KSCE Journal of Civil Engineering (2023) 27(3):1150-1165
DOI 10.1007/s12205-023-0522-1
pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Structural Engineering

# Detection and Location of Steel Structure Trestle Surface Cracks Based on Consumer-grade Camera System

**Chunbao Xiong**[a], **Sida Lian**[a], and **Wen Chen**[a,b]

[a]School of Civil Engineering, Tianjin University, Tianjin 300072, China
[b]College of Science and Technology, Agriculture University of Hebei, Baoding 071001, China

**ARTICLE HISTORY**

Received 31 March 2022
Revised 27 June 2022
Accepted 7 December 2022
Published Online 21 January 2023

**KEYWORDS**

Structural damage identification
Semantic segmentation
Crack localization
Computer vision
Steel structure trestle

**ABSTRACT**

Because the steel structure trestle has been in service under heavy load for a long time, the steel structure trestle is prone to cracks around the welds or bolt holes, which can lead to structural collapse in severe cases. Aiming at the characteristics of stable and high-quality images obtained by the unmanned consumer-grade camera monitoring system, this paper proposed structure health monitoring (SHM) system which is based on consumer-grade camera. The SHM system can identify crack damage and locate steadily in long term, which provides the technical support of practical application in intelligent SHM system. The method first performed edge detection on the trestle structure, followed by pixel-level semantic segmentation and crack localization. Canny edge detection algorithm was used to identify trestle structures in the camera image. The panorama trestle structure was divided into areas of suitable size, and the camera focused on each divided area one by one. Then the improved DeepLab V3+ model was trained by constructing global and local datasets. Then the improved DeepLab V3+ model was used to perform pixel-level semantic segmentation on the trestle images of the divided regions. Finally, based on the Speeded Up Robust Features and combined with the image, a panorama crack location output method was proposed. The system was used to test a section of a trestle in a coal mining industrial park, and the system showed that the method could efficiently and accurately identify and locate the crack damage.

## 1. Introduction

Because of its advantages of light weight and large payload capacity, the steel structure trestle has been widely used in various industrial parks, especially the mining industry. Being worked in a harsh environment, the steel structure trestle has also been in service under heavy load conditions for a long time. At the same time, since it is often used for long-distance load transmission, the excessively long structure length also makes it difficult to maintain the trestle structure, which brings a great burden to manual visual inspection. In summary, steel structure trestle is prone to cracks during long-term service, especially at the bolt joints and weld locations where the force is concentrated. These cracks adversely affect the safety and structural life of steel structure trestle. If these cracks cannot be discovered and dealt with in time when the cracks become worse, it will eventually cause great loss of lives and properties. Therefore, it is necessary to carry out SHM on steel structure trestle. In recent years, with the increase in the service life of civil engineering and construction and the improvement of people's safety awareness, SHM has attracted attention in a growing number. Through real-time monitoring of structure response and service status, the current status of the structure can be quantitatively evaluated. SHM can realize real-time continuous observation of the service status of the structure, thereby improving the safety and reliability of the service, and reducing the cost of maintenance and inspection of the structure (Li et al., 2016; Dizaji et al., 2021; Ngeljaratan et al., 2021).

At present, the crack damage identification of civil engineering buildings such as steel structures mainly relies on manual visual inspection. This detection method also requires the detection workers to have relevant knowledge. Moreover, it's a subjective traditional method relying on the empiricism (Li et al., 2016). This method depends on manual visual inspection, when it is

CORRESPONDENCE Sida Lian liansida@tju.edu.cn School of Civil Engineering, Tianjin University, Tianjin 300072, China

difficult for construction workers to operate or they are reluctant to work on the structures, it will cause omission of detail cracks. This will make the test results incomplete and cause safety hazards. At the same time, this method poses hidden dangers to the life safety of workers and results in low efficiency and high cost (Samantaray et al., 2018).

As the computer technology and deep learning (DL) algorithms continuously develop in recent years, research on intelligent algorithms has once again become a hot issue. The development of computer hardware effectively supports the powerful computing power required by the algorithm. At the same time, the algorithm is constantly updated and improved, gradually improving many of its own shortcomings. Computer vision (CV) technology has been recognized as a key method for improving and developing civil engineering structure detection and monitoring technology recently. To a certain extent, it can allow computers to replace the human to perform damage detection for structures. Image processing methods based on CV have begun to be used in the recognition of local damage in steel structures. A large number of scholars have done a lot of research on the application of CV technology to SHM, Je-Keun Oh a et al. proposed a robot vision system for checking the safety status of bridges, which can automatically photograph and monitor the safety of bridge deck structures through robots (Oh et al., 2009). Lim et al. proposed a crack detection system that used the Laplacian of Gaussian (LoG) algorithm to detect cracks, and obtained a global crack map through camera calibration and robot positioning (Lim et al., 2014). Prasanna et al. proposed a novel automatic crack detection algorithm STRUM (Spatially Adjusted Robust Multiple Features) classifier, and demonstrated the results of real bridge data using the latest robotic bridge scanning system (Prasanna et al., 2016). Prasanna et al. proposed a CV algorithm in terms of the damage characteristics of cracks and spalls (Jahanshahi et al., 2016). A Miyamoto proposed an automatic bridge SHM system that combined information and communication technology to manage the life cycle of existing short and medium span bridges (Miyamoto et al., 2019). Yang et al. proposed an explicit modeling and they used data structures to solve structural dynamics problems, on which real-time structural safety is monitored through video (Nagarajaiah and Yang, 2017). The above classic CV studies depend on images taken under ideal conditions, with less interference around the damage and better recognition of obvious cracks in the image, but their recognition effect is relatively limited under complex shooting environments. It is easy to be disturbed by the complicated background, so that the location of the crack cannot be accurately identified, or all pixel features of cracks cannot be extracted completely.

Thanks to the robust development of DL algorithms in recent years, many advanced algorithms based on DL have been proposed in the field of CV. Convolutional neural network (CNN) and DL algorithm have been gradually applied by many scholars in the field of civil engineering SHM due to their wide applicability (Simonyan and Zisserman, 2014; Lecun et al., 2015; Wu et al., 2017). For example, Hoskere et al. introduced a novel engineering application of SHM in a new type of civil infrastructure. The method set a new binomial loss function to improve the accuracy of the training network and used Monte Carlo to convey the uncertainty of the model (Spencer et al., 2019). Rafiei et al. collected environmental vibration response of the structure by the sensor to evaluate the global and local health status of the structure system (Rafiei and Adeli, 2018). The model combines synchronously compressed wavelet transform, fast Fourier transform and unsupervised deep Boltzmann machine. Bao et al. proposed a random decrement technique which allows estimation of modal parameters without directly measuring input, and applied this technique to structural modal analysis and SHM (Bao et al., 2019). DF Karypidis et al. used a distributed optical fiber system to monitor the strain distribution along the steel bar, based on the analysis of the natural frequency of data obtained from the accelerometer (Karypidis et al., 2019). The preliminary results of the study show that the semi-supervised Deep Auto-encoder algorithm (DAE) can successfully quantify the failure of transverse cracks in reinforced concrete beams subjected to three-point loads. Nahata et al. proposed an autonomous damage detection model based on CNN. Under the application of the VGG16 transfer learning model, the best results are obtained with a learning rate of 1e-5, and the final training accuracy reached 97.85% and 89.38% (Nahata et al., 2019). Han et al. achieved good application results in cracks identification and location of a large steel structure by the images obtained by UAV simulated inspection, which provided a good reference for engineering applications (Han et al., 2022). Zhang et al. systematically summarizes the recent research and application of DL-based CV technology in the field of damage detection, and discusses the problems that need to be solved and future research directions (Zhang et al., 2022). In addition, many scholars applies CV technology to structural damage recognition in various directions, and achieves obvious application effects (Cha et al., 2018; Rubio et al., 2019; Sen et al., 2019; Yeum et al., 2019; Le et al., 2021).

However, the rapid development of many CV theories has not effectively promoted their practical engineering applications, and there is still a large gap between them. The surface cracks of the structure are difficult to detect in time by manual visual inspection. With the development of consumer-grade camera technology, the cost is low and the image quality is getting better and better. It can obtain images stably and clearly, and at the same time, the remote control rotation technology which means making the camera rotate horizontally by remote control, so that a single camera can monitor the structure of a large area. Camera monitoring can carry out long-term and real-time SHM of civil engineering structures in a large area under the premise of a small number of cameras. It is especially suitable for the steel structure trestle whose structural length is often too long, which can effectively reduce the monitoring cost and improve the efficiency and accuracy.

Most of the current research is aimed at materials with relatively uniform color distribution, such as asphalt, concrete and other materials with large cracks (Bao et al., 2019; Jesus et al., 2019;

Hoskere et al., 2020; Shahbaznia et al., 2020; Sun et al., 2020; Han et al., 2021; Wan et al., 2022). Some progresses are made in crack damage identification of steel structure (Dorafshan et al., 2018; Yang et al., 2018; Dung et al., 2019; Dong et al., 2021), however these researches rarely considered the practical application in civil engineering of steel trestle structure SHM on which is focused in this study. Therefore, this paper proposed a DL-based CV integration method to perform real-time long-term SHM on trestle structures. Steel structure trestle is mostly used in coal mining transportation, industrial raw material transportation and other industrial parks. The service scene where the trestle structure is located (mostly used in mountain mining areas or the trestle is built high) is not conducive to the arrangement of sensors. The no-fly zone of the industrial park cannot use drones for SHM, and the overly long structure is also not conducive to regular manual inspections. According to the characteristics of the trestle service environment, this paper proposed a trestle SHM method with low cost, simple operation and automatic performance. The integrated method can quickly and automatically perform long-term and uninterrupted SHM on the surface cracks of the trestle structure in real time without relying on the arrangement of a large number of sensors. The DL-based CV ensemble method in this paper uses a consumer-grade camera system to perform long-term SHM for the entire life cycle of the trestle structure, avoiding inconvenient regular inspections and low cost. It can detect the surface cracks at the concentrated stress of the trestle structure in the early stage of damage to the greatest extent. The set of SHM systems considered how to identify the tiny cracks in the distance, as well as how to prevent interference of rust background to crack identification. The camera monitoring system proposed in this paper provided a technical reference for the application of intelligent SHM of steel trestle structure. The system first used the Canny edge detection algorithm to extract the structural boundary of the partial image obtained by the camera. The image within the boundary of the trestle structure was the only focus, the extracted structural image was divided into several small areas of appropriate size, and then the camera focused on the small areas one by one. 10 improved DeepLabv3+ models were trained by using two datasets, global and local datasets, and after training these models operate joint decision-making. The model was built on the backbone of ResNet18, and a parallel channel attention mechanism was introduced, which effectively eliminated confounding factors such as rust and welds. The improved CV algorithm can perform pixel-level semantic segmentation of cracks more efficiently and accurately. Finally, SURF-based panoramic image stitching and path iterative scanning were used to locate cracks and the approximate location of cracks on the steel structure trestle. This set of SHM systems has been experimentally verified on a long-term service steel structure trestle. The flow chart of SHM of steel trestle structure by using the camera system is as follows.

