Case Studies in Construction Materials 21 (2024) e03953

Contents lists available at ScienceDirect

Case Studies in Construction Materials

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

Quantitative characterization of surface defects on bridge cable based on improved YOLACT++

Hong Zhang [a], Jiangxia He [b], Xiaogang Jiang [b], Yanfeng Gong [c], Tianyu Hu [a], Tengjiao Jiang [d,*], Jianting Zhou [a]

[a] State Key Laboratory of Mountain Bridge and Tunnel Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[b] School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[c] School of Shipping and Naval Architecture, Chongqing Jiaotong University, Chongqing 400074, China
[d] Department of Structural Engineering, Norwegian University of Science and Technology, Rich. Birkelands vei 1A, Trondheim 7491, Norway

ARTICLE INFO

Keywords:
Cables
Defect segmentation
Defect quantitative characterization
YOLACT++
Cylindrical surface correction

ABSTRACT

The safety and reliability of cables are directly linked to the safe operation of bridges as crucial load-bearing components. The accuracy and efficiency of current methods are still insufficient to segment and quantitatively characterize surface defects on cables. This paper proposes a novel and efficient method for the refined segmentation and quantitative characterization of bridge cable surface defects based on an improved you only look at coefficients++ (YOLACT++) model. For defect segmentation, several enhancements have been made to the YOLACT++ model, including incorporating the convolutional block attention module (CBAM), optimizing the anchor box generation mechanism, and introducing the smoother Mish activation function, which enhances both the accuracy and speed of defect detection. For quantitative characterization, the method adopts surface correction algorithms, pixel statistics, and crack skeleton extraction, resulting in a more accurate representation of defect areas and the length and width of cracks. Compared to the baseline model, the optimized model achieves a 3.58 % improvement in mean average precision (mAP) and an inference speed of 25.74 frames per second (FPS). The results show that the error is within 10 % compared with the manually measured area, which offers a more objective and comprehensive foundation for cable safety assessment.

1. Introduction

Large-span cable-stayed bridges and suspension bridges have become essential infrastructures in modern engineering due to their large span ranges and beautiful structural forms. The cables of these bridges act as critical load-bearing structures, directly supporting the weight of the bridge deck, traffic loads, and wind loads [1,2]. However, prolonged exposure to high temperatures and humidity, heavy loads, and high-speed traffic volumes can lead to premature damage to long-lived cables. Therefore, early testing and evaluation of bridge cables are critical to ensure safe maintenance of bridge structures.

Reducing the load-bearing capacity of the cables may seriously threaten the bridge safety [3,4]. In-service cables are primarily affected by external and internal damage. External damage includes surface scratches, cracks, and other surface defects. These surface defects often lead to internal damage. Long-term exposure to rain and sunlight can cause surface cracks on the cables, which may

* Corresponding author.
E-mail address: tj.jiang@outlook.com (T. Jiang).

https://doi.org/10.1016/j.cscm.2024.e03953
Received 3 August 2024; Received in revised form 14 October 2024; Accepted 5 November 2024
Available online 5 November 2024

gradually lead to internal steel wire corrosion and fractures, exacerbating the decline in the cable bearing capacity [5–7]. Therefore, early identification of potential surface defects on cables to prevent further internal damage is also essential to ensure the safe maintenance of bridge structures [8].

Currently, there are several methods for detecting surface defects on cables [9,10]: (1) Manual inspection: manual visual inspection using a high-resolution telescope, which is widely used because of its simple equipment but inefficient and time-consuming. Manual inspection using a bridge inspection vehicle is more efficient and has a lower cost, but the equipment is complex, and there is a high risk associated with high-altitude operations [11–13]. (2) The drone and cable-staying robot inspection: the detection method using drones and cable-climbing robots based on visual technology is flexible and easy to control. However, it is challenging to capture minor surface defects, and the obtained cable images require manual judgment, resulting in low efficiency and time-consuming processes. (3) Sensor techniques: the sensor methods adopt [14], for example, ultrasonic [15], acoustic emitter [16], and magnetic leakage techniques [17], show good performance for internal damage detection, such as corrosion and fracture of steel wires inside the cables. However, the above sensor methods generally have problems such as low efficiency and high cost, and it is challenging to meet the needs of large-scale and normalized surface defect detection of the cables. Recently, researchers have used computer vision to identify and quantitatively characterize target areas in images [18]. These techniques are mainly classified into two categories: traditional machine vision methods [19,20,18] and deep learning-based machine vision methods [19,20]. Conventional machine vision uses classical image processing algorithms with machine learning to recognize bridge cable images automatically. Deep learning-based machine vision methods often use target recognition and segmentation algorithms to process the cable images automatically.

