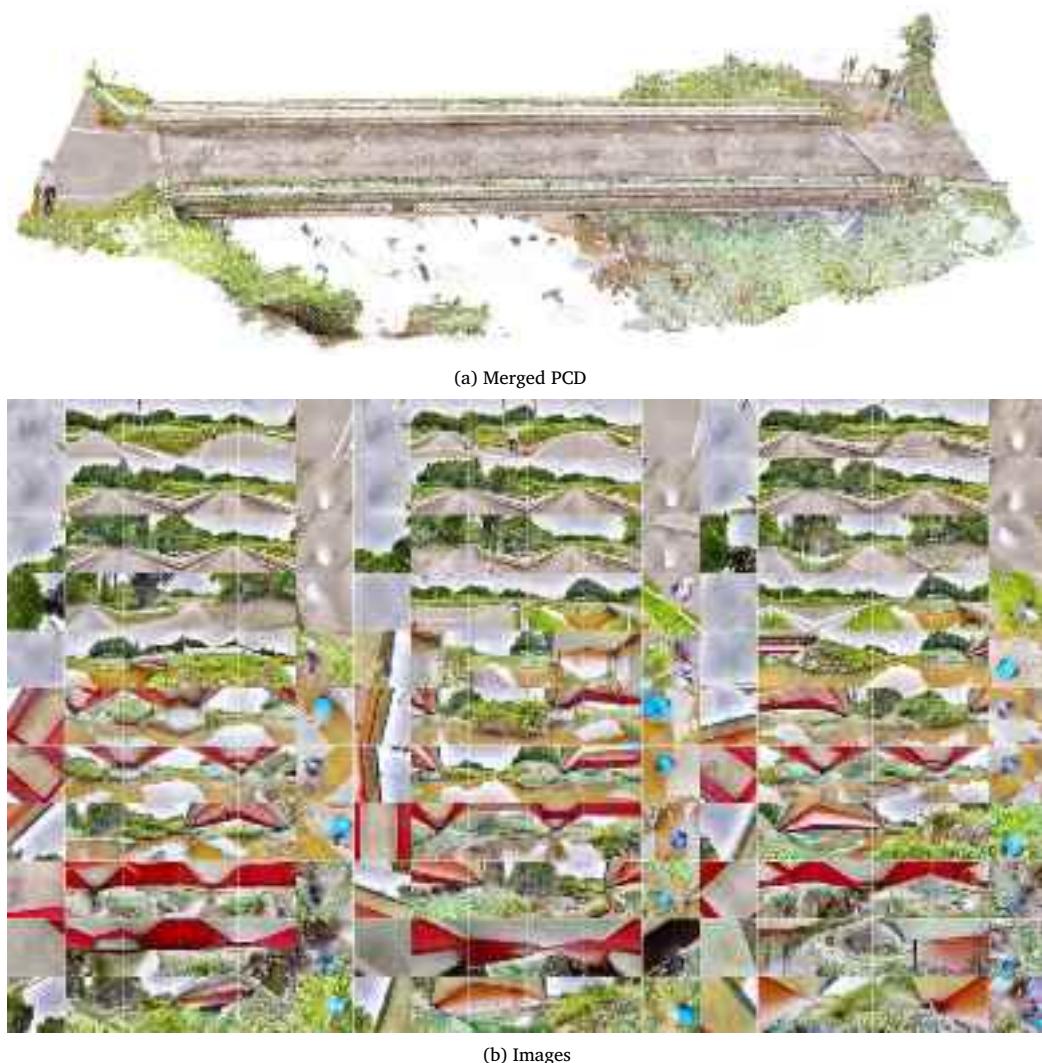
#### 4.3. Image selection and background filtering

The girder was selected as a detection object where corrosion tends to occur. Points in the segmented PCD that belong to the girder were selected and separated according to their scan indices. Subsequently, the 198 images were processed using the workflow elaborated in Fig. 6. The result of image selection is shown in Fig. 11 (false positive rate is 5.71% and false negative rate is 0). All 128 images containing the girder were correctly identified without any omissions. Four images lacking the necessary components were mistakenly selected due to the misclassifications of noisy points near the girder during PCD SS, which were then projected onto the four images. However, the total area of the wrongly filtered regions was limited, placing no significant burden on the subsequent damage detection task.

The proposed workflow also achieves excellent results in image background filtering, as shown in Fig. 12. The girder was accurately extracted from the background and undesired component categories. Notably, the generated masks accurately separated small component parts such as drainage pipes, indicating the remarkable performance of the workflow.

#### 4.4. Damage detection

To achieve corrosion detection, two open-source bridge corrosion datasets [56,57] were selected for model training. There are 954 annotated images in total and all images in each dataset were split 80% training and 20% testing by random sorting. Both the original SAM and the proposed SAM-LoRA were trained using the training samples. As



**Fig. 9.** Merged PCD and images of Kawadura Bridge extracted from e57 file.

**Table 6**

Inputs, outputs, and time costs of major steps.

Major step	Input	Output	Time cost (min)
Data extraction	E57 file	33 PCD files and 198 images	1
PCD SS	Merged PCD with 3,259,744 points	Segmented PCD with five labeled classes	18
Image selection and background filtering	Girder PCD and 198 images	132 images with background filtered	31
Damage detection	132 background-filtered images	68 images with corrosion detected	4
Damage location registration	68 image masks and 33 PCD files	Complete bridge PCD with damage registered	12

**Table 7**

Metadata of the prepared images and PCD.