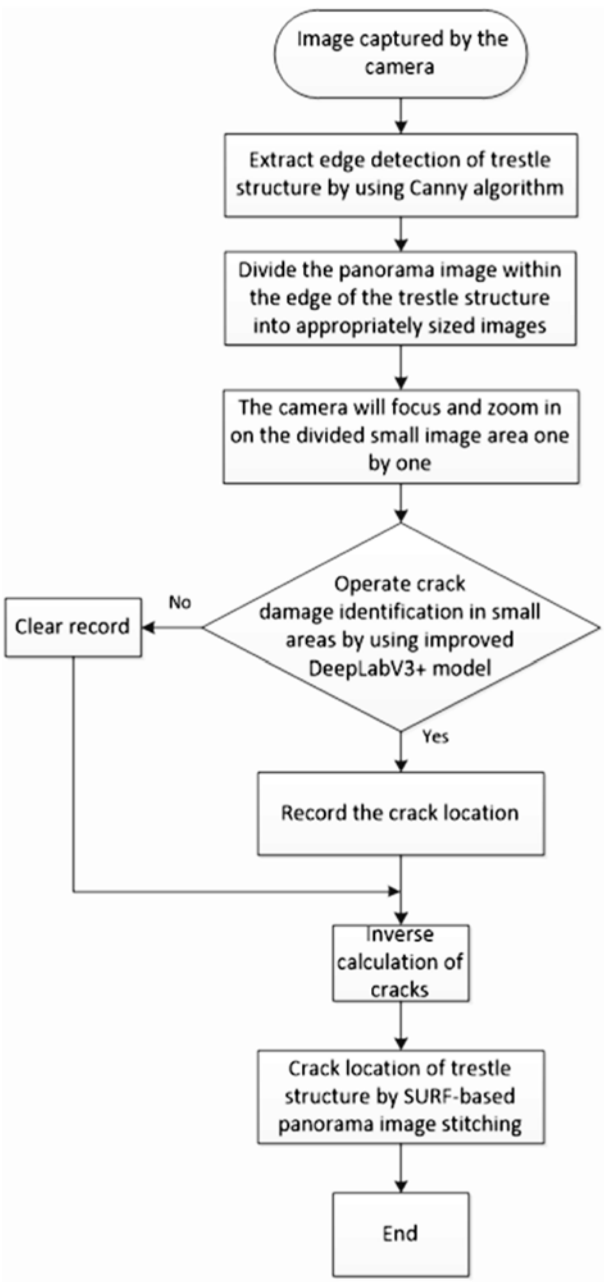


**Fig. 1.** Camera-Based SHM Process of Trestle Structure

## 2. Crack Identification Operation Process

### 2.1 Panorama Preprocessing

The edge of the trestle structure is extracted by using the Canny algorithm. The Canny edge detection algorithm is a multi-level edge detection algorithm. In general, the purpose of edge detection is to significantly reduce the data size of the image while preserving the original image properties. The Canny algorithm is a standard algorithm for edge detection and is still widely used in research.

Firstly, Gaussian filtering was used to remove background interference with fast changing frequency, and then Canny algorithm was used to identify the trestle structure area. It can be seen from Fig. 2 that after the above operations, the Canny algorithm can identify the boundary of the trestle structure. After

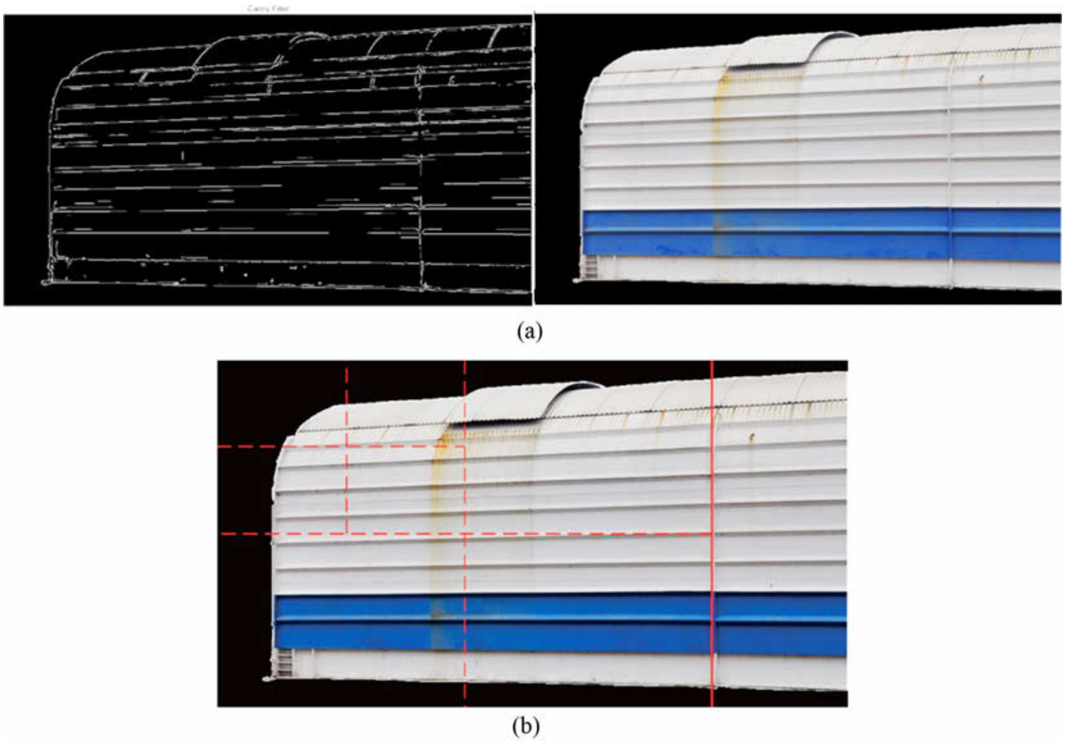


**Fig. 2.** Division of Trestle Panoramic Image: (a) Image Extraction within the Boundary of the Trestle Structure, (b) Image Area Division Method

the panoramic edge of the trestle structure is extracted, only focusing on the image within the boundary of the target structure, it is divided into small images of suitable size, which is convenient for further processing by the CV algorithm. For the trestle structure in this paper, the area division method is shown in Fig. 2. The specific division method was based on the size of the monitored trestle structure, the distance between the camera and the connecting structure, and the pixel details of the obtained image. The structure division method will be different according to different scales. For example, super-pixel segmentation and other methods should be used for larger structures. However, the trestle structure in this study is not very large, and it is more efficient to directly divide the area for crack identification. Specifically, as shown in Fig. 2, each section of the trestle is divided into 3 major sections, and each section is subdivided into 4 to 64 uniform areas (the specific division size is determined by the monitoring accuracy).

Compared with the overall structure of the trestle, the cracks may be very small. At present, in the application of crack identification and location engineering for large and complex structures, in order to identify cracks in distant images, some clustering analysis algorithms are often used to divide the panoramic image into several small areas, such as super-pixel segmentation. Object detection methods are then used on these small areas to identify small areas where cracks may exist. However, this operation will increase the complexity of identifying preprocessing of the cracks and increase the computational cost. In the engineering application of this manuscript, since the camera is far away from the crack during panoramic photography, the crack damage is small, and the tested object detection method cannot detect the possible crack area. At the same time, each section of the steel structure trestle is not very huge, and the number of small areas divided by the panorama is relatively small.

Therefore, this paper directly zooms in and focused on each area after dividing the panorama area, and used the improved DeepLab V3+ model below to perform pixel-level semantic segmentation for this area to identify cracks. After testing, this operation allowed the algorithm to reduce the complexity of image preprocessing when identifying cracks. Under the premise of meeting the accuracy required for crack identification, it saved the calculation cost, and improved the efficiency of the algorithm for crack identification of trestles.

## 2.2 Improved Image Semantic Segmentation Algorithm

In actual engineering applications, the dataset is not very large. Therefore, this paper introduced the transfer learning strategy (Bao et al., 2019), and used the parameters of the machine learning model that has been initially trained as the initialization parameters of this research. The DeepLab V3+ (Zhang et al., 2021) model is built on the basis of the ResNet-18 (He et al., 2016) model, and based on this, the pixel-level segmentation of the cracks in the stress concentration of the steel truss structure was carried out. Most of the crack pictures which used in this paper are actual shots of a steel trestle structure that has been in service for many years. It is of great practical significance to simulate the long-term monitoring environment of the camera. After long-term service of the trestle structure, a large amount of shallow rust appeared on the surface, which caused great interference in the identification of cracks. Therefore, it is necessary to effectively eliminate these interferences when identifying the cracks in the structure. At the same time, the total number of crack-like pixels is very low compared with the number of background pixels. How to effectively extract these pixels also remains as an important issue.

If the training target is a complex network, firstly it will cause too long training time, and the computing power of computer hardware in engineering applications will be difficult to meet the requirements of complex networks. Secondly, the project instance dataset is relatively small, and the complex network is likely to be difficult to converge. To sum up, in view of the structural characteristics of the long-term service steel trestle and the actual requirements of engineering applications, this paper introduced a set of two weak classifiers to carry out collaborative learning and training, and achieve joint decision-making. The two classifiers performed global and local image learning and training respectively, so that the semantic segmentation algorithm can more accurately identify the crack damage of the steel trestle in the image.