In the traditional approach, Chen et al. [21] employed a threshold image difference method to detect surface defects on the stay cable. The approach involved analyzing the grayscale and gradient features of the stay cable images and using a Gaussian mixture model to extract the region of interest. The threshold image difference method was then applied to compare the detection image with the original image, identifying defects based on the maximum connected area. This method achieved an accuracy rate of over 90 %. However, the recognition accuracy for minor defects was relatively low. Jiang et al. [22,23] proposed a novel vision-based line--tracking algorithm to identify and track cable movement using coarse line search and subpixel edge and centerline detection techniques. The excellent robustness was demonstrated in many field tests under different complex backgrounds. Ho et al. [24] adopted the median filter and histogram equalization to map images to the principal component analysis subspace. The false detection rate gradually tends to zero with the increase of training samples, but it is inclined to detect scratch-type damage. Li et al. [25] applied particle swarm optimization to optimize the punish factor and kernel parameters of the SVM model, achieving a classification accuracy of 96.25 % in their experiments. These conventional methods can achieve better segmentation results when the image quality is high, the defect texture is clear, and the gray-scale features are apparent, but there are some limitations for defect segmentation in other complex scenes.

In the deep learning approach, Li et al. [26] proposed a defect recognition method for the cable-stayed based on Gabor wavelet transform and improved regional convolution neural network (RCNN). Data preprocessing was performed using median filtering, mean standardization method, and Retinex algorithm. The Gabor wavelet transform was used to extract the features of surface defects of the stay cable, and the feature maps were applied to a faster RCNN for recognition, and the recognition accuracy was 94.14 %. Ehtisham et al. [27] employed ten convolutional neural networks trained on 9000 images to classify defects in wooden structures. In particular, the Inception-V3 network achieved 99 % accuracy, precision, recall, and F1-score in classifying wood cracks. Jiang et al. [19,20] proposed a visual inspection and damage diagnosis method for bridge rivets based on a conventional neural network (CNN). A novel multiscale moving window searching technique was proposed to solve the challenge of small-size object identification in high-resolution images, enhancing the global and local vision balance, reducing pixel redundancy, and improving the identification rate to 96.3 %. He et al. [28] enhanced feature extraction for defects by incorporating a multi-scale attention module into the YOLOv4 network. The implementation of an oriented bounding box segmentation strategy reduced background interference, enabling the recognition of multiple types of bridge damage. The method achieved a mean precision of 82.49 % and a mean intersection over union of 64.28 %. However, the network model is relatively larger, resulting in a slower inference speed. Dong et al. [29] employed the multi-layer feature association fusion network architecture for intelligent segmentation of asphalt pavement cracks. The architecture integrated a feature coupling encoder, a dual-branch feature association module, and a hybrid multi-layer perceptron architecture with a lightweight feature decoder. The method achieved segmentation accuracy, intersection over union (IoU), and inference speed of 94.39 %, 89.46 %, and 21.87 FPS, respectively. Xu et al. [30] developed an efficient cable-stayed defect detection method using computer vision and diameter measurement techniques. Enhanced cable surface images undergo threshold segmentation via improved local grey contrast enhancement and the enhanced maximum correlation method. The method achieved recall ratios of 80.4 % for type-I defects and 85.2 % for type-II defects on cable surfaces. Dong et al. [31] proposed the StyleGAN network to generate high-quality, realistic crack images. The introduction of dilated convolutions in the Road-Seg-CapsNet network ensured clearer edges for the identified cracks, enabling the recognition of various complex forms of asphalt pavement cracks. The method achieved a mAP of 94.2 % and a minimum accuracy of 90.3 % in measuring crack areas.Wang et al. [32] developed a method to detect small tunnel cracks using the anchor-free algorithm. The YOLOX-x network was optimized with added semantic enhancement modules, achieving an average detection precision of 85.8 %. However, multiple duplicate prediction boxes are formed when recognizing the same crack, which consumes many computational resources. Chen et al. [33] based on the Swin-Unet network, fusing the skip attention module and the residual Swin Transformer block to capture the pavement crack region better. The average detection accuracy and recall rate are 85.96 % and 86.12 % on the Crack500 dataset, respectively. Zhang et al. [34] proposed an improved U-net network pavement crack segmentation method, fusing the VGG16 and upsampling convolution modules as the backbone network for feature extraction, with an accuracy and recall of 87.4 % and 88.7 %. However, the crack detection accuracy is reduced for different light, occlusion, and breakage. Huang et al. [35] introduced a three-stage automated approach to determine the width of small concrete cracks, utilize

