buildings

MDPI

Article

Bridge Surface Defect Localization Based on Panoramic Image Generation and Deep Learning-Assisted Detection Method

Tao Yin [1,2], Guodong Shen [2], Liang Yin [2] and Guigang Shi [1,*]

[1] School of Civil Engineering, Anhui Jianzhu University, Hefei 230009, China; flitetest@163.com
[2] Anhui Transport Consulting and Design Institute Co., Ltd., Hefei 231299, China; shenguodong_jgy@163.com (G.S.); yliang_jgy@163.com (L.Y.)
* Correspondence: shiguigang@ahjzu.edu.cn

Abstract: Applying unmanned aerial vehicles (UAVs) and vision-based analysis methods to detect bridge surface damage significantly improves inspection efficiency, but the existing techniques have difficulty in accurately locating damage, making it difficult to use the results to assess a bridge's degree of deterioration. Therefore, this study proposes a method to generate panoramic bridge surface images using multi-view images captured by UAVs, in order to automatically identify and locate damage. The main contributions are as follows: (1) We propose a UAV-based image-capturing method for various bridge sections to collect close-range, multi-angle, and overlapping images of the surface; (2) we propose a 3D reconstruction method based on multi-view images to reconstruct a textured bridge model, through which an ultra-high resolution panoramic unfolded image of the bridge surface can be obtained by projecting from multiple angles; (3) we applied the Swin Transformer to optimize the YOLOv8 network and improve the detection accuracy of small-scale damages based on the established bridge damage dataset and employed sliding window segmentation to detect damage in the ultra-high resolution panoramic image. The proposed method was applied to detect surface damage on a three-span concrete bridge. The results indicate that this method automatically generates panoramic images of the bridge bottom, deck, and sides with hundreds of millions of pixels and recognizes damage in the panoramas. In addition, the damage detection accuracy reached 98.7%, which is improved by 13.6% when compared with the original network.

Keywords: bridge inspection; UAV; damage detection; 3D reconstruction; deep learning

check for updates

Citation: Yin, T.; Shen, G.; Yin, L.; Shi, G. Bridge Surface Defect Localization Based on Panoramic Image Generation and Deep Learning-Assisted Detection Method. *Buildings* **2024**, *14*, 2964. https://doi.org/10.3390/buildings14092964

Academic Editor: Hugo Rodrigues

Received: 26 July 2024
Revised: 28 August 2024
Accepted: 30 August 2024
Published: 19 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Bridges are inevitably prone to defects and cracking under long-term loads, long-age phenomena of concrete, and prestressing losses with time [1,2]; therefore, damage detection and assessment for bridges constitute essential tasks in bridge safety maintenance, the routine activities to which bridge management authorities must allocate significant human and financial resources. Although conventional manual inspection methods play a crucial role in bridge inspection, their drawbacks, including low efficiency, high costs, and high risk, remain inherent limitations. In recent years, rapid advancements and the widespread application of various UAVs and robots as well as the proliferation and cost reduction of technologies such as light detection and ranging (LiDAR), stereo cameras, and inertial navigation systems have facilitated the widespread adoption of UAVs and robotic technologies across various industries. In the field of bridge defect detection, the utilization of UAVs to access complex, narrow, high-altitude, and hazardous environments and the automated acquisition of images have enabled the inspection of locations such as high piers, tall towers, cables, and bridge undersides, where traditional manual inspections are difficult. All of these methods are regarded as having promising application prospects. Concurrently, rapid advancements and applications of visual analysis methods such as image processing and deep learning have been witnessed in recent years. In the realm of

structural detection, research on employing these methods for automated defect identification and analysis has garnered considerable attention. Consequently, this study proposes a method utilizing UAVs to capture images of bridge surfaces and apply deep learning-based visual recognition techniques to identify typical bridge defects, thereby achieving rapid defect detection as an alternative to manual inspection methods.

The rapid acquisition of complete and high-precision images of bridge surfaces for the detection of damage is key in fieldwork, serving as the basis for deciding whether the damage can be accurately recognized. However, because bridges generally cross mountains and rivers and are large and locally complex, it is difficult to collect complete data. Research has applied robots for photography to collect images depicting bridge damage. For instance, Hou et al. employed a climbing robot to inspect the cables of a large-span cable-stayed bridge, using four cameras mounted on the robot to detect surface damage on the cables [3]. Wang et al. developed a magnetic adhesion-based tracked inspection robot equipped with magnetic adhesion and crawling traction modules, enabling the robot to crawl adaptively on the surfaces of steel structure bridges with varying curvatures [4]. Qin et al. designed a novel hexapod climbing robot and established a kinematic model for the exterior inspection of large bridges [5]. While these climbing robots can be used for relatively simple surfaces or steel structures, challenges persist in crawling on complex and non-magnetic surfaces such as concrete. The use of UAVs for damage detection has also gained significant attention in recent years. Compared to various types of robots, UAVs offer lower costs and greater flexibility, with almost no limitations on the geometry of the detection area [6,7]. For example, Jang et al. proposed a hybrid image-scanning method based on UAVs to capture and detect cracks in concrete structures [6]; the method combines a hybrid of both vision and infrared thermography images, which turned out to be able to improve crack detectability while minimizing false alarms. Tang et al. proposed a 3D modeling method that combines oblique photography with inclined photography so that clearer textures, more complete lines, and higher accuracy can be observed for damage detection [7]. The aforementioned studies demonstrate that UAVs can serve as convenient tools for acquiring bridge surface images. However, most of these studies have focused on localized positions on structures such as bridges and did not comprehensively inspect the entire surface. For bridge surface defect detection, a complete inspection of the entire bridge is necessary. Therefore, the first aspect of our research is studying UAV-based photography methods for each part of a bridge such that the sequential images captured within the limited field of view of the UAV's camera can provide complete coverage of the surface of each part of the bridge.

Recognizing damage in massive, high-resolution bridge surface images through visual analysis is another focus of bridge inspection. With the rapid advancement of computer vision technology in recent years, these methods have increasingly been applied in structural damage assessment [8] and detection [9,10]. For example, Cataldo et al. [8] proposed a low-cost vision-based measurement system for assessing the structural damage caused by seismic loads on buildings, and the results demonstrated that the method has a high potential for extracting and analyzing modal parameters. Concrete cracking is the most common structural defect, and because cracks can be used to improve the finite element model for more accurate analysis of bridge performance, crack detection is considered crucial. Some of the earliest research involved the application of such methods for crack detection. Li et al. [9] proposed a Two-RTSs system for UAV hovering accuracy and applied Faster R-CNN network for crack detection; the achieved accuracy, recall, and F1 score were 92.03%, 96.26%, and 94.10%, respectively. Li et al. [10] proposed a robust and direct crack detection method based on geometric correction and calibration algorithms to overcome the problem of inaccurate crack identification due to the deviation of the viewing angle, and the results showed that the four-point laser-based method has a higher accuracy in crack-width identification compared with the lens imaging conceptual method, with a measurement accuracy of more than 95%. Zhang et al. [11] achieved higher accuracy with their proposed CrackNet network, which lacks any pooling layers and predicts each pixel individually,

enabling pixel-level recognition accuracy. Ni et al. proposed a CNN-based dual-scale detection method to detect wide and thin cracks and a sub-pixel crack-width measurement method based on Zernike moments to detect the width of fine cracks, enabling the accurate detection of cracks with pixel sizes less than 5 [12]. In addition to crack detection, research on the detection of other structural defects, such as concrete spalling, corrosion, and loose bolts, has also been conducted. Various deep learning models such as Faster R-CNN, you look only once (YOLO) v3, and SSD networks [13] have gradually been applied to achieve the high-precision automatic detection of bolt loosening [14], corrosion [15], and various types of concrete damage [16]. For example, Xu et al. proposed a nested attribute-based meta-learning method for the identification of 10 representative defect types, overcoming the low robustness of traditional networks on limited datasets [17]. The above studies have demonstrated the high accuracy, efficiency, and application potential of vision-based damage detection methods in structural inspection; however, these learning-based methods are also characterized by high dependence on their training sets and the use of precision-designed superior networks. A substantial number of results in existing studies have been obtained based on small-batch datasets built by researchers, some of which use images collected in the laboratory, which does not necessarily represent the real surface conditions of in-service bridges. Meanwhile, with the rapid development of deep learning networks, many outstanding novel networks have been designed and proposed. Thus, it is also important to train and test the performance of state-of-the-art networks on datasets containing a large number of real, in-service bridge surface images in order to evaluate whether they can be applied for the detection of multiple damage types.

Based on the above review, this study proposes a bridge inspection method that applies UAVs to capture multi-angle photographs to obtain a complete surface image of a bridge, from which an overall 3D model of the bridge can be reconstructed. It then applies a deep learning-based damage recognition algorithm for damage detection. The contributions of this study are as follows: (1) A UAV-based image-capturing method is proposed for various bridge sections to collect close-range, multi-angle, and overlapping bridge surface images. (2) A 3D reconstruction technique is introduced, utilizing multi-view images to create a textured model of the bridge. This allows for the generation of ultra-high resolution, panoramic, unfolded images of the bridge surface through multi-angle projections. (3) A bridge damage dataset was established based on images collected from in-service bridges, using a Swin Transformer to optimize the YOLOv8 network to improve the network's detection accuracy for small-scale damage. Sliding window segmentation was then applied to automatically identify the damage in the ultra-high resolution panoramic image.