In order to further eliminate the influence of the shallow corrosion around the cracks of the steel trestle structure, an improved parallel channel attention mechanism was introduced. It is based on the channel attention mechanism SE module (Qi et al., 2021). Thanks to the above operation, the convolutional neural network can learn the feature relationship in the channel dimension and record the fuzzy features of the feature map. By learning the feature information of the channel dimension in the network, the performance of image segmentation in complex scenes was improved, and the problem of automatic semantic segmentation of complex images was solved.

### 2.2.1 ResNet Model

With continuous deep research on neural networks, breakthroughs have been made in performance of deep convolutional networks. An increasing number of scholars use deep convolutional networks to solve image classification tasks. The deep convolutional network further forms a multi-level image feature fusion with multi-layer end-to-end training and learning, and the number of features can also be strengthened by continuous superposition. The puzzle of image segmentation in related fields can be better solved by deep convolutional networks, which illustrates the importance of deep convolutional networks. However, too many network layers will reduce the training efficiency of the network, and an overly complex network structure will greatly increase the demand for computing power. In severe cases, the gradient will disappear, and the algorithm will eventually fail to converge, but part of the convergence problem can be solved by regularization. What's more, the deep convolutional network still has the problem of network degradation. As the number of network layers increases, which means the depth increases, the accuracy of network training and verification no longer improves or even decreases. This is because excessive network layers will continuously lead to increased training error of the network. If a deep convolutional network is built based on a shallow network, all the training error of the deep network should be equal to that of the shallow network in this case when all the later added layers will be the same as the previous added layer. Therefore, the degradation problem of deep convolutional networks can be addresses by optimizing the network construction. In 2015, in order to solve the problem of deep network optimization, Kaiming et al. proposed the Deep residual networks (ResNet) (He et al., 2016). In the deep residual networks (ResNet) instead of letting the network directly fit the original mapping, it fit the residual mapping. After that, many scholars built networks such as ResNet50 and ResNet101 to solve the puzzles in the fields of image detection and classification, and achieved great application effects (Ahmed et al., 2020; Zhang et al., 2022). A simple identity mapping was realized by jumping connection, that is, by optimizing $F(x)$, such that $F(x) \to 0$, so that $x \to F(x) + x \to x$, and finally reaching an identity mapping relationship. The network structure of ResNet is shown in Fig. 3.

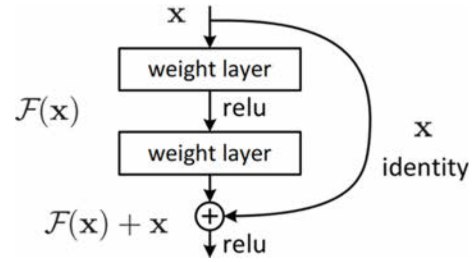


**Fig. 3.** Network Structure (He et al., 2016)

### 2.2.2 DeepLab V3+ Model Training

DeepLab (Lin et al., 2013) is a model proposed by Google that combines the advantages of deep convolutional networks (DCNNs) and traditional probabilistic graph models to obtain better classification processing results. DeepLab introduces atrous convolution and atrous spatial pyramid pooling (ASPP) structures. This improvement can effectively solve the problems of too small perception fields and reducing the details of the picture by down sampling. Atrous convolution obtains different receptive fields through different dilation rate parameter settings, in another words, it obtains multi-scale information, which plays an important role in visual tasks. Therefore, cavity convolution can expand the receptive field arbitrarily without introducing additional parameters. Nowadays many scholars have introduced cavity convolution into semantic segmentation research.

With continuous efforts, DeepLab semantic segmentation series models have successively launched DeepLab V1, DeepLab V2, DeepLab V3. Among them, DeepLab V3+ belongs to its latest model structure. DeepLab V3+ retains the original atrous convolution and ASSP layer communication, in addition to implements image multi-scale information fusion by introducing an encoder and decoder (encoder-decoder) structure. Using the Xception model as its basic network ensures the computational efficiency and robustness in processing image semantic segmentation problems. Outstanding application results have been achieved in many research fields.

### 2.2.3 Joint Decision-Making under Collaborative Learning

In practical engineering applications, because of being restricted by non-open engineering data, the data set cannot be too large, and the computer performance is also limited accordingly. Thus it is difficult to learn and train a relatively complex network. The above unfavorable factors can easily lead to the unsatisfactory training effect of the complex network, and the algorithm is prone to non-convergence. Therefore, this study formed a strong classifier by combining several weak classifiers into a cluster. Because of the complex structure of the neural network for image semantic segmentation, the numerous parameters and the complex training process, it has high requirements for hardware equipment. Therefore, serial training methods were used in the process of training weak classifiers, so that they can run better on computers with lower performance. The clusters of these serially trained weak classifiers were used to make joint decisions to reduce the need for computer hardware in practical engineering applications. At the same time, it can better solve practical engineering application problems, and perform precise semantic segmentation for cracks in the steel truss structure. What's more, when training each individual weak classifier, it is noticed that many background pixels will be included in the pixels whose output labels are cracks, which will cause the poor recognition. Based on the above description, the specific operation of the weak classifier cluster constructed in this study is as follows.

Firstly, the Stacking collaborative learning framework was introduced to train a model for combining various weak classifiers. In other words, multiple models were trained, and the output obtained from the previous training of each model was used as the input again to train a new model, and finally an output is obtained. Secondly, a new image training set was constructed through different random sampling methods for an image data set, and then the different weak classifiers are obtained. This operation is similar to the uniform sampling strategy with replacement in the Bagging method. In order to reduce the volume of the algorithm and lightweight the algorithm, the weight calculation method of each classifier also adopted the weight calculation method in the Bagging method, that is, the weight of each weak classifier is equal. This is conducive to the learning and training of the classifier cluster and will not affect the final semantic segmentation effect too much. Finally, referring to the training process of the Boosting method, and each weak classifier is generated.

The dataset setting was improved based on the above ideas. Firstly, in order to enhance the global search capability, a Global Dataset was prepared. 150 crack images of the steel trestle structure taken on the spot (ie the initial dataset) were used to perform data enhancement operations, randomly crop, color jitter and they were rotated, finally the initial dataset was expanded to 4,500 sheets. Fig. 4 shows a part of the Global Dataset. The floating range of the zoom factor is [0.8, 1.5], and the floating range of horizontal reflection and rotation is [-30, 30]. Secondly, in order to enhance the recognition accuracy of structural cracks and strengthen local search capabilities, Local Dataset was prepared. The floating rectangular window was used to pre-identify and cut out the crack concentrated areas in the image, and the floating rectangular window was used to pre-identify and cut out the crack concentrated areas in the image, and the data that adopted Global Dataset was used to enhance operation to expand the dataset to 4500 sheets. Fig. 5 shows part of the Local Dataset. The pixel sizes of the above two dataset images were uniformly divided into 960*720.

According to the above two improved datasets and network training methods, a total of 10 weak classifiers were trained. Among them, a total of 5 weak classifiers numbered Global 1 to Global 5 were trained using the Global Dataset, and they were combined into a weak classifier cluster, aiming to perform a global search on the image. The remaining 5 weak classifiers, numbered Local 1 to Local 5, were trained using Local Dataset, aiming to enhance the local search capability of images.

Semantic segmentation was performed on the damage image

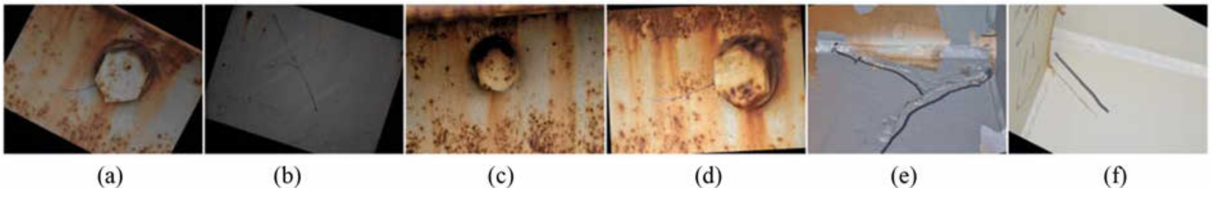

Fig. 4. Schematic Diagram of the Global Dataset: (a) Example 1 of Global Crack Image, (b) Example 2 oF Global Crack Image, (c) Example 3 of Global Crack Image, (d) Example 4 of Global Crack Image, (e) Example 5 of Global Crack Image, (f) Example 6 of Global Crack Image

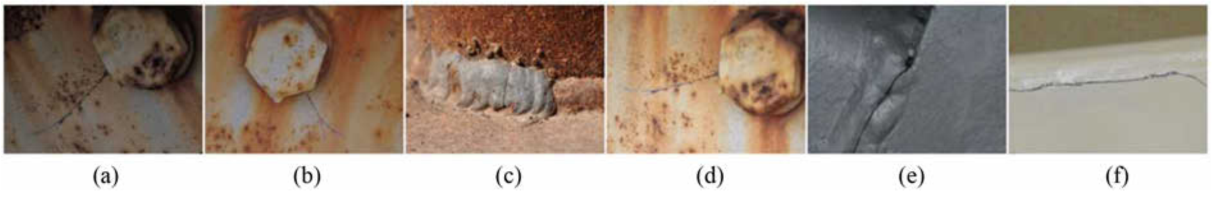

Fig. 5. Schematic Diagram of Local Dataset: (a) Example 1 of Local Crack Image, (b) Example 2 of Local Crack Image, (c) Example 3 of Local Crack Image, (d) Example 4 of Local Crack Image, (e) Example 5 of Local Crack Image, (f) Example 6 of Local Crack Image

of the steel structure trestle through the following two steps.

In the first step, Global 1 to Global 5 were designed to enhance the global search capability of the network. By taking the intersection of the results of Global 1 to Global 5, the crack concentrated areas in the image were divided and cropped.

The second step, Local 1 to Local 5 were used to perform pixel-level segmentation on the cropped image, identify and extract the pixel to which the crack belongs.

The location of the crack was extracted through two Global Datasets training, and then the crack pixel-level semantic segmentation of the Local Dataset was performed through the weak classifier to achieve the pixel identification and extraction of the crack. In this way, it can take both local and global search capabilities into account, so it better located and extracted cracks. Besides, it also improved the accuracy of semantic segmentation, and reduced training time. Moreover, computing efficiency was increased and hardware requirements were reduced.

