Structures 78 (2025) 109075

Contents lists available at ScienceDirect

Structures

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

Crack detection and 3D visualization of crack distribution for UAV-based bridge inspection using efficient approaches

Yahui Qi [a], Pengzhen Lin [a,*], Guojun Yang [b], Tao Liang [a]

[a] School of Civil Engineering, Lanzhou Jiaotong University, Lanzhou, 730050, China
[b] School of Civil Engineering, Lanzhou University of Technology, Lanzhou, 730050, China

ARTICLE INFO

Keywords:
Concrete bridges
Crack detection
Unmanned aerial vehicle
ROI extraction
Deep learning
3D reconstruction

ABSTRACT

In order to improve the detection accuracy and efficiency of bridge crack detection models, while addressing the challenge of crack localization, this paper proposes an efficient Unmanned Aerial Vehicle (UAV)-based concrete bridge crack detection framework. The framework includes depth-based Regions of Interest (ROI) extraction, an improved YOLOv11 crack detection model, the SeaFormer lightweight crack segmentation model, an image quality assessment model, a pseudo-crack removal algorithm, the conversion of pixel values to actual values, and an efficient crack detection scheme. Comparative testing with mainstream models demonstrates the advantages of the proposed models in detection accuracy, localization accuracy, and lightweight design. Additionally, a multi-view 3D reconstruction scheme is proposed, offering lower memory and time requirements while improving performance. Combined with the aforementioned crack detection models, it achieves 3D reconstruction of bridge structures and visualization of the 3D distribution of cracks. In tests involving images of cracks from the piers of Zhongshan Bridge in Lanzhou, the crack identification accuracy reaches 93.2%, with an F1 score of 87.7% and a recall rate of 82.7%. The crack segmentation accuracy is 93.66%, and the Intersection over Union (IoU) is 90.17%. The results show that the proposed bridge crack detection framework delivers high lightweight performance and detection efficiency while maintaining high accuracy, making it more suitable for deployment on mobile devices such as UAVs for crack detection in bridges, towers, and other structures.

1. Introduction

As the service time of bridges increases, internal damage and degradation lead to a continuous reduction in the structural load-bearing capacity. Cracks are a direct reflection of concrete structure damage and are one of the key aspects in bridge inspection. Additionally, cracks accelerate the corrosion of reinforcing bars, reduce the service time of the structure, and directly affect the strength and stability of concrete structures [1–3]. Therefore, accurately identifying bridge cracks and performing maintenance is of great significance.

With the continuous development of computer technology, methods for detecting concrete cracks using digital image technology are gradually being applied [4–9]. Initially, threshold segmentation [10,11], edge detection algorithms [12–14], and percolation methods [15,16] were used to detect cracks. However, these methods have low automation levels and poor detection results, making it difficult to effectively identify and segment cracks on complex concrete surfaces. In recent years, deep learning has rapidly developed in the domain of object detection and has been applied to concrete crack detection. Since it only relied on learning target features from training data to efficiently identify and segment the target, it was highly suitable for detecting concrete cracks with complex backgrounds. Li et al. [17] proposed a bridge crack detection method based on Faster R-CNN and UAV, determining the optimal distance range for stable UAV imaging. The use of pyramid networks with multi-scale feature fusion can further improve the accuracy of crack recognition and segmentation [18]. Fully Convolutional Networks (FCNs), due to their exceptional segmentation capabilities, are widely applied for crack segmentation in concrete images [19]. In addition, Zoubir et al. [20] proposed an end-to-end crack segmentation framework that combined UNet, Gabor filters, and convolutional block attention modules, using the complementary advantages of the spatial and frequency domains to enhance crack feature extraction and reduce background interference. Ding et al. [21] established a full-field scale for UAV gimbal cameras and combined it with transformers to achieve accurate detection and quantification of concrete cracks without reference markers. Deep learning-based crack detection algorithms mainly include crack recognition algorithms, crack classification algorithms, and crack segmentation algorithms. Among them,

* Corresponding author.
E-mail addresses: 13240041@stu.lzjtu.edu.cn (Y. Qi), pzhlin@mail.lzjtu.cn (P. Lin), yanggj403@163.com (G. Yang), 13230041@stu.lzjtu.edu.cn (T. Liang).

https://doi.org/10.1016/j.istruc.2025.109075
Received 22 February 2025; Received in revised form 3 April 2025; Accepted 28 April 2025
Available online 20 May 2025
2352-0124/© 2025 Institution of Structural Engineers. Published by Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

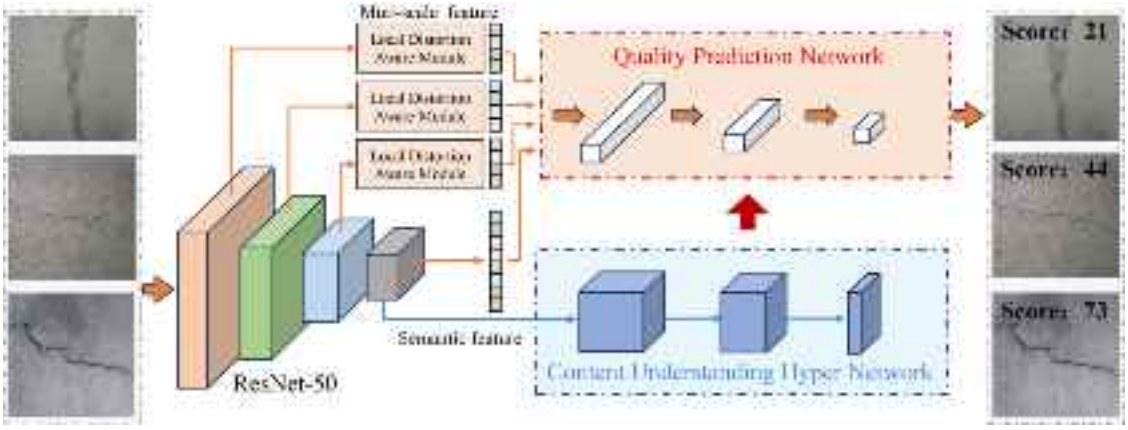

Fig. 1. Image quality assessment network.

detection algorithms can only locate cracks by generating rectangular bounding boxes. Classification algorithms can categorize images with cracks and images without cracks. Crack segmentation algorithms can extract cracks from images. It is important to note that recognition and classification algorithms usually need to be combined with segmentation algorithms to achieve better results. For example, INAM et al. [22] proposed an integrated crack detection and segmentation algorithm that combined YOLOv5 and U-Net, which improved the efficiency of bridge crack detection. Currently, deep learning-based concrete bridge crack detection predominantly relies on UAVs for image acquisition. To enable the integration of the model into UAVs for real-time detection, high-precision and lightweight crack detection algorithms have become a technological trend and development need [23,24]. Therefore, it is essential to strengthen research on crack detection models focusing on high accuracy, lightweight modeling, and detection efficiency [25].

The above methods detect cracks only on two-dimensional images, making it difficult to precisely locate cracks in real engineering applications, and the cracks information visualization is not intuitive. With the development of 3D reconstruction technology, it has been increasingly applied in the engineering field [26–31]. Li et al. [32] based on secondary development of BIM technology, established a bridge information model to visualize the bridge structure and underwater apparent damage information, allowing maintenance and management personnel to view the health and damage status of the bridge and take timely repair actions. However, this method relied on bridge drawings and could not accurately reflect the real conditions of the bridge. Therefore, a 3D reconstruction method based on UAVs and Structure from Motion (SfM) algorithms was proposed, demonstrating its advantages in terms of equipment cost, measurement time, and reconstruction performance [33]. Liu et al. [34] proposed a method to project cracks onto a meshed 3D surface, correcting perspective and geometric distortions on non-planar surfaces. Although a number of studies have been conducted on 3D reconstruction of bridge structures and the visualization of 3D crack distribution, previous 3D reconstruction methods required long processing times, consumed large amounts of memory, and produced poor reconstruction results.

To address the above issues, this paper proposes a lightweight and efficient framework for concrete bridge crack detection and a solution for 3D visualization of crack distribution on bridges, with the main contributions including: (1) A ROI extraction method based on depth information is introduced, which eliminates interference from complex outdoor backgrounds. (2) The YOLOv11 model is introduced and improved based on crack features, enhancing its ability to recognize and locate cracks. At the same time, a lightweight bridge crack segmentation model based on SeaFormer is developed, which outperforms mainstream lightweight CNN models in terms of segmentation speed and model size. By combining the constructed model with image preprocessing algorithms, crack detection is performed on Zhongshan Bridge in Lanzhou. The results show that the proposed models and framework are feasible and have promising applications in practical engineering. (3) To further improve detection efficiency, an efficient detection scheme is proposed in which cracks are first identified, and then the crack bounding boxes are extracted for segmentation mapping. This significantly reduces the non-crack pixels involved in the segmentation, resulting in at least a 70% reduction in segmentation time and computational load compared to directly segmenting the ROI area. (4) A 3D reconstruction scheme is constructed based on the Colmap–Patchmatchnet algorithm [35], which achieves excellent reconstruction results while being at least 2.5 times faster and consuming half the memory compared to the best existing methods [36]. By combining 3D reconstruction and image registration methods, the visualization of the 3D distribution of bridge cracks is achieved.

2. Crack detection framework

2.1. Image preprocessing

Image quality assessment: During the image acquisition process with the UAV, issues such as image shift, focus loss, and exposure anomalies often arise due to the UAV's vibrations, natural wind, and the camera's intrinsic limitations, which significantly impact subsequent crack segmentation and measurement. Therefore, it is essential to conduct an assessment of the acquired images. Traditional image quality assessment methods are mainly based on metrics like mean squared error and image sharpness, which are limited in scope and struggle to assess the quality of concrete crack images with complex backgrounds.