Res-Unet for crack identification and distinguish between genuine cracks and false positives based on fractal dimension. This method achieved precise characterization of detected cracks with a maximum width of 0.1 mm. However, the computational efficiency for crack detection is low. Hou et al. [36] transferred concrete crack feature weights to a cascade mask region conventional neural network for training. Crack sizes were extracted using skeleton extraction and domain point search algorithms to identify and quantify surface cracks on cable-stayed bridges. The method achieved recognition accuracy of 99.6 %, IoU of 74.3 %, and inference speed of 7.8 FPS. The total mean pixel width error was 6.1 %, and the pixel length error was 7.3 %. Gaur et al. [37] employed image processing and fusion-based deep convolutional neural network methods to identify and classify six types of concrete defects. The classification precision for crazing and cracks was 95 % and 98 %, respectively. The overall accuracy for defect classification was 98.31 %, and F1-score was 98 %. These deep learning-based methods can improve the recognition accuracy of crack defect targets. However, they reduce detection speed, consume large computational resources, and have low accuracy in quantitatively characterizing the detected defects. Therefore, there is an urgent need to develop a real-time method for detecting and characterizing surface defects on cables, which can achieve efficient, accurate, and economical detection and quantitative characterization of surface defects on cables.

This study enhances cable surface defect segmentation and characterization to address the above limitations using an improved YOLACT++ model. This involves incorporating the CBAM, optimizing the anchor box generation mechanism, and introducing the smoother Mish activation function, enabling precise and rapid surface defect identification. To identify defect areas, surface correction algorithms, defect area estimation, skeleton extraction, and crack pixel length and width calculations are used to characterize surface defects accurately. The significance lies not only in achieving a refined segmentation and quantitative characterization of surface defects on cables but also in demonstrating the benefits and potential of computer vision for bridge health monitoring.

Using these improved methods as a foundation, this paper further explores the issue of surface defect detection in suspension cables. Through an investigation of the background and research status of suspension cables, the problems existing in the current recognition of surface defects in suspension cables are revealed. Based on the improved YOLACT++ model, this paper achieves the recognition and fine characterization of surface defects on cable-stayed cables in complex backgrounds. This study improves the accuracy of defect recognition and provides new technical support for bridge inspection and maintenance, with significant scientific value and potential engineering applications. Additionally, it provides a strong foundation for bridge safety assessment and management, promoting further advancements in the related field.

2. Method

The follow-up research can be divided into three parts as follows. Section 2 describes data collection and crack sample augmentation, the detailed procedure for optimizing the YOLACT++ model, and the quantitative characterization of the defect region. Section 3 implements the ablation experiment of the optimized model and compares it with other image processing algorithms, and the defect quantitative characterization is also evaluated. Finally, the conclusion is made in Section 4.

The workflow of the cable surface defect detection and quantitative characterization is shown in Fig. 1, including dataset enrichment, improved YOLACT++ model, and characterization of defect regions. The structure of the improved YOLACT++ model is shown in Fig. 2, and the specific improvements of the model are described in red font.

2.1. Dataset establishment

2.1.1. Data collection

The data sources for this study include photos of surface defects on disassembled abandoned cables in a laboratory environment taken with a convenient mobile camera and photos from several cable-stayed bridges in southwestern China, captured using a cable-staying robot. The images have a uniform resolution of 640×832 pixels and were taken at a distance of 0.1–0.15 m. Approximately 900 images were collected, including various scenarios such as uneven lighting conditions, background interference, and varying degrees of picture blurring [38], as illustrated in Fig. 3.

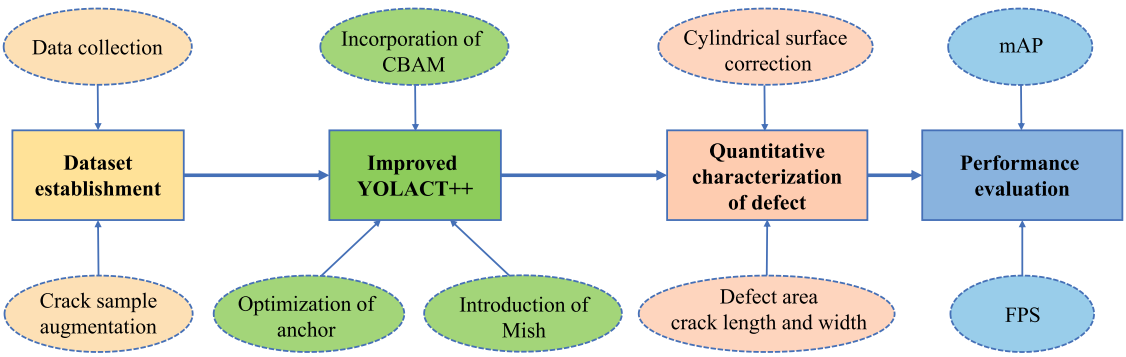


Fig. 1. Flowchart of the cable surface defect detection and quantitative characterization.