The structure of this article is outlined as follows: Section 1 introduces the research background. Section 2 presents the framework of the proposed method and provides detailed explanations of its core methodologies and technologies. Section 3 discusses the application of the proposed method to detect defects in a concrete bridge. Section 4 concludes the article.

2. Framework of the Proposed Method

This section introduces the framework of the proposed method, which can be divided into three parts (Figure 1): the overall bridge image acquisition method based on UAV close-range photography; the refinement model of the 3D reconstruction method based on stereo geometry; and the identification of typical bridge damage based on improved YOLOv8.

Section 2.1 discusses the overall image acquisition method based on UAV close-range photography, focusing on ways to select appropriate photography angles and flight paths for each part of the bridge. Section 2.2 introduces the proposed 3D bridge reconstruction method based on the structure from motion (SFM) and multi-view stereo (MVS) algorithms, focusing on accurately setting the camera position for image sequences with a finite overlap rate. Section 2.3 describes the establishment of a dataset based on in-service bridge defect images as well as improving and training state-of-the-art object detection networks to achieve the automatic recognition of typical bridge defects.

When applying the proposed method for the detection of bridge surface damage, the flow is also shown in Figure 1. Firstly, the method in Section 2.1 is used to control the UAV to take photos of each part of the bridge with a certain speed, distance, and photo overlap rate in a complete coverage style to obtain close-range and high-resolution images. Then, the method proposed in Section 2.2 is used to reconstruct the 3D model of the bridge from the original image, and the ultra-high resolution panoramic image of the bridge surface is projected, which can fuse a large number of fragmented images into a complete panoramic image. Finally, the network proposed in Section 2.3 is used to recognize the damage in the panoramic image and label the location of the damage to complete the detection of the bridge surface damage.

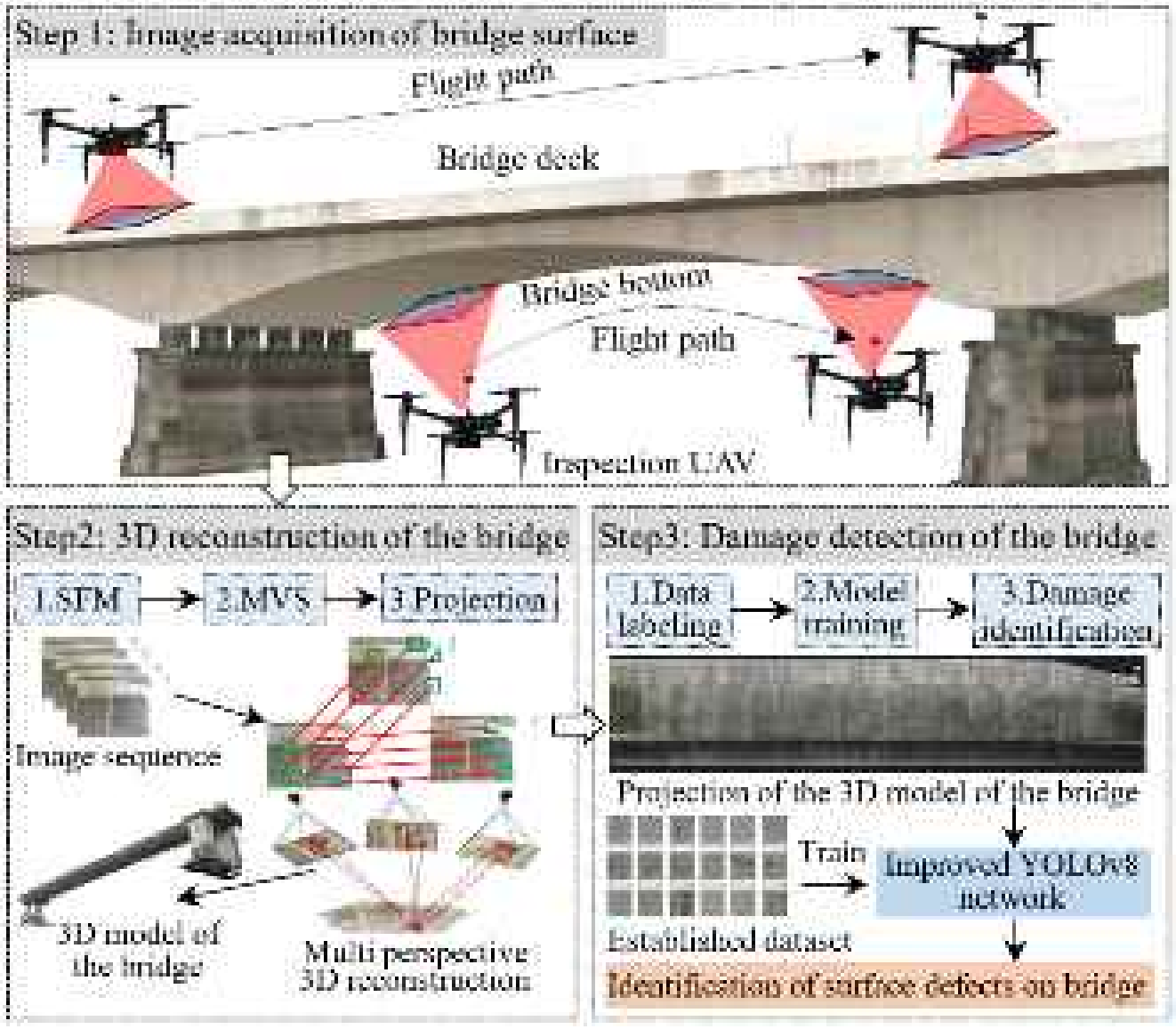


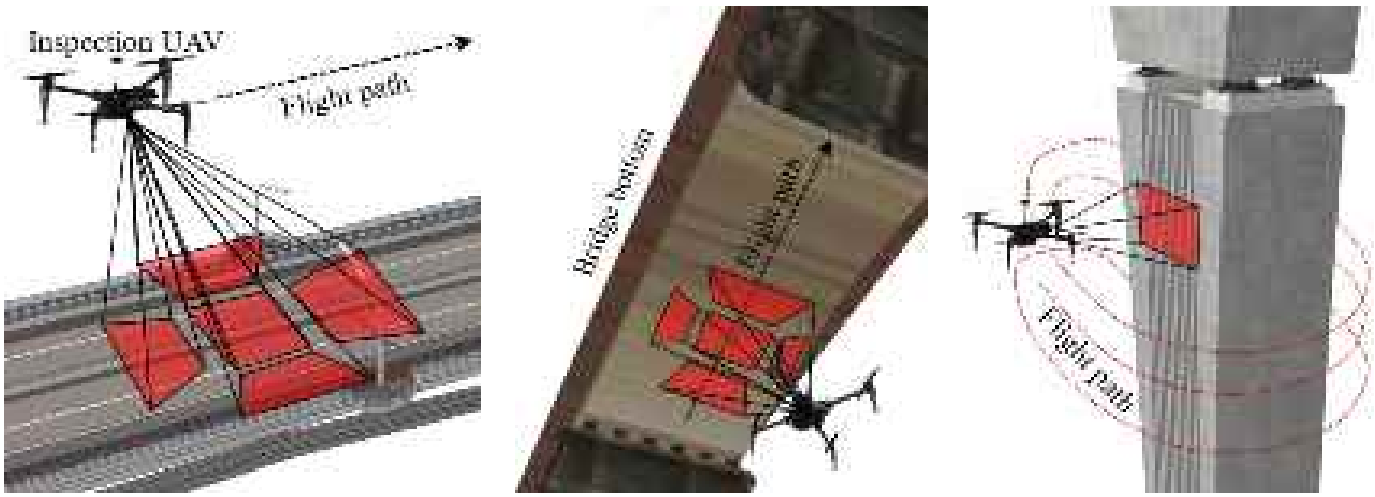
Figure 1. Framework of the proposed method.

2.1. Close-Range Photography

When applying UAVs to capture images around bridges, it is recommended that the UAVs have the following characteristics: (1) As there are obstacles and limited headroom around bridges, especially at the bottoms of small bridges, the UAVs should have obstacle-sensing and obstacle-avoidance capabilities in multiple directions. (2) When flying around bridges, especially underneath, it is difficult for UAVs to receive GPS positioning signals.

Therefore, they should have other positioning capabilities such as visual positioning. (3) The camera carried by the UAV should have high resolution and small distortion to enhance the efficiency of data acquisition. Considering the above conditions, this study adopted the M300RTK equipped with a P1 camera from DJI Company, Shenzhen, China, to acquire bridge surface images. This UAV has six-direction vision and laser-ranging functions, enabling it to perceive obstacles even in complex environments. The P1 camera was used for aerial mapping, which not only has low distortion but also has 45 million effective pixels and can acquire images with a resolution of 8192 × 5460 pixels.

After determining the UAV to be used, it was necessary to determine the appropriate photography method for different bridge sections. Bridges are generally divided into two major sections: the upper structure, including bridge towers, cables, and bridge decks, and the lower structure, mainly including the bottom and piers of the bridge. As the upper structure is generally in an open environment with fewer obstacles, and the UAV can receive sufficient GPS signals for accurate positioning, we adopted a capture method similar to aerial photogrammetry, which involves setting up parallel routes covering the whole bridge using a route planning tool. To capture multiple perspectives, the UAV takes photos from multiple angles, including directly below, in front, at the back, on the left side, and on the right side. In photographing the lower structure, it is difficult to plan a route for automatic photography, as there is typically no GPS signal underneath the bridge. Thus, we mounted the camera on top of the UAV, such that it could take images directly upward or tilted upward while flying under the bridge. Specifically, three route types were used to cover the range of the bridge bottom: several parallel routes distributed directly below the bridge to capture images at a 90° angle vertically from the bottom upwards and two tilted routes on the left and right sides to capture images at a 45° angle tilted toward the bridge. The piers around the bridge were photographed with reference to the method used for UAVs when photographing buildings. The photography methods for the bridge deck, bottom, and piers are shown in Figure 2.