Therefore, this article employed an adaptive supernetwork architecture to assess image quality [37] (The structure is shown in Fig. 1), which mainly consisted of three components: image feature extraction, formulation of image quality perception rules, and image quality prediction. First, the foundational model ResNet50 extracts semantic features from the images, after which the supernetwork formulates quality perception rules based on the extracted image features. The quality prediction target network accepts multi-scale feature inputs from the images, generates weights for the network based on the specific performance of the images, and perceives both global and local complex distortions. Since the model can adaptively generate prediction weights based on the image input, it performs exceptionally well when assessing images in outdoor environments. After extensive testing on real bridge crack images, distorted images consistently score below 50, while images with clear cracks score above 50, meeting the requirements for pixel-level segmentation. Therefore, the threshold is set at 50, with images scoring above this threshold classified as high-quality. The specific image acquisition process is illustrated in Fig. 2.

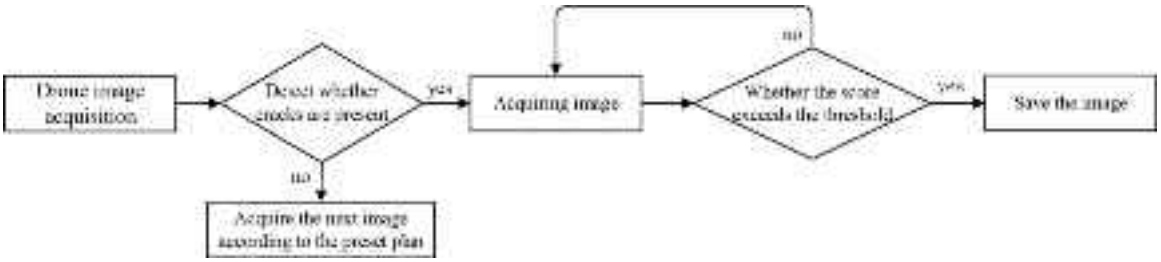

Fig. 2. Image acquisition flowchart.

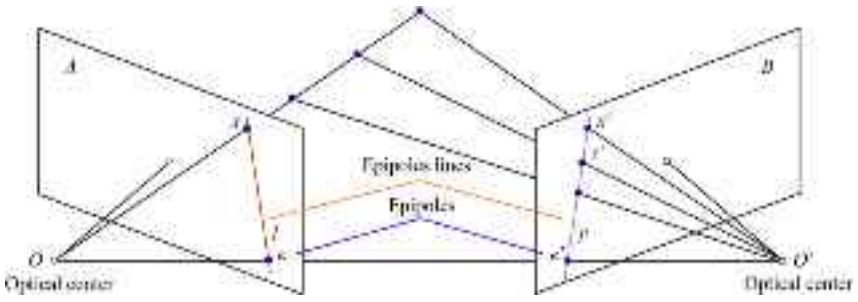

Fig. 3. Depth information calculation.

ROI extraction based on depth information: Previous studies often detected cracks directly from raw images, resulting in extensive background involvement in the computation. This approach not only increases computational complexity but also leads to the misclassification of linear objects, such as power lines and tree branches, as cracks. In recent years, 3D point cloud segmentation has been employed to extract ROI [38]. However, this method involves complex processes such as 3D reconstruction, point cloud segmentation, and data fusion. Therefore, this paper proposes an ROI extraction method based on depth information, where a single UAV equipped with a gimbal camera is used to capture images, from which the relative 3D depth information of each pixel is obtained. This enables the extraction of the foreground ROI, significantly improving the overall crack detection speed and accuracy. As illustrated in Fig. 3, for two images A and B captured from distinct viewpoints, a point $\mathbf{x}$ in image A is projected onto a corresponding epipolar line $\mathbf{l}'$ in image B. This epipolar line $\mathbf{l}'$ represents all potential correspondences in image B for the point $\mathbf{x}$ in image A. The relationship between the point and its corresponding epipolar line is governed by the fundamental matrix F, expressed as:

$$\mathbf{l}' = F\mathbf{x} \tag{1}$$

In image B, the point $\mathbf{x}'$ lies on the corresponding epipolar line $\mathbf{l}'$, satisfying the constraint $\mathbf{x}'^T\mathbf{l}' = 0$, which indicates that $\mathbf{x}'$ must be on $\mathbf{l}'$. By substituting the expression for $\mathbf{l}'$ from the previous equation:

$$\mathbf{x}'^T(F\mathbf{x}) = 0 \tag{2}$$

$\mathbf{x} = \begin{bmatrix} x & y & 1 \end{bmatrix}^T$ represents the homogeneous coordinates of a point in the first image.

$\mathbf{x}' = \begin{bmatrix} x' & y' & 1 \end{bmatrix}^T$ represents the homogeneous coordinates of the corresponding point in the second image.

This formula maps the point $\mathbf{x}$ in image A to a line (epipolar line) in image B, meaning that the corresponding point $\mathbf{x}'$ in image B must lie on this epipolar line.

Features are extracted and matched from multi-view images using the SIFT algorithm. To eliminate erroneous matches, the Random Sample Consensus (RANSAC) algorithm is employed. The fundamental matrix F is computed using the refined set of matching points. Once F is determined, the corresponding epipolar line in image B for a point in image A can be derived. A sliding window matching method is then applied along the epipolar line to identify the best-matching point, enabling precise pixel correspondences across different viewpoints. Finally, the principle of triangulation is utilized to compute the relative 3D depth information for each image.

As shown in Fig. 4, multi-view images of the target object are captured (Fig. 4(a)). Based on the aforementioned principles, the depth information for each image is calculated (Fig. 4(b)). Due to the significant depth difference between the background and foreground (Fig. 4(c)), Kernel Density Estimation (KDE) is applied to fit the kernel density function of the depth data. By computing the second derivative of the density function, the points where it equals zero, i.e., the inflection points, are identified. These points typically represent significant changes in the distribution. The two inflection points with the highest Y-values are selected as thresholds to automatically extract the foreground region, where the density is relatively concentrated. Subsequently, digital image processing techniques are applied to further process the results and obtain the ROI image (Fig. 4(d)). The detailed procedure is described in Section 4.2.

2.2. Crack detection models

In deep learning-based crack detection algorithms, using only a crack recognition algorithm is insufficient for segmenting cracks, while solely relying on a crack segmentation algorithm slows down the detection speed. To address this, the framework in this paper combines both recognition and segmentation algorithms, improving detection efficiency. First, images are captured and the ROI is extracted. The cracks are then identified, and the crack target area is input into the segmentation model for segmentation. After segmentation, pseudo-cracks are removed, and the crack length and width are calculated (Using the central axis transformation method, which is not described in detail). Finally, the crack image is mapped back to the original image based on the location information of the identified bounding boxes. The detailed process is shown in Fig. 5.

YOLOv11 introduces significant improvements in architecture and training methods over previous YOLO versions, integrating an optimized model structure, enhanced feature extraction techniques, and optimized training methods, resulting in superior speed, accuracy, and efficiency. Therefore, YOLOv11 is selected in this study and further improved based on the elongated features of cracks. The YOLO head is augmented with an Explicit Visual Center(EVC) module [39], which

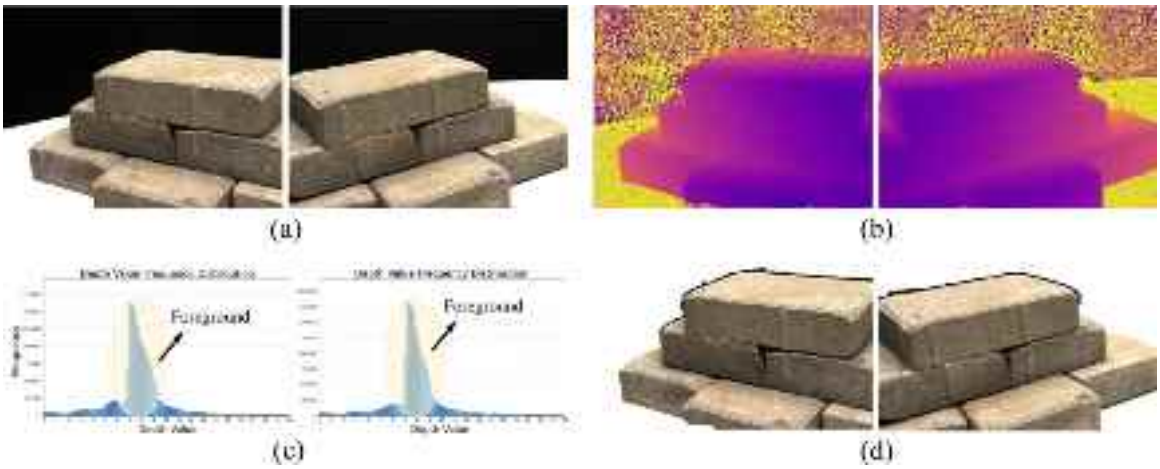

Fig. 4. Depth information calculation. (a) Original image; (b) Depth map; (c) Frequency plot; (d) ROI map.

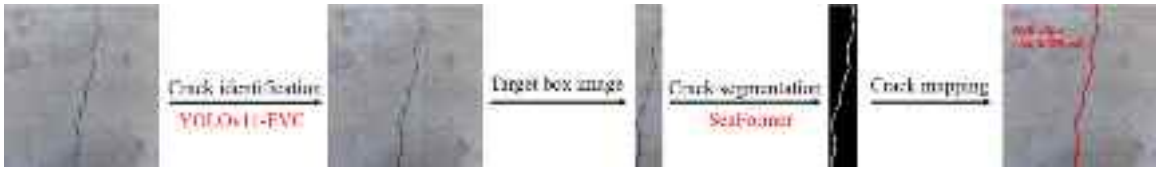

Fig. 5. Crack identification and segmentation flowchart.

