Construction and Building Materials 299 (2021) 123896

Contents lists available at ScienceDirect

Construction and Building Materials

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

A UAV-based machine vision method for bridge crack recognition and width quantification through hybrid feature learning

Xiong Peng [a], Xingu Zhong [a,*], Chao Zhao [b], Anhua Chen [a], Tianyu Zhang [a]

[a] Hunan University of Science and Technology, Xiangtan, Hunan 411201, China
[b] Hunan Provincial Key Laboratory of Structures for Wind Resistance and Vibration Control & School of Civil Engineering, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

HIGHLIGHTS

- UAV-based machine vision system applied on bridge crack recognition and width quantification.
- Crack width quantification through object-distance measurement and measured plane correction.
- Crack detection and crack image recorded of inspection frame based on R-FCN network.
- High-accuracy pixel-level crack segmentation method based on Haar-AdaBoost and local threshold.
- Verification of the UAV-based method for crack detection on a real bridge.

ARTICLE INFO

Article history:
Received 20 January 2021
Received in revised form 29 April 2021
Accepted 6 June 2021
Available online 15 June 2021

Keywords:
Bridge crack
Unmanned Aerial Vehicle
Machine vision
Crack width recognition
Hybrid feature learning

ABSTRACT

Bridge crack width is an important indicator to assess and evaluate the health condition of the bridge. In this paper, we have proposed a UAV-based machine vision method for bridge crack recognition and width quantification through hybrid feature learning. Firstly, we have configured a UAV system that can obtain bridge crack image (effective pixel 7952x5304) and GPS position, calculate image resolution, and correct measured plane simultaneously. Then, the crack recognition method combining the R-FCN network and Haar-AdaBoost suited for UAV imagery recognition is proposed, which can make full use of advanced features, shape features and gray features of bridge cracks. The time cost of our method is about 0.2 s for crack detection per one 7952×5304 pixels images and 4.5 s in pixel-level segmentation per one 1000×1000 pixels bounding box. Additionally, the real bridge crack widths are calculated and quantified by the ranging method using corresponding object distance data. Finally, a case study of the Xiangjiang-River bridge inspection is carried out to demonstrate the effectiveness of the proposed method, achieving above 90% precision in the real bridge crack width quantification.

1. Introduction

Bridge is a major transportation infrastructure, and the crack width is an important indicator for assessing and evaluating the health condition of the bridge. In China, regarding to standards, it is mandatory to detect and quantify the bridge cracks that above 0.2 mm width [1]. In US and UK, a bridge should be inspected every 1–3 years according to the standards [2–3]. Generally, the bridge-inspection vehicle is applied as the manual operation platform for crack width measurement shown in Fig. 1. Nevertheless, some lingering issues such as traffic hindering, restriction by bridge type, and high maintenance cost remain when operating by this method. To date, it is warranted to develop a rapid and cost-effective inspection platform to assess the health condition of the bridge.

Recently, the use of unmanned aerial vehicle (UAV) in structural inspection has become common [4–5]. This approach has significant advantages as means to provide faster, cheaper, safer, and more flexible image data acquisition way [6]. Firstly, the rotor UAV are used as inspection platform to take and record images of infrastructure defects. For example, Pereira, Sankarasrinivasana and Hallermann use unmanned aerial vehicle equipped with a high-resolution camera to obtain appearance images of structure [7–9]. Wang [10] uses the UAV to detect the blades of wind turbines. Dorafshan [11] applies UAV for fracture critical inspection

* Corresponding author.

https://doi.org/10.1016/j.conbuildmat.2021.123896
0950-0618/

Fig. 1. Bridge crack inspection based on special vehicle platform.

of steel bridges by crack detection. Furthermore, researchers try to apply UAV for defect parameters measurement. Specifically, Kang configures particular autonomous UAV geotagging by GPS to detect crack by processing collected video data [6]. Kim [12] utilizes UAV for concrete crack identification in conjunction mapping based on known bridge component size. Park proposes a concrete crack detection and quantification approaches using deep learning and structured light [13]. These studies make meaningful attempts to promote the bridge structure inspection based on UAV.

However, the UAV-based bridge crack inspection method has its unique characteristics. (1) the bridge inspection primarily focuses on the crack located in the web and bottom of bridge superstructure, which requires the equipped camera carried on UAV have enough imaging field; (2) in UAV-based bridge inspection, the angle between the measured plane and the imaging plane is not parallel, which needs to be corrected in pixel resolution calculation; (3) in order to quantify crack width, the pixel resolution of each crack image need to be calculated accurately. In previous study, Zhong has studies the feasibility of concrete crack width recognition based on the UAV via pixel resolution calculation [14].

In this paper, we have proposed a UAV-based machine vision system that can accurately measure the width of bridge cracks. This UAV-based system has a special upper platform equipped with Sony ILCE-7RM2 DSLR camera and a three-point laser-finder, which is synchronized with camera shutter for measuring and recording object distance in real-time. However, the UAV-based bridge crack inspection usually generates large amounts of images in complex backgrounds (tens of thousands or even more images may be generated). Therefore, it is necessary to propose an intelligent method to calculate crack width for evaluating the health condition of the bridge.

2. Related work of crack recognition

Crack width is the most important indicator for bridge health condition evaluation and assessment. In recent years, the crack recognition methods based on deep learning using cascaded convolution neural network have been developed rapidly since its first application in civil engineering [15–16]. As for crack recognition method, it has experienced the following stages of development. Firstly, the target recognition network such as SSD [17] Faster-RCNN [18] and YOLO [19] have applied to detect the ROI of cracks. For example, Yang proposes an automatic pavement crack detection network based on the Single Shot MultiBox Detector (SSD) deep learning framework [20]. Ju proposes Crack Deep Network based on Faster-RCNN to detect sealed and unsealed cracks with complex road backgrounds [21]. Park [13] uses YOLO (You Only Look Once) algorithm for real-time crack detection. Those studies have shown the good ability of deep learning method in crack classification and localization.

Then, in order to obtain the accurate parameters of cracks, those pixel-level semantic segmentation networks such as FCN [22] U-Net [23] Deeplab V3 [24,25] have applied to calculate the width and length of the crack. To be specific, Zhang describes the Crack-Net for automated pavement crack detection [26]. Ren proposes image-based concrete crack detection in tunnels using deep fully convolutional networks [27]. Kim presents a novel crack assessment framework for concrete structures using Mask R-CNN [28]. Ai proposes automatic pixel-wise detection of evolving cracks on rock surfaces in video data [29]. Those studies show high accuracy of crack segmentation if pixel-level semantic segmentation networks are applied.

However, the pixel-level semantic segmentation networks applied on UAV-based bridge crack quantification have three limitations: (1) the prediction mean or maximum width is quite different from the ground truth [30]; (2) a tremendous amount of time is required to build pixel-level data set [31]. Therefore, those methods of combining neural network with adaptive image scanning are proposed. In those approaches, crack image is divided into blocks, which is judged by CNN or semantic segmentation networks [32]. Specifically, Dung use deep fully convolutional neural network for concrete crack detection [33,34]. Jang propose deep learning–based autonomous concrete crack evaluation method through hybrid image scanning [35]. The application of these algorithms significantly improved the accuracy degree of crack segmentation [30]. Notably, in order to solve the high time cost of making crack segmentation dataset, Kang proposes a hybrid pixel-level method for concrete crack segmentation and quantification in complex background, which has been achieved good accuracy, while the time cost is significantly reduced [31].

The UAV-based machine vision method for bridge crack recognition has its unique characteristics: (1) there is a lot of redundant background information in the bridge crack inspection process, so it is necessary to propose a network to detect the crack target for image data acquisition in real-time; (2) crack width is an important index to bridge condition evaluation, which needs to be quantified accurately, the segmentation method needs to reflect the real crack width accurately; (3) the background of bridge crack image is complex, those pixel-level segmentation methods based on neural network which need a large amount of sample data requiring manual label, while crack dataset making is time-consuming.

To address those problems, we propose a novel method for bridge crack recognition and width quantification using hybrid fea-

tures, including the advanced features from R-FCN network [36] to locate the crack target in real-time, and the edges features from Haar-AdaBoost [37] to segment the crack in pixel level. The usefulness of this method and system is demonstrated in a case study of the Xiangjiang-River bridge inspection for data acquisition, crack recognition and crack width quantification.

3. Methodology

To achieve reliable and accurate bridge crack width quantification, we have proposed a UAV-based machine vision method, which is introduced in the following subsections. This framework involves three main tasks: (1) crack image data acquisition, (2) crack recognition (localization and extraction) and (3) crack width calculation, as shown in Fig. 2.

3.1. Image data acquisition method

To realize bridge crack identification based on UAV airborne imagery, we configure the UAV-based machine vision system to obtain image data shown in Fig. 3. This system involves three main parts:

- special UAV with upward platform, consists of six-rotor UAV (with diameter = 1000 mm, height = 540 mm, and load weight = 4200 g), and the upward three-axis stabilizing and mounting platform, and IMU self-stabilization device etc.
- machine-vision measurement system, consists of Sony ILCE-7RM2 camera (with effective pixel 7952x5304, and mass = 582 g), three-point laser rangefinder (made by 3D printing, mass = 105 g) synchronized with the camera shutter, used to measure the distance between the bridge appearance and the camera, and the well-trained crack target detection network built-in UAV.
- GPS and transmission system, including GPS (for locating the real-time position of UAV), wireless remote controller (for controlling UAV attitude), high-definition transmission.

3.1.1. Crack image and object-distance data

As shown in Fig. 4, the UAV-based bridge crack image acquisition process is introduced. The mentioned above system is applied to capture 7952 × 5304 pixels image. As shown in Fig. 6, the frames of bridge appearance image are obtained by the DSLR camera, which are transmitted back to the high-definition transmission monitor on the ground. If the frame of bridge image is detected as crack by the built-in target detection network, the camera shutter and rangefinder recorder would be triggered, the crack image and its corresponding object-distance and GPS data would be recorded and saved into an SD card at the same time.

3.1.2. Pixel resolution and angle data

As shown in Fig. 5, the three laser axes are AB, AC and AD. Points B, C and D are the corresponding spots of the laser on the measured plane. X-Y plane is the camera CMOS plane, AB is perpendicular to X-Y plane, and AD is on the Y-Z plane. The X-Y plane is parallel to the imaging plane of the camera, and the length of AB is the measured object distance shown in Fig. 6(a). The angle α, β, and γ are set in 3D printing design, $AB = l_3, AC = l_2, AD = l_1$, and $\vec{K}$ is the normal vector of the image plane. Assuming point A is the origin of the coordinates, the coordinate of point C is set as $C(X_C, Y_C, Z_C)$. The coordinates of the three laser spots (shown in Fig. 5) can be obtained using geometric relations, followed by the following equations:

Which gives $a = l_2\sqrt{\sin^2\alpha - \frac{(\cos\beta - \cos\partial\cdot\cos\gamma)^2}{\sin^2\gamma}}$, $b = \frac{\cos\beta - \cos\alpha\cdot\cos\gamma}{\sin\gamma} l_2$, $c = l_2\cos\alpha$.

Therefore, the corrected angle of measured plane shown in Fig. 6(c) is calculated by Eq. (1):

$$\cos\theta = \frac{\vec{n}\cdot\vec{k}}{|\vec{n}||\vec{k}|} = \frac{al_3\sin\gamma}{\sqrt{[(c-l_1)l_3\sin\gamma - b(l_3\cos\gamma - l_1)]^2 + [a(l_3\cos\gamma - l_1)]^2 + [al_3\sin\gamma]^2}} \quad (1)$$

According to [15] the ratio of pixel width to actual width is converted by Eq. (2):

$$J = \frac{(L-f)\cdot d}{f\cdot D}\cdot\cos\theta \ \text{mm/pixel} \quad (2)$$

where L is the distance, f is the focal length, J is the pixel resolution, d is the longer dimension of the image sensor, and D is the number of pixels along the long side of the image sensor, and θ is the angle between the measured plane and camera imaging optical axis.