### 2.2.4 Parallel Channel Attention Mechanism

After a certain period of service of the steel structure trestle, a certain area of shallow rust will appear on the surface of the structure. If the shallow rust appears around the crack, it will probably affect the automatic recognition of the crack by the semantic segmentation algorithm, and the shallow rust will also be misjudged as the crack area. In order to further improve the algorithm's recognition accuracy of cracks in a complex image environment, this paper further introduced the parallel channel attention mechanism on the basis of the above-mentioned improvement measures. It aims to further improve the accuracy of the algorithm in identifying cracks, and can effectively eliminate the adverse effects of shallow corrosion around the cracks.

The Squeeze-and-Excitation (SE) module specifically includes a two-step process of compression and excitation. The compression operation uses a global average pooling (GAP) operation to get the global features of the previous feature map in the channel dimension. The specific operation can be expressed by the following formula.

$$Z = F_{sq}(x) = \frac{1}{h \times w}\sum_{i=1}^{h}\sum_{j=1}^{w}x(i,j) \tag{1}$$

Among them, $Z$ represents the output after the compression process, which means the real number matrix output after the channel dimension is globally averaged and pooled. $F_{sq}$ represents the compression function used in the compression process, which means the GAP function. $x$ represents a collection of two-dimensional feature maps of spatial information dimensions. $h$ and $w$ respectively represent the two spatial information dimensions of the height and width of the feature map.

After the compression process, the SE module learns the importance of global features in each channel after GAP through the excitation process. This process generally consists of a fully connected layer, a ReLU activation function, a fully connected layer and a Sigmoid activation function. Through the two-layer fully connected bottleneck structure, the weight of each channel in the feature map is obtained, and the weighted feature map is used as the input of the next layer of the network. The above process can be expressed by the following formula.

$$S = F_{ex}(Z, W) = \sigma(W_2\delta(W_1 Z)) \tag{2}$$

Among them, $S$ is the output of the excitation process. The excitation function is denoted by $F_{ex}$, and $Z$ is the real number matrix combination output after the channel dimension is globally averaged and pooled. $\sigma$ and $\delta$ represent the activation function Sigmoid and the activation function ReLU, respectively. $W_1$ represents the first fully connected layer weight value set with the dimension $\frac{c}{r} \times c$, $W_2$ represents the second fully connected layer weight value set with the dimension $c \times \frac{c}{r}$, and $r$ represents the dimensionality reduction coefficient. By introducing the dimensionality reduction coefficient, the number of parameters of the SE module can be effectively reduced, and the occurrence of over-fitting can also be prevented to a certain extent.

The last step is to re-calibrate the features. The input feature map is multiplied before the SE module with the output value after the excitation process, and then the output $\tilde{x}$ after the SE module can be obtained. The process of feature re-calibration can be expressed by the following formula.

$$\tilde{x} = s \cdot x \tag{3}$$

The process of re-calibrating the eigenvalues is actually to assign weighted values to multiple maps obtained in the convolutional network. The feature re-calibration process will not make the performance of the convolutional network worse. The worst case of the process is to directly map the feature map of the previous layer to the next convolution kernel. Therefore, after introducing the channel attention mechanism, even if the semantic segmentation performance of the convolutional network cannot be improved, it will not adversely affect the semantic segmentation performance of the network.

Based on the above process steps, the output obtained through the SE module can be expressed by the following formula.

$$\tilde{x}_{SE} = \sigma\left\{W_2\delta\left[W_1\frac{1}{h \times w}\sum_{i=1}^{h}\sum_{j=1}^{w}x(i,j)\right]\right\}x \tag{4}$$

As mentioned above, the SE module only considers the use of global average pooling in the process of compression, excitation and feature re-calibration to reduce the variance of the estimated value caused by the reduction of high-level features (Lin et al., 2013). Therefore, the SE module is not ideal for the semantic segmentation of small objects and image textures. Aiming at the deficiencies of the SE module, in order to effectively identify the cracks of the steel structure trestle and reduce the impact of shallow corrosion on the segmentation effect, this paper introduced the Parallel Channel Attention Mechanism (PCAM) on the basis of the SE module. To put it simply, a global max pooling (GMP) compression operation is connected in parallel on the SE module, which worked together with the global tie pooling compression operation. Therefore, when extracting high-level features, the estimated value deviation caused by the convolutional network parameters can be reduced, so that the network can better perform accurate semantic segmentation of small targets and image texture information.

One process of the parallel channel attention mechanism is the same as the SE module, that is, the output $\tilde{x}_{SE}$ is obtained through the process described above. It can be expressed by the following formula.

$$\tilde{x}_{PCAM1} = \tilde{x}_{SE} \tag{5}$$

On this basis, another path is connected in parallel, and the second path compression operation uses the global maximum pooling to replace the global average pooling in the SE module, and the rest of the process is the same as the operation in the SE module. The real number matrix $m$ output by the global maximum pooling can be expressed by the following formula.

$$m = max\{x(i,j)\} \tag{6}$$

$W_3$ and $W_4$ are the weight sets of the two fully connected layers in the global maximum pooling channel respectively, and the dimensions of the two fully connected layers are the same. The final output of the global maximum pooling operation of the PCAM can be expressed by the following formula.

$$\tilde{x}_{PCAM2} = \sigma[W_4\delta(W_3 \cdot max\{x(i,j)\})]x \tag{7}$$

In summary of the above steps, the output process of the PCAM can be summarized as the following formula.

$$\tilde{x}_{PCAM} = \tilde{x}_{PCA1} + \tilde{x}_{PCA2} = \sigma\left\{W_2\delta\left[W_1\frac{1}{h \times w}\sum_{i=1}^{h}\sum_{j=1}^{w}x(i,j)\right]\right\}x + \sigma[W_4\delta(W_3 \cdot max\{x(i,j)\})]x \tag{8}$$

The output result of the PCAM is processed by the batch normalizing (BN), which effectively prevents the gradient from disappearing, speeds up the training and learning efficiency of the network, and to a certain extent also prevents unfavorable factors such as over-fitting (Liang et al., 2021). The schematic diagram of the parallel channel attention mechanism is shown in Fig. 6.

Based on the above improvement measures, the flow chart of the proposed improved semantic segmentation algorithm is shown in Fig. 7.

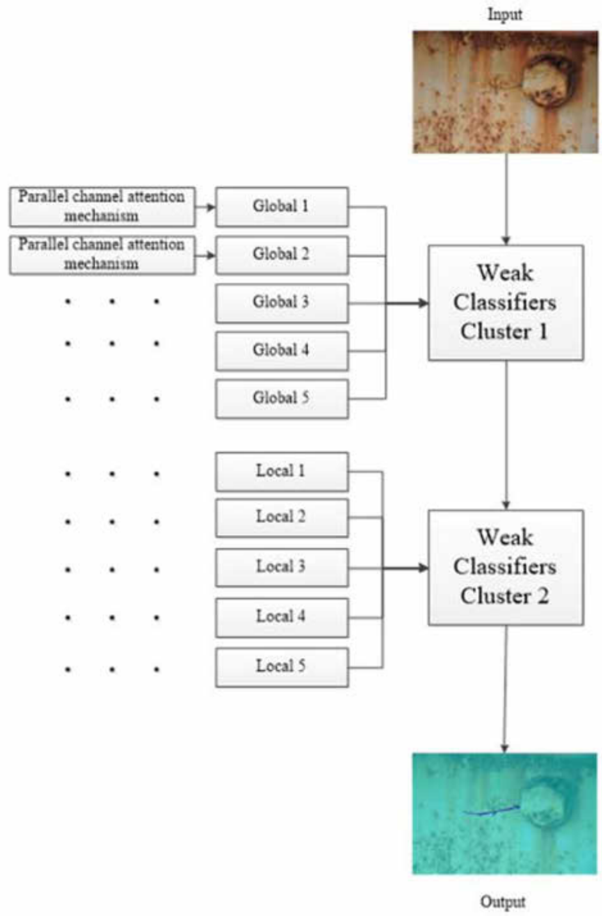


**Fig. 7.** Improved Semantic Segmentation Algorithm Flow Chart

## 2.3 Crack Identification Results and Analysis

### 2.3.1 Dataset Establishment

The datasets used in the training of weak classifiers in this study are all shots of a steel trestle structure engineering site that has been in service for many years. The shooting environment simulates consumer-grade camera monitoring, and a total of 150 images of

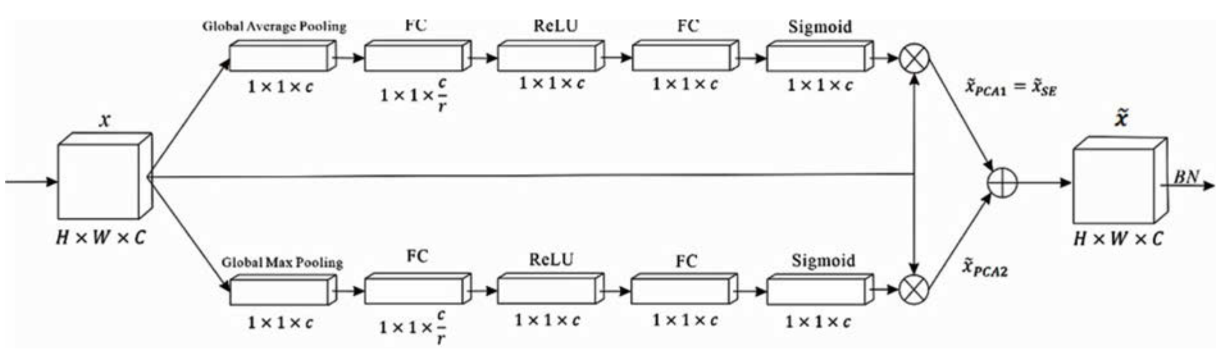


**Fig. 6.** Schematic Diagram of Parallel Channel Attention Mechanism

cracks in the steel structure trestle were taken. The dataset is expanded to 4500 sheets by the data enhancement method described in above, which becomes a Global Dataset. Among them, 65% was used as the training dataset, 20% was used as the verification dataset, and the remaining 15% was used as the test dataset. And the cut out of the crack concentration was used to form a Local Dataset, and the Global Dataset and the Local Dataset were used to train the network. The image size of the above two datasets is set to 960*720.