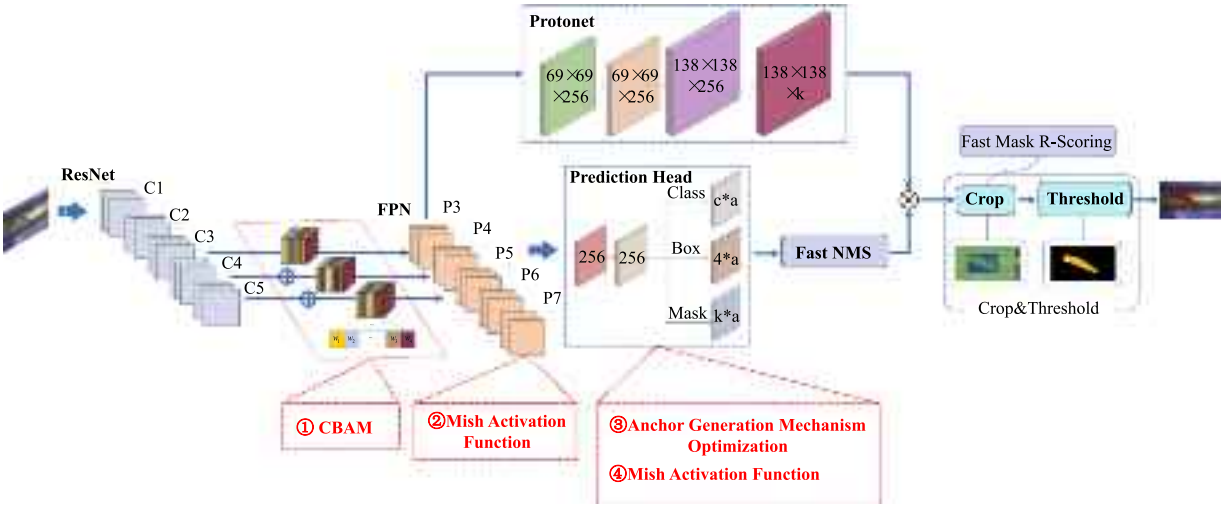


Fig. 2. Network architecture of the improved YOLACT++.

2.1.2. *Crack sample enhancement*

Sufficient training samples are significant for deep learning models to achieve high accuracy. However, the number of crack samples is limited, affecting the learning effect and practical application ability [39]. Therefore, researchers used diverse digital image processing techniques to generate crack masks with various morphologies, improving model adaptability to different cracks [40]. To compensate for the lack of such samples in this study, crack masks with similar characteristics are extracted from various types of existing cables, cylindrical items, and identical bridge structures to construct rich crack mask data. The enhanced crack mask is shown in Fig. 4.

The enhanced crack mask was fused with the healthy cable surface image. The fixed location of the healthy cable area was framed with a solid black line using preset parameters, as shown in Fig. 5. The center point area of the self-adaptive calculated crack mask is marked with a red solid line to ensure it is within the cable area. The final generated surface crack samples of the cable are shown in Fig. 5, successfully solving the issue of insufficient crack samples. The above-generated images and the actual captured images were integrated into 1350 images, and the image size was uniformly converted to 640×832 pixels.

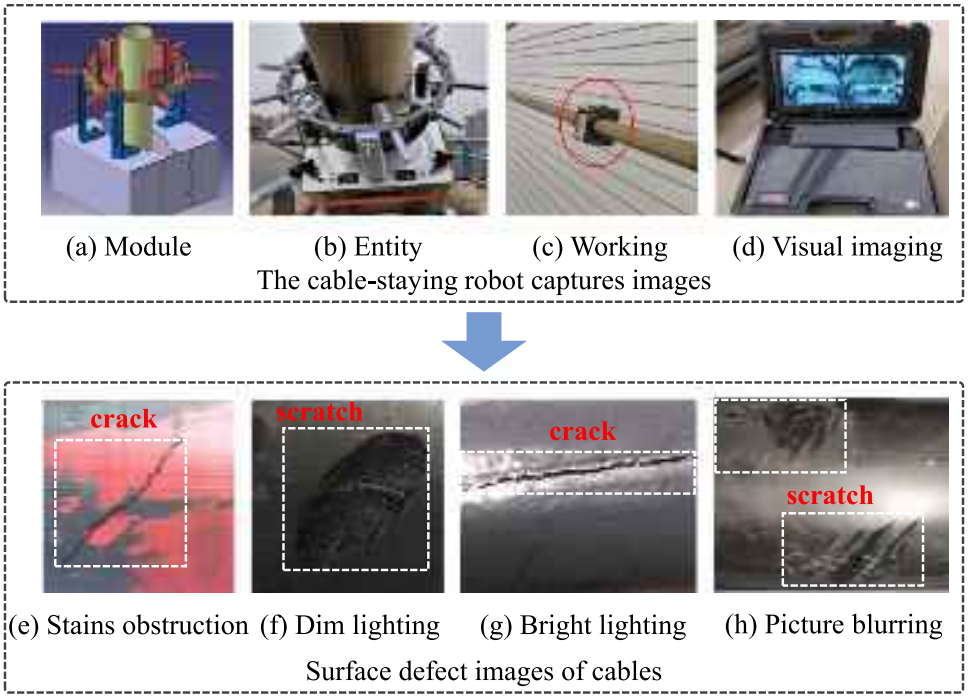


Fig. 3. Example illustrations of cable surface defect image samples.