Number	Total	Class				
		Abutment	Girder	Deck	Parapet	Others
Images	198	63	128	42	45	/
Points	3,259,744	101,263	839,584	265,442	140,758	1,912,697

**Table 8**

SS Results (%) of Kawadura Bridge.

OA	mAcc	mIoU	IoU for each class				
			Abutment	Girder	Deck	Parapet	Others
95.51	91.71	86.37	61.13	95.76	92.91	90.72	91.33

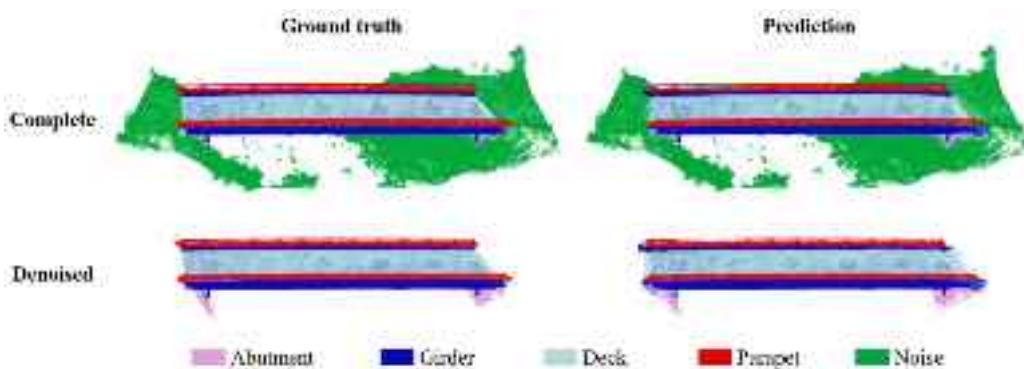


Fig. 10. SS visualization of Kawadura Bridge.

**Table 9**  
Computational efficiency comparison between original SAM and SAM-LoRA.

Method	Total params	Trainable params	Param ratio	Avg epoch time (s)	GPU memory (MB)
Original SAM	93.74M	89.67M	95.66%	763.2	20,682
SAM-LoRA	112.61M	18.87M	16.76%	769.3	20,680

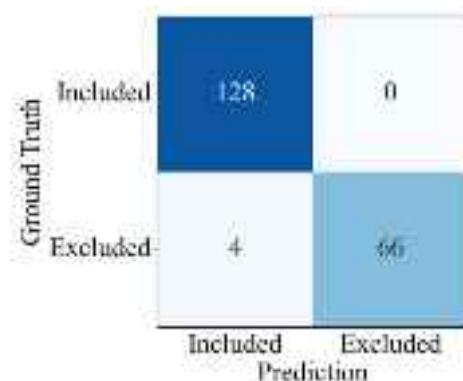


Fig. 11. Confusion matrix of image selection.

shown in [Table 9](#), LoRA freezes the pre-trained SAM weights (89.67M parameters) and introduces trainable low-rank decomposition matrices (18.87M parameters) to capture domain-specific adaptations while preserving foundational capabilities. Furthermore, training convergence analysis shown in [Fig. 13](#) demonstrates that the original SAM model exhibits poor convergence over 50 epochs due to the computational burden of updating all parameters simultaneously, whereas the LoRA approach achieves rapid convergence to superior performance (loss: 0.0164 vs 0.3112) by training only 17% of the combined model parameters, validating the efficiency and practical applicability of this parameter-efficient approach.

The trained SAM-LoRA model was evaluated on the testing samples and the results are shown in [Table 10](#). As presented in [Table 10](#), the overall metrics (i.e., accuracy (90.40%) and mIoU (75.62%)) demonstrate that the parameter-efficient fine-tuning approach can successfully adapt the SAM architecture to the domain-specific task of corrosion detection in steel girder bridges. In particular, the high precision (90.44%) indicates the model's reliability in identifying actual corrosion areas, and the recall value (82.19%) confirms its effectiveness in detecting most corrosion instances. The high specificity (92.61%) further validates the model's robustness in distinguishing between corroded and non-corroded regions.

The images before and after background filtering were fed into the trained SAM-LoRA model for corrosion detection, as shown in [Fig. 14](#). It can be observed that excluding the background significantly

**Table 10**  
Evaluation results (%) of the trained SAM-LoRA model.

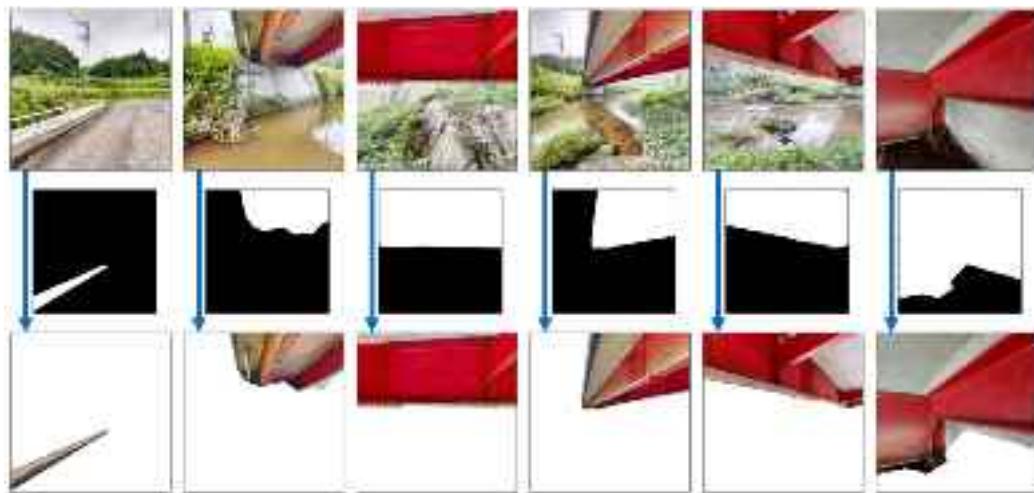
Accuracy	mIoU	F1 Score	Precision	Recall	Specificity
90.40	75.62	85.27	90.44	82.19	92.61

