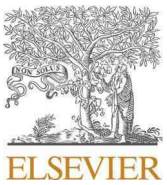
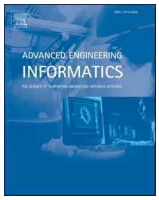
Advanced Engineering Informatics 65 (2025) 103373



Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Full length article

An efficient 2D-3D fusion method for bridge damage detection under complex backgrounds with imbalanced training data

Wen-Jie Zhang [a], Hua-Ping Wan [a,*], Michael D. Todd [b]

[a] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China
[b] Department of Structural Engineering, University of California, San Diego, 9500 Gilman Dr. 0085, La Jolla, CA 92093-0085, USA

ARTICLE INFO

Keywords:
Bridge inspection
Deep learning
Region of interest extraction
2D-3D data fusion
Damage segmentation and location

ABSTRACT

Existing bridge structures are inevitably affected by various adverse environments and loads during routine operations, which accelerates structural damage and highlights the necessity of conducting bridge inspections. Because of their cost-effectiveness and non-contact capabilities, computer vision methods applied to images from unmanned aerial vehicle (UAV) survey campaigns are promising ways to conduct bridge inspections. Bridge images captured by UAVs often contain numerous complex background pixels due to the small size of damage. Additionally, the existing damage datasets used for training suffer from a severe inter-class imbalance, which significantly affects the accuracy of damage recognition. This study proposes a 2D-3D fusion method for bridge damage segmentation and localization, effectively identifying damage under complex backgrounds with imbalanced data. First, a 3D reconstruction method is introduced to reconstruct bridge point clouds and generate depth maps from different viewpoints. Second, an RGB-D segmentation model is presented to extract the region of interest from images by integrating 2D and 3D information. Third, an improved DeepLabv3 + model is developed to segment damage and integrate it with point clouds for three-dimensional visualization. Field experiments are conducted on a multi-span simply supported girder bridge to validate the effectiveness of the proposed method. The ROI extraction model achieves an F-measure of 98.85%, and the damage segmentation model attains a mAP of 82.21%. Additionally, the 3D visualization result indicates areas of interest (e.g., wet spot, cavities, and spalling) on the cover girder, providing valuable guidance for bridge maintenance. These findings demonstrate the effectiveness and practicality of the proposed method in bridge inspection.

1. Introduction

Bridges are crucial components of transportation infrastructure and are inevitably subjected to adverse environmental conditions and loads, which can accelerate structural damage and reduce their lifespan. Multiple-class surface damage, such as cracks, spalling, and cavities, serve as the external manifestations of the bridge condition and suggest a deterioration in their load-bearing capacity. Statistically, approximately half of the bridges in the United States have been in service for over 50 years, posing a significant safety hazard [1]. Therefore, routine inspections of bridges are essential. Traditional bridge inspection typically involves manual visual assessment and large inspection equipment, leading to high costs, low efficiency, and significant safety risks.

In recent years, automated bridge inspection utilizing intelligent algorithms and sensor devices has attracted significant attention [2]. Inspectors can employ UAVs to enhance efficiency, minimize repetitive field labor, and improve safety during bridge inspections. Specifically, the automated bridge inspection process involves three components: data collection, damage detection, and damage localization [3]. The existing UAVs can handle most data collection tasks in bridge inspections [4–9]. This study will focus on damage detection and localization for bridges, assuming an image collection survey campaign has been conducted.

Computer vision-based damage detection is a prominent research area capable of automatically identifying structural defects in digital images. Compared to other inspection techniques, such as ultrasonic guided wave analysis [10], computer vision offers advantages in cost-effectiveness, wide field of view, and intuitive display of results. Early computer vision-based research focus on integrating image processing techniques with machine learning models to detect structural damage, using methods such as threshold segmentation [11,12] and support vector machines (SVM) [13]. However, these methods require manually

* Corresponding author.
E-mail address: hpwan@zju.edu.cn (H.-P. Wan).

https://doi.org/10.1016/j.aei.2025.103373
Received 28 December 2024; Received in revised form 7 April 2025; Accepted 14 April 2025
Available online 21 April 2025

pre-designed features sensitive to specific damage, limiting their generalization and robustness. Additionally, the above methods typically detect only one specific damage, making it difficult to generalize to multiple types of bridge damage, which restricts their practical application in bridge inspection.

With the rapidly advancing development of deep learning algorithms and computational power, deep learning-based segmentation methods have been widely used. Unlike feature-based methods, deep learning employs an end-to-end, data-driven training mechanism, making it more adaptable to data variations and consequently has better robustness to different types of damage. Therefore, numerous studies have utilized various deep learning-based segmentation networks, such as the Vision Transformer (ViT) [14], U-Net [15,16], Swin Transformer [17], and Fully Convolutional Networks (FCN) [18,19], for damage identification. Liu et al. [14] proposed a ViT Net for cable surface disease detection, achieving end-to-end localization of defective regions by combining differential convolution and attention mechanisms. Zoubir et al. [15] developed a U-Net based framework for crack semantic segmentation, utilizing the complementary advantages of the spatial and frequency domains to enhance crack extraction accuracy. Huang et al. [17] presented a Swim Transformer model for multi-category surface damage segmentation of bridge members by introducing a CARAFE upsampler and transfer learning to significantly enhance the recognition accuracy. Despite the considerable amount of research that has been done, the results of existing damage semantic segmentation tasks are relatively unsatisfactory, with the mean Intersection over Union (mIoU) typically ranging from 0.6 to 0.8 [20]. This is attributed to data complexity, where the performance of models consistently declines as additional variables and contextual elements appear in the datasets.

Deep learning models typically contain millions of network parameters and require large, high-quality datasets for training [21]. To address practical challenges in damage segmentation, the quality of available image data must be improved for higher accuracy and generalization of the segmentation models. Specifically, there are two primary issues with image data that necessitate attention. The first is the issue of complex backgrounds. Most bridges are situated in natural and often complex environments, and bridge images captured by UAVs during bridge inspections often include irrelevant elements, such as rivers, trees, hills, and sky, significantly hindering damage identification. Consequently, current research focuses on developing methods for effective damage segmentation in images with complex backgrounds rather than simple scenes [20]. One approach is to improve the accuracy of image segmentation methods under complex backgrounds [22–26]. Yu et al. [23] proposed a U-Net-based crack segmentation network to address the challenges posed by irregular crack shapes and complex background interference. However, this approach places high demands on model performance and dataset quality, causing existing research on damage segmentation in complex backgrounds to stagnate, with an average mean Intersection over Union (mIoU) of only 0.601 [20]. Another straightforward solution is to limit detection to regions of interest (ROIs) with damage. The identification of ROIs can be achieved through deep learning-based recognition [27–29] or guided by Building Information Modeling (BIM) data [30,31], both of which require additional computational resources and prior knowledge. In particular, Xiao et al. [27] introduced an ROI extraction method combining a point cloud segmentation network with 3D projection to generate bridge images containing only ROIs. This method requires large-scale point cloud computation, which demands substantial computational power. Despite this, integrating three-dimensional (3D) information to aid in ROI extraction remains a promising approach for improving the accuracy of damage segmentation.

Another issue is the inter-class imbalance in the training dataset. The small size of common bridge damage (e.g., cracks, cavities, and spalling) compared to bridge components results in a low proportion of damage pixels in the dataset. For instance, only 0.3% of the total pixel area in the entire dacl10k dataset is labeled as a "crack" [62], leading to an extreme imbalance between positive and negative samples. During model training, the loss values generated by the positive samples tend to be swamped by the negative samples (i.e., the background class), hindering effective parameter optimization. Typical learning models often suffer from "majority bias" in the presence of inter-class imbalance, performing well in most classes but poorly in a few. This outcome is undesirable because the minority group, such as bridge damage, is usually the primary focus [32]. Existing research addresses the inter-class imbalance problem at both the data and algorithmic levels. At the algorithmic level, researchers often introduce improved loss functions to increase the weight of positive sample classes. Commonly used loss functions include weighted cross entropy loss [33], dice loss [34–36], focal loss [37], and Jaccard loss [38]. Rakshitha et al. [34] conducted a comprehensive evaluation of various loss functions and demonstrated that binary focal loss is particularly effective in addressing the class imbalance problem. Nonetheless, it has also been argued that the impact of different loss functions on the detection of small targets (e.g., cracks) is not paramount [39]. Data-level methods focus on processing image data. The first approach involves data sampling, which selects samples based on their difficulty and increases the number of samples in a few classes. These resampling methods can alter the original data distribution and may lead to model overfitting. The second approach is data augmentation, which expands the dataset and enhances the generalization ability of models. The generative adversarial network (GAN) [40–43] and the diffusion model [44,45] have been effectively used as data augmentation methods. Wang et al. [40] proposed a data augmentation module based on the GAN, capable of efficiently generating high-quality multi-class surface defect data. Nevertheless, the training process of the above models is challenging and vulnerable to breakdown, and the generated image data still contains many background pixels, which does not fundamentally resolve the inter-class imbalance problem. The emergence of powerful game engines like Unity 3D and Godot has significantly advanced data synthesis technology, providing a foundation for generating realistic virtual data [46,47]. Constructing bridge models in a virtual environment offers a solution to the imbalance problem between positive and negative samples. However, in complex bridge scenarios, data enhancement strategies based on virtual models are not yet fully developed, and their effectiveness in generating data lacks sufficient experimental validation.

Once the damage is identified, capturing its three-dimensional (3D) location on the bridge surface is crucial for the subsequent maintenance. Consequently, extensive attention has been given to 3D reconstruction-based damage modeling [48–50], which provides valuable information about the location and geometry of the damage. As noted by Spencer Jr. et al. [51], analyzing damage identified at the image level within the context of the global structural background is essential for understanding its relevance to structural safety. Ni et al. [53] developed a 3D reconstruction-based method for locating beam surface damage, achieving meter-level localization accuracy. It is worth noting that most studies on damage detection and localization are compartmentalized, lacking interconnection, which leads to the loss of valuable information. Given the homogeneity of images and point clouds [54], both of which describe the visual characteristics of targets from different dimensions and are interrelated through 2D-3D geometric mapping, integrating 2D images and 3D point clouds offers a promising approach for addressing challenges in damage segmentation and localization. Studies have been conducted to compute camera positions and achieve the fusion of 2D-3D information using techniques such as optical flow [55] and structure-from-motion (SfM) [56]. Liu et al. [52] proposed an automatic concrete crack assessment method based on the structure from motion (SfM) algorithm, which accurately determines the 3D coordinates of cracks. Hattori et al. [55] developed an image-processing-based method for measuring bridge corrosion and utilized the optical flow technique to integrate data into a BIM model. Yamane et al. [56] proposed a method to detect damage from bridge images taken from different directions using SfM and deep learning. However, in the above-mentioned

literature, the fusion of 2D and 3D information generally focuses on the 3D localization of damage and does not enhance damage detection in 2D images. The 3D point cloud model contains substantial spatial information, which can effectively support image-level damage detection, a key focus of this study.