(**a**) (**b**) (c)

Figure 2. Photography methods for various bridge sections. (**a**) Photographing bridge deck. (**b**) Photographing bridge bottom. (**c**) Photographing bridge pier.

The images acquired by the UAV should not only cover the entire surface of the bridge but also have a degree of overlap with respect to each other. This makes it possible to reconstruct a 3D model of the bridge. Thus, the flight speed of the UAV should be maintained in a suitable range during photography such that the rate of image overlap is no less than 60%. The distance between the UAV and its target must also be calculated in

advance, based on the required image resolution. According to the pinhole camera imaging model, this distance should satisfy the following formula:

$$\frac{f}{H} = \frac{size_{sensor}}{L} \tag{1}$$

where L is the length of the bridge covered by the image, H is the object distance, f is the focal length of the UAV's camera, and $size_{sensor}$ is the size of the camera sensor.

2.2. Fine 3D Reconstruction

This section describes a method for constructing a 3D model of the bridge from the acquired images and projecting a panoramic unfolded view of its surface using this model. The 3D reconstruction method based on multiple views is an essential computer vision technique that has been widely researched and applied in recent years, which has been applied in various fields such as reverse engineering, photogrammetry, and aerial mapping [18–20]. In this section, the structure from motion (SFM) algorithm and the multi-view stereo (MVS) algorithm [21] are applied to reconstruct a 3D point cloud and a model of the bridge from the original images acquired with the UAV. The detailed textures in the original image are mapped onto the model to obtain a fine 3D model of the bridge. The 3D reconstruction algorithm applied in this step can be divided into three steps: sparse reconstruction, depth map estimation, and dense reconstruction. The sparse reconstruction stage involves image feature extraction, feature matching, matching optimization, triangulation, attitude estimation, and beam leveling. The image feature extraction stage adopts the A-Sift method, which simulates various affine distortions through establishing a hemispherical space and performs SIFT feature point matching on this basis to achieve affine invariance. The feature matching stage uses the Euclidean distance to compute the detected image features, and as the acquired images are sequential, the matching computation is performed only on the neighboring images. The matching results are verified for possible overlapping image feature associations via geometric calibration, using the RANSAC algorithm to estimate the transformation matrix and verifying the validity of the matching through mapping the projective geometry in the image pairs [22]. After feature extraction and matching, sparse point cloud reconstruction is performed using an incremental reconstruction method. The method first selects an appropriate initial image initialization model and then calculates the bitmap in an image-by-image manner using the PnP method, matching the 2D feature points of the image into the 3D coordinates. At the same time, the triangulation and beam-leveling methods are used; the former is used to extend the point set to increase the scene range, and the latter is used to increase pose redundancy through the use of additional triangulation points to further optimize the overall model's pose and avoid obvious drift phenomena in the reconstruction process. At this point, the setting of the image pose and the reconstruction of the sparse point cloud can be completed. The depth information of all points in the image is further calculated using the MVS algorithm to obtain a dense point cloud, applying the method proposed by Schönberger et al. [21], which is a depth map fusion-based method that generates depth and normal maps for all the registered images, fuses the depth and normal maps into a dense point cloud with normal information, and, finally, fuses the dense surface of the point cloud using Poisson estimation. The next step is to project the pixel details onto the surface of the 3D model through texture mapping of the image, which finally yields a 3D model of the bridge with realistic texture [23]. As the UAV is close to the bridge surface when taking photos, usually in the range of 5 to 10 m, the 3D model is a fine model with highly detailed textures.

After obtaining the 3D model, it is projected from the top, bottom, left, and right sides of the bridge to obtain panoramic unfolded views of the entire surface. The steps of the aforementioned method are illustrated in Figure 3.

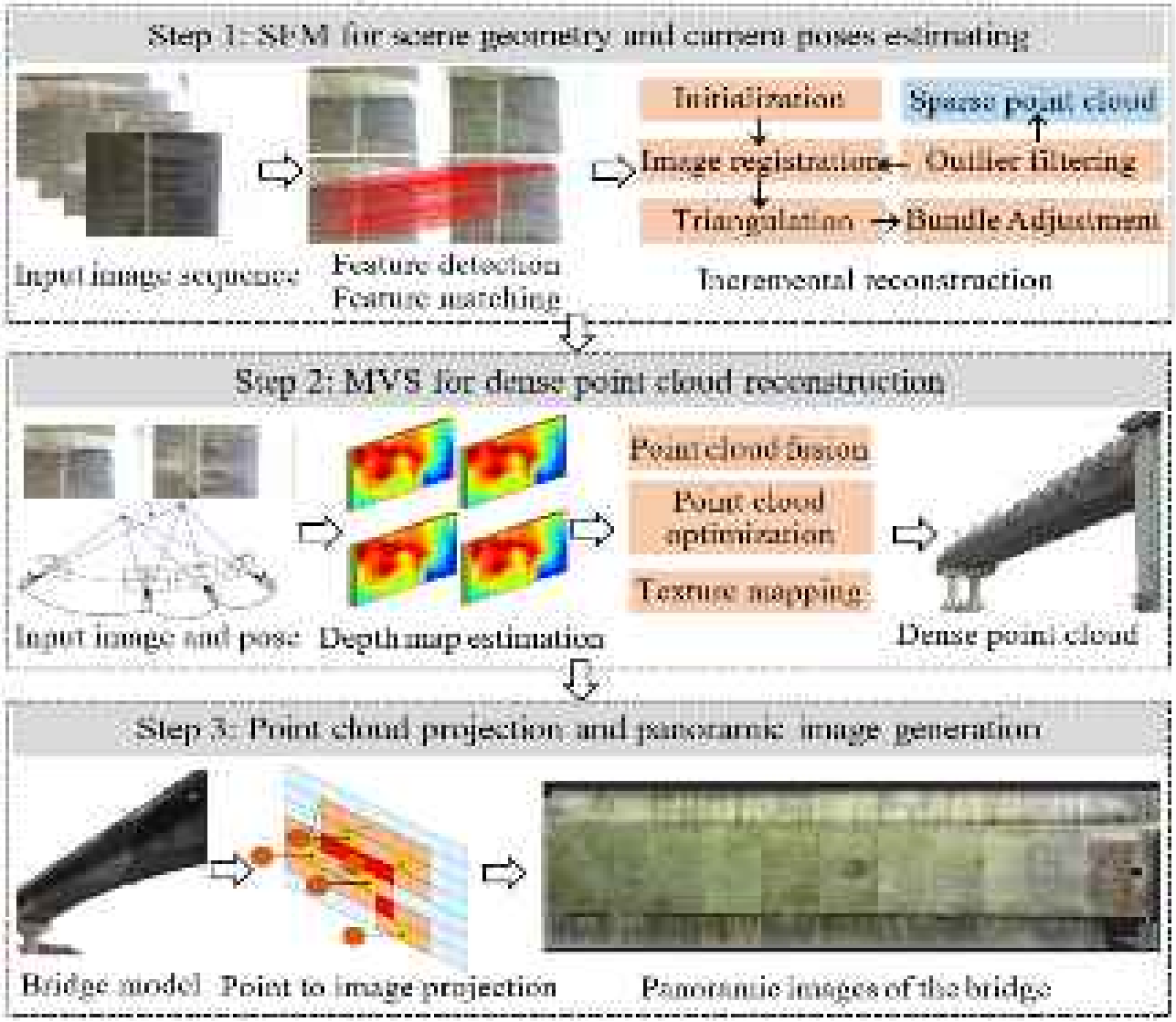


Figure 3. Steps of the 3D reconstruction method for bridges.

2.3. Identifying Bridge Damage

As the most rapidly developing and widely used type of machine vision method, deep learning-based image recognition has become the mainstream technique for structural damage identification. The key aspects of this type of method are establishing datasets and setting up networks, and so, this section introduces these two aspects. Mainstream deep learning networks for image processing are currently divided into three main types: classification networks, object detection networks, and segmentation networks. Object detection and segmentation networks can frame and label objects in an image, where the difference is that segmentation networks segment the pixels of objects from the background. Due to the need for rapid and qualitative inspection in daily inspections, segmentation networks generally have slower inference speeds than object detection networks because of their higher complexity. In this study, an object detection network is used as the identification network for bridge defects [24].

With the rapid development of deep learning networks for detection tasks, the performance of object detectors has greatly improved. Mainstream object detection networks mainly include one-stage and two-stage detectors, with the former framing detection as a "one-step" process and the latter defining it as a "coarse-to-fine" process. Representative two-stage detector networks include Faster R-CNN and Feature Pyramid Networks

(FPNs) [25], which split the detection task into two steps: localization followed by recognition. Localization retains as much useful foreground information as possible, filters out the background information that is not useful for the subsequent task, and then distinguishes the foreground information. As the foreground information is distinguished, this classification network is finer. A one-stage detector relies on a network output containing both coordinates and a tensor with category confidence, greatly simplifying the two-stage framework and leaving the localization and classification to the RPN, thus improving the inference speed and simplifying the training steps. The YOLO series networks are typical two-stage detectors, and for early networks of this type, the inference speed is fast, but the accuracy is insufficient [26–28]. With the rapid development and maturity of these algorithms, the YOLO series networks have gone through eight generations, and the latest network, YOLOv8, has been significantly improved in terms of accuracy and speed. Therefore, we selected the YOLOv8 network for bridge damage identification.