enhances the overall corrosion detection performance. In the absence of background removal, complex textures such as stains, shadows, and dense vegetation can be mistakenly identified as corrosion, leading to a higher rate of false positives. By removing the background, the detection model can focus more precisely on the areas of interest, thereby minimizing the influence of irrelevant artifacts and improving the reliability of corrosion identification. Moreover, background exclusion reduces noise within the feature space, allowing the model to better align its predictions with the distribution learned during training. This focused detection approach not only enhances the robustness of the corrosion detection system but also ensures higher consistency across different testing scenarios with varying background complexities.

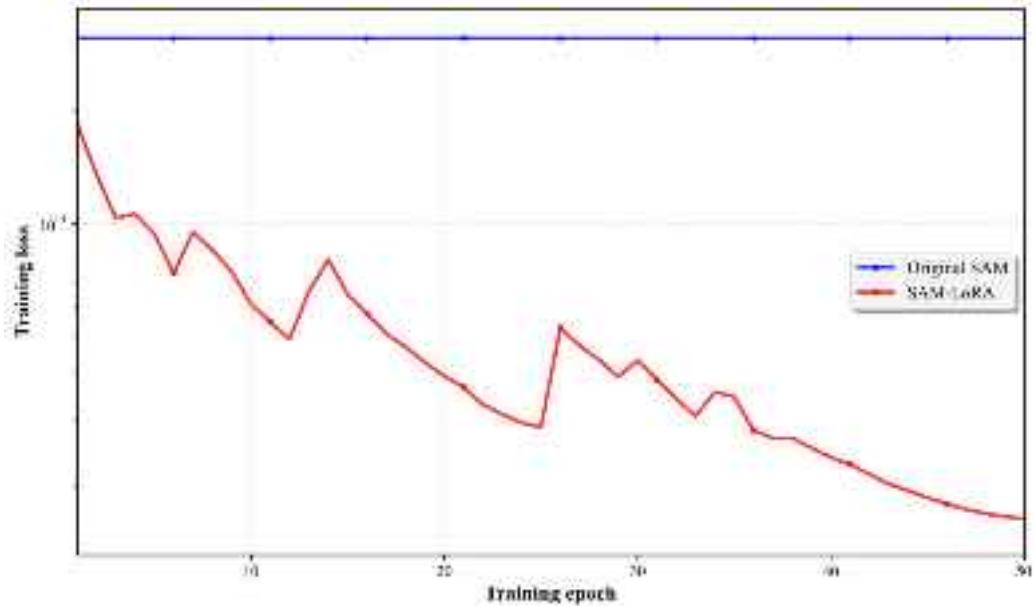
#### 4.5. Damage location registration

The detected corrosion was accurately projected onto the bridge PCD using the registration workflow detailed in [Section 3.5](#). The complete PCD, incorporating all identified corrosion regions, is illustrated in [Fig. 15](#). To facilitate a clear visual distinction, the entire girder structure was rendered in pink, while areas exhibiting corrosion were highlighted in red. As depicted in [Fig. 15](#), the corrosion was meticulously mapped onto various parts of the steel girder, including the side and bottom surfaces, both the inner and outer faces, as well as longitudinal and transverse beams. This comprehensive coverage demonstrates the robustness of the registration process across complex geometries. Furthermore, close-up views of the registered PCD reveal a remarkable alignment between the projected corrosion features and those observed in the original image data. The corrosion patches not only coincide spatially with the visual indicators from the images but also exhibit a high degree of consistency in terms of shape, scale, and spatial distribution. This strong correspondence validates the effectiveness and accuracy of the proposed damage location registration method for corrosion documentation and visualization on 3D models.

The bridge PCD with registered corrosion serves as a versatile foundation for a wide range of downstream applications, including structural condition assessment, maintenance planning, DT construction, and long-term deterioration tracking. In this study, to enhance



**Fig. 12.** Examples of image background filtering.



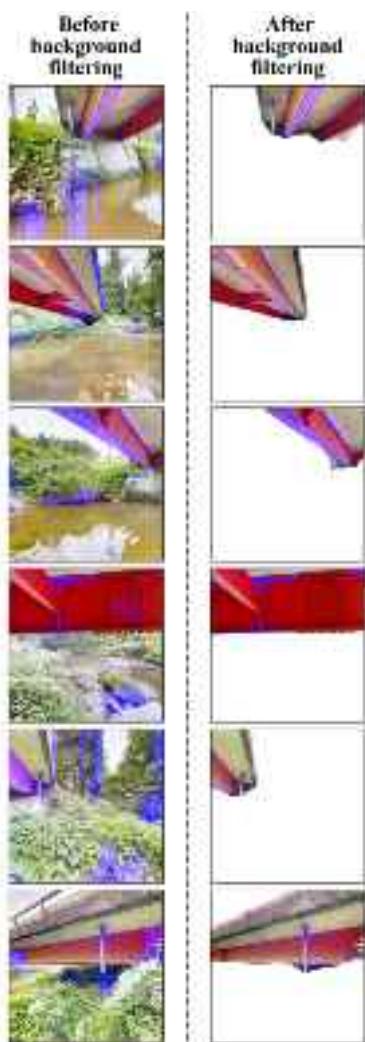
**Fig. 13.** Convergence analysis for original SAM and SAM-Lora.

the accessibility and clarity of corrosion visualization, the PCD was uploaded to Cesium—a widely adopted web-based 3D geospatial platform [58]. This enabled intuitive, browser-based interaction with the corrosion data in a geospatial context, as illustrated in Fig. 16.