In this study, a vision-based method using 2D-3D data fusion is proposed for bridge damage segmentation and localization, which effectively identifies damage under complex backgrounds with imbalanced data. First, the bridge pier point clouds are reconstructed using a 3D reconstruction technique to generate depth maps from various viewpoints. Next, an RGB-D-based ROI extraction model is employed to remove complex backgrounds from bridge images by integrating 2D and 3D information. Then, an improved segmentation model is proposed to segment bridge damage and obtain the segmentation masks. Finally, the 3D visualization of bridge damage is achieved by combining the camera poses with the segmentation masks. In this study, 2D-3D data are systematically integrated to exploit the homogeneity between images and point clouds, achieving pixel-level interactions through 2D-3D

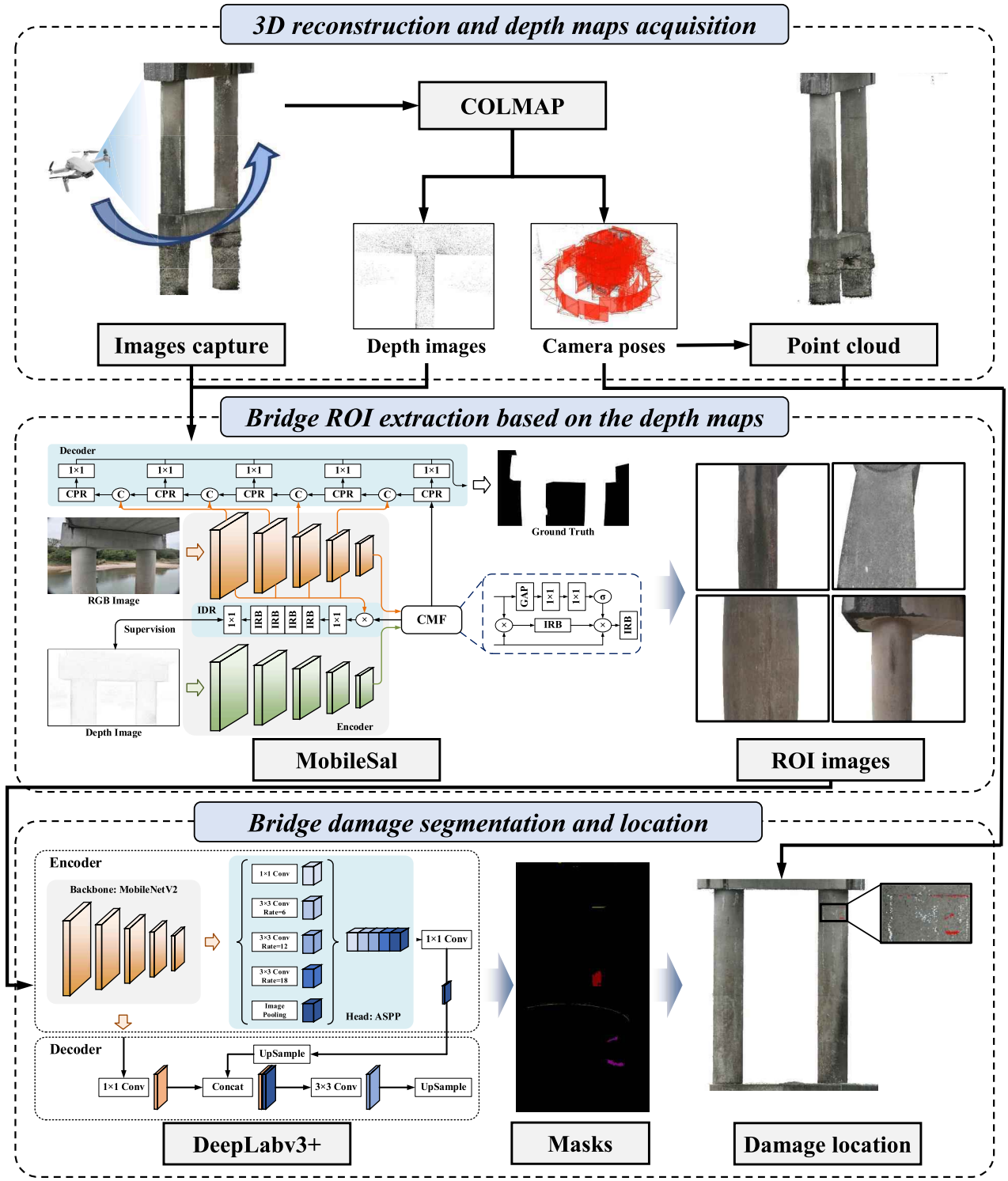


Fig. 1. Flowchart of the proposed method for detection of bridge damage.

geometric mapping. Here, 3D point cloud data assist in the ROI extraction of 2D images, and 2D damage segmentation information is integrated into the 3D model for visualization. Specifically, the contributions of this work are threefold: (1) a ROI extraction model integrating 2D RGB images and 3D depth information is proposed to efficiently extract bridge pixels from complex backgrounds. The depth information is innovatively introduced into 2D image processing through a projection method, which avoids the computational burden of processing large-scale point clouds and improves the accuracy of damage segmentation; (2) an improved DeepLabv3 + model was developed for bridge damage segmentation, incorporating copy-and-paste data augmentation and focal loss to address the significant inter-class imbalance encountered in practical damage detection, thereby significantly enhancing segmentation accuracy; and (3) a camera pose-based damage localization method is presented, utilizing the hidden points removal algorithm to compute visible points in the point cloud, enabling the three-dimensional visualization of bridge damage.

2. The proposed approach

2.1. Overview

The framework of the proposed bridge damage segmentation and localization method is illustrated in Fig. 1. It comprises three main components: 3D point cloud reconstruction and depth map generation, extraction of the bridge image ROIs, and segmentation and localization of bridge damage. Initially, the COLMAP [57] is employed to reconstruct the point cloud of the bridge pier and generate depth maps from various viewpoints. Subsequently, the MobileSal RGB-D segmentation network is introduced to extract the ROIs of bridge images by integrating 2D and 3D information. Finally, an improved DeepLabv3 + model is developed to identify and localize bridge damage using camera poses.

2.2. 3D reconstruction and depth map acquisition

For the bridge inspection task described in this study, the inspector uses UAVs to capture comprehensive images of the bridge. These images, which include the spatial features of the scene and information about bridge damage, are used to reconstruct the 3D point cloud and for subsequent damage detection. The open-source COLMAP software [57] is employed to perform the 3D reconstruction of the bridge piers. The key parameters of COLMAP are summarized in Table 1, while other parameters are configured using the default settings.

The bridge point cloud contains extensive spatial information, which effectively supports the image ROI extraction. However, the large volume of point cloud data imposes a significant computational burden when operated directly. To address this, the study projects the point cloud into 2D space and converts its spatial information into a depth map, facilitating subsequent ROI extraction.

The 3D point cloud is projected onto the pixel plane, with the coordinate transformation relationship illustrated in Fig. 2. Initially, the world coordinate system (U, V, W) is transformed into the camera coordinate system (X, Y, Z) using the external parameters, which can be described as

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} K_1 & K_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U \\ V \\ W \\ 1 \end{pmatrix} \tag{1}$$

where K_1 and K_2 represent the external parameter matrix, which are derived from the camera pose obtained by the COLMAP software during the reconstruction process. Then, the camera coordinate system (X, Y, Z) is transformed into the image coordinate system (x, y), which can be expressed as

$$Z\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_c & 0 & 0 & 0 \\ 0 & f_c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2}$$

$$\begin{pmatrix} x_{cor} \\ y_{cor} \end{pmatrix} = \begin{pmatrix} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)x + 2p_1 xy + p_2(r^2 + 2x^2) \\ (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)y + 2p_2 xy + p_1(r^2 + 2y^2) \end{pmatrix} \tag{3}$$

where f_c represent the focal length of the camera, and r^2 is equal to $(x^2 + y^2)$, and k_1, k_2, k_3, p_1, and p_2 represent the distortion parameters, which are obtained from the calibration. Finally, the points on the pixel coordinate system (u, v) can be obtained by

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & u_o \\ 0 & f_y & v_o \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{cor} \\ y_{cor} \\ 1 \end{pmatrix} \tag{4}$$

where f_x and f_y denote the ratio of pixel to physical size in the x and y direction, respectively, and u_o and v_o represent the coordinates of the pixel origin in the image coordinate system.

In summary, the points in the 3D point cloud are projected onto the 2D image, with the corresponding depth information Z serving as the pixel values. To facilitate subsequent model training, the depth maps are saved as grayscale images. Fig. 3 illustrates several typical depth maps.

The RGB and depth images collectively constitute the RGB-D image dataset, which integrates color and depth information while avoiding the computational burden of processing large-volume point clouds. This integration provides a solid foundation for subsequent data processing.

Table 1
The parameters of COLMAP software.

Settings		Parameters
Feature extractor	Camera model	RADIAL
	Feature extraction	SIFT
	num_threads	−1
	use_gpu	1
	max_image_size	3200
	max_num_features	8193
Sift Matching	Feature matching	Exhaustive matching
	max_distance	0.7
	max_num_matches	32,768
Reconstruction	Quality	High

2.3. Bridge image ROI extraction

Bridge images often contain complex background pixels, making the bridge not the primary object, which can hinder the identification of bridge damage. This study uses RGB-D images to achieve fast and accurate extraction of ROI by combining 2D and 3D information, thereby eliminating background interference in bridge damage segmentation.

The proposed image ROI extraction model MobileSal [58] is depicted in Fig. 4, which mainly consists of the Encoder, the Feature Fusion module, and the Decoder. The Encoder consists of two branches: an RGB stream and a depth stream, each used to extract features from the input RGB and depth images. In RGB stream, the MobileNetV2 serves as the backbone network, executing five stages of feature extraction and producing five feature maps: C_1, C_2, C_3, C_4, and C_5. The network structure of the depth stream parallels that of the RGB stream but with fewer convolutional blocks, as depth maps contain less semantic information than RGB images, thereby reducing computational complexity. The output feature maps of the five stages in the depth stream are denoted as D_1, D_2, D_3, D_4, and D_5.

The depth images reveal the spatial information of RGB images, aiding in distinguishing foreground objects from the background, particularly in complex environments. Therefore, effective fusion of RGB and depth features is crucial for accurate ROI extraction. To ensure the high efficiency of the model, only the RGB feature map C_5 and the depth feature map D_5 are fused. The feature fusion module used in this

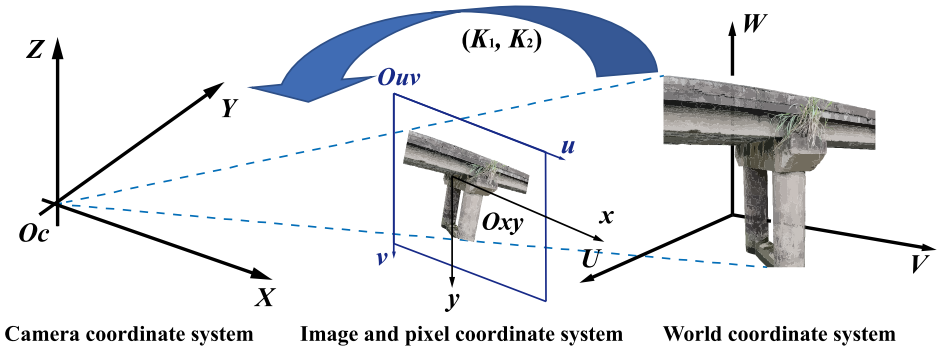


Fig. 2. Coordinate system transformation relationship.