### 2.3.2 Experimental Parameter Settings

This study is based on Matlab 2020a to train the DeepLab V3+ model. The computer hardware parameters used are as follows: CPU: Intel(R) Core(TM)i7-9750H CPU @ 2.70 GHz, RAM: 64.0 GB, GPU: NVIDIA GeForce GTX 2080.

The enhanced data was used set to train the DeepLab V3+ network. The training parameter settings of a single weak classifier are shown in the following table.

### 2.3.3 Evaluation Index

The essence of semantic segmentation task is classification task. It's just that the object of the conventional classification task is the object in the image, and the object of the semantic segmentation is the pixel point in the image. Confusion matrix can be used to value the result of classification. True/False means the prediction is right or wrong, and Positive/Negative means the prediction result. For example, True Positive (TP) in this engineering application indicates that the predicted result is a crack pixel point and is a correct prediction, that is, the crack pixel point is correctly detected. False Negative (FN) indicates that the predicted result is not a crack pixel, but this is a wrong prediction, that means the crack pixel is not detected correctly. Similarly, False Positive (FP) indicates that the prediction result is a crack pixel, but it's a wrong prediction, and True Negative (TN) indicates that the prediction result is not a crack pixel and is a correct prediction.

The evaluation indicators used in the experiment contain the accuracy and mean intersection over union (mIoU). The above indicators can relatively intuitively give the comprehensive performance of image segmentation. Specifically, it can be defined by the following formula.

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

The precision of the prediction results represents the proportion of all samples predicted to be cracks whose true value is also a crack.

$$Recall = \frac{TP}{TP+TN} \tag{10}$$

Recall represents the correct rate of precision of all true crack samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

Accuracy is the simplest metric used to mark the proportion of correct pixels to the total pixels.

$$mIoU = \frac{1}{k+1}\sum\nolimits_{i=0}^{k}\frac{TP}{FN+FP+TP} \tag{12}$$

Mean intersection over union (mIoU) is a recognized algorithm evaluation standard. It calculates the ratio of intersection and union of two sets. In the field of semantic segmentation, the true value and the predicted value are the manifestations of the two sets, where $k+1$ is the number of categories.

**Table 1.** Training Parameter Settings of a Single Weak Classifier

| Parameter name | Parameter value |
|---|---|
| Optimization algorithm | Adam |
| Learn Rate Schedule | Piecewise |
| Learn Rate Drop Period | 10 |
| Learn Rate Drop Factor | 0.3 |
| Initial Learn Rate | 1e-3 |
| L2 Regularization factor | 0.005 |
| Minibatch Size | 8 |
| Maxepochs | 128 |
| Shuffle' | every-epoch |
| Validation Frequency | 302 |
| Validation Patience | 4 |

### 2.3.4 Analysis of Experimental Results

In order to verify the effectiveness of the image semantic segmentation algorithm of the improved DeepLab V3+ model, the accuracy and the mean intersection over union were used as the predictive evaluation indicators. And the results obtained by the improved DeepLab V3+ model's image semantic segmentation algorithm was compared with the standard full convolutional

**Table 2.** Comprehensive Comparison of Network Performance

| Network model | Training set | | Testing set | |
|---|---|---|---|---|
| | Accuracy | mIoU | Accuracy | mIoU |
| FCN | 0.8885 | 0.4738 | 0.8289 | 0.3552 |
| SE-FCN | 0.9162 | 0.5068 | 0.8562 | 0.3871 |
| Single DeepLab V3+ Net | 0.8413 | 0.4783 | 0.8067 | 0.4458 |
| Joint Decision DeepLab V3+ Nets | 0.9437 | 0.5406 | 0.9285 | 0.5064 |
| Improved DeepLab V3+ Nets | 0.9883 | 0.5825 | 0.9647 | 0.5369 |

network, the fully convolutional network with the SE module and the single DeepLab V3+ model to verify the effectiveness of improved algorithm measures. The specific comparison results of each algorithm are shown in Table 2.

It can be seen from the information in Table 2 that after the introduction of weak classifier clusters for joint decision-making and parallel channel attention, the improved DeepLab V3+ model has a corresponding improvement in performance when dealing with crack recognition tasks in steel structure trestle images. Compared with the standard FCN, the accuracy of the training set and test set are increased by approximately 9.98% and 13.58% respectively, and mIoU has been increased by approximately 10.87 and 18.17%, respectively. Compared with the FCN with the introduction of the SE module, the accuracy of the training set and the test set are increased by about 7.21% and 10.85%, and the mIoU has been increased by about 7.57% and 14.98% respectively. Compared with a single DeepLab V3+ network, the accuracy of the training set and test set are increased by approximately 14.70% and 15.8%, respectively, and mIoU has increased by approximately 10.42% and 9.11%, respectively. Compared with the DeepLab V3+ network that only introduces a joint decision-making mechanism, the accuracy of the training set and test set are increased by approximately 4.46% and 3.62%, respectively, and mIoU has increased by approximately 4.19% and 3.05% respectively.

Although the use of a set of weak classifiers for joint decision-making and the introduction of a parallel channel attention mechanism will increase the number of network parameters correspondingly and extend the training time. However, performance indicators such as the accuracy of image pixel segmentation and the average intersection ratio increased significantly after the introduction of improved measures. This improvement is particularly important in the pixel-level recognition task of cracks with complex image background information. In the experiment, the actual segmentation effect of each network for the crack image of the steel structure trestle is shown in Fig. 8. The cracks near the bolts are manually drilled through steel. Manual cracks simulate the characteristics of steel structure cracks, such as narrow, small and metallic luster. Cracks in welds are naturally formed.

It can be seen from Fig. 8 that when Single DeepLab V3+ Net is used to perform pixel segmentation on the crack image of the steel structure trestle, the effect is the worst, and the shallow rust around many cracks will also be judged as crack pixels, causing false alarms. When FCN is used to split the crack image semantics of this structure, the effect is not ideal, and it will also be interfered by certain background factors. At the same time, it can be seen from the second line of image segmentation effect that FCN still has incomplete segmentation of crack pixels. The FCN introduced with the SE module alone is better than FCN in the actual segmentation effect, but still has problems such as being affected by the image background and incomplete segmentation of cracked pixels. Although the segmentation accuracy has been improved, it still does not reach the ideal state. When the joint decision-making improvement mechanism of weak classifiers is introduced separately, the division effect is further improved, and the crack-like pixels can be almost completely segmented. However, the result is still affected by shallow corrosion, and some of the corrosion will be falsely reported as cracked pixels. Therefore, whether the SE module is introduced separately or the weak classifier cluster joint decision-making improvement mechanism is introduced separately, the actual segmentation effect of crack pixels is not ideal.

Finally, the comprehensively improved DeepLab V3+ Nets

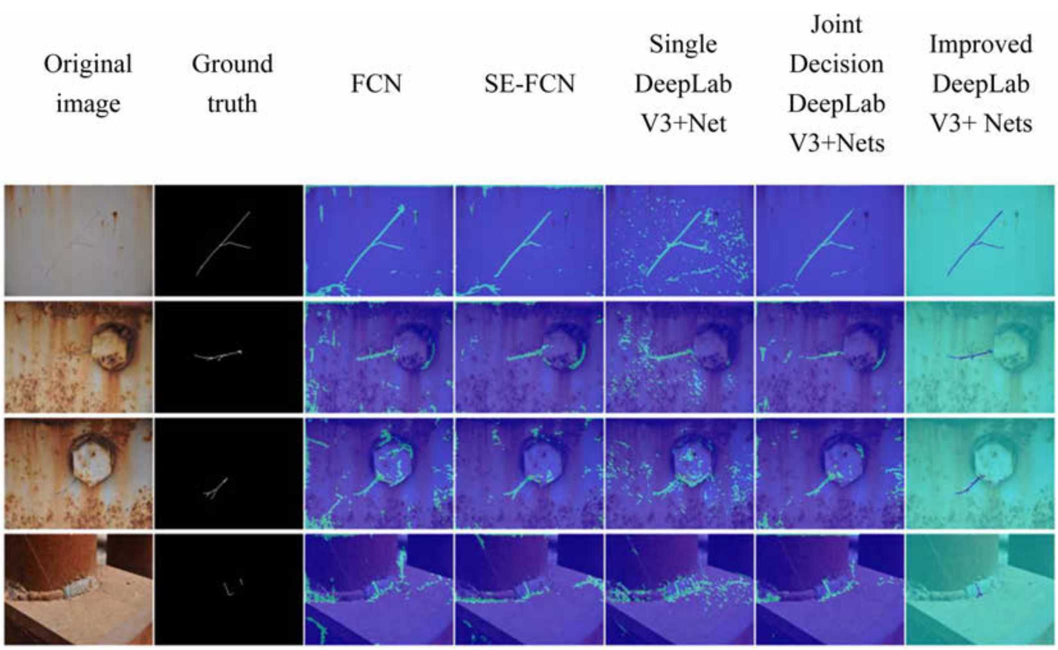


Fig. 8. The Actual Segmentation Effect of the Crack Image of the Steel Structure Trestle of Each Network

can accurately segment the crack structure of the steel structure trestle at the pixel level, and can effectively avoid the interference of the background factors of the shallow cracks, and has a significant improvement in the segmentation of the background. Improved DeepLab V3+ Nets can almost completely identify all crack pixels, so as to effectively judge the damage degree of the stress-concentrated parts of the steel structure trestle. The actual crack pixel segmentation is ideal. This also proved the effectiveness of the measures of Improved DeepLab V3+ Nets. When faced with crack images of complex background information, it can completely