A comparison between Figs. 8 and 16 clearly demonstrates the accurate geo-registration of the bridge PCD. The uploaded PCD is seamlessly aligned within the base map, confirming the fidelity of spatial transformation and integration. By integrating the bridge PCD and this platform, inspectors and engineers can conveniently navigate and examine the corrosion conditions of the bridge in a realistic geographic environment. Critical information such as the exact location, extent, and distribution of corrosion on structural components – including longitudinal beams and their side faces – can be intuitively explored. This integration significantly improves situational awareness and decision-making efficiency in infrastructure management workflows.

#### 4.6. Discussion

Corrosion was selected as the damage type for validation in order to clearly illustrate the proposed workflow under the most demanding conditions. Among various types of structural damage, corrosion presents the greatest complexity in terms of spatial location and geometry. It typically occurs on steel girders, which encompass diverse structural features such as flat plates, slender beams, curved edges, and irregular joints. This geometric diversity imposes the most significant challenges for the entire workflow, particularly for 2D–3D bidirectional projection and damage detection tasks. Corrosion's irregular geometry, distributed occurrence, and variation in surface orientation create a rigorous test case for evaluating the robustness, adaptability, and precision of the proposed methodology. Given that the methodology performs effectively under such geometrically and texturally challenging conditions, it is reasonable to infer that its performance would be potentially

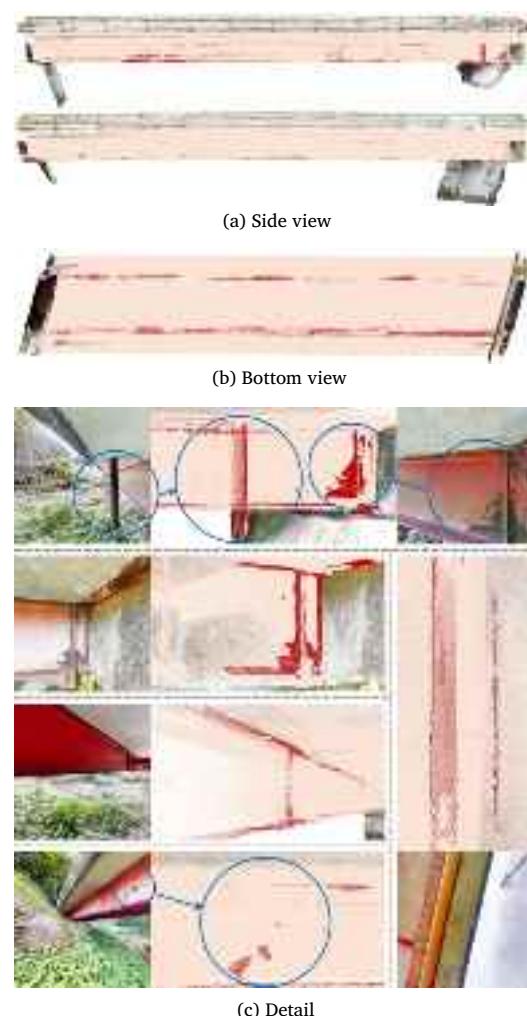


**Fig. 14.** Examples of corrosion detection before and after image background filtering.

improved – or at least maintained – when applied to other damage types with simpler geometric characteristics, such as cracks or concrete spalling, which are typically located on relatively planar surfaces.

From a results perspective, the proposed methodology exhibits outstanding performance in real-world scenarios. While the PCD generated by SfM-based methods cannot be integrated into geospatial platforms, the proposed approach makes it possible. Additionally, from a process perspective, it brings great convenience to bridge inspection and enhances overall operational efficiency. The advantages of the methodology include:

- (1) The data collection process is not only straightforward and efficient, but also capable of simultaneously capturing high-quality images and PCD. Compared to other popular bridge inspection methods (e.g. UAV-based approaches), the use of a LiDAR 3D camera imposes lower professional requirements on inspectors and enables faster data acquisition. Moreover, the LiDAR 3D camera demonstrates greater reliability in challenging environments, particularly in scenarios where bridges are too low or surrounding vegetation is too dense for UAV operations.
- (2) The accuracy of the 2D–3D bidirectional projections is ensured by the employed data organization form. Since the intrinsic and extrinsic parameters remain fixed for each image, both 3D-to-2D and 2D-to-3D projections are highly precise. Consequently,



**Fig. 15.** Visualization of the bridge PCD with corrosion registered.

image background filtering and damage location registration yield dependable results. In contrast, other data organization forms using PCD and images, such as SfM, involve different intrinsic and extrinsic parameters for each image, leading to inaccurate bidirectional projections.