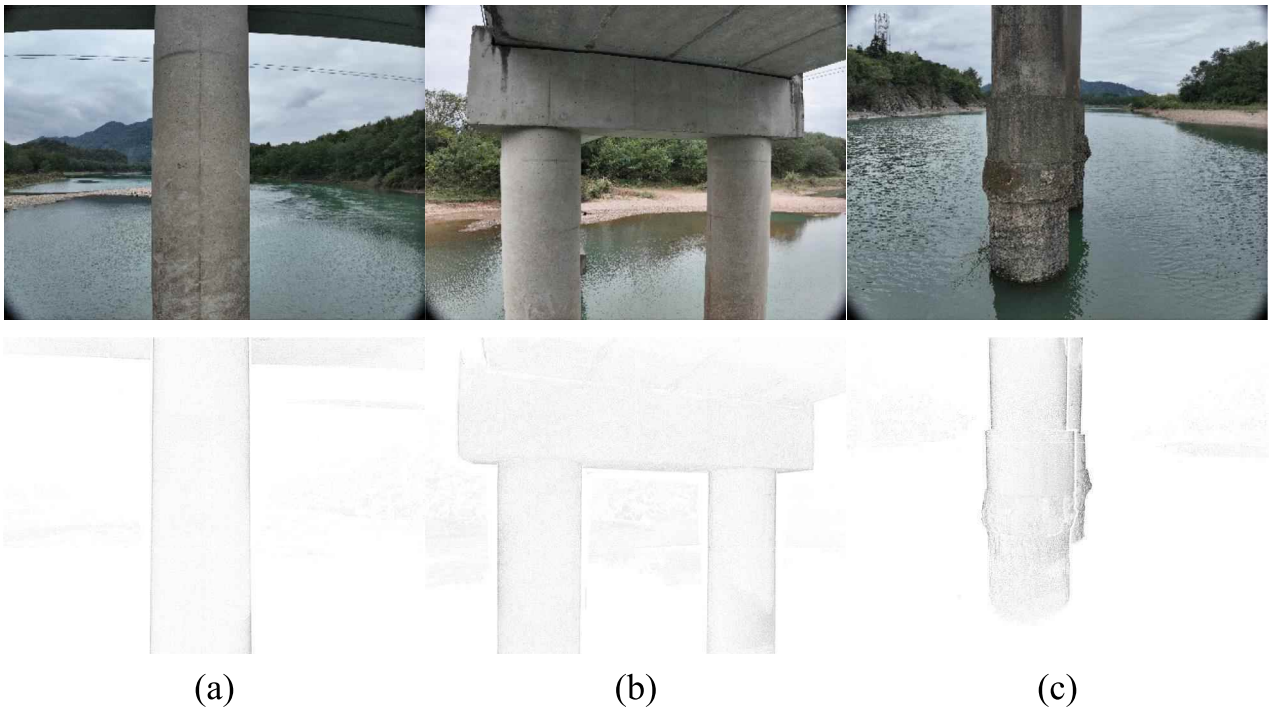


Fig. 3. The generated depth images (gray-scale format).

study is shown in Fig. 4, named the Cross-Modal Fusion (CMF) module. The final fused feature C_5^d is generated by combining RGB features, which provide semantic information, with depth features, which approximate a priori information about the shape or structure of the ROIs.

The proposed Decoder employs the Compact Pyramid Refinement (CPR) module as its basic unit, utilizing multi-branch depth-wise separable convolution and 1 × 1 convolution to enhance modeling efficiency, as illustrated in Fig. 4. For the accurate image ROI extraction, it is essential to fully utilize both high-level and low-level features. In each stage of the Decoder, feature maps from the decoder and the corresponding encoder stages are first concatenated dimensionally, then fused by the CPR module, thereby aggregating multilevel features from top to bottom. Finally, after 1 × 1 convolutional dimensionality reduction, the Decoder outputs the mask of the image ROI.

The adopted ROI extraction model extensively utilizes 1 × 1 convolution and depth-wise separable convolution, employing the lightweight MobileNetV2 as its backbone. This approach significantly enhances network efficiency and avoids the high computational burden associated with large-scale point cloud processing.

2.4. Bridge damage segmentation and location

Once the ROIs of the bridge images are obtained, the bridge damage can be segmented. An improved DeepLabv3 + model is proposed to recognize bridge defects within the ROIs and generate the corresponding damage masks. Furthermore, spatial localization of bridge damage is crucial for subsequent maintenance. This study proposes a method called Damage-Painting, which projects damage information onto a 3D point cloud using camera positions from 3D reconstruction, enabling 3D visualization of bridge damage.

2.4.1. Damage segmentation based on the improved DeepLabv3 + model

The DeepLabv3 + model [59] is a state-of-the-art semantic segmentation algorithm capable of assigning object class labels to each pixel in an image. As illustrated in Fig. 5, the DeepLabv3 + model is comprised of the Encoder and the Decoder. The Encoder utilizes ResNet101 as the backbone network and incorporates Atrous Spatial Pyramid Pooling (ASPP) to extract image features across various receptive fields, thereby producing image features that integrate multi-scale information. The Decoder processes low-level feature maps from the intermediate layers of the backbone and the output from the ASPP

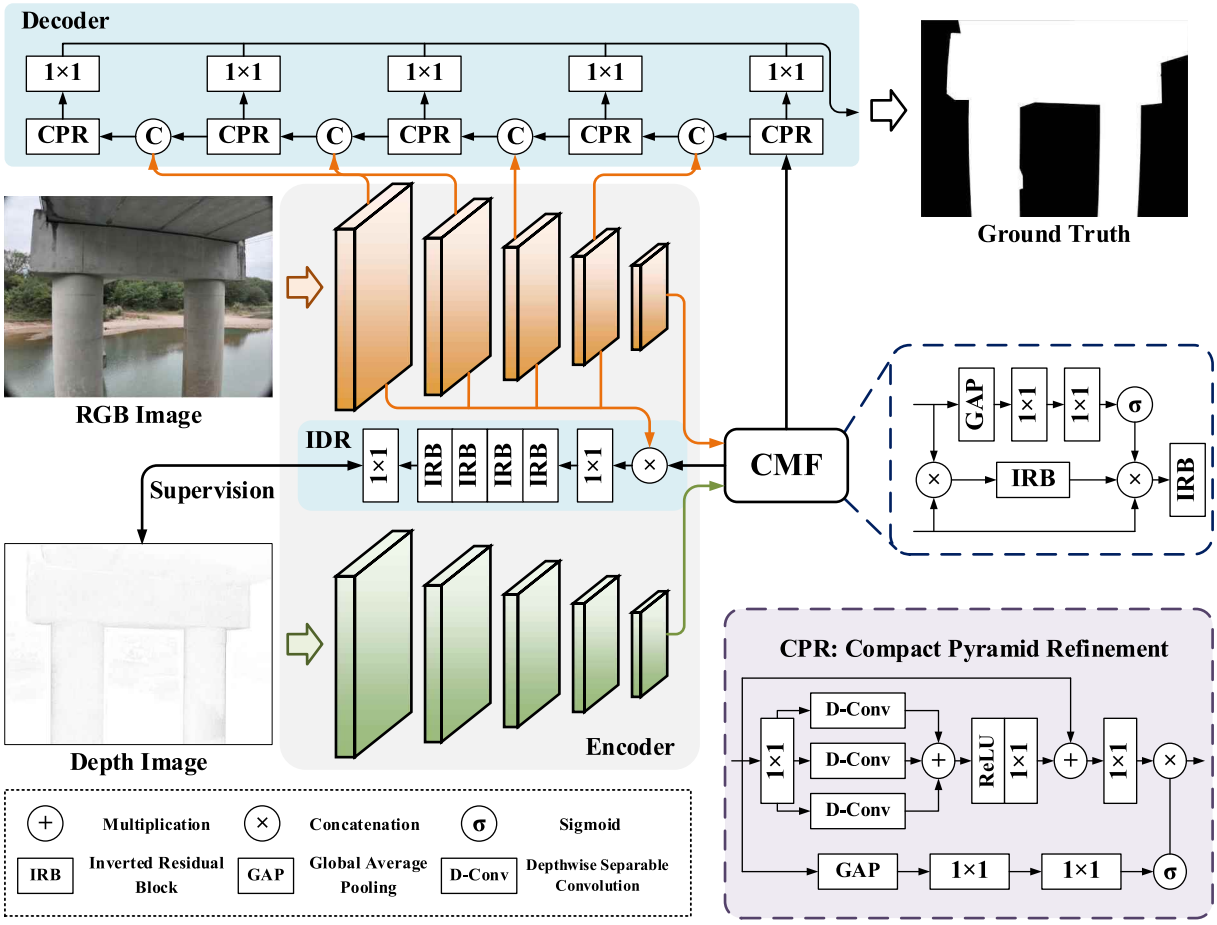


Fig. 4. The structure of MobileSal model.