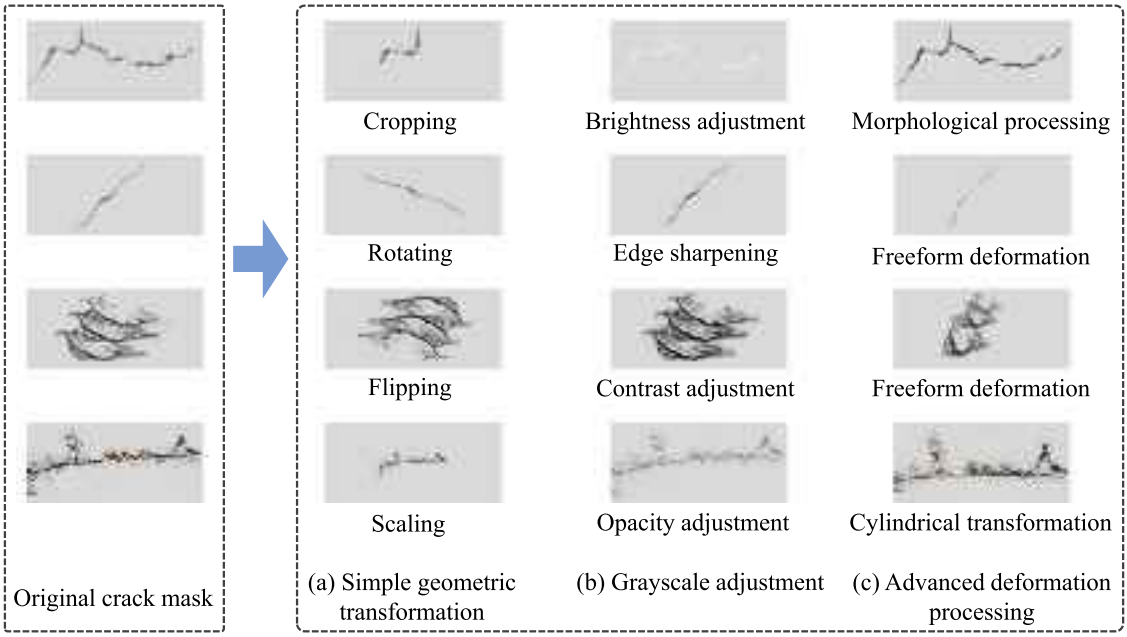


Fig. 4. Example of the enhanced crack mask.

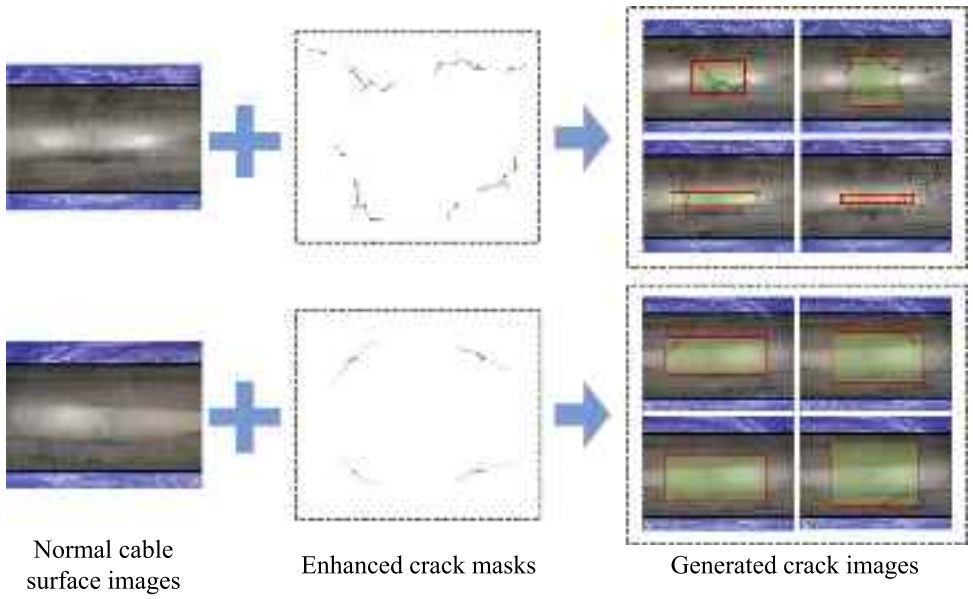


Fig. 5. Effects of crack image fusion.

The Labelme [41] software was then used to label the surface scratches and cracks of the cable, producing a COCO [42] format dataset. The dataset was then divided into 80 % for training and 20 % for validation before being fed into the network for training.