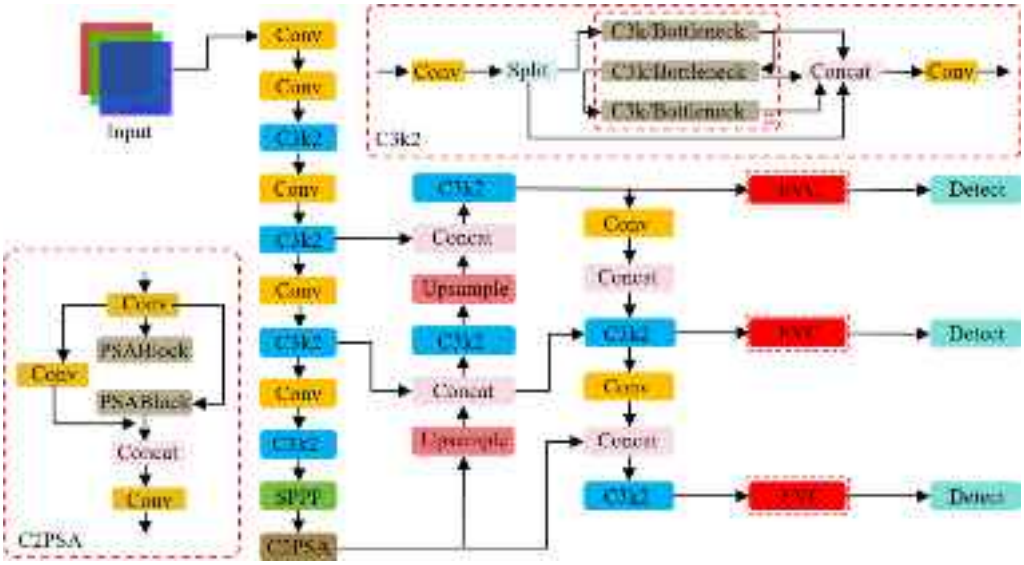

Fig. 6. Diagram of the improved YOLOv11-EVC network.

is capable of extracting long-range semantic information and local features (Fig. 6). The module consists of a lightweight multi-layer perception(MLP) module and a learnable visual center module connected in parallel. The lightweight multilayer perceptron module consists of a depthwise separable convolution module and a channel-based MLP module, followed by channel scaling and DropPath operations, which are used to capture global long-range dependencies and enhance the model's generalization ability and robustness. The learnable visual center can learn and aggregate local features from the input image, thereby improving the model's ability to detect cracks. The specific structure of the EVC module is illustrated in Fig. 7.

The proposed detection scheme generates multi-scale images as inputs for the segmentation model. Transformer-based segmentation models exhibit superior multi-scale image segmentation capability compared to CNNs. However, due to the self-attention mechanism in the Transformer architecture, which incurs high time and space complexity with respect to sequence length, the computational cost and memory requirements are relatively high, making it unsuitable for use in mobile devices such as UAVs. Therefore, a lightweight segmentation model based on the Transformer architecture, SeaFormer, is introduced [40], as shown in Fig. 8. The network first downsamples the image by 1/2, 1/4, and 1/8, then processes it using two branches: the red one is the context branch, and the blue one is the spatial branch. The context branch alternates between MobileNetV2 blocks(MV2s) and SeaFormer layers, using a fusion module to combine the two branches. It extracts weight information using convolution and Sigmoid, multiplies the weight information with the spatial branch, and processes it with a

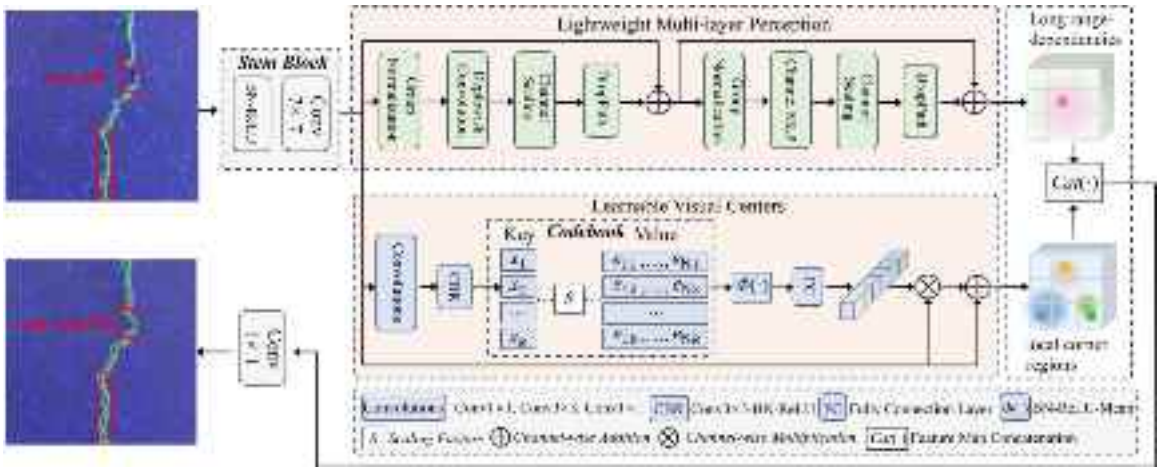

Fig. 7. Structure diagram of the EVC module.

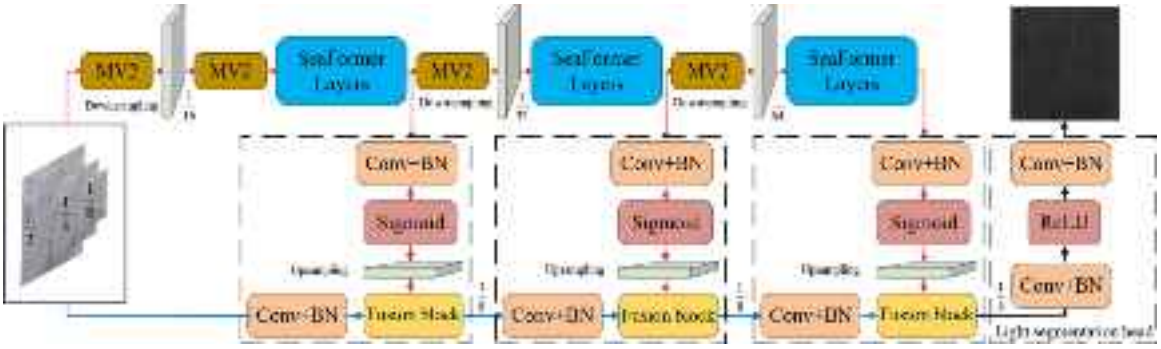

Fig. 8. SeaFormer network structural diagram.

segmentation module after three iterative steps. The SeaFormer layers incorporate a position-aware axial compression attention enhancement module. On the one hand, features are compressed axially and then enhanced using an attention mechanism; on the other hand, features are processed through a convolutional network to enhance local information. The two are then fused to output the enhanced features. Moreover, the model employs a feature-upsampling-based multi-resolution distillation technique, which further reduces the inference latency of the proposed framework, resulting in high detection accuracy and inference speed.

2.3. Image quality assessment and pseudo crack removal

Although deep learning-based crack detection algorithms are not affected by general noise, they struggle to eliminate short cracks that do not impact the structural load-bearing capacity and template seams that resemble cracks in shape. These pseudo cracks are often misidentified as real cracks during detection, and therefore, they need to be filtered out.

Concrete generates many short cracks due to its inherent properties and factors like temperature. These short cracks have negligible impact on the structural load-bearing capacity; therefore, it is necessary to eliminate them during the crack detection process. For short cracks, this paper first traverses the binary image to calculate the length of all connected components, and then removes those whose length is below a certain threshold, as shown in Fig. 9.

In bridge structure crack detection, challenges arise not only from complex backgrounds but also from interference caused by formwork joints and control joints that closely resemble crack patterns. Due to their similarity to cracks, these joints are often mistakenly identified as cracks during algorithmic segmentation. However, these joints do not pose any negative impact on the structure. Therefore, based on the characteristics of formwork joints, which are narrow and straight, we propose the following steps to eliminate them, as shown in Fig. 10.

(1) First, calculate the connected components in the image, representing each with a different color and labeling them.

(2) Fit a straight line to each connected component and calculate the number of intersections between the line and the edges of the component.

(3) If the number of intersections is less than 4, it is considered a pseudo-crack; if the number is greater than or equal to 4, it is considered a crack.

3. Model training and performance comparison

3.1. Model training

Before training the model, hyperparameters were set, and the required dataset was prepared. The dataset was partitioned into training, validation, and test sets in a 7:2:1 ratio. Training was then conducted using a GPU (NVIDIA 3090, 24 GB) and a CPU (Intel Xeon E5-2680 v4). Training was stopped when the model's precision and loss reached an optimal and stable state.