2.3.1. Overview of YOLOv8

The YOLOv8 network is an updated network based on the optimization of the YOLO series networks. Its improvements include the introduction of a new backbone network, the inclusion of anchor-free detection headers, and the use of a new loss function, enhancing the performance and flexibility of the network. The backbone network and neck part of YOLOv8 replace the C3 structure of the earlier YOLOv5 with a C2f structure with richer gradient flow, adjusting the number of channels differently for different scale models. The specific changes are as follows: the kernel of the first convolutional layer was changed from 6×6 to 3×3; all C3 were replaced with C2f modules; more skipping layer connections and additional Split operations were added, thus fine-tuning the model structure and expanding the size of the network from the original four dimensions to five; and a p6 network with a resolution of up to 1280 pixels was added, making it easier to apply the network to different tasks. In the detection head part, the YOLOv8 network changed from the original coupling head to a decoupling head and from the anchor-based method of YOLOv5 to an anchor-free method; this change removes the need to preset the anchor. Only the target center point and the width and height of the feature maps at different scales need to be regressed, reducing time consumption and computing capacity. This can also avoid missed or repeated detection problems caused by unreasonable anchor settings. In the loss function, mainstream object detection networks generally adopt the dynamic allocation strategy, such as YOLOX's simOTA [29] or TOOD network's Task-Aligned Assigner [30]. The dynamic allocation strategy can dynamically adjust the weights according to the progress of training and the characteristics of the samples. In the early stage of training, the model may have difficulty differentiating between positive and negative samples, so it should pay more attention to samples that are easy to misclassify. As training progresses, the model gradually becomes better able to distinguish between samples, slowly reducing the weights of difficult samples while increasing the weights of easy-to-distinguish samples. The dynamic allocation strategy can be adjusted according to the training loss or other metrics, which can be better adapted to different datasets. On the other hand, YOLOv5 still adopts a static allocation strategy, which is usually based on experience and cannot fully utilize sample information, possibly leading to poor training results. YOLOv8 employs the Task-Aligned Assigner, which selects positive samples based on the weighted classification and regression scores, expressed as $t = s^{\alpha} \times u^{\beta}$, where s represents the predicted score corresponding to the annotated category, and u is the intersection over union (IOU) between the predicted box and the ground truth box.

2.3.2. Improvement of YOLOv8

Defects such as fine cracks, small pieces of spalled concrete, and exposed reinforcement challenge the original YOLOv8 network when detecting damage on bridge surfaces. In order to overcome this problem, this study added a module for small-target feature extraction to the original YOLOv8 network, enhancing the network's recognition accuracy

for small lesions. Network strategies and modules for tiny target detection have been proposed in recent years, of which Featurized Image Pyramid (FPN), sliding window segmentation, and attention mechanisms are representative [31]. Notably, a transformer module incorporating self-attention has been widely applied in visual recognition in recent years [32]. In this section, the Swin Transformer module [33], which further optimizes the computational complexity of the original transformer structure, is added to the YOLOv8 network to improve its performance for small-target detection.

The Swin Transformer has a hierarchical design containing four stages, each of which reduces the resolution of the input feature maps and expands the sensory field in a layer-by-layer manner, similar to a CNN. While the original transformer module computes self-attention for N tokens with a computational complexity of $O(N^2)$, the Swin Transformer divides N tokens into N/n groups, each consisting of n tokens, reducing the complexity to $O\left(N \times n^2\right) = O(N)$. However, after group-wise computation, there will be a lack of interaction between groups, posing challenges in extracting global information. Therefore, after each stage, the Swin Transformer integrates and compresses the feature vectors of 2×2 groups, ensuring that the network's receptive field increases gradually with each stage, similar to CNN-based structures. Additionally, through the use of shifted windows, adjacent patches interact with each other. In this section, we attempt to replace the backbone network of YOLO v8 with the Swin Transformer and evaluate the extent of the network's performance improvement on a bridge defect recognition dataset. The modified network architecture is illustrated in Figure 4.

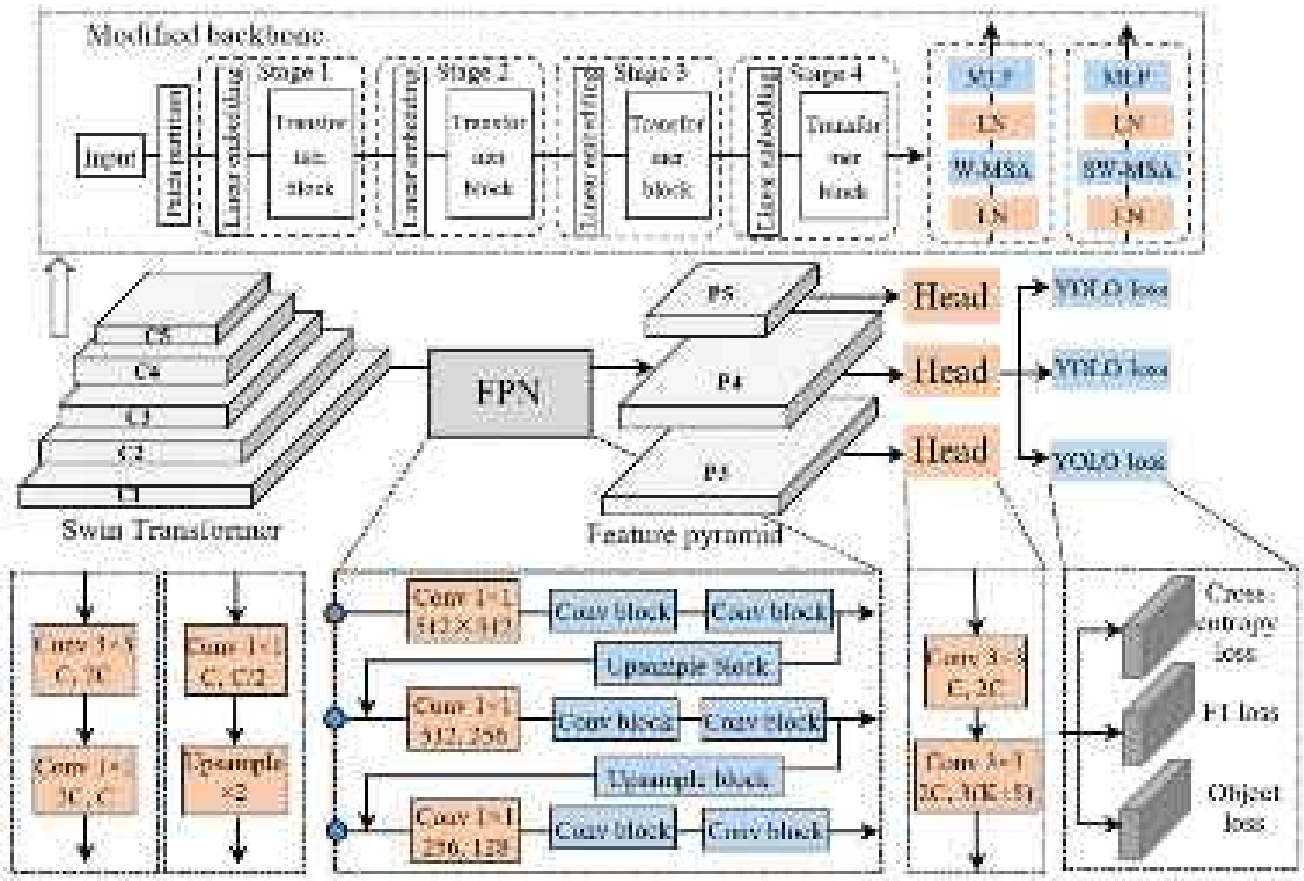
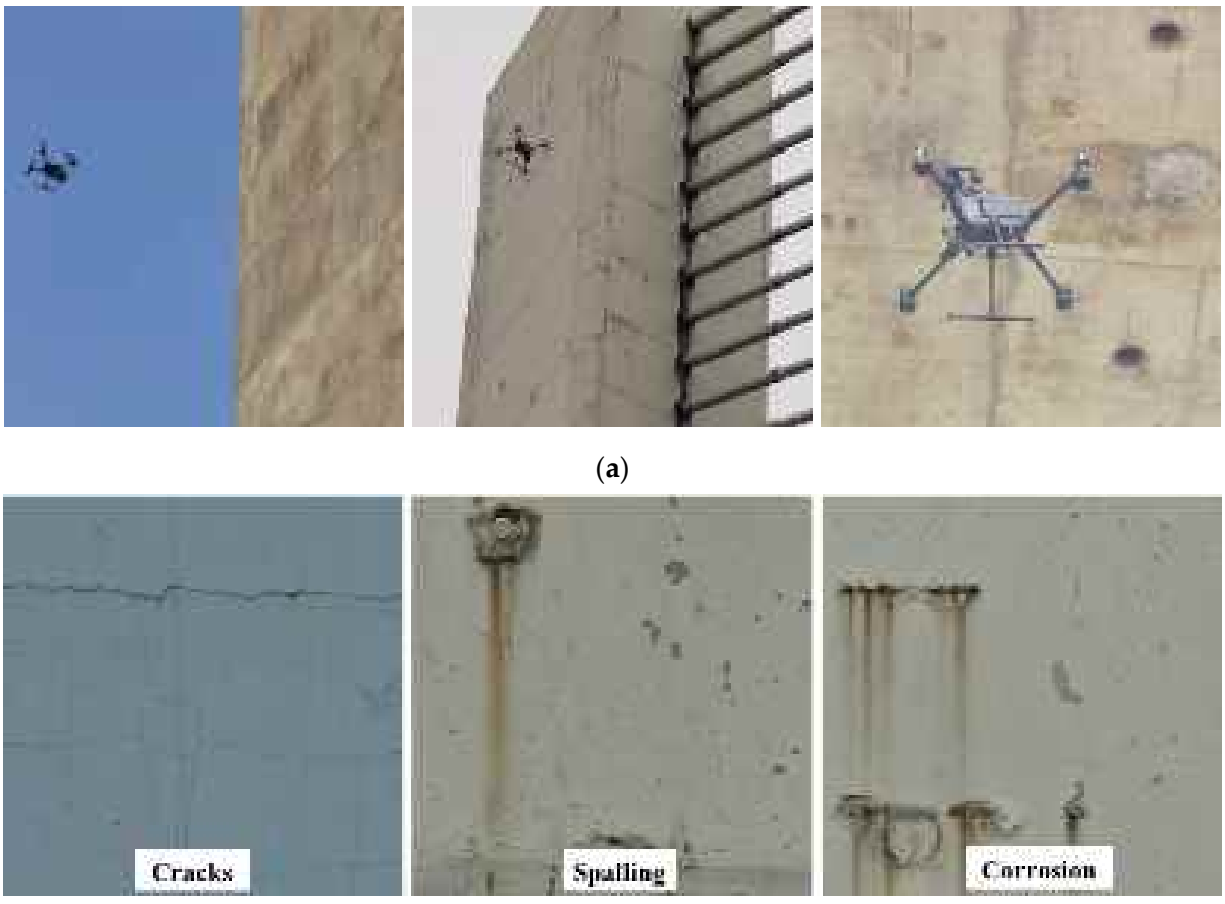