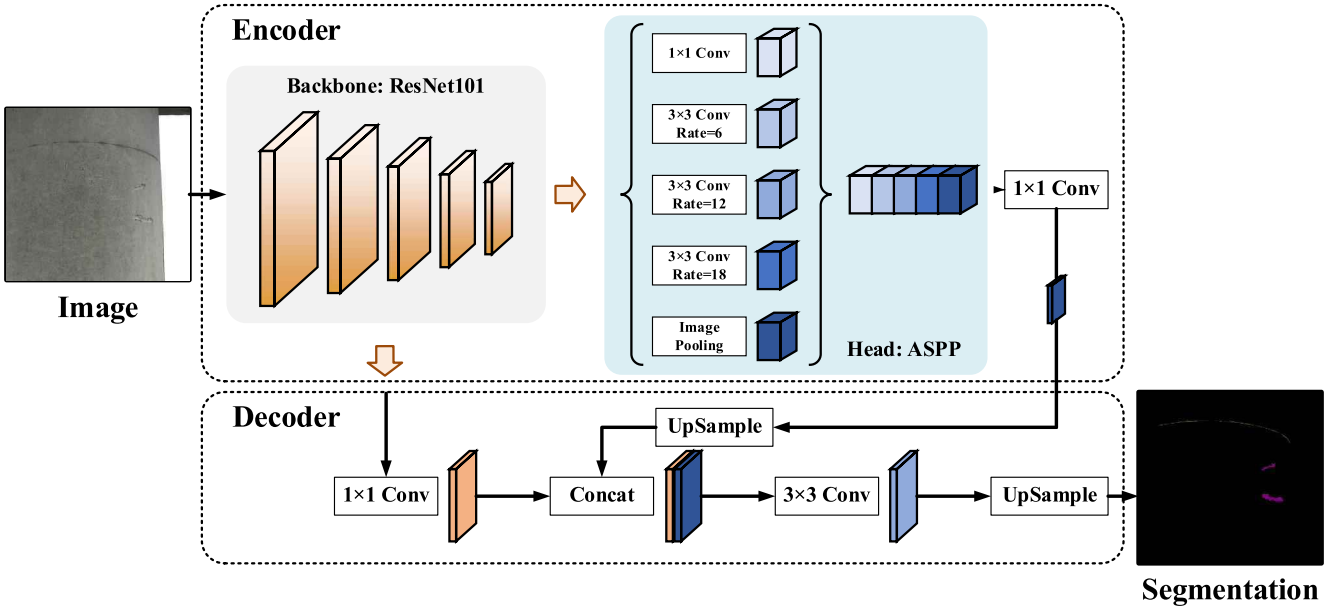


Fig. 5. The structure of the DeepLabv3 + model.

module. Through multiple convolutional and interpolation upsampling operations, it ultimately generates the predicted mask maps.

Bridge damage is a visual representation of the health of a bridge, and its relatively small size compared to the bridge makes it difficult to detect. This characteristic is evident in images, where damage pixels are significantly fewer than background pixels, leading to a substantial inter-class imbalance. The severe imbalance causes the loss to be dominated by the majority class, resulting in overfitting to the majority

class during training and poor detection performance for the minority class. Table 2 shows the percentage of pixels for different classes in the bridge damage dataset used in this study, highlighting the substantial predominance of background pixels over damage pixels and the serious inter-class imbalance.

To address the inter-class imbalance problem, this study employs a data augmentation method known as copy-paste. First, a source image I_1 and a target image I_2 are randomly selected, and the target image I_2 undergoes a random scale transformation. Next, the mask α of the pasted object is computed using the ground truth to extract the pixels from the masked region of the source image I_1. Finally, the extracted pixels are randomly pasted onto the target image I_2. This entire process can be represented as

$$I_{\text{aug}} = \alpha I_1 + (1-\alpha) I_2 \tag{5}$$

where I_{aug} represents the augmented image. Fig. 6 shows some typical augmented images. The copy-paste-enhanced dataset increases the proportion of bridge damage pixels, effectively alleviating the inter-class imbalance problem.

Furthermore, the focal loss [60] is introduced to increase the weight of damage sample loss. Focal loss, derived from the cross-entropy loss, incorporates a dynamic scaling factor that reduces the weight of easily distinguishable samples during training. The focal loss enables the model to concentrate on hard samples, which can be given by

$$\text{FL}(p) = -\beta(1-p)^{\gamma}\log(p) \tag{6}$$

where p denotes the probability that the model predicts the pixel belongs to the foreground, and β represents the weights of different categories, and γ is a pre-set parameter, set to 2 in this study. When p approaches 1, indicating that the sample is easy to distinguish, the modulation factor $(1-p)^{\gamma}$ approaches 0, thereby reducing the loss contribution from these easy samples. Conversely, when a pixel is misclassified and p is very small, the modulation factor $(1-p)^{\gamma}$ approaches 1, minimizing its effect on the loss. Therefore, the introduction of focal loss effectively alleviates the inter-class imbalance problem in semantic segmentation.

2.4.2. Bridge damage localization based on the Damage-Painting method

The improved DeepLabv3 + model enables only two-dimensional image damage segmentation, while three-dimensional localization is crucial for subsequent bridge maintenance. This study proposes the Damage-Painting method to achieve 3D visualization of the bridge damage by integrating the segmentation masks with the camera poses.

To prevent coloring the hidden points, the visible points of the point cloud need to be calculated from the current camera position. For this purpose, the Hidden Points Removal (HPR) algorithm [61] is employed, which consists of two steps: inversion and convex hull construction. Fig. 7 illustrates the point cloud with visible points calculated by the HPR algorithm, where red points indicate visible points from this viewpoint, and blue points represent invisible points.

Subsequently, the point cloud containing only visible points is projected onto the 2D image (shown in Section 2.2) to establish correspondence between the pixel points of the image and the points in the 3D point cloud. Finally, the segmented damage mask is combined with the projected 2D image to determine if the pixel points belong to the damage. The color of the point in the point cloud, corresponding to each pixel identified as damage, is altered, and damage information is appended. The whole process is illustrated in Fig. 8.

Table 2
Pixel ratios for each class in the bridge damage dataset.

Class	Background	Wet spot	Cavity	Crack	Rock pocket	Spalling
Ratio	88.71 %	3.98 %	1.17 %	0.60 %	0.49 %	5.04 %

3. Experimental dataset and evaluation metrics

To validate the proposed bridge damage segmentation and location method, an experimental study is conducted on the piers of an actual bridge.

3.1. Image collection using UAVs

Field tests for this study are conducted on a multi-span simply-supported beam bridge in China, as shown in Fig. 9. The simply-supported beam bridge, has a total length of 205.29 m and a width of 5.5 m, and it was opened to traffic in 2011. Due to river scour and environmental erosion, various bridge piers have developed certain bridge diseases, which are suitable for the research object of this experiment.

The experiment utilizes a DJI Mavic 3E drone for aerial photography, capturing images with an image size of 5280 × 3956 pixels. The drone performs 360° surround shots of multiple piers of the bridge, as depicted in Fig. 10, capturing images from various heights and shooting distances. The overlap rate between neighboring images exceeds 70%. In this study, 145, 163, and 211 images are captured for each of the three piers of the bridge, respectively. These images are used for 3D reconstruction and ROI extraction model training.

3.2. Dataset for damage segmentation

Fig. 11 presents a portion of the dataset used to train the improved DeepLabv3 + model, provided by Flotzinger et al [62]. The dataset contains numerous images of real bridges and their corresponding segmentation masks. For this study, 4571 images of damaged bridges are selected, with 80% used for model training and the remaining 20% for testing. The stratified sampling method is used to divide the training and test sets. The details of the training set are shown in Table 2.

3.3. Evaluation metrics

Following relevant works [58,64], two widely used metrics are employed to evaluate the ROI extraction model. The first metric is the F-measure, which is a harmonic mean of precision and recall. The precision and recall can be calculated by

$$precision = \frac{|M \cap G|}{|M|} \tag{7}$$

$$recall = \frac{|M \cap G|}{|G|} \tag{8}$$

where M and G represent the predicted masks and ground truths, respectively. Then the F-measure for ROI extraction can be obtained by

$$\text{F - measure} = \frac{2precision \cdot recall}{precision + recall} \tag{9}$$

The second metric is mean absolute error (MAE), which is used to evaluate the error between the prediction mask and the ground truth, which is defined as

$$\text{MAE} = \frac{1}{W \cdot H}\sum_{x=1}^{W}\sum_{y=1}^{H}|M(x,y) - G(x,y)| \tag{10}$$

where W and H represent the height and width of the image.

For the bridge damage segmentation model, three metrics are used to evaluate its performance: Pixel Accuracy (PA), Mean Intersection over Union (mIoU), and Frequency Weighted Intersection over Union (FWIoU). The PA metric can be given by

$$\text{PA} = \frac{\sum_{i=0}^{c} p_{ii}}{\sum_{i=0}^{c}\sum_{j=0}^{c} p_{ij}} \tag{11}$$

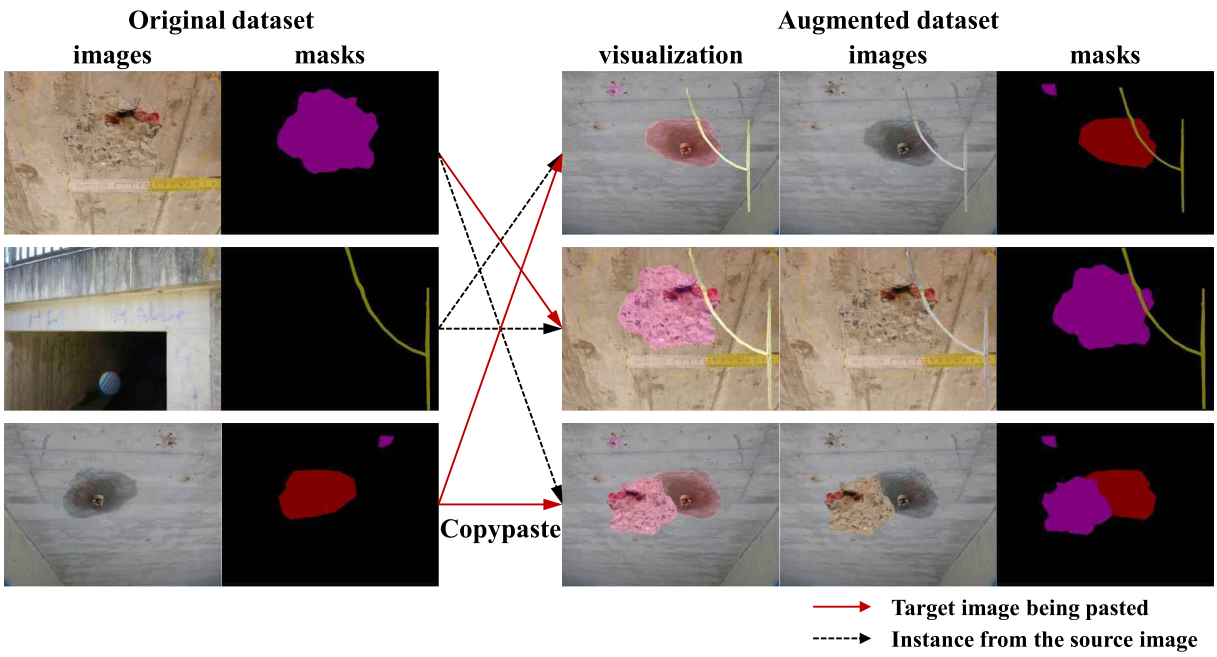


Fig. 6. Images after copy-paste enhancement.