We applied the image quality assessment model to screen publicly available datasets of concrete bridge cracks and constructed a dataset comprising 3272 images of concrete bridge surface cracks for training the YOLOv11-EVC model. This dataset includes 1872 images with a resolution of 1024 × 1024 pixels from the Aft_Original_Crack_DataSet_Second dataset [41], and 1400 images with a resolution of 256 × 256 pixels from the SDNET2018 dataset [42]. Labeling was performed using the LabelImg tool. To accelerate model convergence and enhance training efficiency, a transfer learning approach was employed,

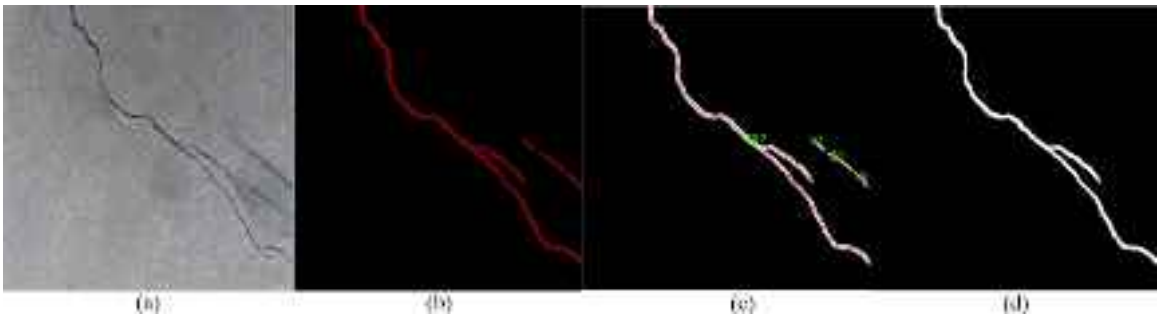

Fig. 9. Short crack removal. (a) Original image; (b) Semantic Segmentation; (c) Connected domain length calculation; (d) Short crack removal.

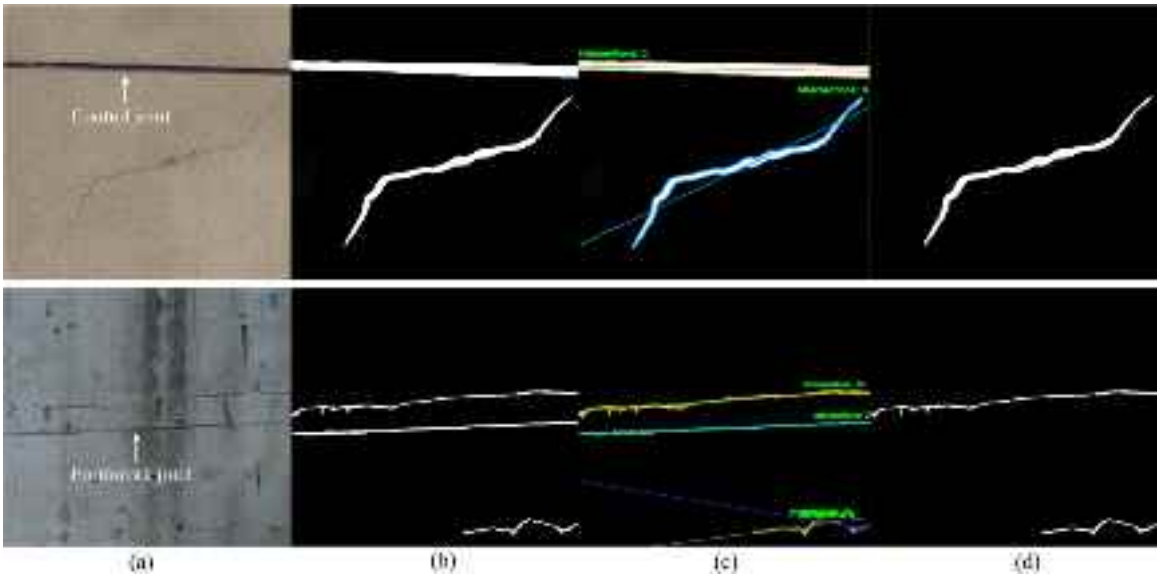

Fig. 10. Pseudo crack removal: (a) Original image; (b) Identify and binarize; (c) Calculate the number of intersections; (d) Remove pseudo crack.

using weights trained on the COCO dataset as the initial weights for YOLOv11-EVC. The model was trained for 300 epochs, with the image size set to 640, a batch size of 8, and the optimizer chosen as SGD. Other hyperparameters were set to their default values.

For training the SeaFormer segmentation model, a crack segmentation dataset was constructed using the previously filtered 1872 crack images from the Aft_Original_Crack_DataSet_Second dataset and 400 images from the SDNET2018 dataset. Labeling was performed using the LabelMe tool, weights pre-trained on the ADE20K dataset were used as the initial weights for the SeaFormer model. The model was trained for 400 epochs using the AdamW optimizer, with a learning rate set to 0.0003, weight decay of 0.01, and a batch size of 16. The best trained weights were also used to construct the segmentation model (see Fig. 11).

3.2. Model performance comparison

The following parameters were calculated using an input image size of 1024 × 1024 and 256 × 256 pixels on a device equipped with an RTX3090 (24 GB) GPU and an Intel Xeon E5-2680 v4 CPU. After training on the dataset constructed in this study, comparisons were made among YOLOv9s, YOLOv10s,YOLOv11s, and the proposed improved model, the results are shown in Table 1. Based on Table 1 and Fig. 12, it can be observed that the model constructed in this study outperforms other models across multiple metrics, demonstrating superior overall performance and being more suitable for crack detection. In the table, P (Precision) represents the proportion of true positive samples among all samples predicted as positive by the model. Higher precision indicates better accuracy in predicting positive samples. R (Recall) refers to the proportion of true positive samples correctly predicted by the model among all actual positive samples. Higher recall signifies fewer missed positive samples by the model. mAP@50 and mAP@50-95 assess the overall performance of the object detection model at different IoU thresholds, with mAP@50-95 being more challenging as it requires higher accuracy in predictions. Params denotes the number of parameters in the model, where a higher number of parameters indicates greater computational resource requirements. FLOPs refers to the number of floating-point operations executed during inference, reflecting the model's computational complexity—higher FLOPs necessitate stronger hardware support.

Table 1
Performance comparison of detection models.

Model	P	R	map50	map50–95	Params	GFLoPs
YOLOv9s	0.876	0.86	0.918	0.76	7 167 475	26.7
YOLOv10s	0.905	0.817	0.906	0.705	8 035 734	24.4
YOLOv11s	0.913	0.839	0.923	0.731	9 413 187	21.3
YOLOv11-EVC	0.93	0.828	0.926	0.762	8 220 555	16.4

Subsequently, mainstream semantic segmentation models were trained using the crack segmentation dataset constructed in this study, and their accuracy was compared with that of the model proposed here. The results demonstrated that the IoU and PA of the model proposed in this study outperformed those of the mainstream semantic segmentation models, as shown in Table 2. IoU (Intersection over Union) measures the ratio of the overlap between two regions (typically the

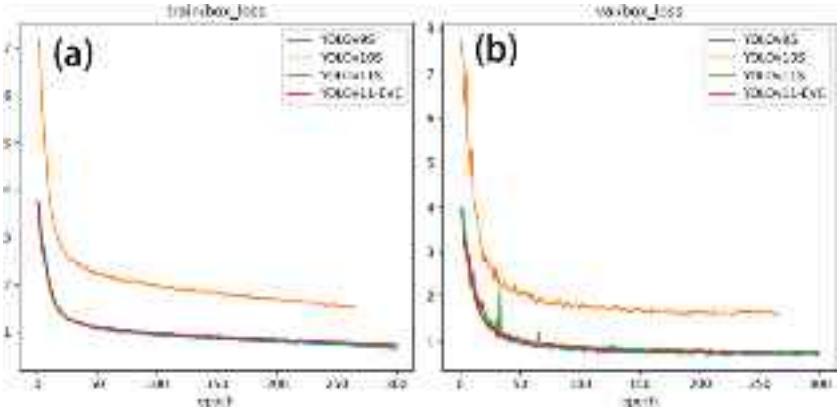

Fig. 11. The loss curves of models. (a) Train loss; (b) Val loss.

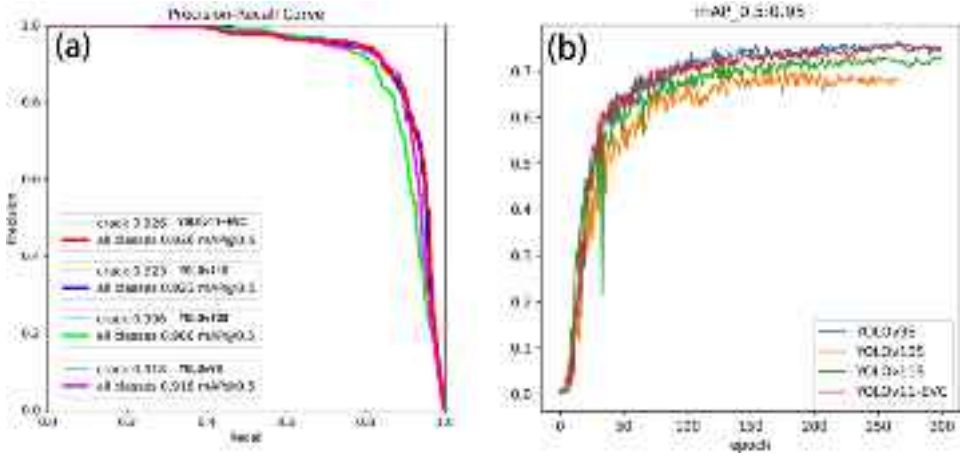

Fig. 12. PR and precision curves. (a) PR curves; (b) Map0.5:0.95 curves.

Table 2
Performance comparison of segmentation models.

Model	Backbone	IOU(%)	PA(%)
Pspnet	Resnet50	77.96	80.94
Deeplabv3+	Xception	83.38	89.40
Improving u-net	Resnet50	86.42	92.01
SeaFormer	SeaFormer	90.21	94.16

Table 3
Comparison of GFLOPs and Params of segmentation models.