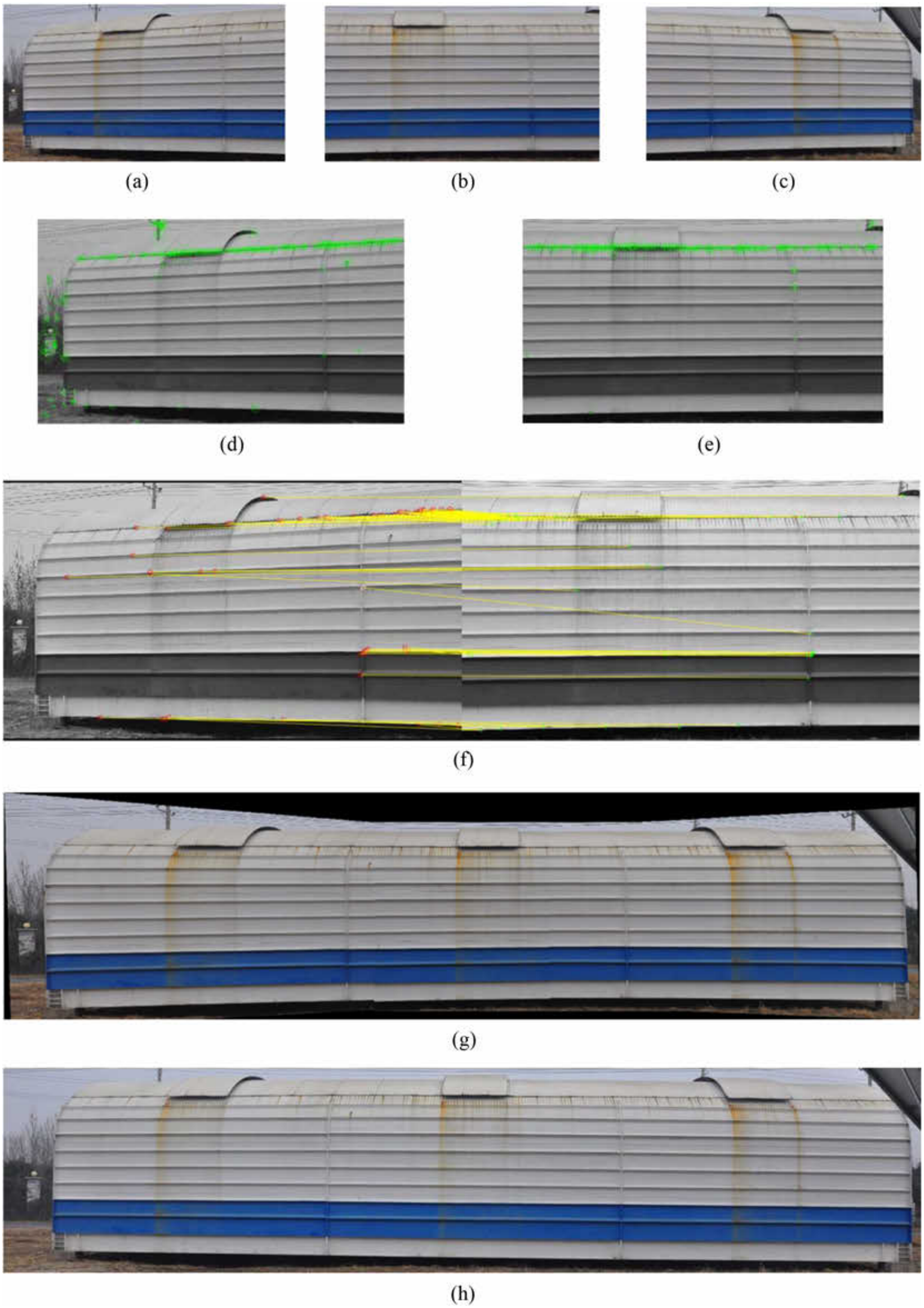

**Fig. 9.** Panoramic Image Stitching of Trestle: (a) Partial Picture 1, (b) Partial Picture 2, (c) Partial Picture 3, (d) Feature Points of Partial Picture 1, (e) Feature Points of Partial Picture 2, (f) Matching of Feature Points, (g) Panoramic Composite Image, (h) Panoramic Image Acquired by the Camera

and accurately segment the cracks at the pixel level, and judge the damage degree of the steel structure trestle where the stress is concentrated.

## 3. Panoramic Image Stitching

In practical engineering applications, consumer-grade cameras can better locate damage and perform maintenance work if they use a larger perceived field of view to acquire images. However, at the same time, the proportion of pixels occupied by damaged cracks will be too small, resulting in a decrease in recognition accuracy. Therefore, the pixels damaged by cracks should occupy the main body, otherwise the performance of computer vision algorithms will be affected.

Aiming at one of the above contradictions, this paper introduced a feature-based panorama image stitching method. Feature detection and stitching have been widely used in computer vision fields, such as object matching and tracking. The principle of this concept is to select certain feature points from the image and analyze the image locally instead of observing the whole image. As long as there are enough detectable interest points in the image, and these interest points are distinct and stable, and thus they can be accurately localized.

Speeded Up Robust Features (SURF, accelerated robust features), is a robust local feature point detection and description algorithm. SURF is an improvement to the Sift algorithm, which improves the execution efficiency of the algorithm and provides the possibility for the algorithm to be applied in real-time computer vision systems. Like the Sift algorithm, the basic path of the SURF algorithm can be divided into three parts: the extraction of local feature points, the description of feature points, and the matching of feature points. The feature points are set according to the number of SURF algorithm parameters, and the algorithm will automatically select the feature points in the image according to the preset number of feature points. The trestle structure image in this study has obvious edge protrusions and some rust. These are all stably obtainable feature points. After many experiments, when the number of feature points is 300, the feature matching effect can be guaranteed while consuming less computing resources, and the matching effect is stable. The panoramic image and its composition are shown in Figs. 9(a) – 9(h).

Figures 9(a) – 9(c) are the partial images of one section of the trestle collected each time the camera rotates. Figs. 9(d) – 9(e) are the feature points of the first two local images of one trestle section obtained by SURF. Fig. 9(f) shows the process of SURF matching the first two local images of the trestle according to the feature points. Fig. 9(g) is the generated result of the trestle panorama based on SURF. Fig. 9(h) is the real panorama of the trestle, which is used to verify whether the effect of Fig. 9(g) meets the requirements of practical engineering applications.

## 4. Crack Localization

The example of this project is the belt conveyor trestle from the No. 2 transfer point of the coal conveying system in an industrial park to the buffer silo, which adopts the modular steel tube space truss structure. Since the end of the project is low from the ground, the pipe truss structure used in this project is a quadrilateral pipe truss structure. Compared with other truss structures, the pipe truss has high bending and torsional rigidity, and the structure is relatively light in weight under the same constraint conditions, and the appearance is neat and tidy. The service life of the trestle structure is expected to be 50 years. The live load of the trestle floor is 3.0 kN/m$^2$, and the standard value of the roof live load is 0.5 kN/m$^2$. This trestle project is composed of 28 sections of modular trestle, and the engineering volume is huge. In this paper, a certain section of the structure in this project is selected for structural damage identification and location verification. This kind of trestle has a long service life and a strong load capacity. Once the crack damage occurs in the concentrated force, it will cause great economic losses and even casualties. Therefore, long-term uninterrupted SHM is required for this type of structure.

According to the analysis of the above experimental results, the crack damage identification method proposed in this paper can effectively identify the cracks of the steel structure trestle. The matching and positioning of crack images is affected by the spatial resolution, so this study used a combination of feature-based panorama stitching and subdivision iterative path methods to locate cracks. The space filling curve is an important approximate representation method. It divides the data space into grids of the same size, and then encodes these grids according to a certain method. Each grid is assigned a unique code and keeps proximity of the space to a certain extent. It means the labels of adjacent grids are also adjacent, and a spatial object consists of a group of grids. In this way, the multidimensional spatial data can be reduced to a one-dimensional space.

The camera cannot monitor the entire surface of the structure while meeting the monitoring accuracy, so it needs to scan the entire surface according to a certain route. After scanning, the related patches can be spliced according to the scanning path to obtain the overall image. The Peano curve can be iteratively subdivided according to the actual monitoring needs. Therefore, the Peano curve is used to stitch the trestle images to locate the cracks on the surface.

The Italian mathematician Peano G invented a curve that fills a square, called the Peano curve. Later, Hilbert made this curve, also known as the Hilbert curve (Jagadish, 1997). The Hilbert-Peano curve is a fractal shape that can be drawn infinitely complex. In the process of iterative generation, it continuously refines small individuals. With the increase of the curve order, the existing two-dimensional image is divided into $4^n$ parts of the same size. The line segments in the figure are actually the lines used to connect the parts. It is characterized by meandering and continuous drawing, which can pass through all points in a certain area on the plane. The Hilbert curve is a fantastic curve. As long as the function is properly selected, a continuous parametric curve is drawn. When the parameter t is within the

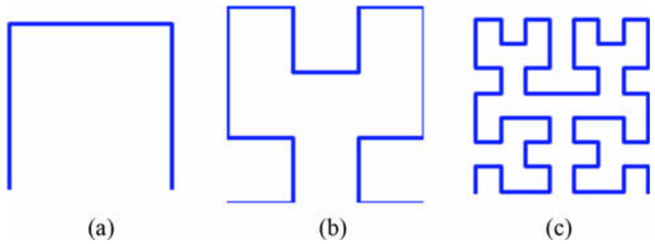


**Fig. 10.** Hilbert Space Filling Curve when n = 1,2,3: (a) n = 1, (b) n = 2, (c) n = 3

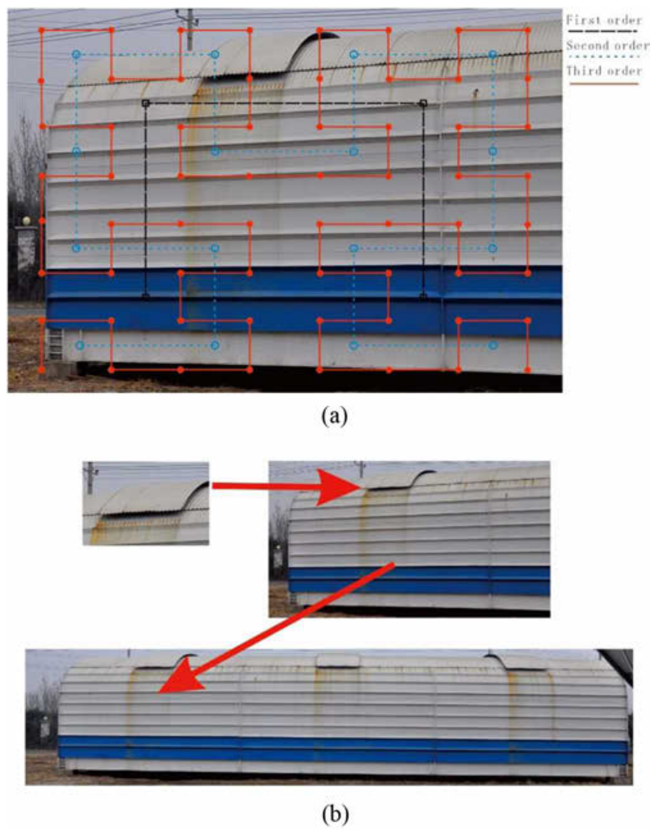


**Fig. 11.** The Panorama Stitching Process Based on SURF Matching: (a) The Diagram of the Path and Imaging Locations, (b) The Location Procedure of Patches

range of 0 and 1, the curve will traverse all the points in the unit square and get a full curve and get a space-filled curve.