- (3) Component recognition and background filtering remain highly robust with image-PCD fusion. However, these tasks become more problematic when relying solely on images. Variations in lighting conditions, such as shadows and highlights, can obscure components and lead to reduced accuracy. In comparison, image-PCD fusion ensures stable performance in component recognition, even under harsh lighting and adverse weather conditions.

This study has some limitations that require future research efforts. The methodology primarily focuses on surface-level visual features for damage detection. While this is effective for identifying superficial issues, it may not be adequate for capturing internal or structural defects that could compromise bridge integrity. To address this shortcoming, PCD can be incorporated into the damage detection step to calculate the depth and volumetric characteristics of the damage, offering a more precise, 3D assessment. Furthermore, integrating complementary techniques (e.g., ultrasonic testing (UT) or infrared thermography (IRT) [59–61]) into the image-PCD fusion framework could significantly enhance the overall diagnostic capability. These additional methods are well-suited for detecting subsurface anomalies and



**Fig. 16.** Bridge PCD registered with corrosion displayed in Cesium.

hidden structural flaws that visual inspection alone might overlook. A feasible roadmap involves three steps: (1) Data co-registration—align UT and IRT data spatially with the absolute-scale PCD using coordinate transformation techniques; (2) Feature-level fusion—extract defect-relevant features such as delamination depth from UT and thermal anomalies from IRT, and integrate them with visual surface damage features through a multimodal learning framework; (3) Decision-level integration—combine outputs from visual and non-visual channels into a unified damage index, which can be directly mapped onto the 3D bridge model for comprehensive structural assessment. By combining these diverse modalities, the resulting system would not only identify surface damage but also provide a comprehensive evaluation of the bridge's condition, thereby supporting more informed maintenance decisions and ensuring long-term structural safety.

Currently, only the PCD captured alongside images serves as the platform for registering damage. This approach is inherently limited in handling dynamic changes in a bridge's condition. Although it is highly efficient for constructing accurate as-is 3D bridge models, its utility for regular inspections is reduced because the process does not facilitate effective damage updating over time. To overcome this limitation, alternative approaches that facilitate the precise alignment of newly captured images with existing 3D models could be employed. Techniques like incremental surface registration or advanced change detection algorithms would enable the system to merge new images or scans with the original 3D models, ensuring that any newly detected damage or structural changes are accurately integrated. These methods would significantly enhance the regular inspection process, offering a more dynamic and reliable tool for long-term bridge health monitoring.

## 5. Conclusion

This paper presented a five-step methodology that leverages image-PCD fusion to remove background elements from images and detect damage from background-filtered images, accurately registering it on the segmented PCD. An experiment conducted on a real bridge demonstrates that all steps in the methodology work seamlessly together, ultimately delivering remarkable results—damage was precisely detected and clearly visualized in a 3D model, effectively supporting inspection and decision-making processes. The methodology is simple yet practical, as most calculations and workflows rely on existing algorithm modules, making it easy to operate and reproduce. Additionally, it offers significant flexibility, with over half of its steps being replaceable to accommodate various scenarios, including different data acquisition equipment, PCD SS and damage detection models.

While the proposed methodology has made noticeable contributions to bridge inspection, its impact can be further enhanced through continuous updates and improvements. Future work could focus on extending and validating the framework for other crucial damage types (e.g. cracks, spalling concrete). Similar to the treatment of corrosion, additional types of damage can be precisely detected and meticulously registered in the 3D model. This would enable the construction of a more comprehensive as-is representation of the structure, providing deeper insights into its current condition and enhancing the effectiveness of various inspection and maintenance tasks. Another key direction for future work is the development of comprehensive functional models (e.g. BIM and DT) using the absolute-scale PCD with registered damage to support advanced management tasks. The availability of absolute-scale PCD provides a geospatially accurate foundation essential for the reliable integration of structural information into BIM and DT systems,

facilitating seamless model alignment, change tracking, and data fusion across platforms. These models would enable real-time monitoring and more effective visualization, analysis, and predictive maintenance, ultimately enhancing decision-making processes and optimizing long-term infrastructure management.

Other future work could focus on damage size estimation using image-PCD fusion to precisely measure necessary parameters such as the length, width, depth, and surface area of detected damage. A geometric processing pipeline previously developed by our group has demonstrated reliable performance in estimating the dimensions of bridge components from PCD [62]. Building on this foundation, the pipeline can be further extended to accurately quantify the size of detected damage by utilizing the damage-registered PCD. This approach would facilitate a more accurate and detailed assessment of structural deterioration. Furthermore, precise damage measurements can contribute to the establishment of high-fidelity FEM models, allowing for realistic simulations of structural behavior. Such advancements would enable reliable and efficient structural health monitoring, support better predictive maintenance strategies, enhance failure risk assessments, and optimize rehabilitation planning.

#### CRediT authorship contribution statement

**Chao Lin:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yu Chen:** Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Kenta Itakura:** Writing – review & editing, Validation, Software, Methodology, Investigation. **Shreejan Maharjan:** Validation, Software. **Pang-jo Chun:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

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 Grant-in-Aid for Scientific Research Grant Numbers 21H01417, 23K22831, and 23H00198, and China Scholarship Council Grant Number 202306210048.

#### Data availability

Data will be made available on request.