Model	GFLOPs	Params(M)
Improving u-net	810.23	29.06
FCN	792.00	49.60
Psanet	778.62	54.07
Deeplabv3+	705.50	43.59
Pspnet	715.04	49.09
Mobilenetv2	158.55	9.82
Mobilenetv3	34.36	3.28
SeaFormer	6.68	8.59

predicted and ground truth bounding boxes) to their union. It quantifies the degree of match between the prediction and the actual ground truth. PA (Pixel Accuracy) is commonly used in image segmentation tasks and indicates the proportion of correctly classified pixels to the total number of pixels. A higher PA value reflects better segmentation accuracy of the model. Table 3 presents a comparison of the FLOPs and Params between the segmentation model proposed in this study and other mainstream segmentation models.

From Tables 2 and 3, it can be seen that the crack segmentation model constructed in this study achieves higher accuracy compared to other mainstream segmentation models, with fewer parameters and lower complexity, resulting in overall superior performance.

4. Crack detection and 3D visualization of crack distribution in an old bridge

The previous section presented the models and processes used for bridge crack detection, along with an analysis of the feasibility and advantages of these methods. To further validate their rationale and effectiveness, the proposed methods were applied to crack detection on the pier of an old bridge.

4.1. Overview of the old bridge

Lanzhou Zhongshan Bridge is a century-old steel bridge, completed in 1907. It is not only a national key cultural heritage site but also still serves as a pedestrian bridge. The bridge spans a total length of 233.5 m, consisting of five spans with a simply supported steel truss structure, as shown in Fig. 13. The single-span length is 46.7 m, and the bridge width is 9.55 m.

In 1954, an arch truss was added to the upper section of the original parallel chord truss system, with the cross-sectional design of the members largely unchanged from the original structure. In 2006, Lanzhou Zhongshan Bridge was designated as a national cultural heritage site. In 2010, the local government reinforced the caisson foundation and piers of the bridge, employing a lifting technique to raise the entire bridge structure by 1.2 m. Simultaneously, maintenance work was

Fig. 13. Zhongshan bridge.

carried out on the deck system and steel beams. The renovation project incorporated various new technologies, techniques, and methods, and was successfully completed in May 2011, effectively preserving the national cultural relic. The bridge not only holds significant historical and cultural value but also stands as a testament to key moments in modern Chinese history, carrying rich regional cultural information and showcasing unique architectural styles and artistic aesthetics. In contemporary society, it has become a cultural landmark and tourist attraction in Lanzhou, playing a crucial role in promoting local economic and social development, preserving historical culture, and enhancing the city's image. To assess the current health of the bridge piers, crack detection was conducted on the four central piers, with significant cracks identified on the two piers near the northern end.

4.2. Crack detection

During the image acquisition process, the DJI Phantom 4 Pro V2.0 UAV was employed to perform a circular flight around the target bridge piers. The UAV was equipped with a gimbal camera with an 8.8 mm focal length and a maximum resolution of 5472 × 3648 pixels. It maintained a constant distance of approximately 3 m from the bridge pier, resulting in the collection of 6 images containing cracks. Based on the principles outlined in Section 2.1, the depth information for each image is calculated, and the frequency of the depth information is statistically analyzed. Kernel Density Estimation (KDE) is then used to estimate the density function(as shown in Fig. 14(b)). The inflection points are identified by finding the zeros of the second derivative of the density function, and the two points with the highest Y-values are selected as thresholds to automatically segment the foreground depth map (as shown in Fig. 14(c)). The foreground depth map is subsequently binarized (as shown in Fig. 14(d)), followed by iterative erosion and dilation operations to remove small noise points and fill the foreground holes. This process results in a binary ROI image (as shown in Fig. 14(e)). Finally, the binarized image is mapped to the original image using a mask to obtain the final ROI region (as shown in Fig. 14(f)).

Since cracks are small objects that occupy only a small portion of the image pixels, directly segmenting the ROI from the original image involves processing a large number of background pixels. Therefore, a sliding window approach was employed to scan and identify the ROI in the original image. The YOLOv11-EVC model was utilized to detect cracks, and the identified bounding boxes were used as input for the SeaFormer model for segmentation. Comparative testing on 100 crack images showed that this method saves at least 70% of the time compared to directly segmenting the original image, while also effectively removing most of the noise. The crack images were then mapped onto the original image based on the bounding box location information obtained from YOLOv11-EVC, as shown in Fig. 15. The collected images were cropped and annotated, then added to the original dataset to form a hybrid dataset. After training and testing, the crack detection and localization precision reached 93.2%, with an F1 score of 87.7% and recall rate of 82.7%. The crack segmentation accuracy was 93.66%, with an IoU of 90.17%.

4.3. Conversion of pixel values to actual values

The above methods accomplish crack identification and segmentation. Then, we use the medial axis transform method to calculate the crack length and width. The resulting values are in pixels, which are not suitable for engineering applications. To address this, this paper derives the relationship between pixel size and actual size using the camera imaging principle, converting the crack feature calculation results from pixels to millimeters. A fixed-focus lens was used during the image acquisition process, so the imaging process can be represented by the lens imaging model, as shown in Fig. 16.

It has the following relationship:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{3}$$

where f is the focal length of the camera, u is the object distance, and v is the image distance. From the imaging model in Fig. 16 and the basic relationship of similar triangles, the relationship between the image size L', the actual object size L, the image distance v, and the object distance u is as follows:

$$\frac{L'}{L} = \frac{v}{u} \tag{4}$$

From Eqs. (3) and (4), it can be derived that:

$$L = \frac{L'(v-f)}{f} \tag{5}$$

Let the physical width of the sensor be S mm, and the resolution of the sensor be R pixels (in the width direction). Then, the physical length per pixel D is given by:

$$D = \frac{S}{R} \tag{6}$$

Let the pixel size of the target object be P. From the above equation, the imaging size of the object on the CMOS sensor is given by:

$$L' = \frac{S}{R} \cdot P \tag{7}$$

Therefore, the actual size of the object L is given by:

$$L = L'D = \frac{PS(u-f)}{Rf} \tag{8}$$

In the above equation, R, S, u, f, and other parameters are known camera intrinsic parameters. Therefore, by measuring the distance u from the drone to the target and the target's pixel value P, the actual size of the target object can be calculated.

To verify the accuracy of the above formula and the impact of camera distortion, multiple sampling tests were conducted using five

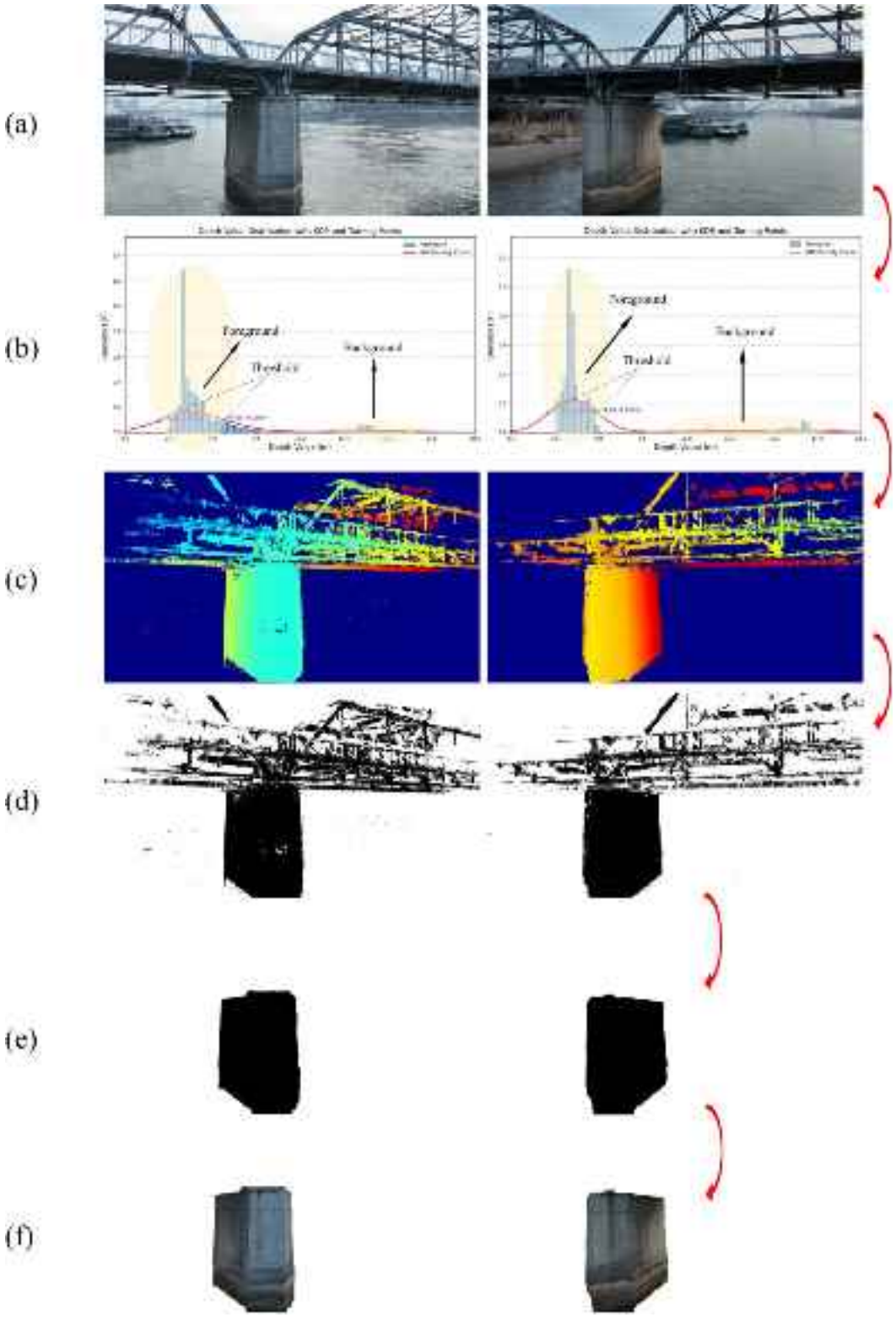

Fig. 14. Steps for ROI extraction based on depth information.

square grids with a side length of 40.60 mm at distances of 1.5 m, 3 m, and 4.5 m, as shown in Figs. 17 and 18.