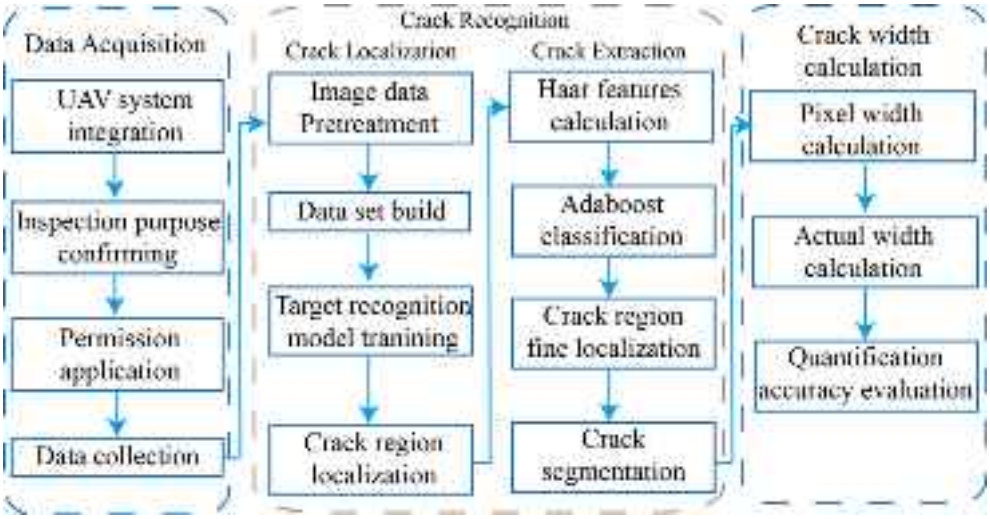

Fig. 2. Framework for bridge crack width recognition.

Fig. 3. UAV-based machine-vision system for bridge crack inspection.

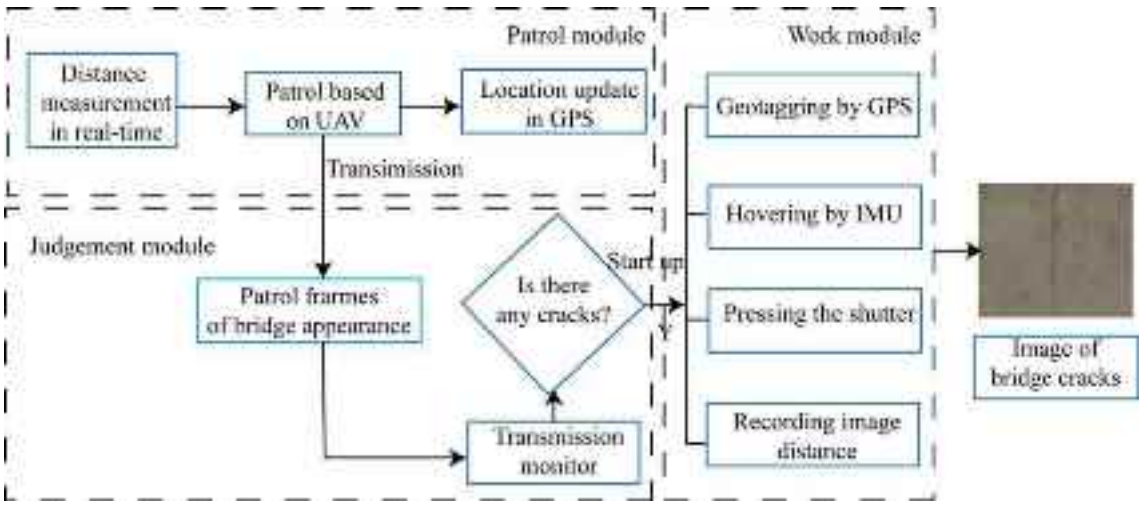

Fig. 4. Bridge crack image acquisition process.

3.2. Crack recognition method

The examples of airborne bridge crack images are shown in Fig. 7, there are 3 forms surface background of bridge inspection crack image with different noises, such as formwork imprint, shadow. To realize bridge crack width calculation, as shown in Fig. 8, a new two-stage recognition method based on hybrid features learning is proposed. In this method, The R-FCN network is the development of Faster-RCNN, which shares convolution network layer more deeply, which is used to detect real-time frame of UAV imagery for recording crack image and reducing a lot of background information. And then the Haar-AdaBoost algorithm [39] is applied to locate the fine cracks from the cropped bounding boxes. Lastly, a local threshold segmentation method is used to segment the crack at the pixel level in this located fine region. In this method, the time cost of data set to build and target recognition model training is much lower, and it has good robustness for crack ROI extraction in complex background.

3.2.1. Crack localization network

Firstly, the R-FCN target recognition network is used for obtaining the boundary frame of bridge crack images to eliminate the background effects, instead of judging directly by pixel-level semantic segmentation network. The R-FCN network is composed

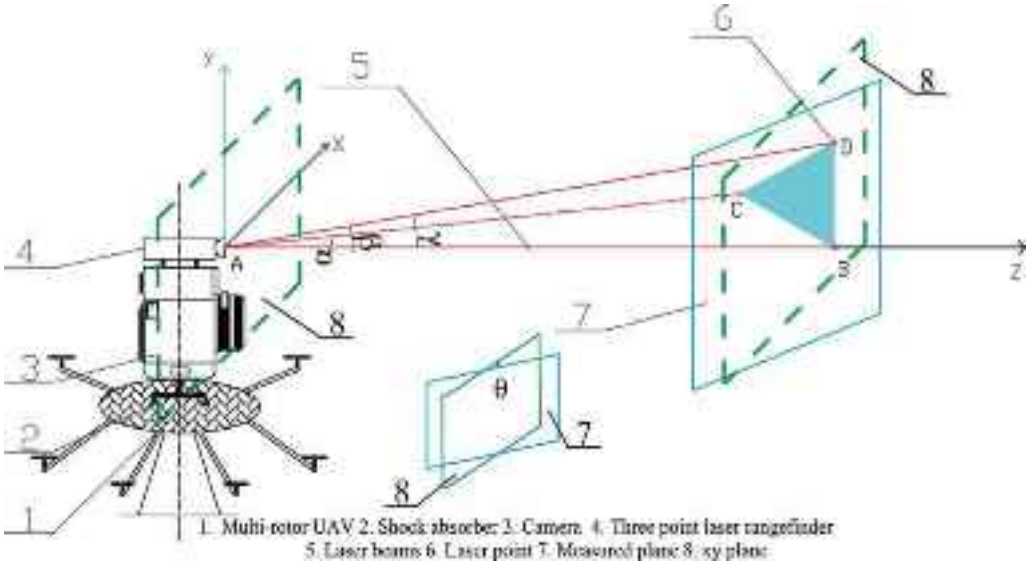

Fig. 5. Principle of three-point laser rangefinder.

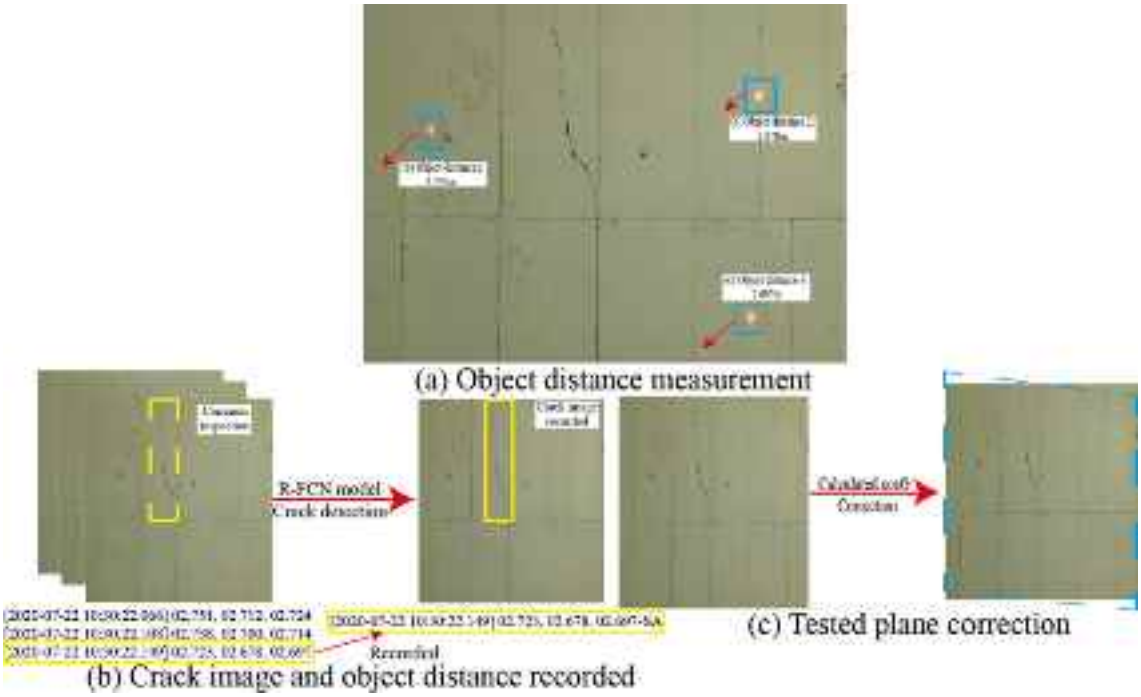

Fig. 6. The process of object distance measurement and tested plane correction.

of basic full convolution network (FCN), regional generation network (RPN) and RoI (region of interest) subnetwork. The FCN network is used to extract image features and the RPN network generates ROI based on these features.

3.2.1.1. Basic FCN. With the development of deep learning theory, there are many deep neural network models, including AlexNet [38] ZF [39] VGG [40] GoogleNET [41] and ResNet [42] among which each network can build different depth network models by designing different weight layers. Although deeper network may bring higher accuracy, it will reduce network training and detection speed. Due to the residual structure does not increase the model parameters, it can effectively alleviate the problem of gradient disappearance and training degradation of deep network training to improve the network convergence performance. Therefore, this paper uses FCN based on ResNet as the basic network.

3.2.1.2. Region Proposal Network (RPN). The input of the RPN network is a convolution characteristic graph. In this paper, three area scales and three aspect ratios are combined to form nine reference windows (32*32, 64*64, 128*128, 0.5, 1, and 2) when the region of interest (RoI) is generated by RPN network. The accuracy of ROI output depends on the prediction of target position by a reference window to realize the accurate identification of airborne bridge cracks.

Fig. 7. The examples of bridge crack with three background types.