As shown in the figure below, the scanning path of each part of the trestle structure is made of the Hilbet-Peano curve, and imaging is performed according to the path nodes. Finally, the region where the crack is located in detail by combining the above-mentioned panorama stitching method based on SURF matching. The panorama stitching process based on SURF matching is shown in Fig. 11.

It can be seen from the above synthesis that the stitching of the structural panorama is performed by using the Speeded Up Robust Features (SURF) and the iterative scanning path of the Hilbet-Peano space filling curve, which can determine the general orientation of the crack damage part on the overall structure of the steel trestle.

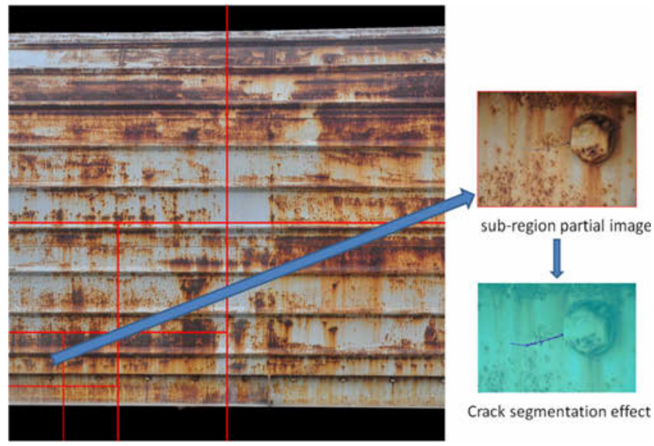


**Fig. 12.** The Detection-Segmentation for Local Crack Monitoring

To verify its feasibility without destroying the structure, real existing cracks from other sources were detected and segmented using region division and improvement DeepLab V3+, as shown in Fig. 12. The camera system can save a lot of computing resources through region division and Canny edge detection algorithm. Then the improved DeepLab V3+ model was used to efficiently identify cracks on the trestle surface. Finally, the SURF algorithm and the Hilbet-Peano space filling curve were used to determine the approximate location of the crack region globally.

In this simulation, a Nikon D90 camera (the camera imaging resolution in this manuscript is 3216*2136 pixels) is used to simulate the work of a consumer-grade camera. The camera was arranged on a vertical line which is 5 meters away from the trestle, and is fixed on a 1.8-meter-high bracket to simulate the working state of the camera. In the simulation, a whole section of trestle is taken as an example (the actual project is made up of 28 sections of this structure, the total span of the truss in this type of bid section is 20 meters, the height is 2.5 meters, and the transverse section size of the truss is 4.9 meters. The rods were made of Q345B hollow steel pipes). The camera rotates to take images sequentially from one end of the trestle to the other. For each acquired image, the curve as shown in Fig. 11 is used to scan, and each order curve represents a different scanning precision. A schematic diagram of the camera's shooting direction and

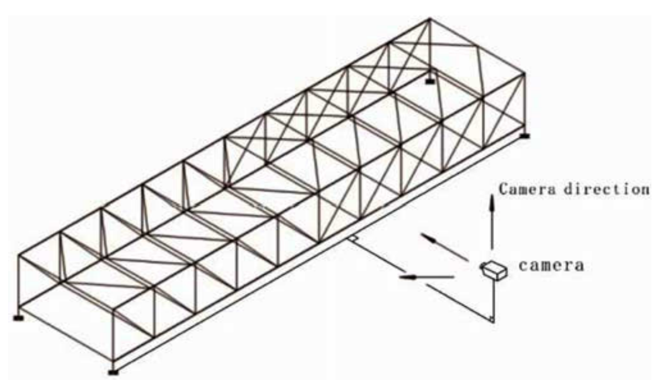


**Fig. 13.** Schematic Diagram of Camera Shooting Direction and Layout

**Table 3.** Comparison of Characteristics between DL-Based CV Method and Ultrasonic Inspection

| | DL-based CV crack detection | Ultrasonic inspection-based crack detection |
|---|---|---|
| 1 | In order to prevent the observation blind spots, it is reasonable to equip each section of the trestle with three cameras to monitor the key stress parts in real time (the coverage area of a single camera is about 25 m long and 3 m high). | The actual structure of the trestle is larger and more complicated. It is necessary to arrange the appropriate number and position of ultrasonic inspection probes according to the actual situation. |
| 2 | The non-contact automatic preliminary identification of the surface cracks in the key stress parts of the trestle can quickly and easily determine whether there are cracks on the surface of the structure. | Contact inspection with an ultrasonic inspection probe enables precise and efficient identification of cracks inside the trestle. |
| 3 | It can perform non-contact automatic long-term uninterrupted SHM in the working state of the trestle. | When a crack occurs in the trestle, it is necessary to suspend the work of the conveyor belt in the trestle, eliminate the interference of vibration, and then accurately inspect and identify the cracked part. |
| 4 | The devices involved in the DL-based CV method are all low-priced consumer-grade devices that can perform SHM on the trestle at a low cost. At the same time, the system can also be used for multi-purpose applications (such as security monitoring). | More professional equipment is required, and the price of the equipment will vary greatly according to the detection range and accuracy of the probe. Appropriate ultrasonic inspection equipment should be selected according to the characteristics of the actual structure. |

arrangement is shown in Fig. 13. The simulation was choosing a position 8 meters away from the trestle on the vertical line of it. The data that identified the smallest crack on the surface of the trestle structure were about 3 mm wide and 6 to 7 cm long. In this recognition result, the camera resolution was set to 3216*2136 pixels in order to simulate the imaging effect of a consumer-grade camera, and the lighting condition is light haze during the day.

## 5. Discussion

It is worth noting that the method proposed in this paper is to meet the production and monitoring requirements of enterprises. This method is the SHM system used in conjunction with the new steel trestle in service at the same time. The SHM of the key stress parts of the steel trestle can be realized at a relatively low cost. Meanwhile, the SHM system has the monitoring characteristics of unmanned automation, full life cycle, and long-term uninterrupted. Therefore, the CV method for relatively inexpensive consumer-grade camera systems was chosen. The DL-based CV method used in this study was designed to automatically and quickly identify cracks on the surface of steel trestles. Therefore, based on the method in this study, the cracks of the trestles can be detected more accurately by other measures (e.g., ultrasonic inspection). Ultrasonic inspection is also one of the most widely used crack damage identification methods. It has many advantages, for example, ultrasonic inspection can judge the damage degree of cracks, and can find cracks inside the structure. The DL-based CV method can make a quick judgement on whether the crack occurs, and its application is more flexible. Therefore, the combination of DL-based CV method and ultrasonic inspection can further improve the accuracy of crack identification on the trestle surface, which will be a very valuable research direction in the future. The characteristics of the DL-based CV method and the ultrasonic inspection in the task of identifying surface cracks on steel trestle are compared as shown in Table 3.

In general, traditional SHM methods, such as ultrasonic inspection, have better accuracy in structural damage detection, and can find possible damage inside the structure and judge the damage degree of the structure. However, due to the particularity of the trestle structure and service environment, the detection method of the traditional SHM method is often limited when dealing with the steel trestle in this study. It led to inconvenient inspections or the results of inspections were interrupted, which can affect the productivity of enterprises using trestles. The ensemble SHM method proposed in this study, according to the structure and service characteristics of the steel trestle, can automate the long-term uninterrupted SHM of the steel trestle at a lower cost. The deployment of this method is more flexible, and it can quickly and effectively determine whether there is a crack on the surface of the steel trestle and perform preliminary positioning, but the judgement of the degree of crack damage is currently not as accurate as the traditional SHM method. If the method in this study is combined with the traditional SHM method, the SHM of the steel trestle can be better performed. And the advantages are embodied in taking into account the speed, convenience and accuracy of monitoring tasks at the same time. This will have good research and application prospects in the future.

## 6. Conclusions

Aiming at various limitations of the service environment of steel structure trestle, a SHM system based on consumer-grade cameras can be used to effectively conduct long-term, stable and efficient SHM of steel trestle structure, which reduced monitoring costs. Small cracks on its surface can be identified autonomously and without contact, which improves the monitoring accuracy. At the same time, the approximate orientation of the crack on the whole structure was located, and the automatic SHM of the trestle structure was realized. The main conclusions are as follows:

1. By introducing the Canny boundary detection algorithm to

extract the image of the steel structure trestle, it can effectively improve the efficiency of the subsequent algorithm for image processing, and improve the accuracy of the subsequent algorithm recognition.

2. As to each section of the steel structure trestle image, it is divided into areas of suitable size, and each area was focused and zoomed by camera to perform crack pixel segmentation. Distant regions can be identified for cracks without loss of pixel resolution. Compared with the traditional manual structural defect inspection method, the efficiency and accuracy of crack identification are significantly improved. At the same time, the size of each trestle structure is not very huge, and there are not many areas divided by the image. Therefore, the efficiency of crack identification by region-by-region magnification is higher than that of object detection algorithms.
3. An improved pixel-level semantic segmentation algorithm was proposed, which can effectively identify the surface cracks of steel structures. A global dataset and a local dataset were constructed by data augmentation, and a total of 10 DeepLab V3+ networks built with ResNet18 as the backbone were trained using these two datasets. The 10 DeepLab V3+ networks were divided into two weak classifiers, so that the algorithm took into account both global and local search capabilities. The introduction improved the performance of the original network with a small increase in the amount of parameters and training time, so that the network model can better complete the task of crack pixel segmentation, and it can run well on consumer-grade computer equipment. At the same time, the difficulty of training a single complex model was avoided and time cost was reduced.
4. During the service process of the trestle structure, the problem of shallow corrosion will inevitably occur, and the crack identification of the trestle will be affected by the shallow-corrosion, which will cause interference. The parallel attention mechanism was introduced to effectively distinguish the shallow corrosion interference, so as to more accurately identify the more dangerous crack damage, and further improve the SHM accuracy of the trestle.
5. Using SURF matching and setting an iterative scan path for the stitching of the structural panorama. The location of crack damage can be roughly located on the overall steel trestle structure. It provided a practical reference method for continuous SHM by using consumer-grade cameras.