At each distance, four images were captured by the drone, and the pixel values of each grid were measured. The average pixel value was taken as the pixel size of the corresponding grid. Then, the actual size of each grid was calculated using Eq. (8), and the results were compared with the true size of the grid for analysis as Table 4.

As shown in Table 4, the overall calculation accuracy is relatively high, with the relative error within 1.6%. The main source of error comes from the measurement error of the distance between the camera lens and the target object. In practice, it is difficult to precisely align the drone's camera lens to a specific distance. Additionally, the calculation error for the edge grids is slightly larger compared to the middle grids, primarily due to camera distortion. The calculation results for

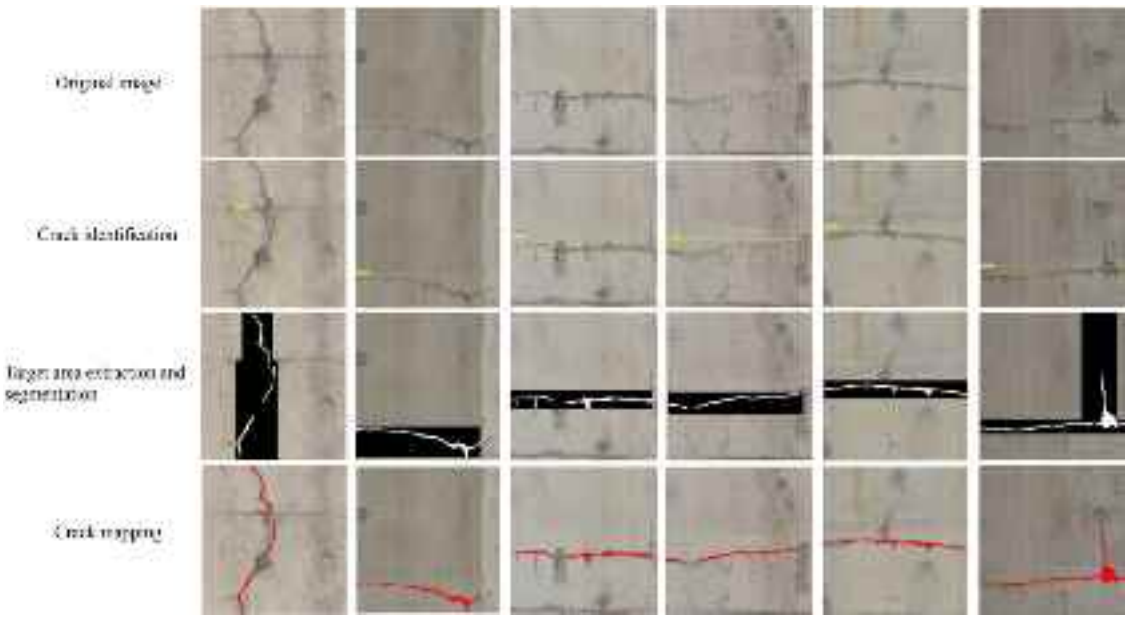

Fig. 15. Bridge pier surface crack identification.

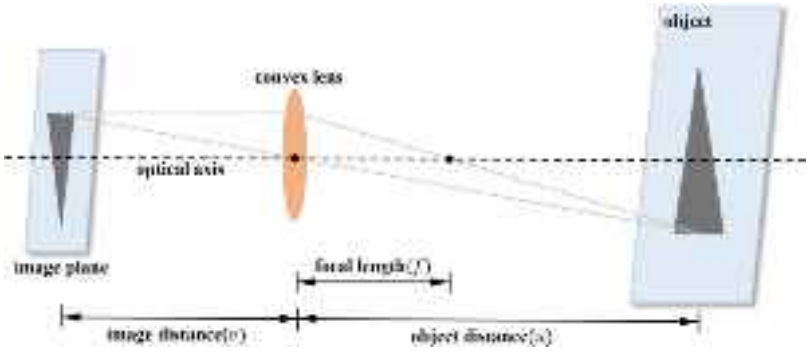

Fig. 16. Lens model.

Table 4
Accuracy and error calculation at different distances.

Grid	1500 mm				3000 mm				4500 mm			
	Pixel value	Computed value	Error	Relative error	Pixel value	Computed value	Error	Relative error	Pixel value	Computed value	Error	Relative error
c1	122	40.994 mm	0.394 mm	0.971%	59.25	39.957 mm	−0.643 mm	−1.584%	40	40.509 mm	−0.091 mm	−0.223%
c2	121.25	40.742 mm	0.142 mm	0.350%	59.50	40.125 mm	−0.475 mm	−1.169%	39.75	40.256 mm	−0.344 mm	−0.847%
c3	121	40.658 mm	0.058 mm	0.144%	60.00	40.463 mm	−0.137 mm	−0.338%	40	40.509 mm	−0.091 mm	−0.223%
c4	121	40.658 mm	0.058 mm	0.144%	61.00	41.137 mm	0.537 mm	1.323%	40	40.509 mm	−0.091 mm	−0.223%
c5	121.5	40.826 mm	0.226 mm	0.557%	60.25	40.631 mm	0.031 mm	0.077%	40	40.509 mm	−0.091 mm	−0.223%

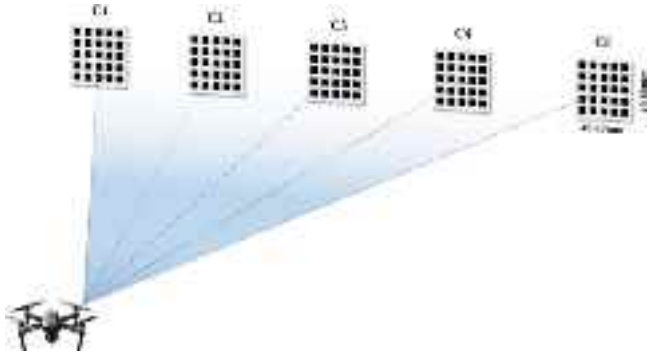

Fig. 17. Accuracy validation.

the middle grid (C3), which is directly facing the camera, are more accurate, with a relative error within 0.35%. Furthermore, a further analysis based on the formula reveals that to improve detection accuracy, the actual size corresponding to each pixel must be reduced. In other words, the key is to either decrease the shooting distance or increase the camera's resolution. However, it is important to note that the closer the distance, the more pronounced the edge distortion effect becomes.

4.4. Visualization of 3D crack distribution

The aforementioned methods perform crack detection on 2D images, but it is difficult to locate the exact position of the cracks in practical engineering applications, and the results are not intuitive. To

Fig. 18. Tests at various distances.

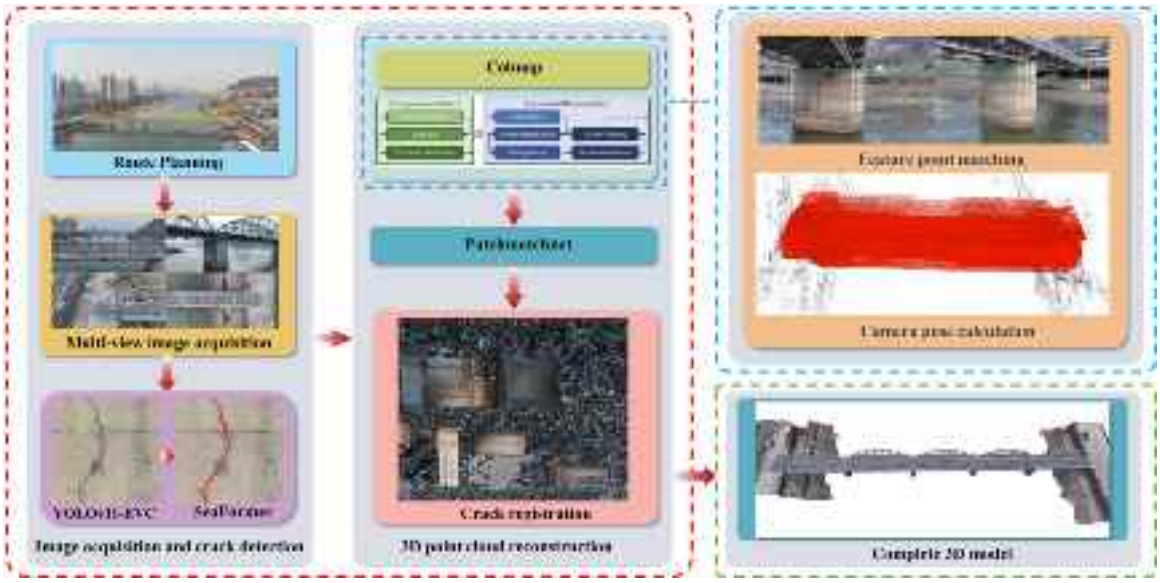

Fig. 19. Visualization method of 3D distribution of bridge cracks based on multi-view 3D reconstruction.

address this, this paper uses 3D reconstruction methods and image registration techniques to visualize the 3D distribution of crack information. A 3D reconstruction method based on the Colmap–Patchmatchnet algorithm is developed, and the crack images are registered to the 3D model to visualize the crack information distribution on the bridge. This process includes multiple steps, such as multi-view image acquisition, crack identification and segmentation, camera pose estimation, sparse 3D reconstruction, dense point cloud reconstruction, and crack image registration. The specific process is shown in Fig. 19.