Figure 4. The structure diagram of the improved network.

2.3.3. Establishment of the Bridge Damage Dataset

To make the dataset used in the network training process as close as possible to the actual situation of the bridges targeted by the network deployment, the images in

the dataset were taken from images of typical damages on several bridges captured by UAVs or digital cameras. The types of damage include three categories: concrete cracks, spalling, and exposed internal reinforcement due to spalling. The bridges from which the images were collected and typical damage images are shown in Figure 5. A 800D camera, from Canon Company, Tokyo, Japan, a M200 UAV and a M300 UAV from DJI Company, Shenzhen, China, were used for data acquisition, and 500, 1000, and 2000 images captured with the three devices, respectively, were selected for manual labeling. Considering the different resolutions of the three devices, the acquired images were first cropped and resized to the same 1280 pixels as the input size of the network. Finally, 3000 images containing the damage were acquired. Several weather conditions, including sunny, cloudy, and foggy days, were also selected to acquire bridge images with different exposures, considering that different weather conditions result in different levels of image exposure and shadow masking.



(**a**)

(**b**)

Figure 5. Establishment of a dataset for typical bridge damage. (**a**) Image acquisition of surface defects on three in-service bridges using UAVs. (**b**) Images of three typical types of bridge damage.

In addition to the 3000 original images containing defects, we expanded the dataset in order to enhance the generalization ability of the trained network through applying data augmentation techniques to introduce random disturbances such as random noise and occlusion. The augmented dataset thus comprised 9000 images with defects. Finally, the dataset was partitioned into training, testing, and validation sets in a ratio of 7:2:1, respectively.

3. In-Service Bridge Damage Detection

To test the practicality and accuracy of the proposed method, it was tested on a three-span concrete bridge using UAV-based data acquisition and the 3D reconstruction of multi-view images, followed by training and testing of the improved network.

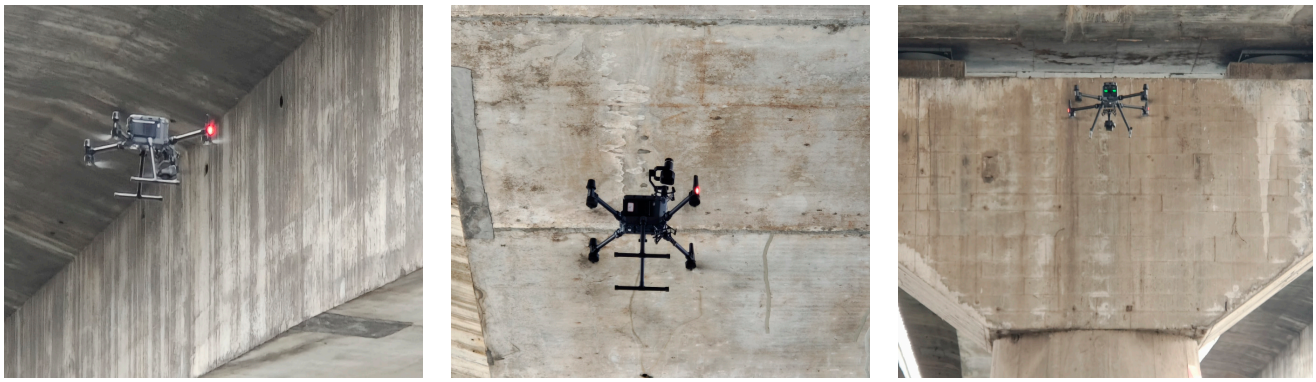
3.1. Overview of the Tested Bridge

We tested a concrete bridge on an urban circular highway with a length of 170 m and a main span of 70 m. The bridge consists of four closely spaced bridges, each of which is 13 m wide, for a total width of 60 m. The bridge crosses a river that is approximately 100 m wide, with an underbelly clearance of 15 m. The bridge has been in service for over 20 years and presents a series of surface defects. Previous manual-based inspection methods showed that the bridge was subjected to long-term loading, deterioration of concrete properties, and loss of prestressing, resulting in extensive surface cracking, concrete spalling, and corrosion of reinforcing steel, some of which were repaired and filled in early on, but some new defects have continued to appear ever since. As the four bridges are closely spaced, and their width exceeds the length of the bridge inspection vehicle's cantilever, the inspection vehicle could not inspect the bottom of the main span. An overview of the bridge is shown in Figure 6.

Figure 6. Overview of the tested bridge.

3.2. Data Collection and 3D Reconstruction

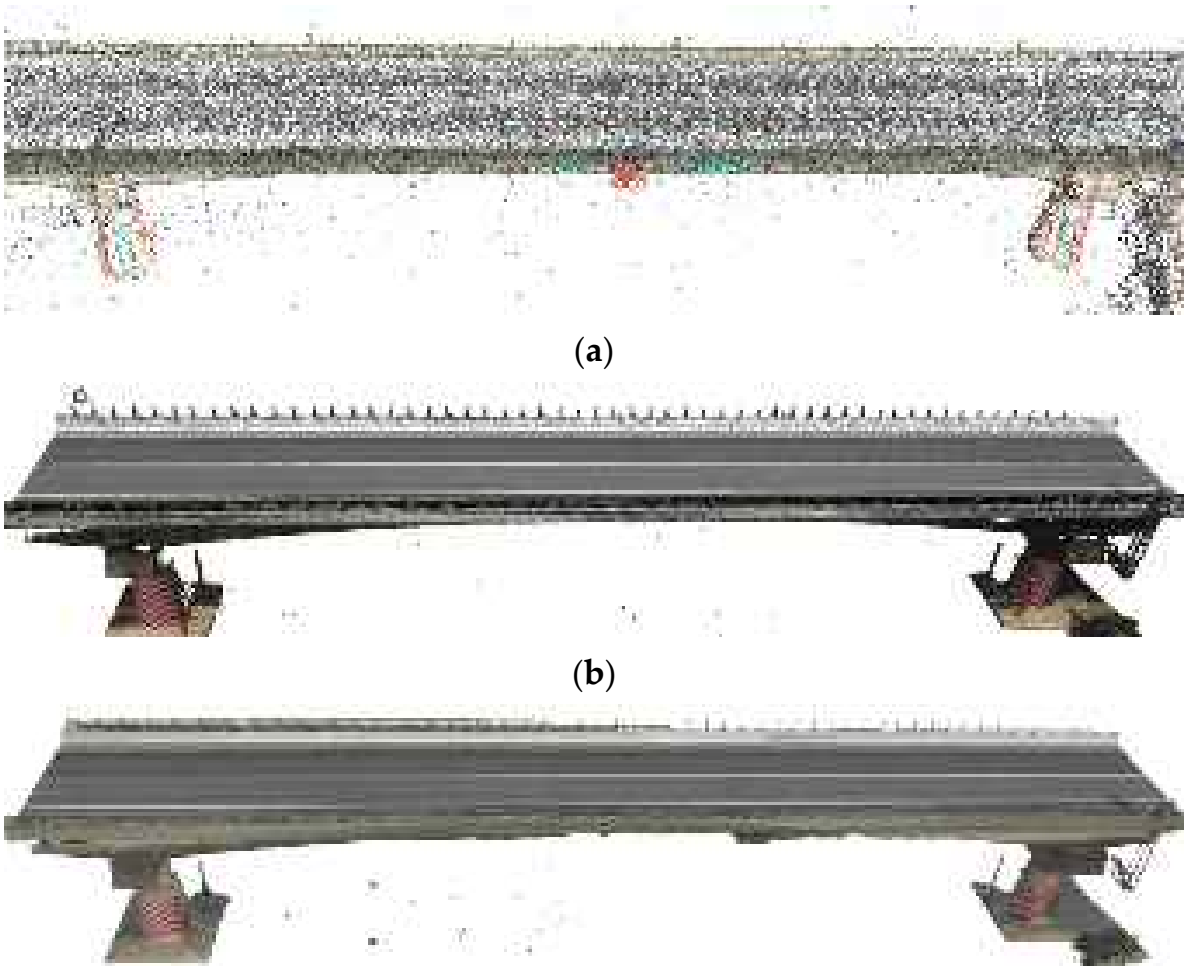
The photography method proposed in Section 2.1 was first used to capture the bridge surface; specifically, three photography methods were used, according to the three main bridge sections: the deck, bottom, and piers. As there were no obstacles around the bridge, the same multi-angle photography method as the aerial photogrammetry was adopted for the deck; that is, five angles of vertical down, 45° tilt-forward, tilt-backward, tilt-left, and tilt-right were photographed at each position. Using a planning tool set, the UAV's route was directly planned to cover the entire bridge with several routes. The overlap of the photos was set to 70%. For the bottom of the bridge and the piers, given the limited clearance height of the bottom, the obstacles, and the lack of GPS signals, the photographs were taken in a completely manual control mode. The photographs of the bottom were similar to those of the deck and were also taken from multiple angles according to multiple parallel routes. The UAV maintained a safe distance of about 5 m from the bottom of the bridge. For the piers, encircling capture was carried out; that is, the UAV flew and took photos around the piers at multiple heights and angles. The photography process of the UAV is shown in Figure 7. The UAV and camera used in the tests are consistent with those introduced in Section 2.1.