#### References

- [1] P. Chun, S. Izumi, T. Yamane, Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine, Computer-Aided Civ. Infrastruct. Eng. 36 (1) (2021) 61–72, <http://dx.doi.org/10.1111/mice.12564>.
- [2] A. Omar, O. Moselhi, Automated data-driven condition assessment method for concrete bridges, Autom. Constr. 167 (2024) 105706, <http://dx.doi.org/10.1016/j.autcon.2024.105706>.
- [3] Y. Okazaki, S. Okazaki, S. Asamoto, P. Chun, Applicability of machine learning to a crack model in concrete bridges, Computer-Aided Civ. Infrastruct. Eng. 35 (8) (2020) 775–792, <http://dx.doi.org/10.1111/mice.12532>.
- [4] S. Inadomi, P. Chun, Spatially aware Markov chain-based deterioration prediction of bridge components using a graph transformer, Computer-Aided Civ. Infrastruct. Eng. 40 (14) (2025) 1932–1955, <http://dx.doi.org/10.1111/mice.13497>.
- [5] P. Chun, T. Yamane, Y. Maemura, A deep learning-based image captioning method to automatically generate comprehensive explanations of bridge damage, Computer-Aided Civ. Infrastruct. Eng. 37 (11) (2022) 1387–1401, <http://dx.doi.org/10.1111/mice.12793>.
- [6] R. Khallaf, M. Khallaf, Classification and analysis of deep learning applications in construction: A systematic literature review, Autom. Constr. 129 (2021) 103760, <http://dx.doi.org/10.1016/j.autcon.2021.103760>.
- [7] Z. He, S. Jiang, J. Zhang, G. Wu, Automatic damage detection using anchor-free method and unmanned surface vessel, Autom. Constr. 133 (2022) 104017, <http://dx.doi.org/10.1016/j.autcon.2021.104017>.
- [8] J. Guo, P. Liu, B. Xiao, L. Deng, Q. Wang, Surface defect detection of civil structures using images: Review from data perspective, Autom. Constr. 158 (2024) 105186, <http://dx.doi.org/10.1016/j.autcon.2023.105186>.
- [9] Y. Xu, X. Tong, U. Stilla, Voxel-based representation of 3D point clouds: Methods, applications, and its potential use in the construction industry, Autom. Constr. 126 (2021) 103675, <http://dx.doi.org/10.1016/j.autcon.2021.103675>.
- [10] B. Wang, Q. Wang, J.C. Cheng, C. Song, C. Yin, Vision-assisted BIM reconstruction from 3D LiDAR point clouds for mep scenes, Autom. Constr. 133 (2022) 103997, <http://dx.doi.org/10.1016/j.autcon.2021.103997>.
- [11] N. Rimella, L. Rimella, A. Osello, Machine learning method for as-is tunnel information model reconstruction, Autom. Constr. 172 (2025) 106039, <http://dx.doi.org/10.1016/j.autcon.2025.106039>.
- [12] Y. Xu, Z. Ye, R. Huang, L. Hoegner, U. Stilla, Robust segmentation and localization of structural planes from photogrammetric point clouds in construction sites, Autom. Constr. 117 (2020) 103206, <http://dx.doi.org/10.1016/j.autcon.2020.103206>.
- [13] R. Huang, Y. Xu, L. Hoegner, U. Stilla, Semantics-aided 3D change detection on construction sites using UAV-based photogrammetric point clouds, Autom. Constr. 134 (2022) 104057, <http://dx.doi.org/10.1016/j.autcon.2021.104057>.
- [14] J. Chen, W. Lu, J. Lou, Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model, Computer-Aided Civ. Infrastruct. Eng. 38 (8) (2023) 1079–1098, <http://dx.doi.org/10.1111/mice.12928>.
- [15] K. Itakura, F. Hosoi, Automated tree detection from 3D lidar images using image processing and machine learning, Appl. Opt. 58 (14) (2019) 3807–3811, <http://dx.doi.org/10.1364/ao.58.003807>.
- [16] K. Itakura, F. Hosoi, Three-dimensional tree monitoring in urban cities using automatic tree detection method with mobile LiDAR data, Artif. Intell. Data Sci. 2 (2) (2021) 1–10, <http://dx.doi.org/10.11532/jscieiii.2.2.1>.
- [17] S. Inadomi, P. Chun, A comparative study of projection-based vs. point-based point clouds segmentation for 3D bridge modeling, Struct. Infrastruct. Eng. (2025) 1–13, <http://dx.doi.org/10.1080/15732479.2025.2494257>.
- [18] N. Saovana, N. Yabuki, T. Fukuda, Automated point cloud classification using an image-based instance segmentation for structure from motion, Autom. Constr. 129 (2021) 103804, <http://dx.doi.org/10.1016/j.autcon.2021.103804>.
- [19] D. Ma, N. Wang, H. Fang, W. Chen, B. Li, K. Zhai, Attention-optimized 3D segmentation and reconstruction system for sewer pipelines employing multi-view images, Computer-Aided Civ. Infrastruct. Eng. (2024) 1–20, <http://dx.doi.org/10.1111/mice.13241>.
- [20] J.-L. Xiao, J.-S. Fan, Y.-F. Liu, B.-L. Li, J.-G. Nie, Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection, Autom. Constr. 158 (2024) 105226, <http://dx.doi.org/10.1016/j.autcon.2023.105226>.
- [21] H. Li, Y. Chen, J. Liu, C. Che, Z. Meng, H. Zhu, High-resolution model reconstruction and bridge damage detection based on data fusion of unmanned aerial vehicles light detection and ranging data imagery, Computer-Aided Civ. Infrastruct. Eng. 39 (8) (2024) 1197–1217, <http://dx.doi.org/10.1111/mice.13133>.
- [22] K. Hu, Z. Chen, H. Kang, Y. Tang, 3D vision technologies for a self-developed structural external crack damage recognition robot, Autom. Constr. 159 (2024) 105262, <http://dx.doi.org/10.1016/j.autcon.2023.105262>.
- [23] L. Ge, A. Sadhu, Deep learning-enhanced smart ground robotic system for automated structural damage inspection and mapping, Autom. Constr. 170 (2025) 105951, <http://dx.doi.org/10.1016/j.autcon.2024.105951>.
- [24] J.K. Chow, K.-f. Liu, P.S. Tan, Z. Su, J. Wu, Z. Li, Y.-H. Wang, Automated defect inspection of concrete structures, Autom. Constr. 132 (2021) 103959, <http://dx.doi.org/10.1016/j.autcon.2021.103959>.
- [25] L. Deng, H. Yuan, L. Long, P. Chun, W. Chen, H. Chu, Cascade refinement extraction network with active boundary loss for segmentation of concrete cracks from high-resolution images, Autom. Constr. 162 (2024) 105410, <http://dx.doi.org/10.1016/j.autcon.2024.105410>.
- [26] H. Chu, L. Deng, H. Yuan, L. Long, J. Guo, A transformer and self-cascade operation-based architecture for segmenting high-resolution bridge cracks, Autom. Constr. 158 (2024) 105194, <http://dx.doi.org/10.1016/j.autcon.2023.105194>.
- [27] H. Chu, P. Chun, Fine-grained crack segmentation for high-resolution images via a multiscale cascaded network, Computer-Aided Civ. Infrastruct. Eng. 39 (4) (2024) 575–594, <http://dx.doi.org/10.1111/mice.13111>.