Due to the numerous structural members in the upper part of the Zhongshan Bridge, the image overlap rate was approximately 80% to ensure the reconstruction quality. The images were captured using the UAV's built-in orbital flight function, resulting in a total of 2464 images for the entire bridge, including 6 images of bridge piers with cracks. Using the aforementioned image quality assessment model, low-quality images with defocus and abnormal exposure were removed, and the remaining 1782 high-quality images were used for 3D reconstruction.

Colmap utilizes the SIFT algorithm for feature extraction and matching, enabling the identification and matching of feature points across images captured by UAVs from different viewpoints [35]. During the pose estimation phase, Bundle Adjustment (BA) is employed to minimize projection errors by optimizing both the camera poses and the 3D point coordinates. The Structure from Motion (SfM) algorithm first provides the initial camera pose estimation and 3D point cloud. The BA algorithm then alternately optimizes the camera pose and 3D point coordinates, using the nonlinear least squares method to refine the camera pose and the Gauss–Newton method to optimize the 3D point coordinates. After several iterations, it obtains the optimal solution. The Patchmatchnet algorithm is a novel, learnable algorithm designed for high-resolution multi-view stereo (MVS) [36]. It has high computational speed and low memory requirements, enabling it to handle higher-resolution images and making it more suitable for resource-constrained devices compared to other methods using 3D cost volume regularization, without significantly compromising accuracy. This algorithm introduces an iterative multi-scale Patchmatch approach within an end-to-end trainable framework for the first time. It enhances the core Patchmatch algorithm by incorporating a novel, learnable adaptive strategy for both propagation and evaluation at each iteration. Despite its simple structure, extensive experiments conducted on the DTU, Tanks & Temples, and ETH3D datasets demonstrate that the algorithm outperforms existing methods in terms of computational efficiency and memory usage.

By combining the strengths of the Colmap and PatchMatchNet algorithms, a novel 3D reconstruction framework is proposed, delivering improved processing speed and reduced memory consumption. Upon completion of the reconstruction, a mesh model is generated, and the surface of the 3D model is unfolded onto a 2D plane. The detected crack information is then registered onto this 2D plane and reconstructed onto the 3D model. Through these steps, the 3D distribution of cracks is visualized (Fig. 20(c)), providing a clearer and more intuitive representation of crack position, width, length, and orientation (Fig. 20(d)). This approach offers a convenient platform for damage management and maintenance. Additionally, the 3D distribution and progression of crack clusters can be analyzed to infer the damage state of the bridge structure, facilitating early warnings and maintenance recommendations.

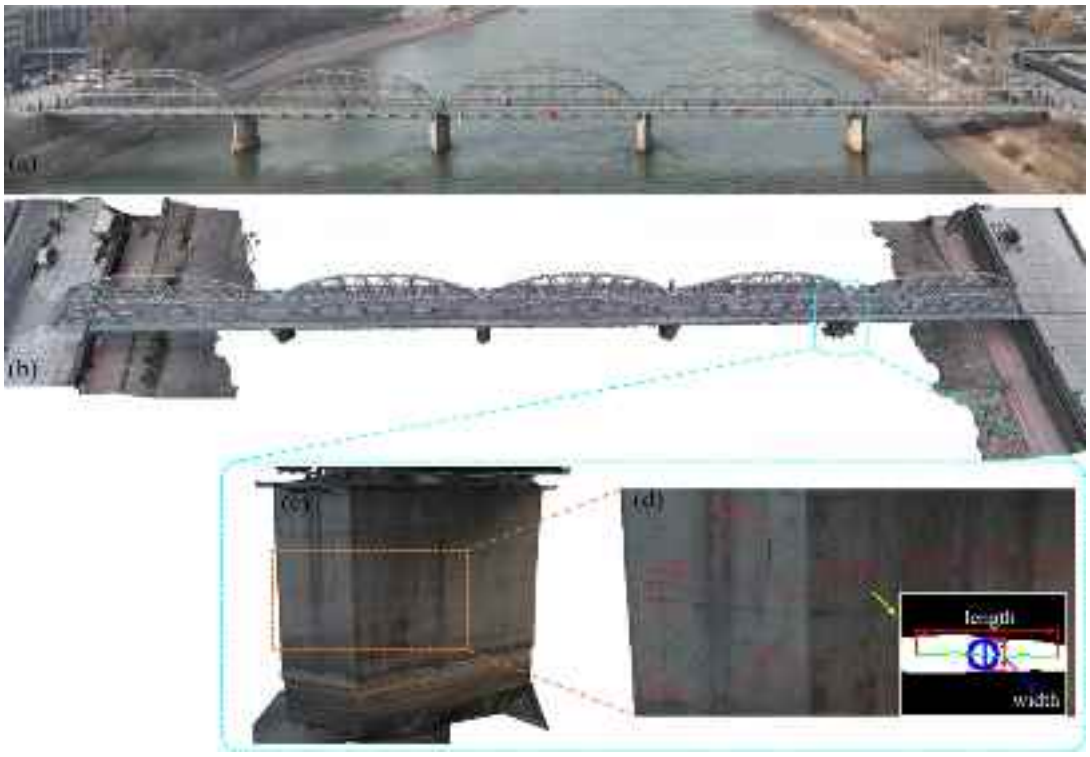

Fig. 20. Visualization of 3D distribution of cracks in Zhongshan Bridge.

5. Conclusion

This paper proposes an efficient bridge crack detection framework that combines lightweight design and high detection accuracy. The framework incorporates an image quality assessment model tailored for field environments and introduces depth-based ROI extraction and pseudo-crack removal methods. To address the challenges of crack localization and visualization, a superior 3D reconstruction scheme and a method for visualizing the 3D distribution of cracks are also proposed. The following conclusions can be drawn from this study:

(1) Compared to previous models, the crack detection model proposed in this paper, based on YOLOv11-EVC and SeaFormer, demonstrates superior performance in terms of detection accuracy, localization accuracy, and detection efficiency. Testing results show a recognition accuracy of 93.2%, an F1 score of 87.7%, a recall rate of 82.7%, a crack segmentation accuracy of 93.66%, and an IoU of 90.17%.

(2) The ROI extraction method based on depth information significantly reduces computational complexity while greatly enhancing crack detection accuracy in complex outdoor environments. In addition, this paper proposes a method where cracks are first identified, and then the bounding boxes are used as input for the segmentation model. After segmentation, the results are mapped back to the original image. This approach saves over 70% of the time compared to methods that directly segment the original image.

(3) The Colmap–Patchmatchnet multi-view 3D reconstruction scheme achieves superior reconstruction results while requiring lower memory and computational resources. By combining 3D surface unfolding with image registration techniques, the 3D distribution of crack information is visualized, making the crack details clearer and more intuitive. This provides valuable insights for bridge load capacity assessment and maintenance.

This paper primarily focuses on the automated detection of concrete bridge cracks using UAVs. In the future, additional damage datasets can be incorporated for training, enabling the identification of surface damage in various types of building structures. Additionally, LiDAR and other equipment can be integrated to achieve 3D depth information detection and visualization of cracks.

CRediT authorship contribution statement

Yahui Qi: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pengzhen Lin:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Guojun Yang:** Writing – review & editing, Supervision, Project administration, Investigation. **Tao Liang:** Writing – original draft, Investigation, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: We would like to submit the manuscript entitled ‘Crack detection and 3D visualization of crack distribution for UAV-based bridge inspection using efficient approaches’, which we wish to be considered for publication in Structures. No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication.

Acknowledgments

This study is supported by the Gansu Provincial Higher Education Industry Support Program Project (2024CYZC-21), and the Gansu Provincial Key R&D Programme-Industrial Projects (23YFGA0042)

Data availability

Data will be available on request.

References

[1] Spencer Jr Billie F, Hoskere Vedhus, Narazaki Yasutaka. Advances in computer vision-based civil infrastructure inspection and monitoring. Engineering 2019;5:199–222. http://dx.doi.org/10.3390/s22103789.

[2] Dong Chuan-Zhi, Catbas F Necati. A review of computer vision–based structural health monitoring at local and global levels. Struct Heal Monit 2021;20:692–743. http://dx.doi.org/10.1177/1475921720935585.

[3] Shu Jiangpeng, Bagge Niklas, Plos Mario, Johansson Morgan, Yang Yuguang, Zandi Kamyab. Shear capacity of a RC bridge deck slab: Comparison between multilevel assessment and field test. J Struct Eng 2018;144:04018081. http://dx.doi.org/10.1061/(ASCE)ST.1943-541X.00020.

[4] Ali Raza, Chuah Joon Huang, Talip Mohamad Sofian Abu, Mokhtar Norrima, Shoaib Muhammad Ali. Structural crack detection using deep convolutional neural networks. Autom Constr 2022;133:103989. http://dx.doi.org/10.1016/j.autcon.2021.103989.

[5] Deng Jianghua, Singh Amardeep, Zhou Yiyi, Lu Ye, Lee Vincent Cheng-Siong. Review on computer vision-based crack detection and quantification methodologies for civil structures. Constr Build Mater 2022;356:129238. http://dx.doi.org/10.1016/j.conbuildmat.2022.129238.

[6] Li Hongxia, Wang Weixing, Wang Mengfei, Li Limin, Vimlund Vivian. A review of deep learning methods for pixel-level crack detection. J Traffic Transp Eng (Engl Edition) 2022;9:945–68. http://dx.doi.org/10.1016/j.jtte.2022.11.003.