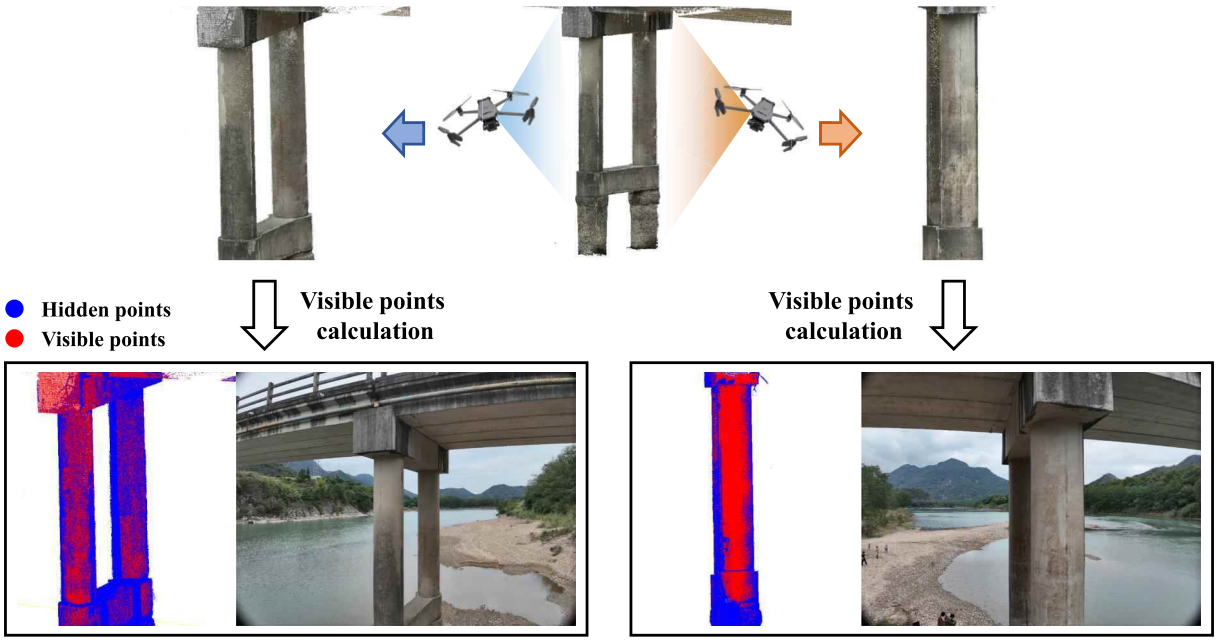


Fig. 7. The visible points under different viewpoints.

where c denotes the number of categories of bridge damage, and p_{ii} represents the number of pixels correctly identified, and p_{ij} indicates the number of pixels that incorrectly identify category i as category j. The mIoU measures the overlap between the predicted masks and the ground truths, providing a comprehensive evaluation of the model. Building on this, the FWIoU assigns weights to each category based on their frequency of occurrence. The above metrics can be computed as

$$\mathrm{mIoU} = \frac{1}{c+1}\sum_{i=0}^{c}\frac{p_{ii}}{\sum_{j=0}^{c}p_{ij}+\sum_{j=0}^{c}p_{ji}-p_{ii}} \tag{12}$$

$$\mathrm{FWIoU} = \frac{1}{\sum_{i=0}^{c}\sum_{j=0}^{c}p_{ij}}\sum_{i=0}^{c}\frac{\sum_{j=0}^{c}p_{ij}p_{ii}}{\sum_{j=0}^{c}p_{ij}+\sum_{j=0}^{c}p_{ji}-p_{ii}} \tag{13}$$

4. Validation results

All experiments in this study are conducted in Python using the PyTorch framework version 2.5.0. The high-performance computer used to implement the program comprises an AMD 7713 processor with 64 cores and an NVIDIA RTX 4090 24 GB GPU.

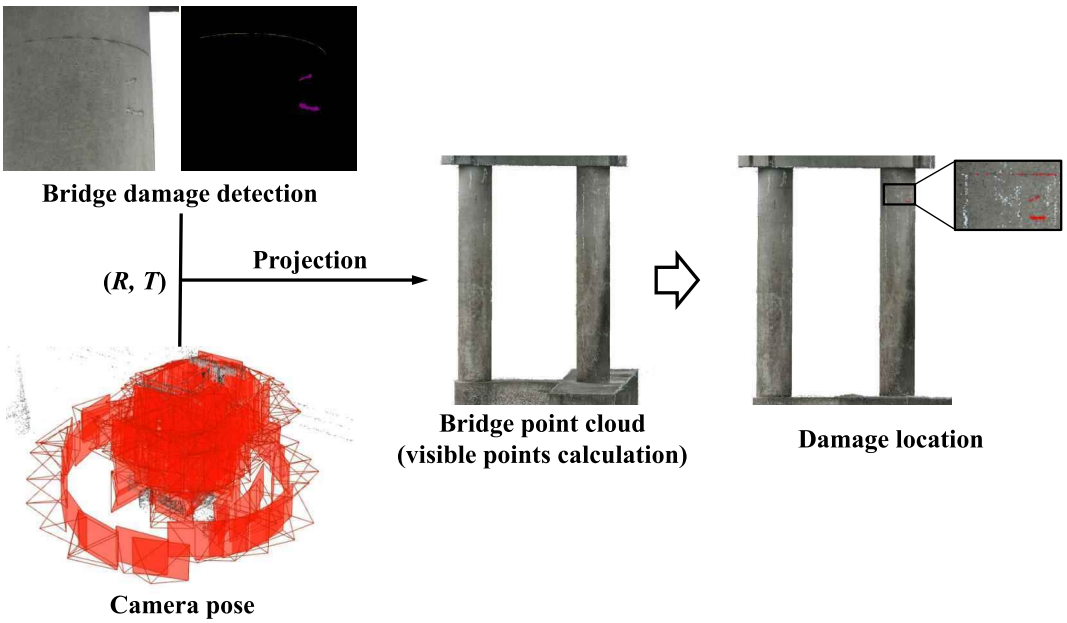


Fig. 8. The process of the Damage-Painting method.

Fig. 9. Multi-span simply-supported beam bridge.

4.1. 3D reconstruction and ROI extraction of bridge piers

The COLMAP software is used to reconstruct the piers of the simply-supported beam bridge, with the resulting point clouds displayed in Fig. 12. Table 3 presents the number and density of points in the reconstructed point clouds. The density is estimated by calculating the average distance between points in the cloud: the smaller the average distance, the denser the point cloud. In this study, the surface texture and other details of the piers are relatively clear, and the point cloud density is enough to meet subsequent processing requirements.

Based on the camera poses obtained from the 3D reconstruction process, the point clouds are projected onto the image coordinate system to obtain the depth maps. The RGB and depth images constitute the RGB-D image dataset, which contains a total of 519 RGB images and corresponding depth images. 80% of the dataset is used to train the ROI extraction model and the remaining 20% is used for testing.

The batch size, initial learning rate, training epochs, and input size for the ROI extraction model are set to 64, 0.001, 60, and 320 × 320, respectively. Two training modes are employed: one using depth maps and the other without. Fig. 13 illustrates the training process with depth images.

Table 4 presents the performance of the ROI extraction model on the test set under different training modes, highlighting the highest F-measure of 98.85% and the lowest MAE of 0.0161. Training with depth image assistance results in a higher F-measure and lower MAE, demonstrating that incorporating depth images enhances the accuracy of ROI extraction. Fig. 14 displays the results of ROI extraction for several bridge images, demonstrating effective extraction of bridge pixels.

Additionally, the MobileSal model utilizing depth images contains only 6.5 million parameters and runs at 68 frames per second (fps) on the RTX 4090 GPU. Compared to the model without depth images, the parameters increase by only 22.6%, meeting the requirement of real-time inference. According to the related study [58], the model can

Fig. 10. UAV image acquisition process.

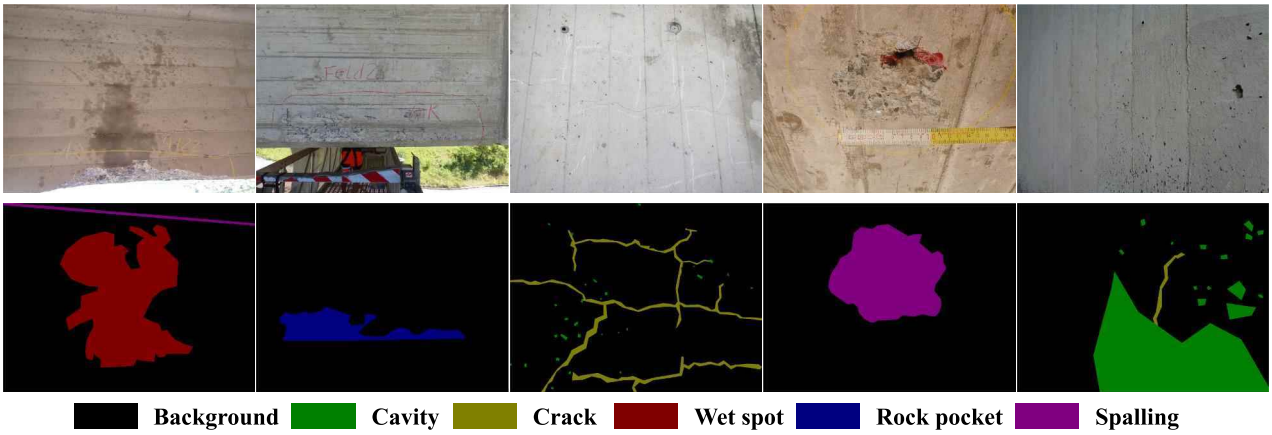


Fig. 11. The dataset for the damage segmentation model [63].

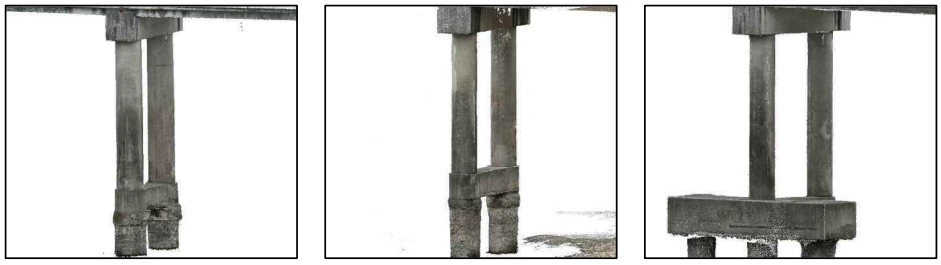


Fig. 12. The reconstructed point clouds of the bridge piers.

Table 3
Quality of the reconstructed point cloud.

Metrics	Bridge pier 1	Bridge pier 2	Bridge pier 3
Points number	2,263,963	2,212,864	2,530,358
Density (10^{-3})	2.7197	3.1834	2.7288

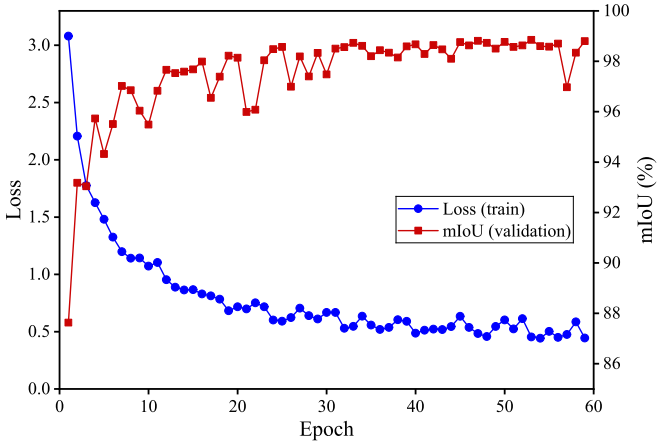


Fig. 13. The training process of the MobileSal model.