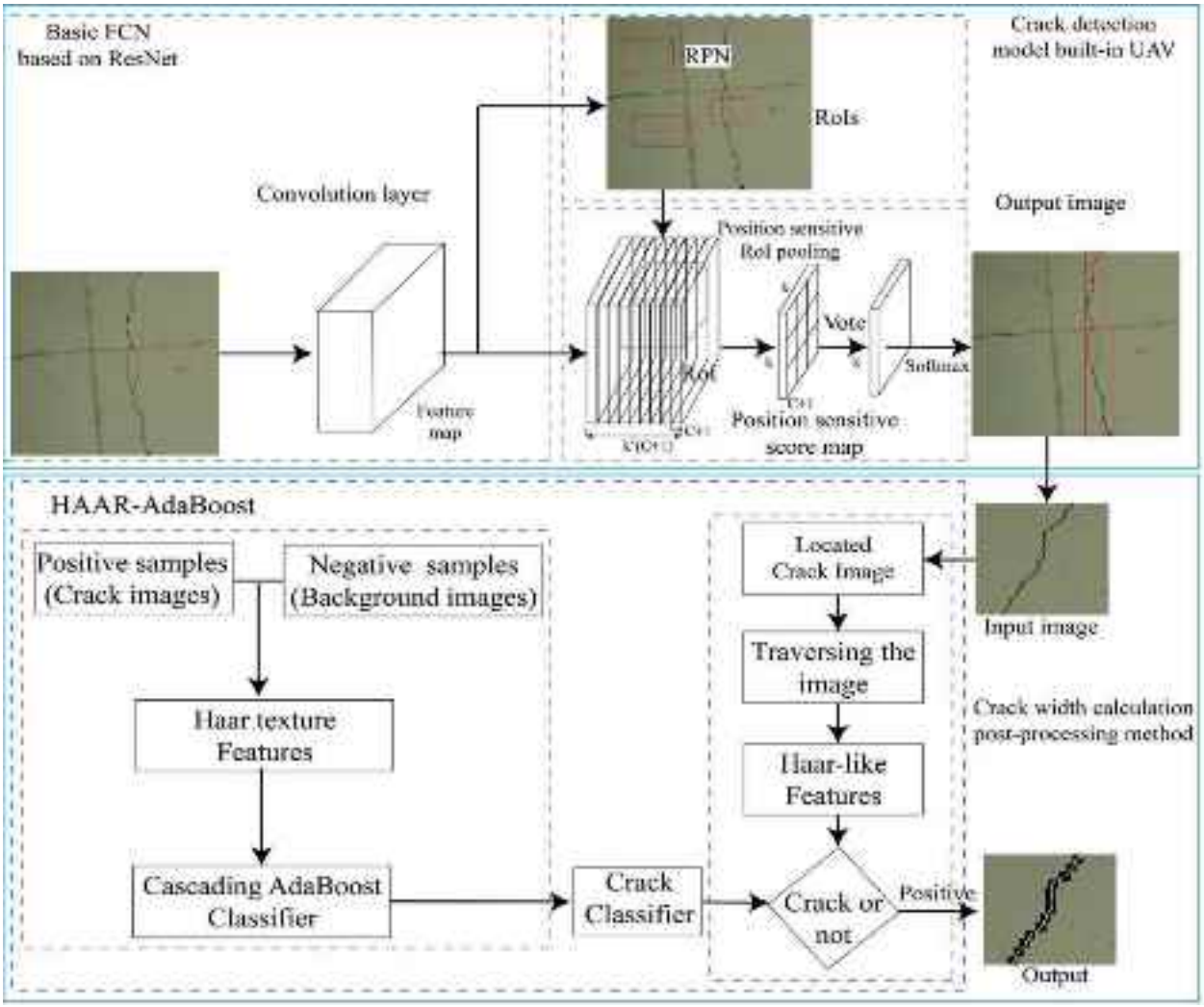

Fig. 8. Recognition framework for bridge crack width.

3.2.1.3. ROI subnet. The position sensitive fractional map the dimension of $k^2(C+1)$ where k^2 means that the ROI is divided into k^2 spatial location by the pooling layer, and C denotes the number of target species generated by the last convolution layer of FCN network. For a ROI with $w \times h$ size obtained by RPN network, it is first divided into $k \times k$ sub-regions, with the size of $w/k \times h/k$ and the pooling response of ROI's (i,j) sub-regions ($0 \leq$ i. j $\leq$ k-1). Then the probability of ROI belonging to each category can be determined by calculating the Softmax response. The category and the position parameter of ROI region is calculated by ROI subnet. is the upper-left coordinate and, and, and are respectively the width and height of the defect area.

3.2.2. Crack extraction

There are three kinds of commonly image texture acquisition technologies, such as Haar [36] LBP [43] and HOG [43] which has been applied in civil engineering in recent year [44,45]. Haar-like feature is a common feature description operator, which reflects the change of image gray level. The extended basic rectangular Haar features are used to constructed the training model as shown in Fig. 9. To realize an accurate detection, a cascading classifier is developed to detect bridge cracks based on Haar-like features. As shown in Fig. 10, the tested bridge crack images are scanned by detection window, which moves and scans whole image with a pre-defined number of pixels, which is judged by the cascading classifier. Finally, the positive detection window of bridge crack is segmented using local threshold.

3.2.2.1. Haar feature calculation. Haar feature value is a weighted subtraction of the sum of the pixel values under the white rectangle and the pixel value under the black rectangle in each feature block. As shown in Fig. 11, by changing the size and position of the feature template, large number of features can be listed in the image sub window to achieve the purpose of identifying objects. In the detection window, the number of general rectangular features is as follows:

$$\min\ XY\left(W+1-W\frac{X+1}{2}\right)\left(H+1-h\frac{Y+1}{2}\right) \tag{3}$$

In order to improve the calculation speed and enhance the real-time performance of the algorithm, the integral image algorithm is introduced to realize the fast extraction of Haar-like features. As shown in Fig. 11, for any point I (x, y) in the image, the integral graph containing all pixels in the upper left corner of the point can be defined as the formula.

3.2.2.2. AdaBoost classifier. AdaBoost is an iterative algorithm, which is realized by changing the data distribution. The weight of each sample is determined according to that the classification of sample in training set and the accuracy of the last overall classification. The boosting is a process from weak learning to strong learning, and AdaBoost is an adaptive boosting algorithm based on boosting. This algorithm cascades a series of weak classifiers to form a strong classifier as shown in Fig. 11.

$$\min\ Sum_D = \prod(x_3,y_3) - \prod(x_1,y_1) - \prod(x_2,y_2) + \prod(x_4,y_4) \tag{4}$$

- Weak classifier
- In this paper, the decision tree model is used as a weak classifier
- Constructing crack image training set, then the Haar feature of each training samples are measured and sorted to get a high-dimensional eigenvector.
- Scanning the ordered eigenvalues, and calculate the following four quantities of each element: the weight and P1 of all positive samples; the weight and N1 of all negative samples; the weight and P2 of positive samples before this sample; the weight and N2 of negative samples before this sample;
- Calculate the classification error of each element

$$\text{Loss} = \min(P2 + (N1 - N2)),\ (N2 + (P1 - P2)) \tag{5}$$

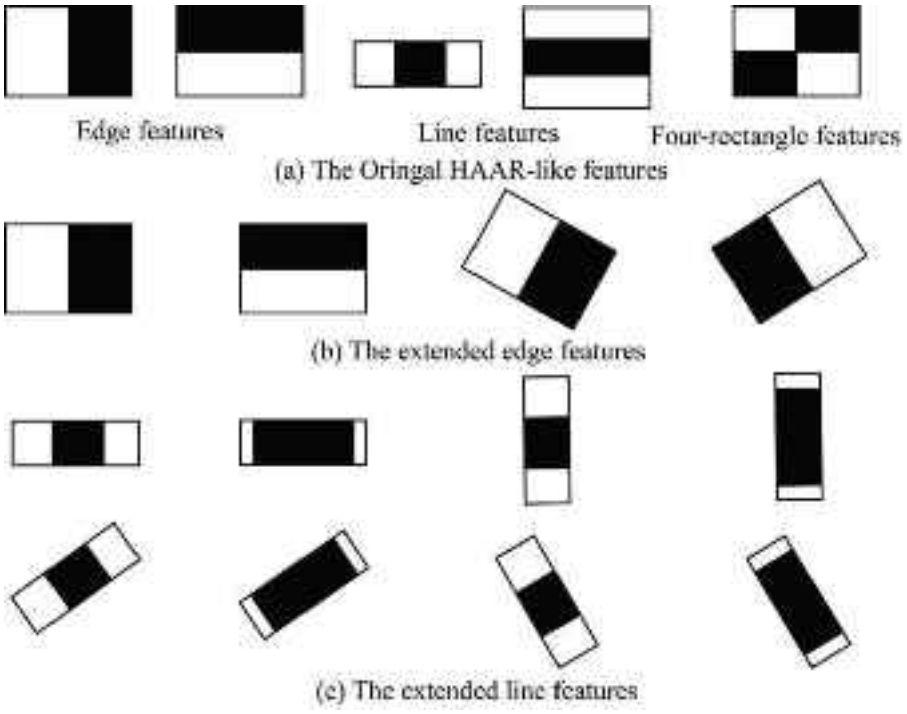

Fig. 9. Harr-like features set.

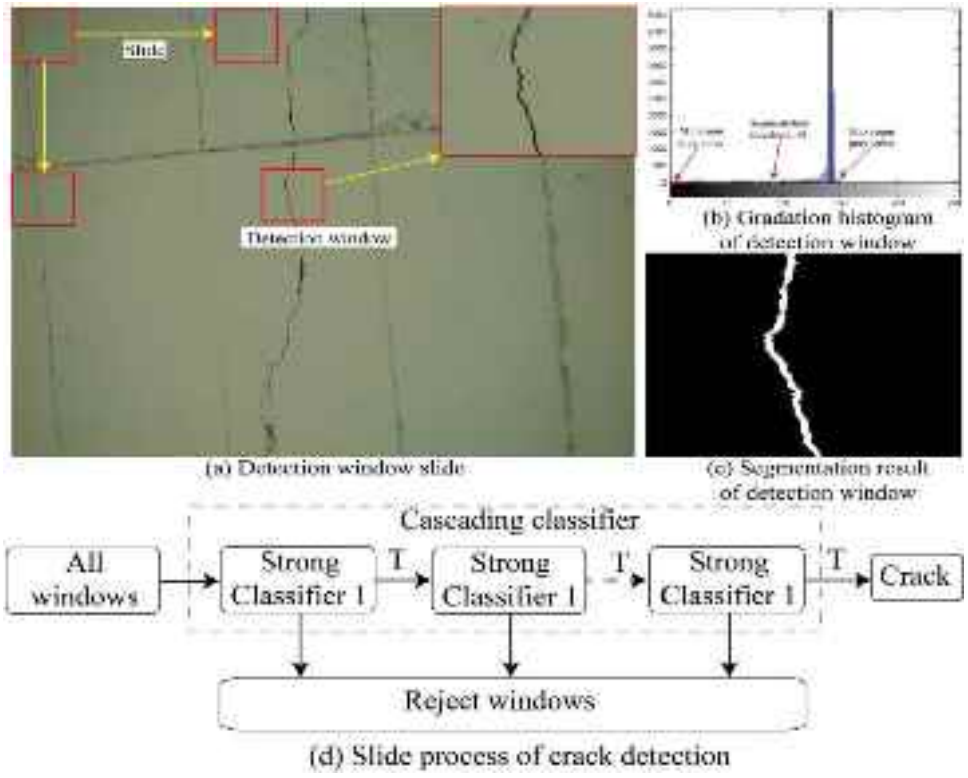

Fig. 10. Framework of bridge crack extraction based on Harr-local threshold.

- Search for the element with the lowest loss value, which is considered as the best threshold value, thus resulting in the best weak classifier.
- Strong classifier

The specific design process of strong classifier is as follows:

1) Giving n training data sets, the number of training iterations (weak classifiers) is set;

2) By iterative training, the optimal weak classifier and error rate e^m are obtained. According to e^m, each weight is updated and the weight coefficient of weak classifier is calculated.

$$\partial_m = \frac{1}{2}\log\frac{1-e^m}{e^m} \tag{6}$$

$$\omega_{m+1}, i = \frac{w_{m,i}}{z_m}\exp(-\partial_m y_i G_m(x_i)), i = 1, 2, \ldots N \tag{7}$$

$$z_m = \sum_{i=1}^{N} \omega \exp(-\partial_m y_i G_m(x_i)) \tag{8}$$

In which, $z_m = \sum_{i=1}^{N} \omega_{i+1}(x_i) = 1$ normalization factor, $w_{m,t}$ is initialize weights.

3) After getting the updated weights, the new samples are fused with the samples of the previous round of error recognition to form a new sample set, and then continue to learn again;

4) Repeat steps 2)–3), stop training when the number of iterations reaches the set value or the error rate is lower than the set value.