- [28] P. Chun, T. Kikuta, Self-training with Bayesian neural networks and spatial priors for unsupervised domain adaptation in crack segmentation, Computer-Aided Civ. Infrastruct. Eng. 39 (17) (2024) 2642–2661, <http://dx.doi.org/10.1111/mice.13315>.
- [29] I. Katsamenis, N. Doulamis, A. Doulamis, E. Protopapadakis, A. Voulodimos, Simultaneous precise localization and classification of metal rust defects for robotic-driven maintenance and prefabrication using residual attention U-net, Autom. Constr. 137 (2022) 104182, <http://dx.doi.org/10.1016/j.autcon.2022.104182>.
- [30] K. Hattori, K. Oki, A. Sugita, T. Sugiyama, P. Chun, Deep learning-based corrosion inspection of long-span bridges with bim integration, Helion 10 (15) (2024) <http://dx.doi.org/10.1016/j.heliyon.2024.e35308>.
- [31] L. Huang, G. Fan, J. Li, H. Hao, Deep learning for automated multiclass surface damage detection in bridge inspections, Autom. Constr. 166 (2024) 105601, <http://dx.doi.org/10.1016/j.autcon.2024.105601>.
- [32] P.A.S. Santos, M.T.M. Carvalho, Multi-class segmentation of structural damage and pathological manifestations using YOLOv8 and segment anything model, Autom. Constr. 172 (2025) 106037, <http://dx.doi.org/10.1016/j.autcon.2025.106037>.
- [33] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 779–788, <http://dx.doi.org/10.1109/cvpr.2016.91>.
- [34] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, P. Dollár, R. Girshick, Segment anything, Proc. IEEE/CVF Int. Conf. Computer Vis. (2023) 4015–4026, <http://dx.doi.org/10.48550/arXiv.2304.02643>.
- [35] P. Chun, H. Chu, K. Shitara, T. Yamane, Y. Maemura, Implementation of explanatory texts output for bridge damage in a bridge inspection web system, Adv. Eng. Softw. 195 (2024) 103706, <http://dx.doi.org/10.1016/j.advengsoft.2024.103706>.
- [36] T. Kunlamarai, T. Yamane, M. Suganuma, P. Chun, T. Okatani, Improving visual question answering for bridge inspection by pre-training with external data of image-text pairs, Computer-Aided Civ. Infrastruct. Eng. 39 (3) (2024) 345–361, <http://dx.doi.org/10.1111/mice.13086>.
- [37] Q. Kong, J. Gu, B. Xiong, C. Yuan, Vision-aided three-dimensional damage quantification and finite element model geometric updating for reinforced concrete structures, Computer Aided Civ. Infrastruct. Eng. 38 (17) (2023) 2378–2390, <http://dx.doi.org/10.1111/mice.12967>.
- [38] Y. Zhang, B. Xia, S. Taylor, High-resolution 3-D geometry updating of digital functional models using point cloud processing and surface cut, Computer Aided Civ. Infrastruct. Eng. 39 (1) (2024) 3–19, <http://dx.doi.org/10.1111/mice.13076>.
- [39] T. Yamane, P. Chun, J. Dang, R. Honda, Recording of bridge damage areas by 3D integration of multiple images and reduction of the variability in detected results, Computer Aided Civ. Infrastruct. Eng. 38 (17) (2023) 2391–2407, <http://dx.doi.org/10.1111/mice.12971>.
- [40] Y. Gao, H. Li, W. Fu, C. Chai, T. Su, Damage volumetric assessment and digital twin synchronization based on LiDAR point clouds, Autom. Constr. 157 (2024) 105168, <http://dx.doi.org/10.1016/j.autcon.2023.105168>.
- [41] ASTM, Committee E57 on 3D imaging systems, 2006, URL <https://www.astm.org/get-involved/technical-committees/committee-E57>. (Accessed 27 July 27 2025).
- [42] D. Huber, The ASTM E57 file format for 3D imaging data exchange, Three Dimensional Imaging Interact. Meas. 7864 (2011) 88–96, <http://dx.doi.org/10.1117/12.876555>.
- [43] C. Lin, S. Abe, S. Zheng, X. Li, P. Chun, A structure-oriented loss function for automated semantic segmentation of bridge point clouds, Computer Aided Civ. Infrastruct. Eng. 40 (6) (2025) 801–816, <http://dx.doi.org/10.1111/mice.13422>.
- [44] T. Xia, J. Yang, L. Chen, Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning, Autom. Constr. 133 (2022) 103992, <http://dx.doi.org/10.1016/j.autcon.2021.103992>.