Fig. 6 shows the number of different defect categories in the training and testing datasets. As illustrated in Fig. 6(a), the training dataset contains 888 crack samples and 1395 scratch samples, while Fig. 6(b) shows the testing dataset with 194 crack samples and 343 scratch samples.

Fig. 7 demonstrates the distribution of defects within the images in the training and testing datasets. In Fig. 7(a), the training dataset contains 2283 defects distributed across 1080 image samples, while Fig. 7(b) shows the testing dataset with 537 defects distributed across 270 image samples. The number of defects per image ranges from 1 to 7. The last bar cluster of Fig. 7(a) contains 4 images, each containing 7 defects. The majority of images in the training dataset contain only one defect, with 473 image samples having a single defect.

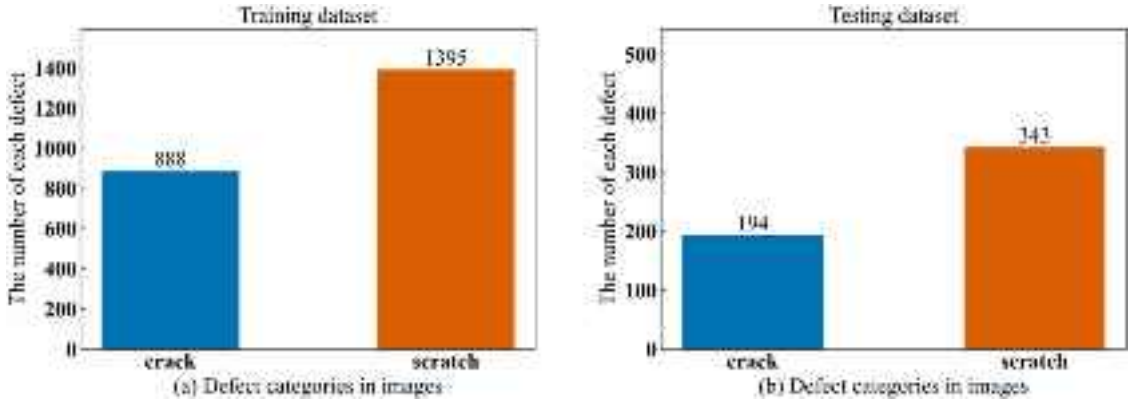


Fig. 6. The number of different defect categories in the training and testing datasets.

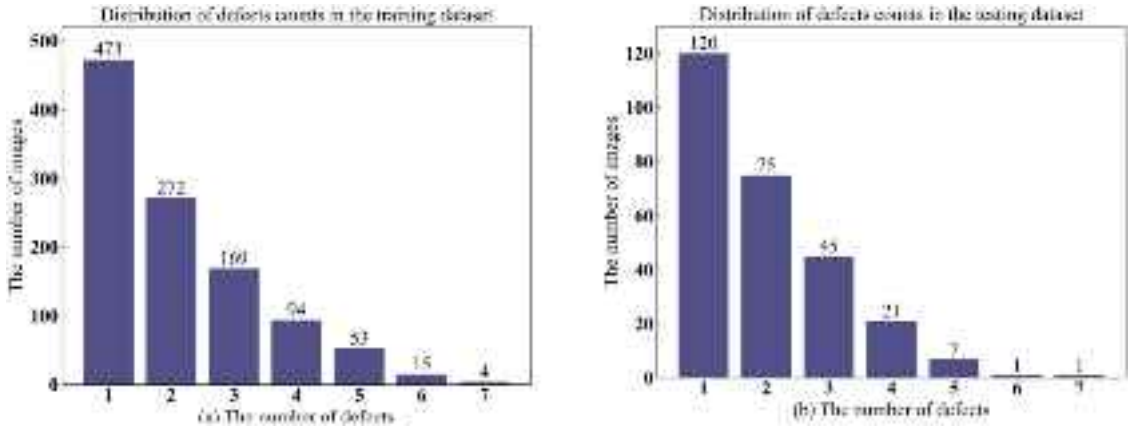


Fig. 7. The distribution of defects within the images in the training and testing datasets.

2.2. *YOLACT++ and its improvements*

2.2.1. *YOLACT++ model*

The YOLACT++ [43] network structure is shown in Fig. 8. A deep residual network, ResNet [44], is used as the backbone network. By introducing a deformable convolution network (DCN) in the [C3, C4, C5] layer, the sampling points can be flexibly offset according to the shape and size of the defects, which enhances the grasping of the defect contour; the feature pyramid network (FPN) can learn richer defect features through multi-scale information fusion, feature reuse, and feature enhancement. The Prediction Head branch