Table 4
The performance of the MobileSal model.

Mode	MAE	F-measure (%)	Parameters (M)	Speed (FPS)
With depth maps	0.0161	98.85	6.5	68
Without depth maps	0.0514	94.73	5.3	89

achieve a real-time inference speed of 43 ms on an Intel i7 8700 K CPU. Therefore, the MobileSal model enables efficient ROI extraction with minimal computational cost.

4.2. Bridge damage segmentation

The training epochs, batch size, initial learning rate, momentum, and weight decay of the improved DeepLabv3 + model are set to 180, 32, 0.01, 0.9, and 0.0005, respectively, and the ResNet-101 model is introduced to serve as the backbone. Fig. 15 illustrates the training process of the improved DeepLabv3 + model. The corresponding curves demonstrate the gradual convergence of the model, with increasing accuracy on the validation set and no signs of overfitting.

Table 5 presents the damage segmentation performance of the improved DeepLabv3 + model on the test set. The experimental results indicate that the model achieves 92.40% PA, 82.21% mIoU, and 87.72% FWIoU, with segmentation accuracy comparable to manual labeling, thereby verifying the feasibility of the proposed damage segmentation model. Additionally, Table 5 provides the IoU values for different classes. The model achieves the highest segmentation performance on the background category, reaching 90.02%, and the lowest segmentation performance on the crack and rock pocket categories. This corresponds to the pixel proportions in the dataset after copy-paste augmentation, as shown in Table 6. The highest proportion of background pixels and the lowest proportion of crack and rock pocket pixels after enhancement affect the segmentation performance of the model across different categories.

To further evaluate the effectiveness of various improvements to the model, an ablation study is conducted, and the results are presented in Table 7. The original DeepLabv3 + model is served as the baseline, with enhancements added incrementally. First, the cross-entropy loss is replaced with focal loss. Compared to the original DeepLabv3 + model, the introduction of focal loss improved PA, mIoU, and FWIoU by 3.65%,

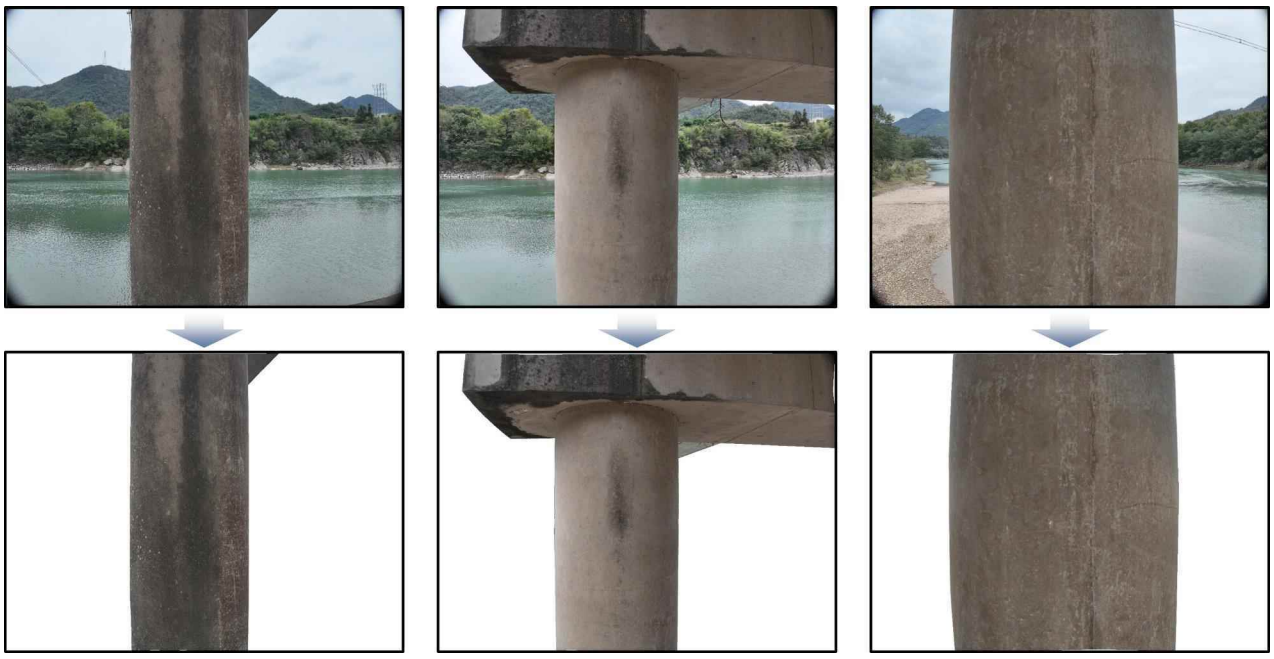


Fig. 14. The ROI extraction of bridge images.

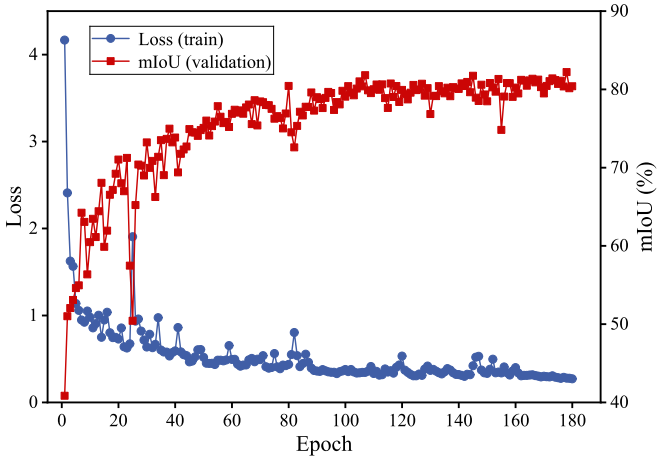


Fig. 15. The training process of the improved DeepLabv3 + model.

2.81%, and 4.1%, respectively, demonstrating that focal loss can enhance segmentation accuracy under inter-class imbalance. However, the segmentation performance remains insufficient for practical application. Subsequently, the dataset is augmented using the copy-paste method, and pixel ratios of the augmented data are shown in Table 6.

Compared with the data before augmentation, the inter-class imbalance problem of the enhanced dataset is alleviated, and the ratio between the least and most pixels is decreased from two orders of magnitude to one. The enhanced dataset is used to train the segmentation model, and the performance of the model is significantly improved, with the PA, mIoU, and FWIoU improving by 5.51%, 50.36%, and 7.14%, respectively. The substantial mIoU improvement indicates effective mitigation of inter-class imbalance, bringing the model performance to an acceptable level. These results demonstrate the effectiveness of the improved DeepLabv3+, particularly the copy-paste data augmentation method. By integrating the above improvements, the performance of the proposed damage segmentation model is notably impressive.

The simply-supported beam bridge images after the ROI extraction are recognized with the improved DeepLabv3 + model, and the results of damage segmentation is shown in Fig. 16. The left two columns display the damage segmentation results for images without the ROI extraction, while the right two columns show the results after the ROI

Table 7
The ablation study of the improved DeepLabv3 + model.

Model	Method	PA (%)	mIoU (%)	FWIoU (%)
A	DeepLabv3+	83.24	29.04	76.48
B	A + focal loss	86.89	31.85	80.58
C	B + copy-paste	92.40	82.21	87.72

Table 5
The damage segmentation performance of the improved DeepLabv3 + model.

Model	IoUs for each class (%)						PA (%)	mIoU (%)	FWIoU (%)
	Background	Wetspot	Cavity	Crack	Rockpocket	Spalling			
Proposed	90.02	86.39	83.95	76.93	77.08	78.87	92.40	82.21	87.72

Table 6
Pixel ratios for each class in the dataset after copy-paste augmentation.

Class	Background	Wetspot	Cavity	Crack	Rockpocket	Spalling
Origin	88.71 %	3.98 %	1.17 %	0.60 %	0.49 %	5.04 %
After	64.79 %	17.22 %	7.78 %	1.97 %	1.25 %	6.99 %

extraction. The images without ROI extraction are interfered by the background pixels and the model misidentifies a substantial quantity of objects in the background as damage. In contrast, the images after ROI extraction eliminate the background pixels, thereby improving the accuracy of damage segmentation. Therefore, ROI extraction effectively improves segmentation accuracy by eliminating complex background interference, allowing the model to focus on the bridge itself and enhancing the robustness of the model to image changes. Fig. 16 illustrates that the original and ROI-processed images differ in detecting bridge pixels, possibly due to the influence of background information. The DeepLabv3 + segmentation model uses the ASPP structure to achieve a deep fusion of global and local features. The original image retains background pixels, which can provide additional information for global features. In contrast, the ROI image removes the background, increasing the reliance of the model on local features and resulting in differences in detection results. The proposed method exhibits some errors in damage segmentation results when compared to the ground truths. For example, in the third row of Fig. 16, the proposed model fails to detect the cavity at the bottom. Additionally, the boundaries between the detected masks and the ground truths also differ significantly, preventing an accurate assessment of the disease size. These issues arise due to the substantial differences between the tested images and the training dataset, leading to the weak generalization ability of the model, which is an urgent need to be addressed in future research.

4.3. Bridge damage location

After obtaining the bridge damage segmentation masks and the reconstructed point clouds from the previous section, the corresponding camera poses are indexed to the damage images. The Damage-Painting method is then used to project the damage into the point cloud, enabling the 3D visualization of the bridge damage.

A portion of the pier from the multi-span simply supported girder bridge is selected for damage localization and 3D visualization, as shown in Fig. 17. Due to regular maintenance by the management, there is minimal surface damage and only minor cracking and spalling observed, which do not compromise the structural safety. However, at the cover girders, which are typically located at the expansion joints and are difficult for maintenance personnel to access, there is a significant issue with water erosion, characterized by numerous wet spots that require attention.