- [45] X. Yang, E. del Rey Castillo, Y. Zou, L. Wotherspoon, Semantic segmentation of bridge point clouds with a synthetic data augmentation strategy and graph-structured deep metric learning, Autom. Constr. 150 (2023) 104838, <http://dx.doi.org/10.1016/j.autcon.2023.104838>.
- [46] Y. Jing, B. Sheil, S. Acikgoz, A lightweight transformer-based neural network for large-scale masonry arch bridge point cloud segmentation, Computer Aided Civ. Infrastruct. Eng. 39 (16) (2024) 2427–2438, <http://dx.doi.org/10.1111/mice.13201>.
- [47] Y. Kim, J. Kim, J. Kim, H. Kim, Superpoint transformer-based bridge component recognition using UAV-mounted LiDAR and synthetic point cloud generation, J. Comput. Civ. Eng. 39 (4) (2025) 04025043, <http://dx.doi.org/10.1061/jccee5.cpeng-6383>.
- [48] C.R. Qi, L. Yi, H. Su, L.J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst. 30 (2017) <http://dx.doi.org/10.48550/arXiv.1706.02413>.
- [49] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation forest, 2008 Eighth IEEE Int. Conf. Data Min. (2008) 413–422, <http://dx.doi.org/10.1109/ICDM.2008.17>.
- [50] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 96, (34) 1996, pp. 226–231, <https://dl.acm.org/doi/10.5555/3001460.3001507>.
- [51] H. Edelsbrunner, D. Kirkpatrick, R. Seidel, On the shape of a set of points in the plane, IEEE Trans. Inform. Theory 29 (4) (1983) 551–559, <http://dx.doi.org/10.1109/tit.1983.1056714>.
- [52] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, Int. Conf. Med. Image Comput. Computer-Assisted Interv. (2015) 234–241, <http://dx.doi.org/10.48550/arXiv.1505.04597>.
- [53] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Proc. Eur. Conf. Comput. Vis. (2018) 801–818, [http://dx.doi.org/10.1007/978-3-030-01234-2\\_49](http://dx.doi.org/10.1007/978-3-030-01234-2_49).
- [54] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: Simple and efficient design for semantic segmentation with transformers, Adv. Neural Inf. Process. Syst. 34 (2021) 12077–12090, <http://dx.doi.org/10.48550/arXiv.2105.15203>.
- [55] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, Lora: Low-rank adaptation of large language models, ICLR 1 (2) (2022) <http://dx.doi.org/10.48550/arXiv.2106.09685>.
- [56] E. Bianchi, M. Hebdon, Corrosion condition state semantic segmentation dataset, 2021, [https://data.lib.vt.edu/articles/dataset/Corrosion\\_Condition\\_State\\_Semantic\\_Segmentation\\_Dataset/16624663](https://data.lib.vt.edu/articles/dataset/Corrosion_Condition_State_Semantic_Segmentation_Dataset/16624663). (Accessed 27 July 2025).
- [57] Z. Ameli, E. Lanis, Corrosion condition rating database, 2023, [https://digitalcommons.library.umaine.edu/student\\_work/30/](https://digitalcommons.library.umaine.edu/student_work/30/). (Accessed 27 July 2025).
- [58] CESIUM: The platform for 3D geospatial, 2025, <https://cesium.com/>. (Accessed 27 July 2025).
- [59] Z. Wang, X. Li, Y. Liu, Y. Lv, M. Li, Enhancing precision of defect 3D reconstruction in metal ultrasonic testing through point cloud completion, Ultrasonics 142 (2024) 107381, <http://dx.doi.org/10.1016/j.ultras.2024.107381>.
- [60] H. Huang, Y. Cai, C. Zhang, Y. Lu, A. Hammad, L. Fan, Crack detection of masonry structure based on thermal and visible image fusion and semantic segmentation, Autom. Constr. 158 (2024) 105213, <http://dx.doi.org/10.1016/j.autcon.2023.105213>.
- [61] C. Li, B. Liu, F. Li, X. Wei, X. Liang, W. He, X. Nie, Intelligent detection and modelling of composite damage based on ultrasonic point clouds and deep learning, Measurement 246 (2025) 116708, <http://dx.doi.org/10.1016/j.measurement.2025.116708>.
- [62] Y. Chen, C. Lin, X. Li, S. Kubo, T. Yamane, P.-j. Chun, Automated dimension estimation of bridge components using semantic segmentation and geometric fitting of point cloud data, Eng. Struct. 342 (2025) 120837, <http://dx.doi.org/10.1016/j.engstruct.2025.120837>.