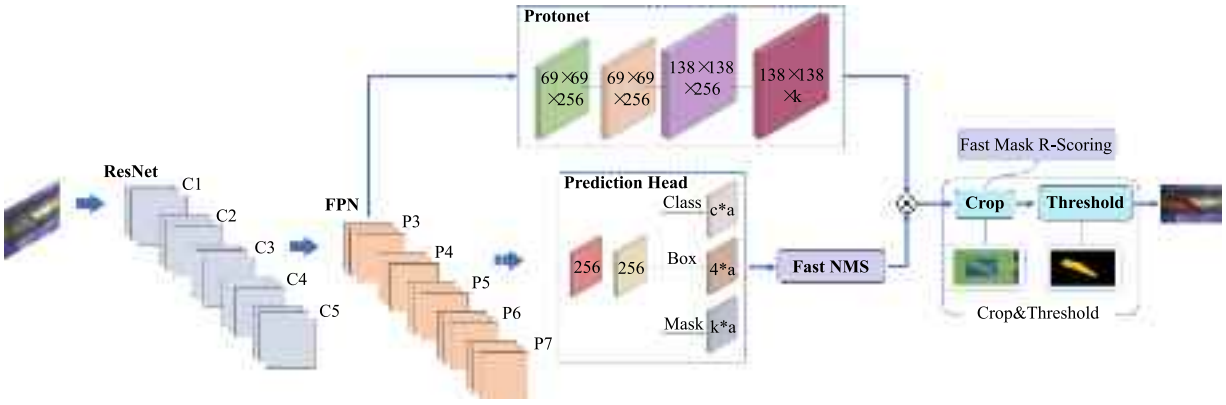


Fig. 8. YOLACT++ network architecture.

generates category confidence, coordinates, and prototype mask coefficients of candidate boxes, enhancing inter-task correlation by utilizing the shared features. However, it generates redundant candidate boxes, which interferes with the judgment of defects. The Protonet branch generates a fixed number of prototype masks of k. Each coefficient is then multiplied element by element with the corresponding prototype mask by matrix multiplication, and it is summed to obtain the final prediction mask for a single image. Although YOLACT++ shows some advantages in the instance segmentation task, it is insufficient for deep semantic information extraction, inefficient in processing redundant candidate boxes, and limited in real-time. Therefore, the above problems were addressed in this study.

2.2.2. Incorporation of the CBAM

Stained rain lines produced by cables under natural conditions can easily interfere with detecting surface defects. This study incorporates the CBAM [45] between the [C3, C4, C5] of ResNet and the [P3, P4, P5] of the FPN to distinguish the deeper semantic information between stained rain lines and accurately extract the key features of the surface defects of the cable. Thus, the network could focus on the defective surface features and suppress other useless information when performing feature fusion, focusing on the local information of interest.

The CBAM is a lightweight hybrid attention module that integrates a channel attention module (CAM) with a spatial attention module (SAM), aiming to enhance the network's fine-grained understanding and utilization of input features. CAM enhances the correlation between different channels, allowing information to be interacted and transferred between channels. In addition, SAM enhances essential features of the spatial region, enabling a more accurate focus on target location and shape. The addition of a small number of parameters brings significant performance improvement and also allows the model to have a faster training speed. The structure is shown in Fig. 9.

A one-dimensional channel attention map M_c *(size* $C\times1\times1$*)* is generated by CAM, expressing the contribution to the target for each channel in the input feature map *F (size* $C\times H\times W$*)*. A two-dimensional spatial attention map M_s *(size* $1\times H\times W$*)* is generated by SAM, expressing the importance of each spatial pixel point to the target for each of the input feature maps *F' (size* $C\times H\times W$*)*. In conclusion, CAM focuses on channel information, while SAM focuses on spatial information. The combination of CAM and SAM allows for the simultaneous consideration of the correlation between channels and the importance of spatial regions. This ultimately enhances the recognition accuracy and generalization ability of the model.

2.2.3. Optimization of the anchor box generation mechanism

The anchor box generation mechanism is a strategy for generating candidate boxes, which aids the model in accurately capturing targets of various sizes and shapes. In YOLACT++, the predefined aspect ratios for prediction heads are 1:1, 1:2, and 2:1. However, the generated candidate boxes struggle to adapt to the varying shapes of defects in cable images, resulting in omissions or misclassifications. Therefore, this paper proposes an optimization method for the anchor box generation mechanism to enhance the model accuracy and generalization in defect detection. The process can be divided into three steps: data binning, evaluation of cluster count, and K-Means++ clustering, as illustrated in Fig. 10.

Step 1: data binning. The aspect ratio data of all training samples containing defects was collected, and an intuitive histogram was constructed, as shown in Fig. 11. The data distribution is wide-ranging, with a precise concentration in the range of 0–2. Further drawing from the original model's preset aspect ratios of 0.5, and 2 were chosen as the bin boundaries. Thus, we define three bins: 0–0.5 (a bin), 0.5–2 (b bin), and greater than 2 (c bin).