## Acknowledgments

Not Applicable

## ORCID

Sida Lian https://orcid.org/0000-0003-3794-979X
Wen Chen https://orcid.org/0000-0001-7433-1329

## References

Ahmed H, La HM, Tran KT (2020) Rebar detection and localization for bridge deck inspection and evaluation using deep residual networks. *Automation in Construction* 120(8):103393, DOI: 10.1016/j.autcon.2020.103393

Bao Y, Tang Z, Li H, Zhang Y (2019) Computer vision and deep learning–based data anomaly detection method for structural health monitoring. *Structural Health Monitoring* (2):401-421

Bull LA, Gardner P, Rogers TJ, Cross EJ, Dervilis N, Worden K (2021) Probabilistic inference for structural health monitoring: New modes of learning from data. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part A Civil Engineering*, DOI: 10.1007/978-1-4842-6825-4_8

Cha YJ, Choi W, Suh G, Mahmoudkhani S, Buyukozturk O (2018) Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering* 33(9):731-747

Dizaji MS, Alipour M, Harris DK (2021) Subsurface damage detection and structural health monitoring using digital image correlation and topology optimization. *Engineering Structures* 230:111712, DOI: 10.1016/j.engstruct.2020.111712

Dong C, Li L, Yan J, Zhang Z, Catbas FN (2021) Pixel-level fatigue crack segmentation in large-scale images of steel structures using an encoder-decoder Network. *Sensors (Basel, Switzerland)* 21(12): 4135, DOI: 10.3390/s21124135

Dorafshan S, Thomas RJ, Maguire M (2018) Fatigue crack detection using unmanned aerial systems in fracture critical inspection of steel bridges. *Journal of Bridge Engineering* 23(10):04018078, DOI: 10.1061/(ASCE)BE.1943-5592.0001291

Dung CV, Sekiya H, Hirano S, Okatani T, Miki C (2019) A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Automation in Construction* 102:217-229, DOI: 10.1016/j.autcon.2019.02.013

Han Q, Liu X, Xu J (2022) Detection and location of steel structure surface cracks based on unmanned aerial vehicle images. *Journal of Building Engineering* 50:104098, DOI: 10.1016/j.jobe.2022.104098

Han Q, Qian M, Xu J, Liu M (2021) Structural health monitoring research under varyingtemperature condition: A review. *Journal of Civil Structural Health Monitoring* 11(3):1-25, DOI: 10.1007/s13349-020-00444-x

He K, Zhang XY, Ren SQ, Sun J (2016) Identity mappings in deep residual networks. European conference on computer vision. *Springer International Publishing*, DOI: 10.1007/978-3-319-46493-0_38

Hoskere V, Narazaki Y, Hoang TA , Spencer BF (2020) MaDnet: Multi-task semantic segmentation of multiple types of structural materials and damage in images of civil infrastructure. *Journal of Civil Structural Health Monitoring* 10:757-773, DOI: 10.1007/s13349-020-00409-0

Jagadish HV (1997) Analysis of the hilbert curve for representing two-dimensional space. *Information Processing Letters* 62(1):17-22, DOI: 10.1016/S0020-0190(97)00014-8

Jahanshahi MR, Chen FC, Joffe C, Masri SF (2016) Vision-based quantitative assessment of microcracks on reactor internal components of nuclear power plants. *Structure & Infrastructure Engineering* 1-14, DOI: 10.1080/15732479.2016.1231207

Jesus A, Brommer P, Westgate R, Koo K, Brownjohn J, Laory I (2019) Modular bayesian damage detection for complex civil infrastructure. *Journal of Civil Structural Health Monitoring* 9:201-215, DOI: 10.1007/s13349-018-00321-8

Karypidis DF, Berrocal CG, Rempling R, Granath M (2019) Structural Health monitoring of RC structures using optic fiber strain measurements: A deep learning approach. 2019 IABSE Congress - New York City - The Evolving Metropolis, DOI: 10.2749/newyork.2019.0397

Le N, Rathour VS, Yamazaki K, Luu K, Savvides M (2021) Deep reinforcement learning in computer vision: A comprehensive survey. *Artif Intell Rev*, DOI: 10.1007/s10462-021-10061-9

Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553): 436, DOI: 10.1038/nature14539

Liang S, Khoo Y, Yang H (2021) Drop-Activation: Implicit parameter reduction and harmonious regularization. *Communications on Applied Mathematics and Computation* 3(2):293-311, DOI: 10.1007/s42967-020-00085-3

Li HN, Ren L, Jia ZG, Yi TH, Li DS (2016) State-of-the-art in structural health monitoring of large and complex civil infrastructures. *Journal of Civil Structural Health Monitoring* 6(1):3-16, DOI: 10.1007/s13349-015-0108-9

Lim RS, La HM, Sheng W (2014) A robotic crack inspection and mapping system for bridge deck maintenance. *IEEE Transactions on Automation Science & Engineering* 11(2):367-378, DOI: 10.1109/TASE.2013.2294687

Lin M, Chen Q, Yan SC (2013) Network in network. Computer Science

Miyamoto A, Kiviluoma R, Yabe A (2019) Frontier of continuous structural health monitoring system for short & medium span bridges and condition assessment. *Frontiers of Structural and Civil Engineering: English Version* 13(3):36, DOI: 10.1007/s11709-018-0498-y

Nagarajaiah S, Yang Y (2017) Modeling and harnessing sparse and lowrank data structure: A new paradigm for structural dynamics, identification, damage detection, and health monitoring. *Structural Control & Health Monitoring* 24(1), DOI: 10.1002/stc.1851

Nahata D, Mulchandani HK, Bansal S, Muthukumar G (2019) Post-earthquake assessment of buildings using deep learning. *arXiv*, DOI: 10.48550/arXiv.1907.07877

Ngeljaratan L, Moustafa MA, Pekcan G (2021) A compressive sensing method for processing and improving vision-based target-tracking signals for structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering*, DOI: 10.1111/mice.12653

Oh JK, Jang G, Oh S, Lee JH, Yi BJ, Moon YS, Lee JS, Choi YJ (2009) Bridge inspection robot system with machine vision. *Automation in Construction* 18(7):929-941, DOI: 10.1016/j.autcon.2009.04.003

Prasanna P, Dana KJ, Gucunski N, Basily BB, La HM, Lim RS, Parvardeh H (2016) Automated crack detection on concrete bridges. *IEEE Transactions on Automation Science and Engineering* 13(2):591-599, DOI: 10.1109/TASE.2014.2354314

Qi F, Xie Z, Tang Z, Chen H (2021) Related study based on otsu watershed algorithm and new squeeze-and-excitation Networks for segmentation and level classification of tea buds. *Neural Processing Letters* 53(3), DOI: 10.1007/s11063-021-10501-1

Rafiei MH, Adeli H (2018) A novel unsupervised deep learning model for global and local health condition assessment of structures. *Engineering Structures* 156(FEB.1):598-607, DOI: 10.1016/j.engstruct.2017.10.070

Rubio JJ, Kashiwa T, Laiteerapong T, Deng W, Nagai K, Escalera S, Nakayama K, Matsuo Y, Prendinger H (2019) Multi-class structural damage segmentation using fully convolutional networks. *Computers in Industry* 112

Samantaray S, Mittal SK, Mahapatra P, Kumar S (2018) An impedance-based structural health monitoring approach for looseness identification in bolted joint structure. *Journal of Civil Structural Health Monitoring*, DOI: 10.1007/s13349-018-0307-2

Sen D, Aghazadeh A, Mousavi A, Nagarajaiah S, Baraniuk R, Dabak A (2019) Data-driven semi-supervised and supervised learning algorithms for health monitoring of pipes. *Mechanical Systems and Signal Processing* 131:524-537, DOI: 10.1016/j.ymssp.2019.06.003

Shahbaznia M, Mirzaee A, Dehkordi MR (2020) A new model updating procedure for reliability-based damage and load identification of railway bridges. *KSCE Journal of Civil Engineering* 24(3):890-901, DOI: 10.1007/s12205-020-0641-x

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *Computer Science*

Sun L, Shang Z, Xia Y, Bhowmick S, Nagarajaiah S (2020) Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *Structural Engineering* 146(5):04020073, DOI: 10.1061/(ASCE)ST.1943-541X.0002535

Spencer BF, Hoskere V, Narazaki Y (2019) Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* 5(2):199-222, DOI: 10.1016/j.eng.2018.11.030

Wan L, Xie X, Wang L, Li P, Liu Y (2022) New damage identification method for operational metro tunnel based on perturbation theory and fuzzy logic. *KSCE Journal of Civil Engineering* 26(1):193-206, DOI: 10.1007/s12205-021-2299-4

Wu S, Zhong S, Liu Y (2017) Deep residual learning for image steganalysis. *Multimedia Tools and Applications* 77(9):10437-10453, DOI: 10.1007/s11042-017-4440-4

Yang X, Li S, Zhang D, Yao J, Zhang F, Na L, Hui L (2018) Identification framework for cracks on a steel structure surface by a restricted Boltzmann machines algorithm based on consumer-grade camera images. *Structural Control and Health Monitoring* 25, DOI: 10.1002/stc.2075

Yeum CM, Choi J, Dyke SJ (2019) Automated region-of-interest localization and classification for vision-based visual assessment of civil infrastructure. *Structural Health Monitoring* 18(3):675-689

Zhang L, Shen J, Zhu B (2022) A review of the research and application of deep learning-based computer vision in structural damage detection. *Earthquake Engineering and Engineering Vibration* (21):1-21, DOI: 10.1007/s11803-022-2074-7

Zhang M, Hu H, Li Z, Chen J (2021) Attention-based encoder-decoder networks for workflow recognition. *Multimedia Tools and Applications* (1):1-23, DOI: 10.1007/s11042-021-10633-5

Zhang Q (2022) A novel ResNet101 model based on dense dilated convolution for image classification. *SN Applied Sciences* 4(9), DOI: 10.1007/s42452-021-04897-7