(**a**) (**b**) (**c**)

Figure 7. Photography schematic diagram of different sections of the bridge using a UAV. (**a**) Bridge side. (**b**) Bridge bottom. (**c**) Bridge pier.

The data collection process took approximately 1.5 h, resulting in a total of 1608 surface images of the target span. Utilizing the 3D reconstruction method proposed in Section 2.2, the original images were reconstructed, yielding a sparse 3D point cloud, a dense point cloud, and a 3D bridge model, as depicted in Figure 8.



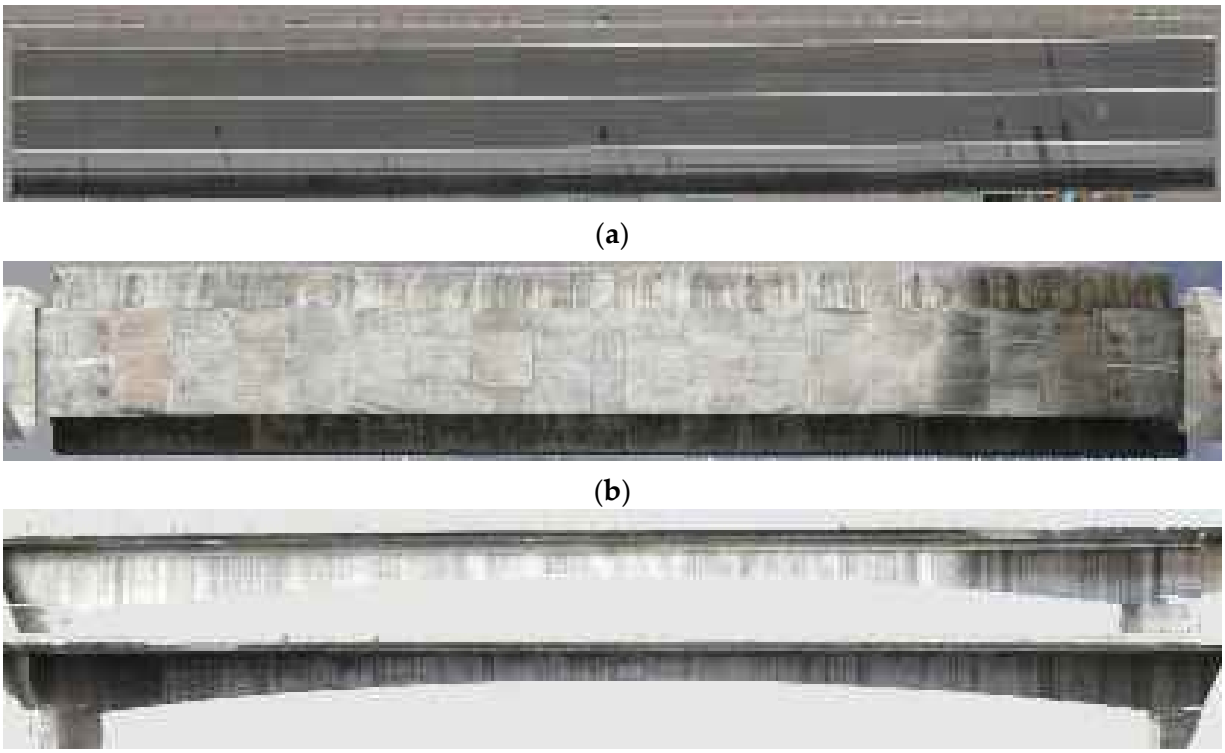
(**a**)

(**b**)

(**c**)

Figure 8. 3D reconstruction results for the tested bridge. (**a**) Sparse point cloud of the bridge. (**b**) Dense point cloud of the bridge. (**c**) 3D model of the bridge.

The model was further projected from multiple directions to panoramically unfold the bridge deck, sides, and bottom, as shown in Figure 9. The resolution of the bridge deck panorama was 26,715 × 5397 pixels, with a true resolution of 3.25 mm/pixel; the resolution of the bottom panorama was 76,241 × 13,011 pixels, with a true resolution of 1.26 mm/pixel; and the resolution of the side panorama was 77,079 × 11,605 pixels, with a true resolution of 1.19 mm/pixel. These results demonstrate that the proposed UAV-based photography method can be used to obtain a complete 3D bridge model in a global range and a panoramic unfolding map at the millimeter scale.



(**a**)

(**b**)

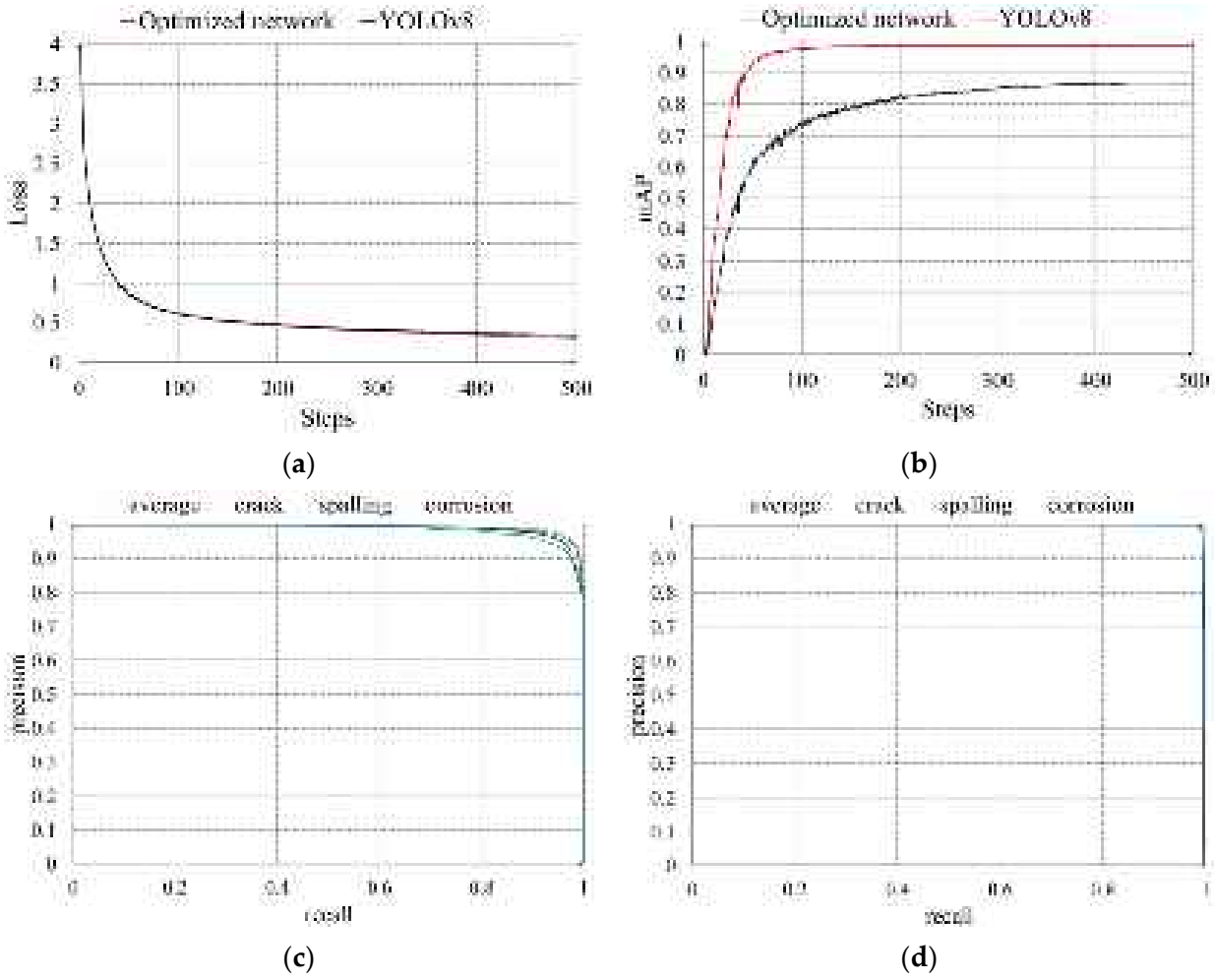
(**c**)

Figure 9. Panoramic unfolded view of the tested bridge. (**a**) Bridge deck. (**b**) Bridge bottom. (**c**) Bridge side.