3.2.2.3. Local threshold determining. After fine locating, most of the regional background in crack image has been removed. Then the local threshold segmentation algorithm is applied to segment bridge crack, which is calculated by the threshold iteration. As shown in Fig. 10 (b), we can find minimum and maximum gray value on the gradation histogram, which are recorded as $I_{\min}$ and $I_{\max}$. The specific calculation process of local segmentation threshold is as follows:

1) The initial segmentation threshold is calculated as:

$$I_0 = (I_{\min} + I_{\max})/2 \tag{9}$$

2) The pixel of image is divided into G_1 and G_2 by threshold I_0, the gray value of G_1 is greater than I_0, otherwise it belongs to G_2;

3) The gray mean values of all pixels in G_1 and G_2 are calculated respectively u_1 and u_2;

4) The segmentation threshold is updated as;

$$I_1 = (u_1 + u_2)/2 \tag{10}$$

5) Repeat steps 2–4 until the change of the mean value of the two adjacent times is less than the limit value ∇I;

Finally, the local segmentation threshold I is determined, shown in Fig. 10 (b).

3.3. Crack width calculation

As shown in Fig. 12 (a), the shortest distance between two corresponding edge lines of crack is defined as crack thickness [33]. Moreover, the two edges of the crack are connected by the pixels with a pixel values1. The crack width at point (x_n, y_m) is defined as the minimum distance between the crack edges at that point. The distance from the pixel point (x_n, y_m) on one crack edge to the other crack edge is:

$$S = \sqrt{(x'_{n-l} - x_n)^2 + (y' - y_m)^2} \tag{11}$$

where x'_{n-l} is the x-coordinate of $(n-l)$ column on the other edge of crack, y' is the y-coordinate corresponding to x'_{n-l}, and l is a value taken between −5 and 5. The minimum S value calculated by Eq. (10) is taken as the crack width of point (x_n, y_m). To quantitatively evaluate the safety status of bridges, the method of actual crack width

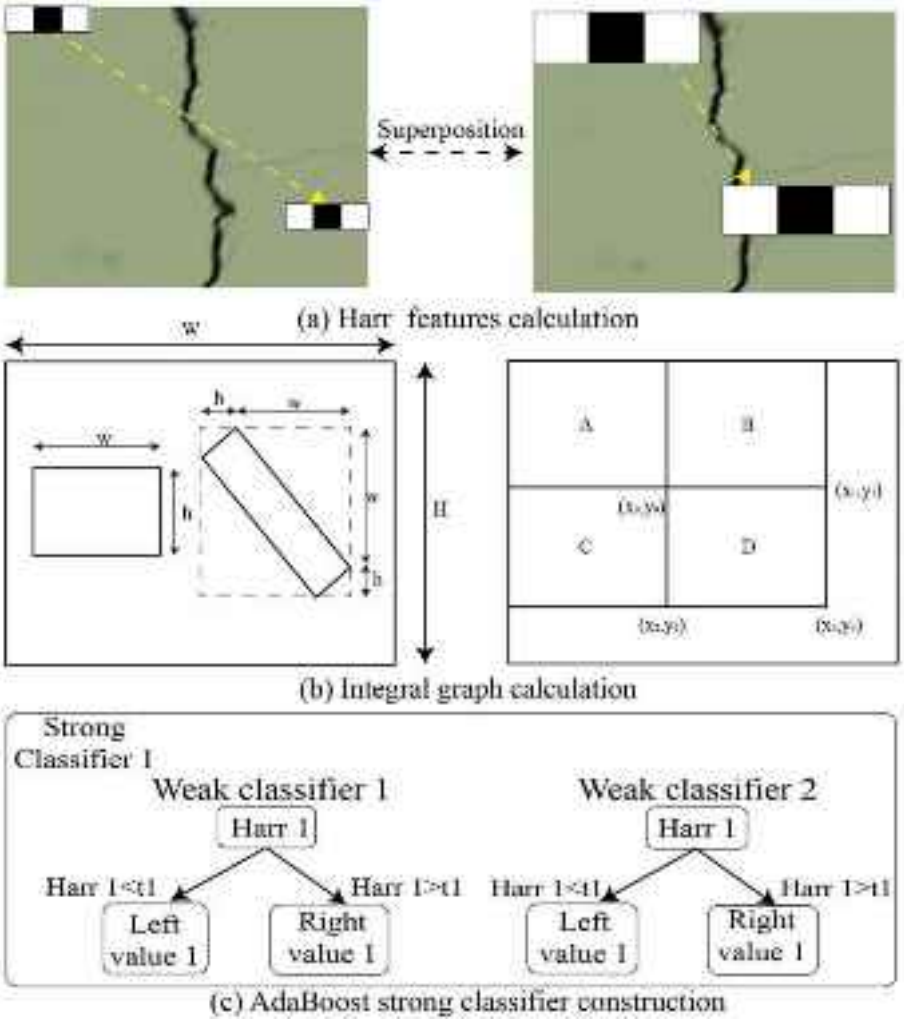

Fig. 11. Haar feature calculation and AdaBoost.

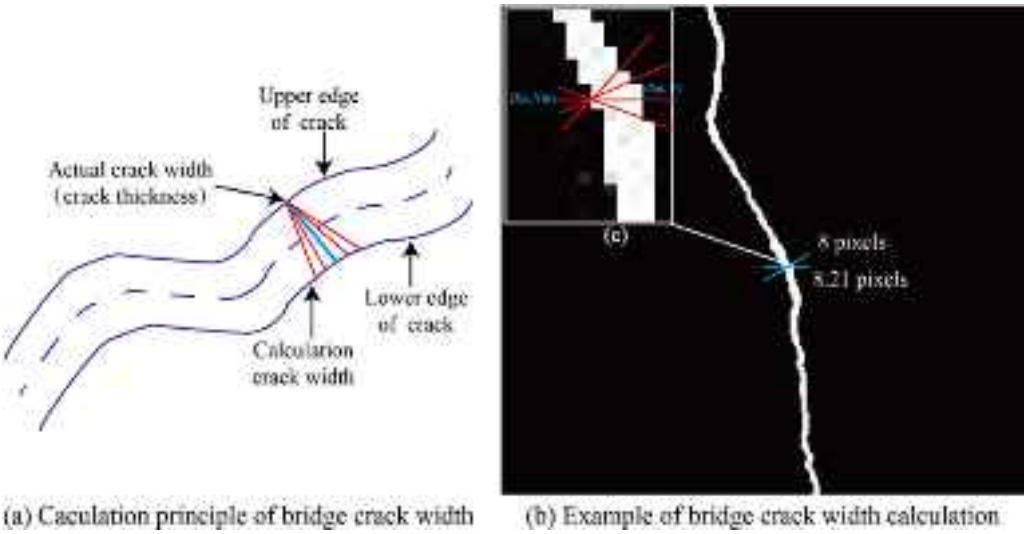

Fig. 12. The crack thickness calculation method.

recognition based on proportional conversion is proposed, which uses the three-point laser rangefinder synchronized with the shutter to measure the real-time distance between the camera and bridge.

4. Case study

This paper has developed a novel UAV-based machine vision method for bridge crack recognition and width quantification through hybrid feature learning. To validate this method, a case study of bridge crack inspection based on UAV is carried out. Then, the proposed method is trained and tested using the collected data set, and the tested result is compared with the well-known networks. Finally, we have carried out a manual crack width measurement based on bridge-inspection vehicle platform to demonstrate the accuracy of this proposed method.

4.1. Data acquisition and data set build

As shown in Fig. 13, the proposed method and system are applied on Xiang-Jiang River bridge located in Hunan province. The equipment parameters of UAV system are shown in Table 1. During this inspection, the operator controls the UAV system at a low speed of 0.6 m/s. There are initially 3548 images collected from airborne imagery in this inspection, which is divided into several categories shown in Table 2.

In order to enhance the robustness of the model, the influence of component position and bridge type on crack identification is comprehensively considered in the crack image data set making. The crack recognition data set contain 3 backgrounds (normal, aged, painted) has been taken with different imaging distances. All the images are annotated with bounding boxes using the graphical image annotation tool "LabelImg" [46]. After data augmentation, there are 3540 images used for training the model as shown in Table 3, and the examples of annotated crack image are shown in Fig. 14.

Table 1
Equipment parameters.

Hardware/System	Specifications/Parameters
UAV	EHM-6 V
Camera	SONY A7R2
Camera lens	80 mm
Effective pixels	7952 × 5304 pixels
CMOS sensor sizes	36 × 24 mm
Laser rangefinder angles	$\partial = 5^{\circ}, \beta = 5^{\circ}, \gamma = 7^{\circ}$

Table 2
Information of bridge crack inspection case.

Items	Flight time	Crack images	Clear images
North approach bridge (20 spans)	20*20mins	2920	2705
Main bridge	6*20 min	872	856
South approach bridge (18 spans)	18*20mins	3553	3506

Fig. 13. The bridge crack inspection process based on UAV.

Table 3
The target recognition image data set of bridge cracks.

Items	No. of images	Size (pixels)	Training	Validation
Normal background	2334	7952*5304	1867	467
Aged background	670	7952*5304	540	130
Painted background	536	7952*5304	428	108
total	3540	7952*5304	2835	705

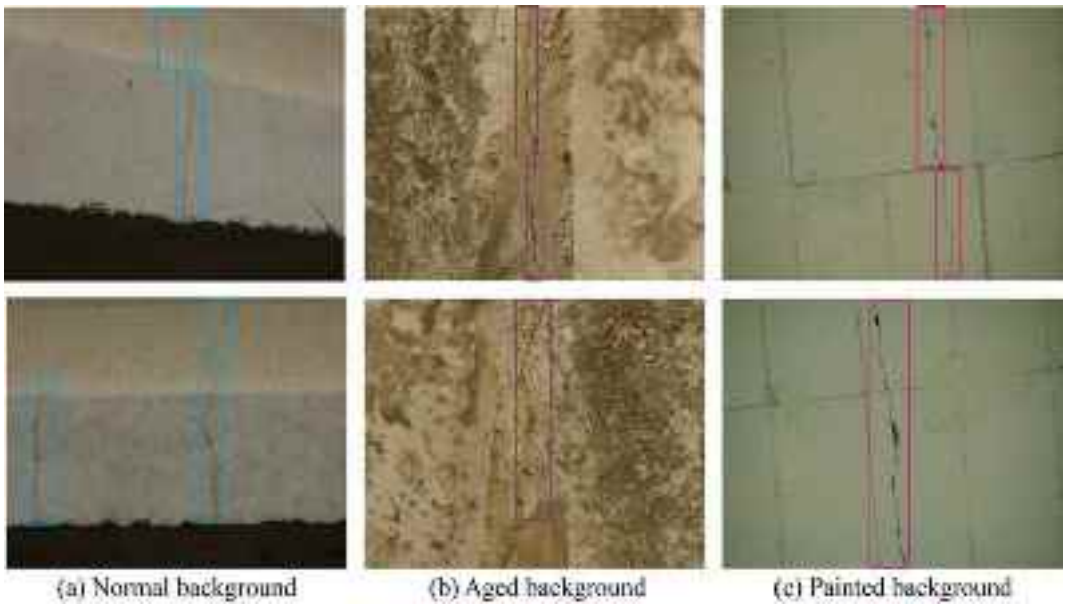

Fig. 14. Annotated examples of target recognition image dataset.

4.2. R-FCN training and validation

The training curve of total accuracy, classification loss of RPN, loss of detector classifier is shown in Fig. 15. In fact, when the iterations are above 1000, the model has been to converge. Furthermore, the result examples of bridge cracks detection with normal background, aged background, and painted background are shown in Fig. 16. As shown in Table 4, this method effectively recognizes and locates bridge cracks under complex conditions such as shadow, concrete spalling, bridge framework imprint, and so on, which can achieve above 95% average precision (AP) for tested images. To realize the crack detection and recorded in real-time [47] the well-trained R-FCN model is built in the UAV transmission system by the partner UAV company.