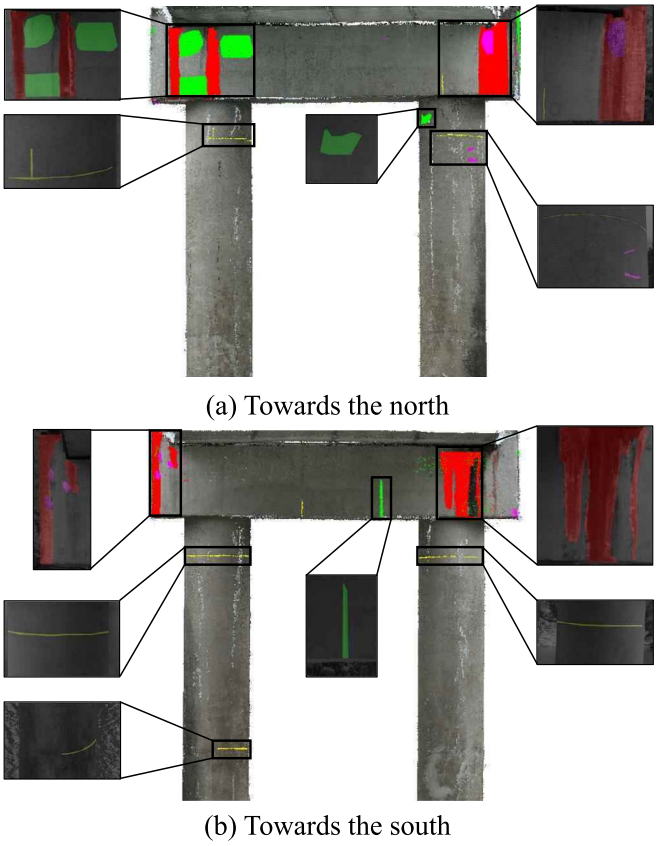


Fig. 17. Three-dimensional localization of multiple types of bridge damage.

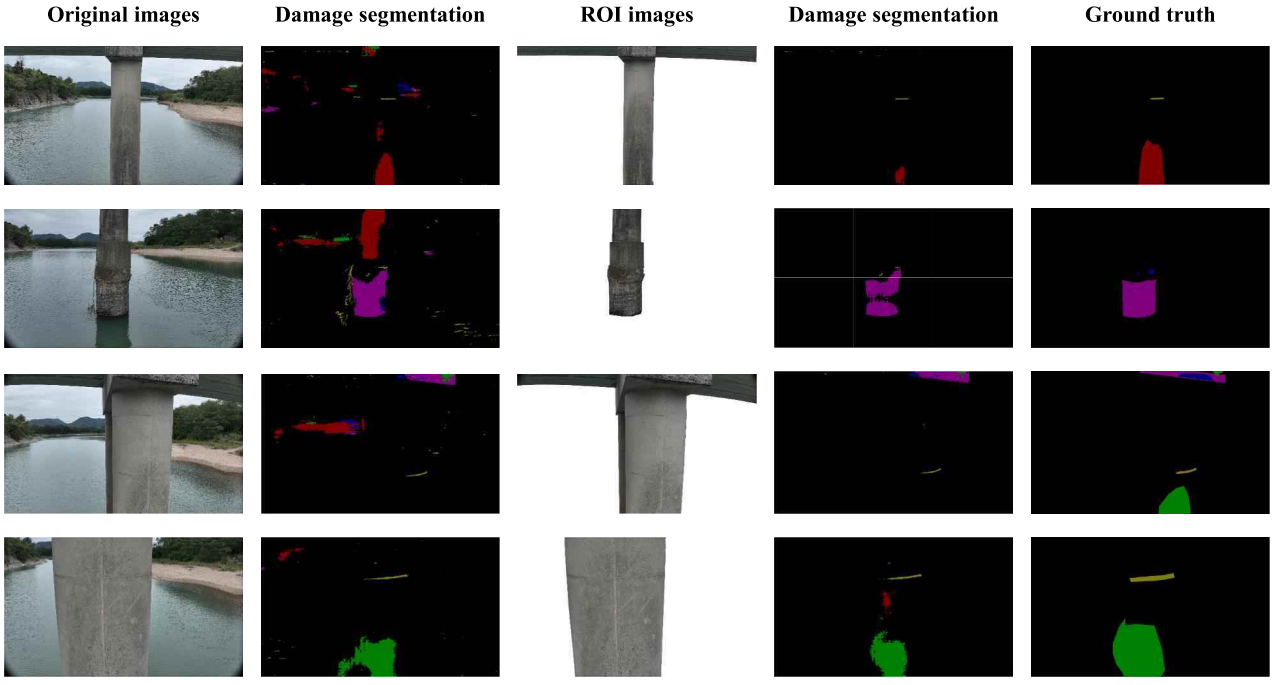


Fig. 16. The effect of ROI extraction on damage segmentation.

5. Discussion

In this study, a novel method for bridge damage segmentation and localization is developed using 2D-3D data fusion, which enables effective identification of bridge damage in the presence of complex backgrounds and imbalanced data. However, several issues warrant further discussion.

First, the proposed improved DeepLabv3 + model addresses the severe inter-class imbalance in the bridge damage dataset by incorporating copy-paste data augmentation and focal loss, which do not alter the model architecture itself. The above improvements are simply two generic modules and can be seamlessly integrated into any semantic segmentation architecture (e.g., U-Net, FCN) with negligible overhead, demonstrating the broad applicability of the proposed improvements.

Furthermore, the effect of ROI extraction on the damage segmentation requires further investigation. For several damage images, the damage segmentation is quantified before and after the ROI extraction, and the results are presented in Table 8. The filtering rate refers to the percentage of incorrectly identified damage pixels removed following ROI extraction. Overall, the ROI extraction process filters out 48.55% of the misidentified pixels, significantly enhancing the accuracy of damage segmentation.

Finally, as shown in Fig. 17, certain areas on the cover beam are labeled as damage but are not detected in the images. In conjunction with Fig. 17(a), it is evident that these "painted" points are caused by leakage from the damaged points on the backside. Due to limitations in the 3D reconstruction method and the quality of the acquired images, the reconstructed point clouds are sparse in some areas, allowing the color leakage from one side to the other, which impairs the 3D visualization of the bridge damage. Future research should focus on developing more efficient 3D reconstruction methods to enhance the quality and density of reconstructed point clouds. Additionally, developing effective surface reconstruction techniques based on 3D point clouds for bridge mesh modeling could effectively address the issue.

Several limitations will be the focus of the future research. First, the methods for image acquisition require further exploration. When using UAVs to collect bridge images in the field, the automatic obstacle avoidance function prevented the capture of detailed images. As the width of concrete cracks is usually in the millimeter scale, the image resolution must be sufficient to display the crack pixels, necessitating the ability of UAVs to approach a suitable working distance. Additionally, the existing drones rely on Global Positioning System (GPS) signals for navigation, but many bridges are located in areas where GPS signals are weak or unavailable. These practical challenges necessitate the development of image acquisition devices capable of capturing detailed images safely and effectively in GPS-free environments.

Second, the proposed damage localization method requires further development. The Damage-Pointing method introduced in this study achieves three-dimensional visualization of bridge anomalies, offering an intuitive display for maintenance managers. However, this study lacks the capability to output the specific three-dimensional coordinates of the damage locations. Future research should focus on constructing a world coordinate system to enable real-time output of these coordinates, thereby providing digital technical support for managers.

Table 8
Effect of ROI extraction on damage segmentation.

Image number	Number of damage pixels in ROI	Number of damage pixels in background	Filtration rate (%)
1	384,092	348,309	47.56 %
2	569,018	142,775	20.06 %
3	323,474	254,483	44.03 %
4	713,939	362,808	33.69 %
5	10,119	383,251	97.43 %
total	2,000,642	1,491,626	48.55 %

6. Conclusions

This paper proposes a method for bridge damage segmentation and localization using 2D-3D data fusion, which effectively identifies damage under complex backgrounds with imbalanced data. The homogeneity between images and point clouds is exploited to achieve pixel-level interaction through 2D-3D geometric mapping, where the 3D point cloud data assisted in the ROI extraction of 2D images, and the 2D damage segmentation information is integrated into the 3D model for visualization. To verify the effectiveness of the proposed method, a comprehensive field experiment is conducted on a multi-span simply supported girder bridge. The results demonstrate that the proposed method offers a promising solution for the segmentation and localization of bridge damage, overcoming the low accuracy of traditional methods in segmenting complex background images and addressing inter-class imbalance in bridge damage datasets. The main conclusions are as follows:

- The proposed ROI extraction model integrates 2D RGB images and 3D depth information to efficiently extract bridge pixels from complex backgrounds, thereby avoiding the high computational burden associated with extensive point cloud processing. Compared to the model without depth images, the proposed method achieves more accurate ROI extraction while maintaining real-time performance.
- The improved DeepLabv3 + model, which incorporates copy-paste data augmentation and focal loss, achieves a segmentation accuracy of 82.21%, surpassing the initial model. The copy-paste data augmentation method addresses the issue of imbalance between background and foreground pixels in the damage dataset, thereby significantly improving segmentation accuracy.
- ROI extraction eliminates the interference of complex backgrounds, allowing the segmentation model to focus more on bridge pixels and thereby significantly improving the accuracy of damage segmentation. The results indicate that ROI extraction effectively filters out most misrecognized background pixels, enhancing the robustness of the model against complex backgrounds.
- The proposed Damage-Painting method enables the projection of the 2D damage segmentation masks onto the 3D point clouds, allowing extensive and complex damage information to be visualized in three dimensions. In the multi-span simply supported girder bridge analyzed in this study, only the cover girder exhibits significant water erosion, cavities, and spalling, providing valuable guidance for maintenance management.

CRediT authorship contribution statement

Wen-Jie Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Hua-Ping Wan:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis. **Michael D. Todd:** Writing – review & editing, Visualization, Supervision, Project administration, Formal analysis.

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 supported by National Natural Science Foundation of China (52422804) and Zhejiang Provincial Natural Science Foundation (LR23E080003).

Data availability

Data will be made available on request.