[7] Tian Yongding, Chen Chao, Sagoe-Crentsil Kwesi, Zhang Jian, Duan Wenhui. Intelligent robotic systems for structural health monitoring: Applications and future trends. Autom Constr 2022;139:104273. http://dx.doi.org/10.1016/j.autcon.2022.104273.

[8] Ai Dihao, Jiang Guiyuan, Lam Siew-Kei, He Peilan, Li Chengwu. Computer vision framework for crack detection of civil infrastructure—A review. Eng Appl Artif Intell 2023;117:105478. http://dx.doi.org/10.1016/j.engappai.2022.105478.

[9] Mohan Arun, Poobal Sumathi. Crack detection using image processing: A critical review and analysis. Alex Eng J 2018;57:787–98. http://dx.doi.org/10.1016/j.aej.2017.01.020.

[10] Hoang Nhat-Duc. Detection of surface crack in building structures using image processing technique with an improved Otsu method for image thresholding. Adv Civ Eng 2018;2018:3924120. http://dx.doi.org/10.1155/2018/3924120.

[11] Talab Ahmed Mahgoub Ahmed, Huang Zhangcan, Xi Fan, HaiMing Liu. Detection crack in image using Otsu method and multiple filtering in image processing techniques. Optik 2016;127:1030–3. http://dx.doi.org/10.1016/j.ijleo.2015.09.147.

[12] Dorafshan Sattar, Thomas Robert J, Maguire Marc. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Constr Build Mater 2018;186:1031–45. http://dx.doi.org/10.1016/j.conbuildmat.2018.08.011.

[13] Luo Jie, Lin Huazhi, Wei Xiaoxu, Wang Yongsheng. Adaptive canny and semantic segmentation networks based on feature fusion for road crack detection. IEEE Access 2023;11:51740–53. http://dx.doi.org/10.1109/ACCESS.2023.3279888.

[14] Hoang Nhat-Duc, Nguyen Quoc-Lam. Metaheuristic optimized edge detection for recognition of concrete wall cracks: a comparative study on the performances of roberts, prewitt, canny, and sobel algorithms. Adv Civ Eng 2018;2018:7163580. http://dx.doi.org/10.1155/2018/7163580.

[15] Yamaguchi Tomoyuki, Hashimoto Shuji. Fast crack detection method for large-size concrete surface images using percolation-based image processing. Mach Vis Appl 2010;21:797–809. http://dx.doi.org/10.1007/s00138-009-0189-8.

[16] Yamaguchi Tomoyuki, Hashimoto Shuji. Improved percolation-based method for crack detection in concrete surface images. In: 2008 19th international conference on pattern recognition. IEEE; 2008, p. 1–4. http://dx.doi.org/10.1109/ICPR.2008.4761627.

[17] Li Ruoxian, Yu Jiayong, Li Feng, Yang Ruitao, Wang Yudong, Peng Zhihao. Automatic bridge crack detection using Unmanned aerial vehicle and Faster R-CNN. Constr Build Mater 2023;362:129659. http://dx.doi.org/10.1016/j.conbuildmat.2022.129659.

[18] Zhao Weijian, Liu Yunyi, Zhang Jiawei, Shao Yi, Shu Jiangpeng. Automatic pixel-level crack detection and evaluation of concrete structures using deep learning. Struct Control Heal Monit 2022;29:e2981. http://dx.doi.org/10.1002/stc.2981.

[19] Li Shengyuan, Zhao Xuefeng, Zhou Guangyi. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Comput - Aided Civ Infrastruct Eng 2019;34:616–34. http://dx.doi.org/10.1111/mice.12433.

[20] Zoubir Hajar, Rguig Mustapha, El Aroussi Mohamed, Saadane Rachid, Chehri Abdellah. Pixel-level concrete bridge crack detection using convolutional neural networks, gabor filters, and attention mechanisms. Eng Struct 2024;314:118343. http://dx.doi.org/10.1016/j.engstruct.2024.118343.

[21] Ding Wei, Yang Han, Yu Ke, Shu Jiangpeng. Crack detection and quantification for concrete structures using UAV and transformer. Autom Constr 2023;152:104929. http://dx.doi.org/10.1016/j.autcon.2023.104929.

[22] Inam Hina, Islam Naeem Ul, Akram Muhammad Usman, Ullah Fahim. Smart and automated infrastructure management: A deep learning approach for crack detection in bridge images. Sustainability 2023;15:1866. http://dx.doi.org/10.3390/su15031866.

[23] Xu Yang, Fan Yunlei, Li Hui. Lightweight semantic segmentation of complex structural damage recognition for actual bridges. Struct Heal Monit 2023;22:3250–69. http://dx.doi.org/10.1177/14759217221147.

[24] Li Yangtao, Bao Tengfei, Huang Xianjun, Chen Hao, Xu Bo, Shu Xiaosong, Zhou Yuhang, Cao Qingbo, Tu Jiuzhou, Wang Ruijie, et al. Underwater crack pixel-wise identification and quantification for dams via lightweight semantic segmentation and transfer learning. Autom Constr 2022;144:104600. http://dx.doi.org/10.1016/j.autcon.2022.104600.

[25] Xu Gang, Yue Qingrui, Liu Xiaogang. Deep learning algorithm for real-time automatic crack detection, segmentation, qualification. Eng Appl Artif Intell 2023;126:107085. http://dx.doi.org/10.1016/j.engappai.2023.107085.

[26] Zhang Jian, Qian Songrong, Tan Can. Automated bridge surface crack detection and segmentation using computer vision-based deep learning model. Eng Appl Artif Intell 2022;115:105225. http://dx.doi.org/10.1016/j.engappai.2022.105225.

[27] Zhu Guijie, Liu Jiacheng, Fan Zhun, Yuan Duan, Ma Peili, Wang Meihua, Sheng Weihua, Wang Kelvin CP. A lightweight encoder–decoder network for automatic pavement crack detection. Comput - Aided Civ Infrastruct Eng 2024;39:1743–65. http://dx.doi.org/10.1111/mice.13103.

[28] Ni FuTao, Zhang Jian, Chen ZhiQiang. Pixel-level crack delineation in images with convolutional feature fusion. Struct Control Heal Monit 2019;26:e2286. http://dx.doi.org/10.1002/stc.2286.

[29] Meng Shiqiao, Gao Zhiyuan, Zhou Ying, He Bin, Djerrad Abderrahim. Real-time automatic crack detection method based on drone. Comput - Aided Civ Infrastruct Eng 2023;38:849–72. http://dx.doi.org/10.1111/mice.12918.

[30] Chen Liujie, Yao Haodong, Fu Jiyang, Ng Ching Tai. The classification and localization of crack using lightweight convolutional neural network with CBAM. Eng Struct 2023;275:115291. http://dx.doi.org/10.1016/j.engstruct.2022.115291.

[31] Zeng Qiqi, Fan Gao, Wang Dayang, Tao Weijun, Liu Airong. A systematic approach to pixel-level crack detection and localization with a feature fusion attention network and 3D reconstruction. Eng Struct 2024;300:117219. http://dx.doi.org/10.1016/j.engstruct.2023.117219.

[32] Li Xiaofei, Meng Qinghang, Wei Mengpu, Sun Heming, Zhang Tian, Su Rongrong. Identification of underwater structural bridge damage and BIM-based bridge damage management. Appl Sci 2023;13:1348. http://dx.doi.org/10.3390/app13031348.

[33] Chen Siyuan, Laefer Debra F, Mangina Eleni, Zolanvari SM Iman, Byrne Jonathan. UAV bridge inspection through evaluated 3D reconstructions. J Bridg Eng 2019;24:05019001. http://dx.doi.org/10.1061/(ASCE)BE.1943-5592.00013.

[34] Liu Yu-Fei, Nie Xin, Fan Jian-Sheng, Liu Xiao-Gang. Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. Comput - Aided Civ Infrastruct Eng 2020;35:511–29. http://dx.doi.org/10.1111/mice.12501.

[35] Schonberger Johannes L, Frahm Jan-Michael. Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 4104–13.

[36] Wang Fangjinhua, Galliani Silvano, Vogel Christoph, Speciale Pablo, Pollefeys Marc. Patchmatchnet: Learned multi-view patchmatch stereo. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 14194–203.

[37] Su Shaolin, Yan Qingsen, Zhu Yu, Zhang Cheng, Ge Xin, Sun Jinqiu, Zhang Yanning. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 3667–76.

[38] Xiao Jing-Lin, Fan Jian-Sheng, Liu Yu-Fei, Li Bao-Luo, Nie Jian-Guo. Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection. Autom Constr 2024;158:105226. http://dx.doi.org/10.1016/j.autcon.2023.105226.

[39] Quan Yu, Zhang Dong, Zhang Liyan, Tang Jinhui. Centralized feature pyramid for object detection. IEEE Trans Image Process 2023. http://dx.doi.org/10.1109/TIP.2023.3297408.

[40] Wan Qiang, Huang Zilong, Lu Jiachen, Yu Gang, Zhang Li. Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation. 2023, http://dx.doi.org/10.48550/arXiv.2301.13156, arXiv preprint arXiv:2301.13156.

[41] Li Liangfu, Sun Ruiyun. Bridge crack detection algorithm based on image processing under complex background. Laser Optoelectron Prog 2019;56:061002. http://dx.doi.org/10.3788/LOP56.061002.

[42] Dorafshan Sattar, Thomas Robert J, Maguire Marc. SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. Data Brief 2018;21:1664–8. http://dx.doi.org/10.1016/j.dib.2018.11.015.