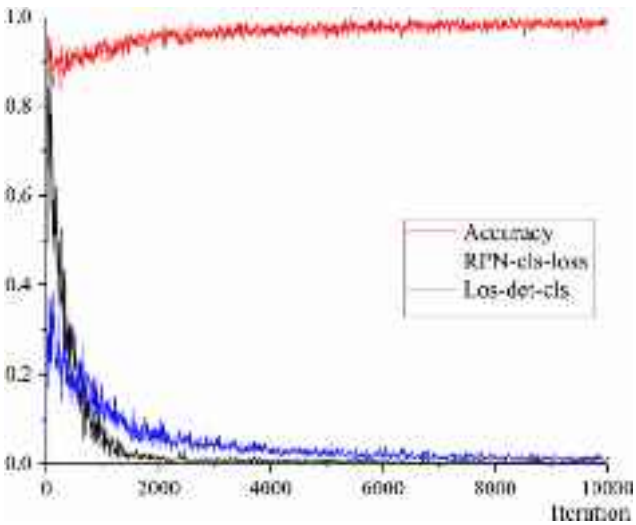

Fig. 15. The loss-iterations curves of the R-FCN network.

4.3. Haar-AdaBoost results

The Haar-AdaBoost data set is formed from the crack images classified and located by the R-FCN network. As shown in Fig. 17, the positive bridge crack samples are divided into three categories according to the above background conditions, which are set as 40×40. The negative samples are mainly typical bridge background, which is also divided into three categories. After manual cropping, there are 1500 positive samples and 1500 negative samples used for training the model. Due to the cracks have continuous texture features, the search step size should not be too large. The detection window moves with a pre-defined number of pixels, which is set to 20 (half of detection window sizes) for movements in both horizontal and vertical directions.

As shown in Fig. 18, the examples of crack image are extracted using the proposed method. In the process of crack segmentation, due to the application of Haar-like features for crack fine locating, the gray distribution in the cropped bounding boxes is almost single. Therefore, the local threshold of gray distribution histogram can be used to segment crack directly. The prediction results show that the Harr-AdABoost method has good robustness and generalization ability for noise interfering, such as concrete spalling and pitted surface.

4.4. Evaluation of hybrid features learning method

To evaluate the performance of the proposed method in crack prediction accuracy and time cost. Compared with the proposed hybrid features learning method, we use the existing segmentation

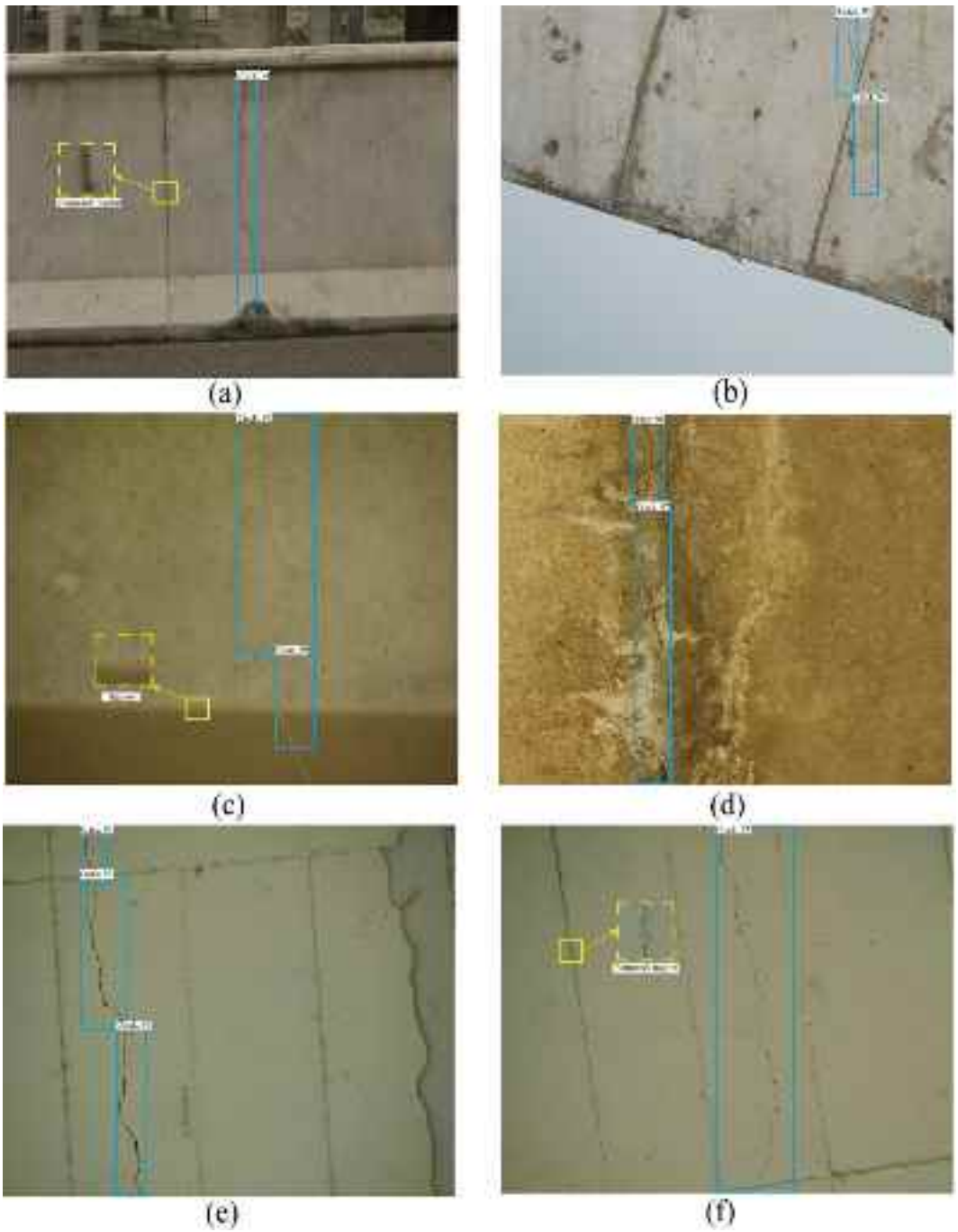

Fig. 16. Examples of R-FCN network recognition results.

Table 4
Performance of the R-FCN model.

Background type	Normal background	Aged background	Painted background
AP	98.8	96.1	95.2

network named Mask R-CNN and U-Net [48] to extract crack, which is trained using the publicly available crack datasets [27]. To evaluate the performance of these methods quantitatively, the 100 test images formed the data set concluding 3 categories (normal background 40 images, aged background 30 images, painted background 30 images) from located cracks area are inputted into the framework, and its size depended on the crapped detection area of R-FCN model shown in Fig. 16. And the examples of tested results are shown in Fig. 19 and Table 5. And the four examples of tested image in crack data set are presented and analyzed in detail in Figs. 20–23 and Tables 6–9.

4.4.1. Crack prediction accuracy evaluation

4.4.1.1. Crack width prediction. Crack width is the most important indicator for evaluating the health condition of the bridge, especially for the concrete bridge structure. As an accurate UAV-based machine vision method, the accuracy of crack pixel width prediction is the most important indicator to evaluate this method. As shown in Figs. 20–23, the crack segmentation parameters include area, average width and thickness of several points are shown in Table 6–9, and the overall accuracy of crack pixel width

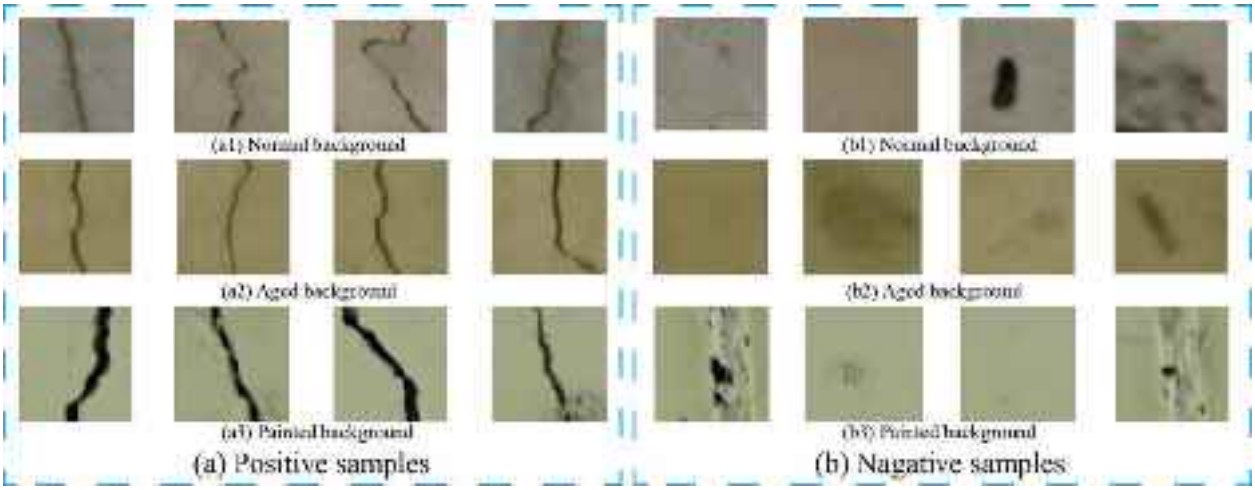

Fig. 17. Positive and negative samples of crack classifier training.

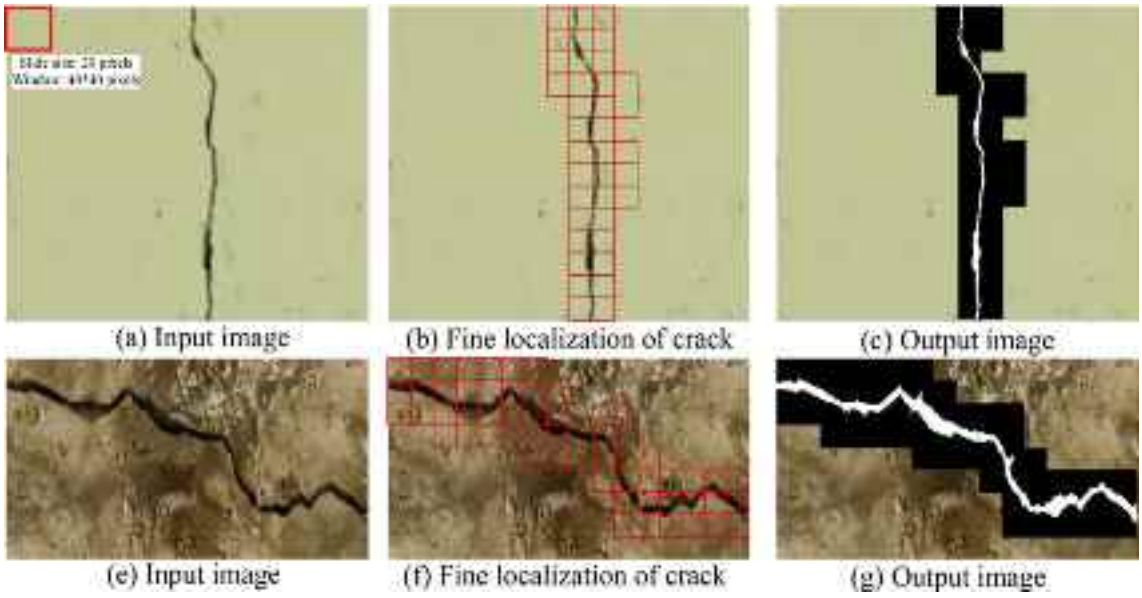

Fig. 18. Segmentation results based on Haar-local threshold method; (a)-(c) is the crack recognition process of painted background and (e)-(g) is aged background.

prediction is shown in Table 5. Those results indicate that the average accuracy of crack width prediction by proposed method is about 90%, which is almost consistent with the ground truth. Also, the accuracy of U-Net in the general situation is about 86% shown in Fig. 20 and Fig. 21, while the crack width segmentation result is quite different in some special cases with blurred edges shown in Fig. 22 and Fig. 23, which is common in the process of UAV-based bridge inspection. On the other hand, the crack width segmentation result of Mask-RCNN is quite large than ground-truth in all the situation, which is not suitable for the crack quantification of UAV-based bridge inspection.