References

[1] ASCE. America's Infrastructure Report Card 2021. Available at https://infrastructurereportcard.org/wpcontent/uploads/2020/12/National_IRC_2021-report.pdf. 2021.
[2] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, S. Nagarajaiah, Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection, J. Struct. Eng. 146 (5) (2020) 04020073.
[3] C.Q. Feng, B.L. Li, Y.F. Liu, F. Zhang, Y. Yue, J.S. Fan, Crack assessment using multi-sensor fusion simultaneous localization and mapping (SLAM) and image super-resolution for bridge inspection, Autom. Constr. 155 (2023) 105047.
[4] S. Jiang, J. Zhang, C. Gao, Bridge deformation measurement using unmanned aerial dual camera and learning-based tracking method, Struct. Control Health Monit. 2023 (1) (2023) 4752072.
[5] S. Jiang, Y. Wu, J. Zhang, Bridge coating inspection based on two-stage automatic method and collision-tolerant unmanned aerial system, Autom. Constr. 146 (2023) 104685.
[6] H.P. Wan, W.J. Zhang, H.B. Ge, Y. Luo, M.D. Todd, Improved vision-based method for detection of unauthorized intrusion by construction sites workers, J. Constr. Eng. Manag. 149 (7) (2023) 04023040.
[7] H. Chu, W. Chen, L. Deng, Cascade operation-enhanced high-resolution representation learning for meticulous segmentation of bridge cracks, Adv. Eng. Inf. 61 (2024) 102508.
[8] J.S. Chou, C.Y. Liu, Optimized lightweight edge computing platform for UAV-assisted detection of concrete deterioration beneath bridge decks, J. Comput. Civ. Eng. 39 (1) (2025) 04024045.
[9] C. Zhang, Z. Lu, X. Li, Y. Zhang, X. Guo, A two-stage correction method for UAV movement-induced errors in non-target computer vision-based displacement measurement, Mech. Syst. Sig. Process. 224 (2025) 112131.
[10] H. Su, X. Xu, S. Zuo, S. Zhang, X. Yan, Research progress in monitoring hydraulic concrete damage based on acoustic emission, J. Intelligent Constr. 9180024 (2023).
[11] V. Pakrashi, F. Schoefs, J.B. Memet, A. Connor, ROC dependent event isolation method for image processing based assessment of corroded harbour structures, Struct. & Infrastructure Eng. 6 (3) (2010) 365–378.
[12] U.A. Nnolim, Automated crack segmentation via saturation channel thresholding, area classification and fusion of modified level set segmentation with Canny edge detection, Heliyon 6 (12) (2020).
[13] J. Chen, D. Liu, Bottom-up image detection of water channel slope damages based on superpixel segmentation and support vector machine, Adv. Eng. Inf. 47 (2021) 101205.
[14] Q. Liu, D. He, Z. Jin, J. Miao, S. Shan, Y. Chen, M. Zhang, ViTR-Net: An unsupervised lightweight transformer network for cable surface defect detection and adaptive classification, Eng. Struct. 313 (2024) 118240.
[15] L. Zhou, Y. Jiang, H. Jia, UAV vision-based crack quantification and visualization of bridges: system design and engineering application, Struct. Health Monit. 14759217241251778 (2024).
[16] H. Zoubir, M. Rguig, M. El Aroussi, R. Saadane, A. Chehri, Pixel-level concrete bridge crack detection using Convolutional Neural Networks, Gabor filters, and attention mechanisms, Eng. Struct. 314 (2024) 118343.
[17] L. Huang, G. Fan, J. Li, H. Hao, Deep learning for automated multiclass surface damage detection in bridge inspections, Autom. Constr. 166 (2024) 105601.
[18] C.V. Dung, Autonomous concrete crack detection using deep fully convolutional neural network, Autom. Constr. 99 (2019) 52–58.
[19] C. Xiang, W. Wang, L. Deng, P. Shi, X. Kong, Crack detection algorithm for concrete structures based on super-resolution reconstruction and segmentation network, Autom. Constr. 140 (2022) 104346.
[20] J. Chen, I. Chan, I. Brilakis, Shifting research from defect detection to defect modeling in computer vision-based structural health monitoring, Autom. Constr. 164 (2024) 105481.
[21] X. Pan, T.Y. Yang, Postdisaster image-based damage detection and repair cost estimation of reinforced concrete buildings using dual convolutional neural networks, Comput. Aided Civ. Inf. Eng. 35 (5) (2020) 495–510.
[22] Y. Xu, J. Zhao, F. Hu, W. Zhai, Y. Xu, Y. Bao, H. Li, A modified U-net for crack segmentation by self-attention-self-adaption neuron and random elastic deformation, Smart Struct. Syst., Int. J. 29 (1) (2022) 1–16.
[23] Y. Yu, Y. Zhang, J. Yu, J. Yue, Lightweight decoder U-net crack segmentation network based on depth-wise separable convolution, Multimedia Syst. 30 (5) (2024) 1–15.
[24] W. Chen, J. Zhang, Efficient and lightweight monitoring network for cracks in complex background regions based on adaptive perception, Autom. Constr. 166 (2024) 105614.
[25] Y. Xu, Y. Fan, H. Li, Lightweight semantic segmentation of complex structural damage recognition for actual bridges, Struct. Health Monit. 22 (5) (2023) 3250–3269.
[26] S. Bang, S. Park, H. Kim, H. Kim, Encoder-decoder network for pixel-level road crack detection in black-box images, Comput. Aided Civ. Inf. Eng. 34 (8) (2019) 713–727.
[27] 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.
[28] H. Kim, Y. Narazaki, B.F. Spencer Jr, Automated bridge component recognition using close-range images from unmanned aerial vehicles, Eng. Struct. 274 (2023) 115184.
[29] X. Liang, Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization, Comput. Aided Civ. Inf. Eng. 34 (5) (2019) 415–430.
[30] J. Chen, D. Liu, S. Li, D. Hu, Registering georeferenced photos to a building information model to extract structures of interest, Adv. Eng. Inf. 42 (2019) 100937.
[31] J. Chen, X. Yu, Q. Li, W. Wang, B.G. He, LAG-YOLO: efficient road damage detector via lightweight attention ghost module, J. Intelligent Constr. 2 (1) (2024) 1–10.
[32] Z. Liu, P. Wei, Z. Wei, B. Yu, Y. Tian, J. Jiang, W. Cao, J. Bian, Y. Chang, Handling inter-class and intra-class imbalance in class-imbalanced learning, arXiv preprint arXiv:2111.12791, 2021.
[33] Y. Zhang, B. Chen, J. Wang, J. Li, X. Sun, APLCNet: Automatic pixel-level crack detection network based on instance segmentation, IEEE Access 8 (2020) 199159–199170.
[34] R. Rakshitha, S. Srinath, N.V. Kumar, S. Rashmi, B.V. Poornima, Integrated pixel-level crack detection and quantification using an ensemble of advanced U-net architectures, Results Eng. 103726 (2024).
[35] D.H. Kang, Y.J. Cha, Efficient attention-based deep encoder and decoder for automatic crack segmentation, Struct. Health Monit. 21 (5) (2022) 2190–2205.
[36] C. Zhang, C. Chang, M. Jamshidi, Concrete bridge surface damage detection using a single-stage detector, Comput. Aided Civ. Inf. Eng. 35 (4) (2020) 389–409.
[37] H. Wang, J. Xie, J. Fu, C. Zhang, D. Chen, Z. Zhu, X. Zhang, Rapid acquisition and surface defects recognition based on panoramic image of small-section hydraulic tunnel, Underground Space 21 (2025) 270–290.
[38] J. Dong, N. Wang, H. Fang, Q. Hu, C. Zhang, B. Ma, D. Ma, H. Hu, Innovative method for pavement multiple damages segmentation and measurement by the Road-Seg-CapsNet of feature fusion, Constr. Build. Mater. 324 (2022) 126719.
[39] T. Siriborvornratanakul, Pixel-level thin crack detection on road surface using convolutional neural network for severely imbalanced data, Comput. Aided Civ. Inf. Eng. 38 (16) (2023) 2300–2316.
[40] G. Wang, Z. Yang, H. Sun, Q. Zhou, Z. Yang, AC-SNGAN: Multi-class data augmentation for damage detection of conveyor belt surface using improved ACGAN, Measurement 224 (2024) 113814.
[41] D. Mazzini, P. Napoletano, F. Piccoli, R. Schettini, A novel approach to data augmentation for pavement distress segmentation, Comput. Ind. 121 (2020) 103225.
[42] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, H. Omata, Generative adversarial network for road damage detection, Comput. Aided Civ. Inf. Eng. 36 (1) (2021) 47–60.
[43] Y. Que, Y. Dai, X. Ji, A. Kwan Leung, C. Zheng, Y. Tang, Z. Jiang, Automatic classification of asphalt pavement cracks using a novel integrated generative adversarial network and improved VGG model, Eng. Struct. 277 (2023) 115406.
[44] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei, F. Zhou, A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data, IEEE Trans. Ind. Inf. (2024).
[45] W. Shen, D. Zeng, Y. Zhang, X. Tian, Z. Li, Image augmentation for nondestructive testing in engineering structures based on denoising diffusion probabilistic model, J. Build. Eng. 89 (2024) 109299.
[46] K. Li, Y. Li, S. You, N. Barnes, Photo-realistic simulation of road scene for data-driven methods in bad weather, Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, 491-500.
[47] Q. Lu, Y. Jing, X. Zhao, Bolt loosening detection using key-point detection enhanced by synthetic datasets, Appl. Sci. 13 (3) (2023) 2020.
[48] A. Khaloo, D. Lattanzi, K. Cunningham, R. Dell'Andrea, M. Riley, Unmanned aerial vehicle inspection of the Placer River Trail Bridge through image-based 3D modelling, Struct. Infrastruct. Eng. 14 (1) (2018) 124–136.
[49] M. Rahman, H. Liu, M. Masri, I. Durazo-Cardenas, A. Starr, A railway track reconstruction method using robotic vision on a mobile manipulator: A proposed strategy, Comput. Ind. 148 (2023) 103900.
[50] R. Lu, C. Rausch, M. Bolpagni, I. Brilakis, T.C. Haas, Geometric accuracy of digital twins for structural health monitoring. Structural integrity and failure, IntechOpen, London, UK, 2020.
[51] B.F. Spencer Jr, V. Hoskere, Y. Narazaki, Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering 5 (2) (2019) 199–222.
[52] Y.F. Liu, S. Cho, B.F. Spencer Jr, J.S. Fan, Concrete crack assessment using digital image processing and 3D scene reconstruction, J. Comput. Civ. Eng. 30 (1) (2016) 04014124.
[53] Y. Ni, J. Mao, H. Wang, Z. Xi, Z. Chen, Surface damage detection and localization for bridge visual inspection based on deep learning and 3D reconstruction, Struct. Control Health Monit. 2024 (1) (2024) 9988793.
[54] H.P. Wan, W.J. Zhang, Y. Chen, Y. Luo, M.D. Todd, An efficient three-dimensional point cloud segmentation method for the dimensional quality assessment of precast concrete components utilizing multi-view information fusion, J. Comput. Civ. Eng. 39 (3) (2025) 04025028.
[55] K. Hattori, K. Oki, A. Sugita, T. Sugiyama, P.J. Chun, Deep learning-based corrosion inspection of long-span bridges with BIM integration, Heliyon 10 (15) (2024).
[56] 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, Comput. Aided Civ. Inf. Eng. 38 (17) (2023) 2391–2407.

[57] J.L. Schonberger, J.M. Frahm, Structure-from-motion revisited, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 4104-4113.
[58] Y.H. Wu, Y. Liu, J. Xu, J.W. Bian, Y.C. Gu, M.M. Cheng, MobileSal: Extremely efficient RGB-D salient object detection, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2021) 10261–10269.
[59] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Proceedings of the European conference on computer vision (ECCV), 2018, 801-818.
[60] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. PP(99) (2017) 2999–3007.
[61] S. Katz, A. Tal, R. Basri, Direct visibility of point sets, ACM Trans. Graph. (2007).
[62] J. Flotzinger, P.J. Rösch, T. Braml, dacl10k: benchmark for semantic bridge damage segmentation, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, 8626-8635.
[63] Phiyodr. dacl10k-toolkit. GitHub repository, 2024. Available at: https://github.com/phiyodr/dacl10k-toolkit.
[64] H. Chen, Y. Li, Three-stream attention-aware network for RGB-D salient object detection, IEEE Trans. Image Process. 28 (6) (2019) 2825–2835.