3.3. Bridge Damage Identification Results

After obtaining the panoramic unfolding maps of each face of the bridge, the next step was to automatically identify damage in the panoramic unfolding maps using the method described in Section 2.3. For the improved network to recognize bridge damages, it was first necessary to train the network on the typical damage dataset. To verify that the designed network had a higher detection accuracy than the original network, both the designed network and the original YOLOv8 network were trained on the created dataset, and the final accuracy achieved by both networks was compared. Because the YOLOv8 network has four scales, from small to large, and a larger scale means more layers in the network and higher accuracy, the largest scale of YOLOv8-p6-x was chosen as the base network.

Network training was performed on a computer equipped with an Nvidia RTX3090 GPU, an i7 13,700 K CPU, and 64 GB of RAM, and the framework used for training was Ultralytics. The parameters set during network training were as follows: The optimizer was set to istochastic gradient descent, the base learning rate was set to 0.01, the base weight decay was 0.0005, the optimizer momentum was 0.937, the batch size was set to 16, the learning rate schedule was set to linear, the number of training epochs was 500, and the input size was 1280 × 1280 pixels. The parameter variation curves of the two networks during training are shown in Figure 10.



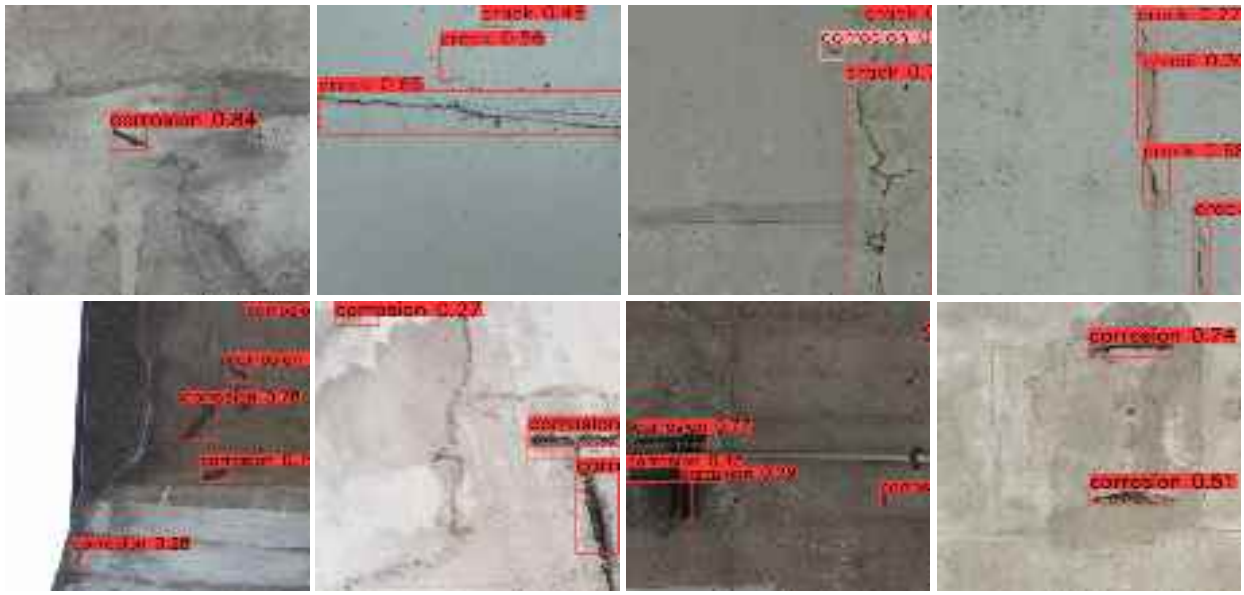
(**a**) (**b**)

(**c**) (**d**)

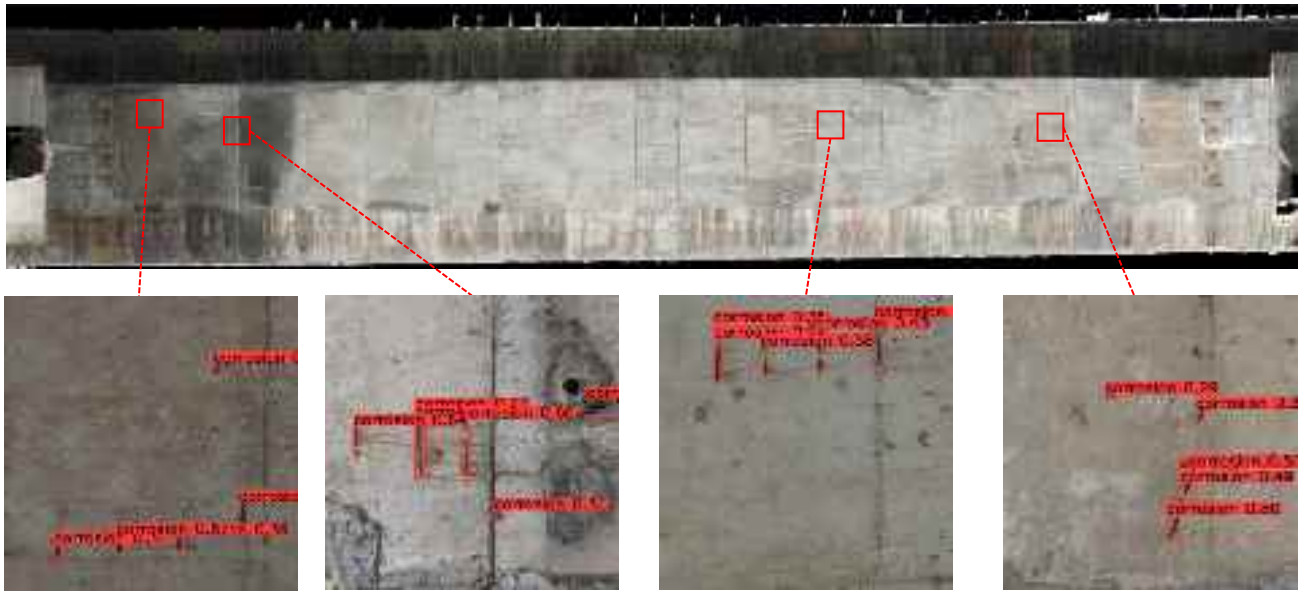
Figure 10. Parameter variation curves of the two networks during training. (**a**) Loss curves of the two networks. (**b**) mAP curves of the two networks. (**c**) P-R curves of YOLOv8. (**d**) P-R curves of the optimized network.

The loss curve and accuracy curve trends of the training process show that both networks were correctly trained and eventually stabilized. After stabilization, the original YOLOv8 network achieved a mAP of 0.869, with accuracies of 0.897 for cracks, 0.808 for spalling, and 0.903 for rebar exposure. The improved network achieved a mAP of 0.987, with accuracies of 0.990 for cracks, 0.981 for spalling, and 0.991 for rebar exposure. These results show that the improved network improved the average accuracy of damage detection relative to the original network by 13.6%, proving that the improved network performs better. Additionally, the precision–recall curves of the two networks indicate that the improved network exhibited higher precision.

After the network was correctly trained, it was applied for damage detection in the panoramic bridge images. Because the panoramic images had high resolution, and it was thus not possible to input them directly into the network, a sliding window with the same size as the network input was used to sequentially slide and segment the panoramic images to obtain a sequence of sub-images. These sub-images were input into the network for damage detection. The recognition results on the sub-images are shown in Figure 11a, and after all the sub-images were recognized correctly, they were stitched together in sequence to obtain a panoramic image of the bridge with damage markers, as shown in Figure 11b. Using this panoramic image, it is possible to quickly count the number of damage points on the bridge and to analyze the distribution density of the damages in each section, providing a basis for analysis of the health status of the bridge.



(**a**)



(**b**)

Figure 11. Damage detection results in sub-images and panoramic images. (**a**) Damage detection results for sub-images cut by sliding windows in panoramic images. (**b**) Damage detection in panoramic image of bridge bottom.