4.4.1.2. IoU prediction. To quantify the accuracy performance of crack pixel level segmentation method, the statistical indicator, IoU (Intersection over Union) is used to compared with the ground-truth. The IoU is defined as [41]:

$$C(X_C, Y_C, Z_C) \tag{12}$$

where TP and FP are the positive and negative samples predicted by the model respectively and FN is a positive sample predicted to be negative by the model. The examples of the comparative segmentation results are presented in Figs. 20–23. And all the 100 tested results from the crack image data set is formed in Table 5 showing that the proposed method achieves about 90% average IoU, while the Mask R-CNN and U-Net achieve 59% and 72% IoU respectively. Therefore, the test result demonstrate that our method has higher precision ability in complex crack background obtained by UAV.

However, the crack images of UAV-based bridge inspection are affected by weathering, UAV vibration, illumination and so on, the edges between the crack and background usually are unclear shown in Fig. 22(f) and Fig. 23(f). As shown in Fig. 22 and Fig. 23, our method has better robustness in special cases with blurred edges, which can also achieve above 85% accuracy, while the U-Net and Mask R-CNN extraction method is not effective for crack segmentation in UAV-based special situation. The test result demonstrate that our method uses R-FCN to detect and locate crack region instead of segmentation directly, which has higher precision ability and robustness in complex crack background.

4.4.2. Time cost in crack detection and data set make

As a real-time crack detection and recording system for UAV, the time cost of crack detection and analysis is important. As

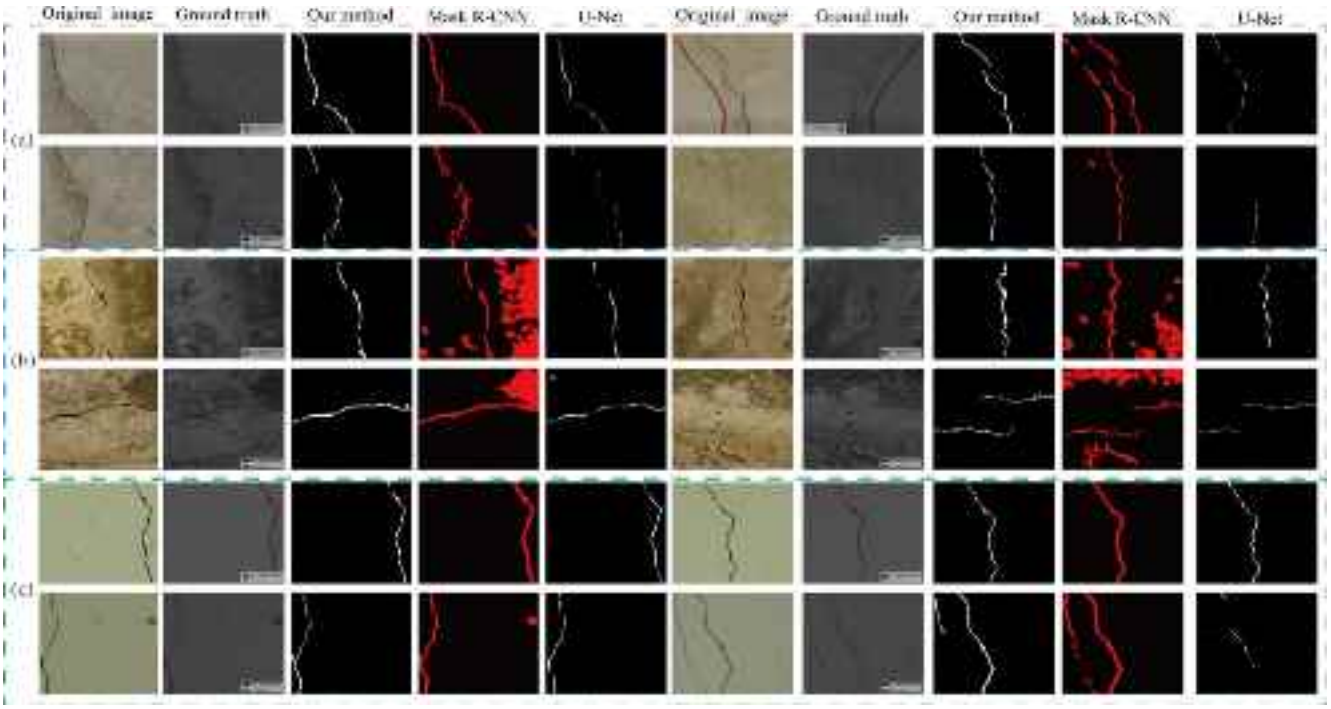

Fig. 19. Comparison of segmentation results for our method, U-Net and Mask R-CNN, (a) normal background, (b) Aged background and (c) painted background.

Table 5
Comparison of crack segmentation accuracy for proposed method, U-Net and Mask R-CNN.

Methods	Average area accuracy (%)	Average width accuracy (%)	IoU (%)
Proposed method	0.89	0.92	0.90
U-Net	0.67	0.79	0.72
Mask R-CNN	0.55	0.56	0.59

shown in Table 10, the time cost of our method is about 0.2 s in crack detection in 7952 × 5304 pixels images and 4.5 s in pixel-level segmentation per one 1000 × 1000 pixels bounding box, and the time cost of U-Net and Mask-RCNN is 1.2 s and 2.6 s respectively per one 1000 × 1000 pixels images on NVIDIA GTX1080Ti platform.

Although the total time cost of our method is not the lowest, the time cost of our proposed method is also two-stage. In the stage of crack target detection, the R-FCN network can quickly detect and locate cracks to meet the requirements of real-time crack detection

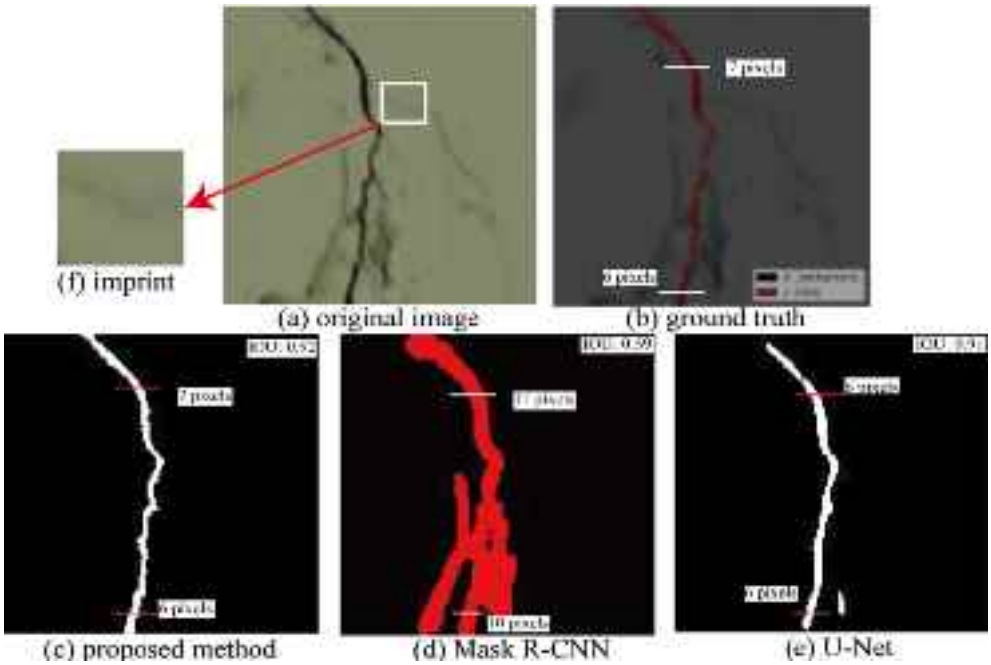

Fig. 20. Recognition results of painted background image with clear crack edge.

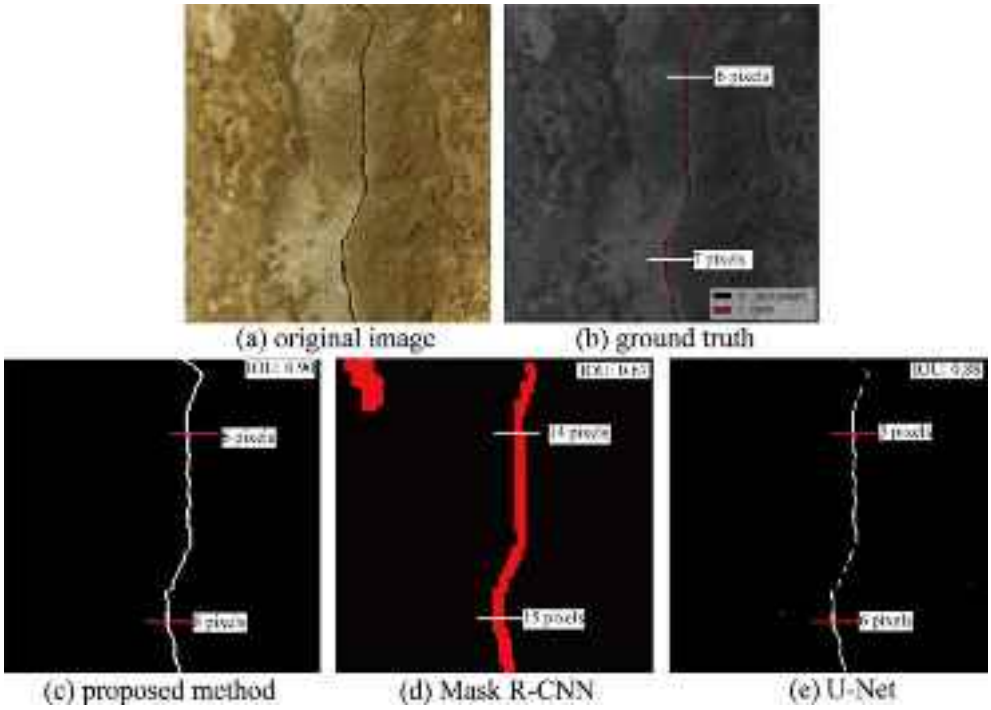

Fig. 21. Recognition results of aged background image with clear crack edge.

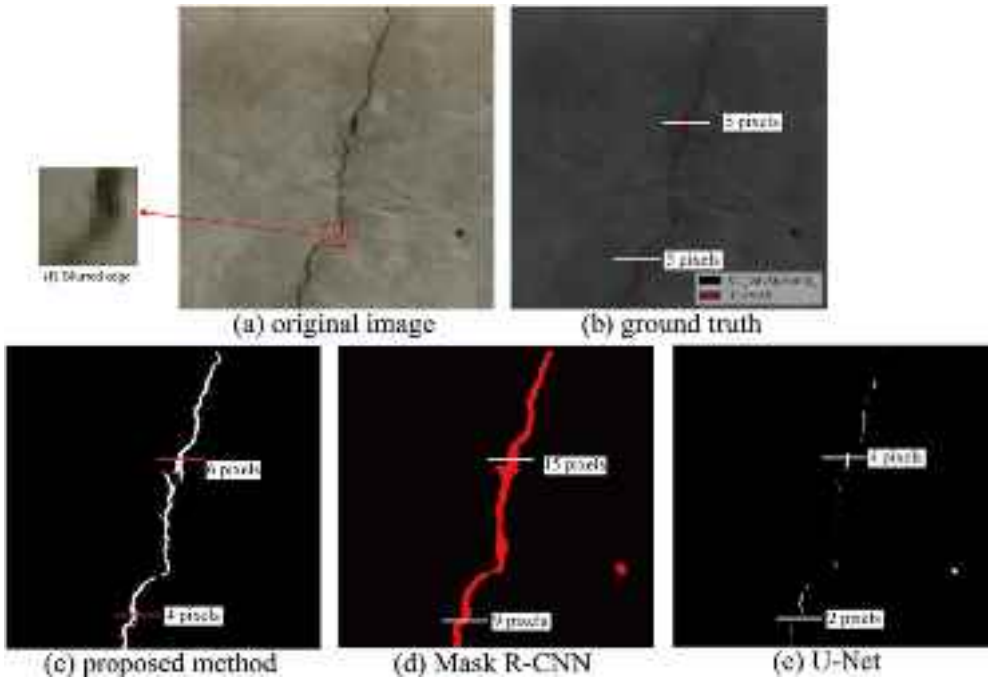

Fig. 22. Recognition results of normal background image with unclear crack edge.

based on UAV platform. And the stage of crack segmentation and analysis based on Haar-AdaBoost and local threshold is a post process to ensure the accuracy. However, the U-Net and Mask-RCNN extraction methods are single-stage pixel-level semantic segmentation algorithms, and their direct time cost cannot meet the requirements of UAV platform for rapid detection of bridge cracks.

On the other hand, the time cost of ground truth making by drawing bounding boxes is only about 10 s in labelImg platform, while the Mask R-CNN and U-Net using pixel-level marking need above 30mins in labelme platform. The proposed method can reduce the time cost of data set making, improve efficiency, and provide practical application conditions.

4.5. Crack quantification accuracy verification

To verify the crack width quantification accuracy of the proposed method based on UAV, the bridge-inspection vehicle is used as the manual operation platform to measure the crack width in

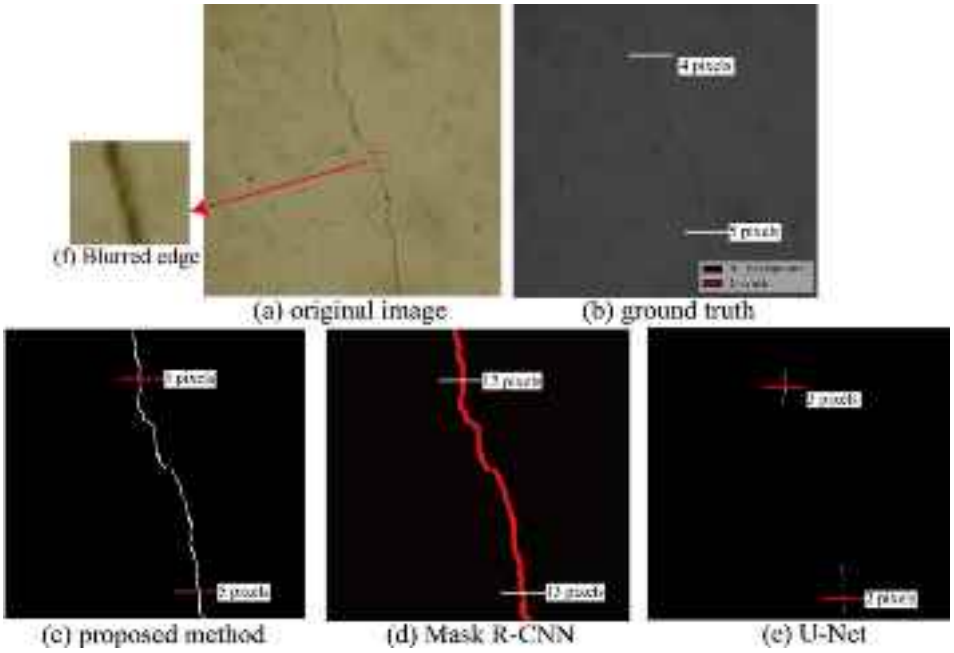

Fig. 23. Recognition results of aged background image with unclear crack edge.

Table 6
Segmentation results of Fig. 20.

Methods	Crack area pixels	IoU	Average width	Crack width of point 1	Crack widths of point 2
Proposed method	584 pixels	0.92	3.79 pixels	7 pixels	6 pixels
Mask R-CNN	1875 pixels	0.59	8.56 pixels	11 pixels	10 pixels
U-Net	578 pixels	0.91	4.3 pixels	6 pixels	6 pixels
Ground truth	523 pixels	/	3.51 pixels	7 pixels	6 pixels

Table 7
Segmentation results of Fig. 21.

Methods	Crack area pixels	IoU	Average width	Crack width of point 1	Crack width of point 2
Proposed method	4513 pixels	0.90	6.02 pixels	6 pixels	8 pixels
Mask R-CNN	7220 pixels	0.67	12.29 pixels	14 pixels	15 pixels
U-Net	2784 pixels	0.88	4.92 pixels	5 pixels	6 pixels
Ground truth	4400 pixels	/	6.21 pixels	6 pixels	7 pixels

Table 8
Segmentation results of Fig. 22.

Methods	Crack area pixels	Average width	Crack width of point 1	Crack width of point 2
Proposed method	8209 pixels	5.6 pixels	6 pixels	4 pixels
Mask R-CNN	17,558 pixels	12.71 pixels	12 pixels	9 pixels
U-Net	2525 pixels	2.82 pixels	4 pixels	2 pixels
Ground truth	8979 pixels	6.08 pixels	5 pixels	5 pixels

Table 9
Segmentation results of Fig. 23.

Methods	Crack area pixels	Average width	Crack width of point 1	Crack width of point 2
Proposed method	4070 pixels	4.32 pixels	5 pixels	5 pixels
Mask R-CNN	11,311 pixels	12.08 pixels	12 pixels	13 pixels
U-Net	372 pixels	2.82 pixels	3 pixels	2 pixels
Ground truth	3724 pixels	4.23 pixels	4 pixels	5 pixels

Table 10
Training and testing information and result of average IoU.

Parameters	Ours	U-Net	Mask R-CNN
Preparation training data Time cost of sample making	Bounding box 10 s	Pixel mask 30 min	Pixel mask 30 min
Testing time	0.2 s + 4.5 s	1.2 s	2.6 s

the bottom slab of the 12th span hollow slab girder of the south approach bridge. As shown in Fig. 24, the total length of crack 1# is 11.2 m, and the total length of crack 2# is 9.8 m. Moreover, the crack width measurement instrument is applied to measure the crack 1# and crack 2# every 50 cm. As shown in Fig. 24, seven points on the two through cracks are selected for the verification experiment of quantifying bridge crack width, the manual measurement results are shown in Fig. 24(c).

On the other hand, the proposed UAV-based method and system are applied on verify experiments. The real-time object distance of inspection image is recorded, and the corresponding pixel resolution and measured plane angle are calculated by Eq (1) shown in Table 11. The pixel widths are calculated by Eq (11), and the real crack widths are calculated by Eq (2), as listed in Table 12 and Fig. 24(c). Therefore, the bridge crack recognition system results based on the UAV-based machine vision system have little difference compared with the manual measurement, which indicates that it is feasible to identify the crack width of UAV bridge.

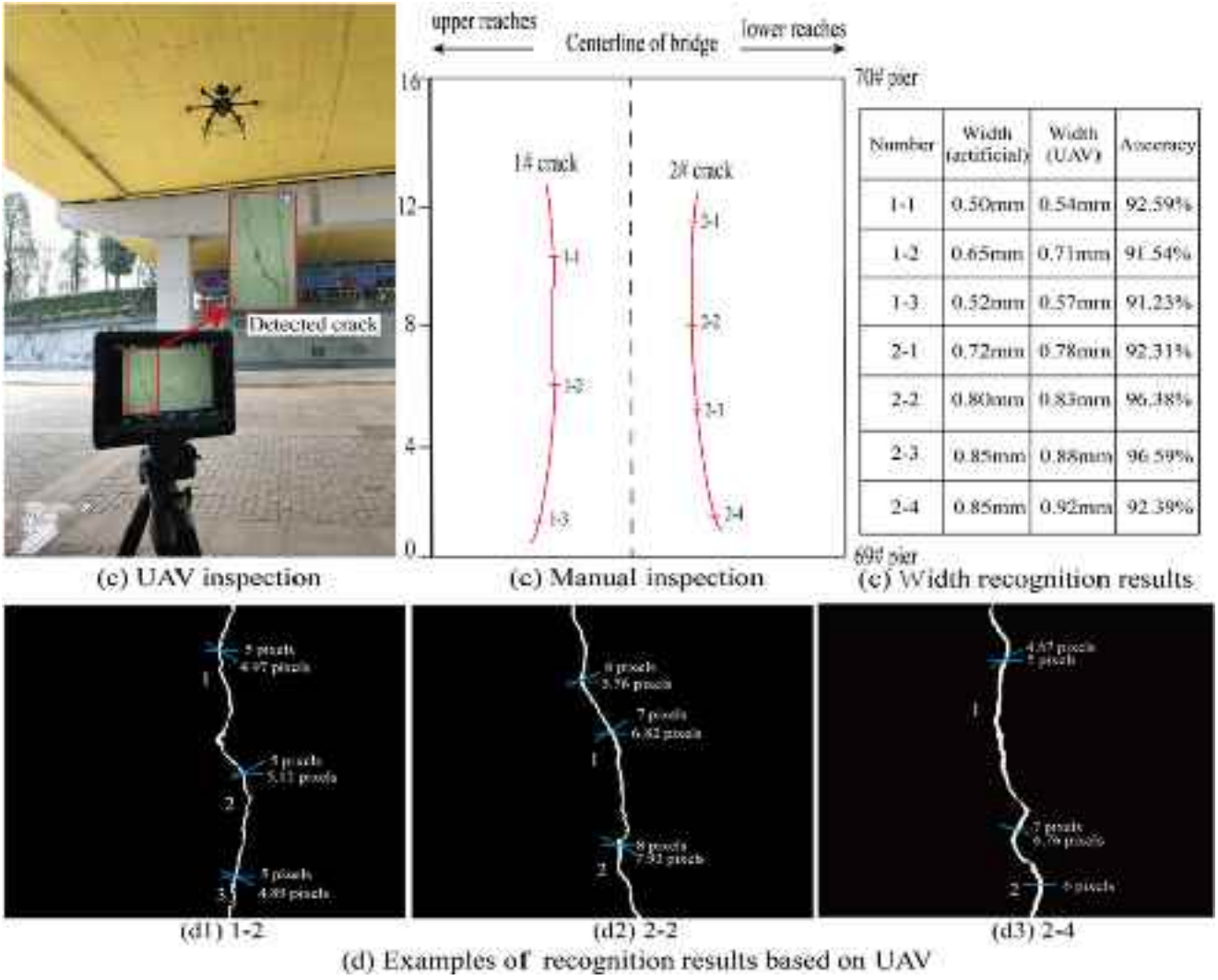

Fig. 24. Comparison of bridge inspection results for our method and manually measurement.

Table 11
Image distance measurement data.

Items	1-1	1-2	1-3	2-1	2-2	2-3	2-4
Distance AB	2.644	2.621	2.517	2.291	2.169	2.635	2.508
Distance AC	2.614	2.589	2.484	2.260	2.153	2.572	2.492
Distance AD	2.630	2.610	2.494	2.072	2.086	2.602	2.503
cosθ	0.98	0.97	0.97	0.93	0.89	0.93	0.99
Pixel resolution (mm/p)	0.142	0.139	0.134	0.116	0.105	0.133	0.136

Table 12
Examples of bridge crack width recognition.