4. Conclusions and Future Work

This study proposed a bridge damage detection method based on UAV photography, multi-view 3D reconstruction, and deep learning-based object detection. Specifically, we used close-range bridge surface images captured by a UAV from multiple angles in a covering manner as the data source, which allowed for the establishment of a refined 3D model of the bridge that could be projected to obtain a panoramic, unfolding image of the bridge surface. We then adopted a deep learning-based object detection method to rapidly and automatically identify typical damage on the global scope of the bridge using the obtained panoramic map. The proposed method contains the whole process from data acquisition to preprocessing and damage identification, and the method can be used as a guide in the inspection of similar bridges to enhance the inspection efficiency and reduce

the cost. In addition, unlike existing methods, the proposed method does not directly use the acquired original images for damage detection but reconstructs a large number of fragmented images into a complete panoramic view of the bridge surface, which avoids omissions and enables the location of the damage to be known from the panoramic view. Specific conclusions are as follows:

(1) To quickly acquire detailed, high-resolution surface images, we propose using a UAV to photograph bridges from multiple angles in close range, providing overlapping images with full coverage of the bridge. The results of photographing in-service bridges proved that the proposed method allows millimeter-level images of complete bridge surfaces to be acquired, which can be used for damage analysis;

(2) To generate panoramic images of bridge surfaces, we used multi-view stereo algorithms to reconstruct a 3D model of the bridge based on a large number of original images. The bridge was then panoramically unfolded through projection of the model. The test results proved that the method can panoramically unfold the bridge surface at the millimeter scale, and the resolution of the resulting panoramic image can be up to hundreds of millions of pixels. In the tested bridge, the resolution of the panorama at the bottom of the bridge was 1.26 mm/pixel, and the size of the image amounted to 76,241 $\times$ 13,011 pixels;

(3) To automatically detect damage, the state-of-the-art YOLOv8 network was improved using a Swin Transformer. The results of training on a bridge damage dataset and testing on data obtained from an in-service bridge demonstrated that the improved network achieved a 13.6% increase in bridge damage recognition accuracy when compared with the original network, reaching 98.7% in the bridge test. After damage identification, the panorama presents the location of each instance of damage on the bridge surface.

The proposed damage identification method is expected to supplement or even replace existing manual-based damage detection methods for many similar concrete bridges, but drawbacks and limitations of the present method have also been identified in its application, including non-automatic data acquisition, limitations on the types of bridges to be detected, and long inspection times. First, due to the influence of weak GPS signals and multiple obstacles at the bottom of bridges, the UAV still needs to rely on the control of skilled manipulators when performing inspections at the bottom of bridges, and such manipulation skills are not possessed by most workers, which makes it difficult to popularize the use of UAV-based bridge damage inspections in a large number of applications. With the extensive research on automated UAV-based inspection techniques in GPS-constrained environments (e.g., indoors), it has become possible to make UAVs fly automatically in such environments, so future work will investigate the use of sensors such as LIDAR, stereo vision, and simultaneous localization and mapping (SLAM)-based algorithms to make UAVs fly automatically at the bottom of bridges and get rid of the dependence on manual control. Secondly, the current proposed method was only trained and tested on concrete bridges, but the applicability of the proposed method could be expanded through the consideration of more bridge types (e.g., steel bridges) and damage types (e.g., corrosion and cable damage). In addition, since 3D reconstruction-based methods require a large amount of computational time, the currently proposed method cannot achieve fast solving and output of the detection results. Investigating deep learning-assisted 3D reconstruction methods, which utilize learning-based methods to directly match images and generate depth maps, will expectedly lead to a reduction in computational time, and this work will also be a future focus of research.

Author Contributions: T.Y., conceptualization, methodology, validation, writing—original draft, and writing—review and editing; G.S. (Guodong Shen), conceptualization and validation; L.Y., validation and writing—review and editing; G.S. (Guigang Shi), methodology, supervision, resources, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The research presented was financially supported by the Key technology projects in the transportation industry (No.: 2021-MS1-055) and Anhui provincial natural science research project—major project (No.: KJ2019ZD53).

Data Availability Statement: The data supporting this study's findings are available from the corresponding author upon reasonable request.

Conflicts of Interest: Authors Tao Yin, Guodong Shen and Liang Yin were employed by the company Anhui Transport Consulting and Design Institute Co., Ltd. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yang, J.; Guo, T.; Li, A. Experimental investigation on long-term behavior of prestressed concrete beams under coupled effect of sustained load and corrosion. *Adv. Struct. Eng.* **2020**, *23*, 2587–2596. [CrossRef]
2. Bonopera, M.; Chang, K.C. Novel method for identifying residual prestress force in simply supported concrete girder-bridges. *Adv. Struct. Eng.* **2021**, *24*, 3238–3251. [CrossRef]
3. Hou, S.; Dong, B.; Wang, H.; Wu, G. Inspection of surface defects on stay cables using a robot and transfer learning. *Autom. Constr.* **2020**, *119*, 103382. [CrossRef]
4. Wang, Y.; Zhang, X.; Zhang, M.; Sun, L.; Li, M. Self-compliant track-type wall-climbing robot for variable curvature facade. *IEEE Access* **2021**, *10*, 51951–51963. [CrossRef]
5. Qin, Y.; Dong, S.; Pang, R.; Xia, Z.; Zhou, Q.; Yang, J. Design and kinematic analysis of a wall-climbing robot for bridge appearance inspection. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2021; Volume 638, p. 012062. [CrossRef]
6. Jang, K.; Kim, N.; An, Y.K. Deep learning–based autonomous concrete crack evaluation through hybrid image scanning. *Struct. Health Monit.* **2019**, *18*, 1722–1737. [CrossRef]
7. Tang, Z.; Peng, Y.; Li, J.; Li, Z. UAV 3D Modeling and Application Based on Railroad Bridge Inspection. *Buildings* **2023**, *14*, 26. [CrossRef]
8. Cataldo, A.; Roselli, I.; Fioriti, V.; Saitta, F.; Colucci, A.; Tatì, A.; Ponzo, F.C.; Ditommaso, R.; Mennuti, C.; Marzani, A. Advanced video-based processing for low-cost damage assessment of buildings under seismic loading in shaking table tests. *Sensors* **2023**, *23*, 5303. [CrossRef]
9. Li, R.; Yu, J.; Li, F.; Yang, R.; Wang, Y.; Peng, Z. Automatic bridge crack detection using Unmanned aerial vehicle Faster, R.-C.N.N. *Constr. Build. Mater.* **2023**, *362*, 129659. [CrossRef]
10. Li, J.; Li, X.; Liu, K.; Yao, Z. Crack identification for bridge structures using an unmanned aerial vehicle (UAV) incorporating image geometric correction. *Buildings* **2022**, *12*, 1869. [CrossRef]
11. Zhang, A.; Wang, K.C.; Fei, Y.; Liu, Y.; Tao, S.; Chen, C.; Li, J.Q.; Li, B. Deep learning–based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet. *J. Comput. Civ. Eng.* **2018**, *32*, 04018041. [CrossRef]
12. Ni, F.; Zhang, J.; Chen, Z. Zernike-moment measurement of thin-crack width in images enabled by dual-scale deep learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 367–384. [CrossRef]
13. Han, X.; Chang, J.; Wang, K. You only look once: Unified, real-time object detection. *Procedia Comput. Sci.* **2021**, *183*, 61–72. [CrossRef]
14. Jiang, S.; Zhang, J.; Wang, W.; Wang, Y. Automatic inspection of bridge bolts using unmanned aerial vision and adaptive scale unification-based deep learning. *Remote Sens.* **2023**, *15*, 328. [CrossRef]
15. Jiang, S.; Wu, Y.; Zhang, J. Bridge coating inspection based on two-stage automatic method and collision-tolerant unmanned aerial system. *Autom. Constr.* **2023**, *146*, 104685. [CrossRef]
16. Bao, Q.; Xie, T.; Hu, W.; Tao, K.; Wang, Q. Multi-type damage localization using the scattering coefficient-based RAPID algorithm with damage indexes separation and imaging fusion. *Struct. Health Monit.* **2023**, *23*, 1592–1605. [CrossRef]
17. Xu, Y.; Bao, Y.; Chen, J.; Zuo, W.; Li, H. Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. *Struct. Health Monit.* **2019**, *18*, 653–674. [CrossRef]
18. Hu, F.; Zhao, J.; Huang, Y.; Li, H. Structure-aware 3D reconstruction for cable-stayed bridges: A learning-based method. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 89–108. [CrossRef]
19. Ma, Z.; Liu, S. A review of 3D reconstruction techniques in civil engineering and their applications. *Adv. Eng. Inform.* **2018**, *37*, 163–174. [CrossRef]
20. Pintore, G.; Mura, C.; Ganovelli, F.; Fuentes-Perez, L.; Pajarola, R.; Gobbetti, E. State-of-the-art in automatic 3D reconstruction of structured indoor environments. *Comput. Graph. Forum* **2020**, *39*, 667–699. [CrossRef]
21. Schönberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
22. Li, Z.; Shan, J. RANSAC-based multi primitive building reconstruction from 3D point clouds. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 247–260. [CrossRef]
23. Dundar, A.; Gao, J.; Tao, A.; Catanzaro, B. Fine detailed texture learning for 3d meshes with generative models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 14563–14574. [CrossRef] [PubMed]

24. Padilla, R.; Netto, S.L.; Da Silva, E.A. A survey on performance metrics for object-detection algorithms. In Proceedings of the 2020 IEEE International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242. [CrossRef]
25. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
26. Yu, Z.; Shen, Y.; Shen, C. A real-time detection approach for bridge cracks based on YOLOv4-FPM. *Autom. Constr.* **2021**, *122*, 103514. [CrossRef]
27. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
28. Zhang, Y.; Liu, C. Network for robust and high-accuracy pavement crack segmentation. *Autom. Constr.* **2024**, *162*, 105375. [CrossRef]
29. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430. [CrossRef]
30. Feng, C.; Zhong, Y.; Gao, Y.; Scott, M.R.; Huang, W. Tood: Task-aligned one-stage object detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; IEEE Computer Society. pp. 3490–3499. [CrossRef]
31. Patil, S.M.; Raut, C.M.; Pande, A.P.; Yeruva, A.R.; Morwani, H. An efficient approach for object detection using deep learning. *J. Pharm. Negat. Results* **2022**, *13*, 563–572. [CrossRef]
32. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef]
33. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.