Items	C1	C2	C3
Airborne image pixels(p)	5.12	7.93	6.76
Airborne image width (mm)	0.71	0.83	0.92
Manual measurement width (mm)	0.65	0.80	0.85
Absolute tolerance (mm)	0.06	0.03	0.07
Relative accuracy (%)	91.54	96.38	92.39

5. Conclusions

In this paper, we have developed a UAV-based machine vision method for bridge crack recognition and width quantification. The following findings and conclusions are offered:

1. The UAV-based machine vision system with a special upward platform equipped Sony ILCE-7RM2 DSLR camera has been applied to bridge crack measurement, which can obtain crack image (effective pixel 7952x5304), image distance and GPS position simultaneously. By bridge crack inspection frame detection and crack pixel resolution calculation, the bridge crack image and corresponding analysis data has been recorded and saved into an SD card at the same time.
2. We have proposed a bridge crack recognition and width quantification method combining neural network and traditional feature learning in view of the complex bridge background. In this method, the bridge crack can be detected and recoded based on a built-in R-FCN network in real-time, and the bridge crack can be segmented in pixel-level with high-accuracy based on Haar-AdaBoost learning and local threshold. Based on the large number of UAV bridge inspection flights, the crack target recognition data set including 3540 images (7952×5304 pixels) and the Haar-like feature learning data set including 3000 images are established. The bridge crack width recognition based on the proposed method is shown to have good performance, in which the AP of R-FCN is more than 95%, and the IoU of Haar-local threshold segmentation method is about 90%. The time cost of our method is about 0.2 s for crack detection per one 7952×5304 pixels images and 4.5 s in pixel-level segmentation per one 1000×1000 pixels bounding box.
3. The verification experiment of the UAV-based method for crack detection on a real bridge has been carried. Compared with the manual detection results, the accuracy of proposed bridge crack width recognition method is more than 90%, showing good reliability to bridge health condition evaluation.

CRediT authorship contribution statement

Xiong Peng: Methodology, Software, Data curation, Writing - original draft. **Xingu Zhong:** Supervision, Investigation, Project administration, Funding acquisition. **Chao Zhao:** Software, Writing - review & editing. **Anhua Chen:** Supervision. **Tianyu Zhang:** Software, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

This study is supported by National Natural Science Foundation of China (Grant No. 51678235), to which the authors are grateful.

References

[1] Research Institute of highway science. Standards for technical condition evaluation of highway bridges. Communications Press, 2011.
[2] AASHTO, Manual for Maintenance Inspection of Bridges, AASHTO, Washington, DC, 1970.
[3] G. Sterritt, Review of Bridge Inspection Competence and Training Project Rep. Research Project UG637, UK Bridges Board, London, 2009.
[4] T. Rakha, A. Gorodetsky, Review of unmanned aerial system (UAS) applications in the built environment: towards automated building inspection procedures using drones, Autom. Constr. 93 (2018) 252–264, https://doi.org/10.1016/j.autcon.2018.05.002.
[5] S. Yoon, H.J. Jung, I.H. Kim, Bridge Inspection and condition assessment using Unmanned Aerial Vehicles (UAVs): Major challenges and solutions from a practical perspective, Smart Struct. Systems 24 (5) (2019) 669–681, https://doi.org/10.12989/sss.2019.24.5.669.
[6] D. Kang, Y.-J. Cha, Autonomous UAVs for Structural Health Monitoring Using Deep Learning and an Ultrasonic Beacon System with Geo-Tagging, Comput.-Aided Civ. Infrastruct. Eng. 33 (10) (2018) 885–902, https://doi.org/10.1111/mice:2018.33.issue-1010.1111/mice:12375.
[7] N. Hallermann, G. Morgenthal, Unmanned Aerial Vehicles (UAV) for the Assessment of Existing Structures [C], IABSE Symposium Report 101 (2013), pp. 1–8 (14), doi: 10.2749/222137813808627172.
[8] F.C. Pereira, C.E. Pereira. Embedded Image Processing Systems for Automatic Recognition of Crack using UAV [J], 48 IFAC-Papers On Line, 2015, pp. 16–21 (10), doi: 10.1016/j.ifacol.2015.08.101.
[9] S. Sankarasrinivasan, E. Balasubramanian, K. Karthik, U. Chandrasekar, R. Gupta, Rishi Guptac, Health monitoring of civil structures with integrated UAV and image processing system, Procedia Comput. Sci 54 (2015) 508–515, https://doi.org/10.1016/j.procs.2015.06.058.
[10] L. Wang, Z. Zhang, Automatic detection of wind turbine blade surface cracks based on UAV-taken images, IEEE Trans. Ind. Electron. 64 (9) (2017) 7293–7303, https://doi.org/10.1109/TIE.2017.2682037.
[11] D. Sattar, R.J. Thomas, M. Marc, Fatigue crack detection using unmanned aerial systems in fracture critical inspection of steel bridges, J. Bridge Eng. 23 (10) (2018) 04018078, https://doi.org/10.1061/(ASCE)BE.1943-5592.0001291.
[12] H. Kim, J. Lee, E. Ahn, S. Cho, M. Shin, S.-H. Sim, Concrete crack identification using a UAV incorporating hybrid image processing, Sensors 17 (9) (2017) 2052, https://doi.org/10.3390/s17092052.
[13] S.E. Park, S.H. Eem, H. Jeon. Concrete crack detection and quantification using deep learning and structured light[J]. Construction and Building Materials, 252. doi: 10.1016/j.conbuildmat.2020.119096.
[14] X. Zhong, X. Peng, S. Yan, M. Shen, Y. Zhai, Assessment of the feasibility of detecting concrete cracks in images acquired by unmanned aerial vehicles, Autom. Constr. 89 (2018) 49–57, https://doi.org/10.1016/j.autcon.2018.01.005.
[15] Y.-J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Comput. Aided Civil Infrastructure Eng. 32 (5) (2017) 361–378, https://doi.org/10.1111/mice:2017.32.issue-510.1111/mice:12263.
[16] Y.J. Cha, W. Choi, G. Suh, Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, Comput.-Aided Civ. Infrastruct. Eng. 33 (4) (2018) 1–17, https://doi.org/10.1111/mice.12334.
[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg. SSD: Single Shot MultiBox Detector. arXiv:1512.02325. doi: 10.1007/978-3-319-46448-0_2
[18] S. Ren, K. He, R. Girshick, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031.
[19] J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You Only Look Once: Unified, Real-Time Object Detection. arXiv:1506.02640. doi: 10.1109/CVPR.2016.91
[20] J. Yang, Q. Fu, M. Nie, Road crack detection using deep neural network with receptive field block, IOP Conference Series Materials Science and Engineering 782 (2020), https://doi.org/10.1088/1757-899X/782/4/042033 042033.
[21] H. Ju, W. Li, S. Tighe, J. Zhai, Z. Xu, Y. Chen, Detection of sealed and unsealed cracks with complex backgrounds using deep convolutional neural network, Autom. Constr. 107 (2019) 102946, https://doi.org/10.1016/j.autcon.2019.102946.
[22] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2017), https://doi.org/10.1109/CVPR.2015.7298965.
[23] O. Ronnebergerl, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv: 1505.04597v1, 18 May 2015. doi: 10.1007/978-3-319-24574-4_28.
[24] L.C. Chen, G. Papandreou, I. Kokkinos, et al., DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, https://doi.org/10.1109/TPAMI.2017.2699184.
[25] L.C. Chen, G. Papandreou, F. Schroff, et al. Rethinking Atrous Convolution for Semantic Image Segmentation, 2017. https://arxiv.org/abs/1706.05587
[26] A. Zhang, K.C.P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network, Comput.-Aided Civ. Infrastruct. Eng. 32 (10) (2017) 805–819, https://doi.org/10.1111/mice:2017.32.issue-1010.1111/mice:12297.

[27] Y. Ren, J. Huang, Z. Hong, W. Lu, J. Yin, L. Zou, X. Shen, Image-based concrete crack detection in tunnels using deep fully convolutional networks, Constr. Build. Mater. 234 (2020) 117367, https://doi.org/10.1016/j.conbuildmat.2019.117367.
[28] B. Kim, C. Soojin. Image-based concrete crack assessment using mask and region-based convolutional neural network. Structural Control and Health Monitoring 26.8(2019):e2381.1-e2381.15. doi: 10.1002/stc.2381.
[29] D. Ai, G. Jiang, S.K. Lam, Automatic pixel-wise detection of evolving cracks on rock surface in video data, Autom. Constr. 119 (2020) 103378, https://doi.org/10.1016/j.autcon.2020.103378.
[30] A. Ji, X. Xue, Y. Wang, X. Luo, W. Xue, An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement, Autom. Constr. 114 (2020) 103176, https://doi.org/10.1016/j.autcon.2020.103176.
[31] D. Kang, S.S. Benipal, D.L. Gopal, Y.-J. Cha. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. 2020, 118, 103291. doi: 10.1016/j.autcon.2020.103291.
[32] W. Wang, M. Wang, H. Li. Pavement crack image acquisition methods and crack extraction algorithms: A review. Journal of Traffic and Transportation Engineering (English Edition), 2019. http://dx.doi.org/CNKI:SUN:JTTE.0.2019-06-001.
[33] C.V. Dung, L.D. Anh, Autonomous concrete crack detection using deep fully convolutional neural network, Autom. Constr. 99 (2019) 52–58, https://doi.org/10.1016/j.autcon.2018.11.028.
[34] C.V. Dung, H. Sekiya, S. Hirano, T. Okatani, C. Miki, A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks, Autom. Constr. 102 (2019) 217–229, https://doi.org/10.1016/j.autcon.2019.02.013.
[35] K. Jang, N. Kim, Y.K. An, Deep learning-based autonomous concrete crack evaluation through hybrid image scanning, Struct. Health Monit. 18 (5-6) (2019) 1722–1737, https://doi.org/10.1177/1475921718821719.
[36] J. Dai, Y. Li, K. He, J. Sun. R-FCN: Object Detection via Region-based Fully Convolutional Networks. arXiv:1605.06409 [cs.CV]. https://arxiv.org/pdf/1605.06409
[37] P.A. Viola, M.J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. doi: 10.1109/CVPR.2001.990517.
[38] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90, https://doi.org/10.1145/3065386.
[39] M.D. Zeiler, R. Fergus. Visualizing and Understanding Convolutional Networks. European Conference on Computer Vision. Springer International Publishing, 2013. https://arxiv.org/abs/1311.2901.
[40] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014). https://arxiv.org/abs/1409.1556.
[41] R.U. Khan, X. Zhang, R. Kumar, Analysis of ResNet and GoogleNet models for malware detection, J. Comput. Virol. Hacking Techniques 15 (1) (2019) 29–37, https://doi.org/10.1007/s11416-018-0324-z.
[42] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. (2016). https://arxiv.org/abs/1512.03385
[43] X. Wang, T.X. Han, S. Yan. An HOG-LBP human detector with partial occlusion handling. 2009 IEEE 12th International Conference on Computer Vision IEEE, 2010. https://ieeexplore.ieee.org/document/5459207.
[44] Y.J. Cha, K. You, W. Choi, Vision-based detection of loosened bolts using the Hough transform and support vector machines, Autom. Constr. 71 (2016) 181–188, https://doi.org/10.1016/j.autcon.2016.06.008.
[45] L. Ramana, W. Choi, Y.J. Cha, Fully automated vision-based loosened bolt detection using the Viola-Jones algorithm, Structural Health Monitoring 18 (2) (2019) 422–434, https://doi.org/10.1177/1475921718757459.
[46] M Lin, C. Chen, C. Lai. Object detection algorithm based AdaBoost residual correction Fast R-CNN on network. the 2019 3rd International Conference 2019. doi: 10.1145/3342999.3343013.
[47] W. Choi, Y.-J. Cha, SDDNet: real-time crack segmentation, IEEE Trans. Ind. Electron. 67 (9) (2020) 8016–8025, https://doi.org/10.1109/TIE.4110.1109/TIE.2019.2945265.
[48] K. He, G. Gkioxari, D. Piotr. Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017. https://arxiv.org/abs/1703.06870.