Step 2: evaluation of cluster count. The silhouette coefficient method is used to determine the optimal number of clusters for a bin and c bin, where a higher silhouette coefficient indicates better clustering effectiveness. The relationship between the silhouette coefficients and the number of clusters for a bin and c bin is shown in Fig. 12. The silhouette coefficient for a bin reaches its maximum when the number of clusters is 2, while for c bin, it peaks at 4 clusters. However, increasing the number of clusters also increases computational resources. Therefore, it is determined that the optimal number of clusters for a bin and c bin is both 2.

Step 3: K-Means++ clustering. Based on the optimal cluster counts determined in Step 2 for a bin and c bin, the K-Means++ algorithm was applied separately. The results in Fig. 13 indicate that the cluster centers for a bin are located at 0.15 and 0.39, while for c bin, they are located at 2.84 and 8.98. Meanwhile, the clustering results for b bin follow the original model's preset anchor box values set to 1. The summarized results of the above three steps are presented in Table 1.

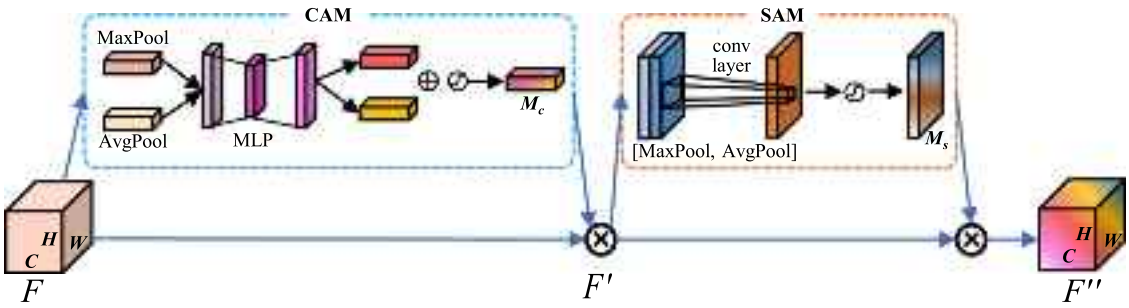


Fig. 9. CBAM structure.

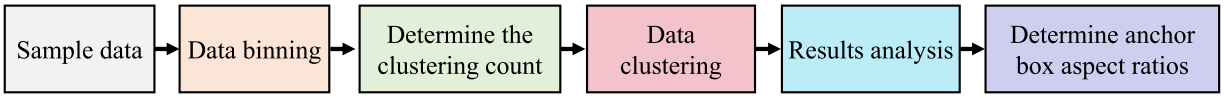


Fig. 10. Basic flowchart of optimized anchor box generation mechanism.

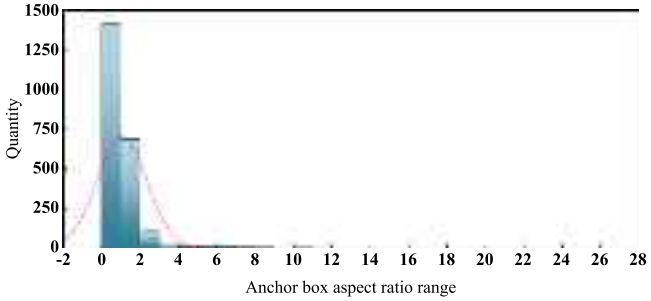


Fig. 11. Example of anchor box sample data.

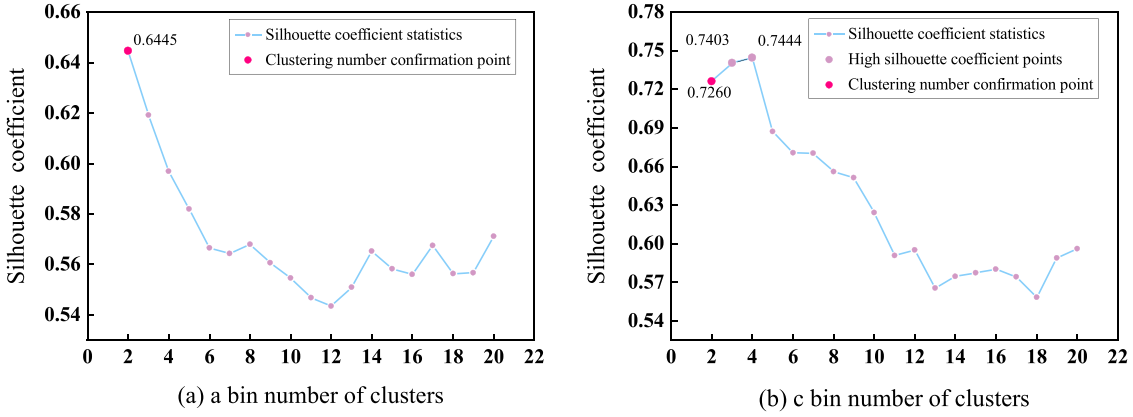


Fig. 12. Computing clustering number via silhouette coefficient method.

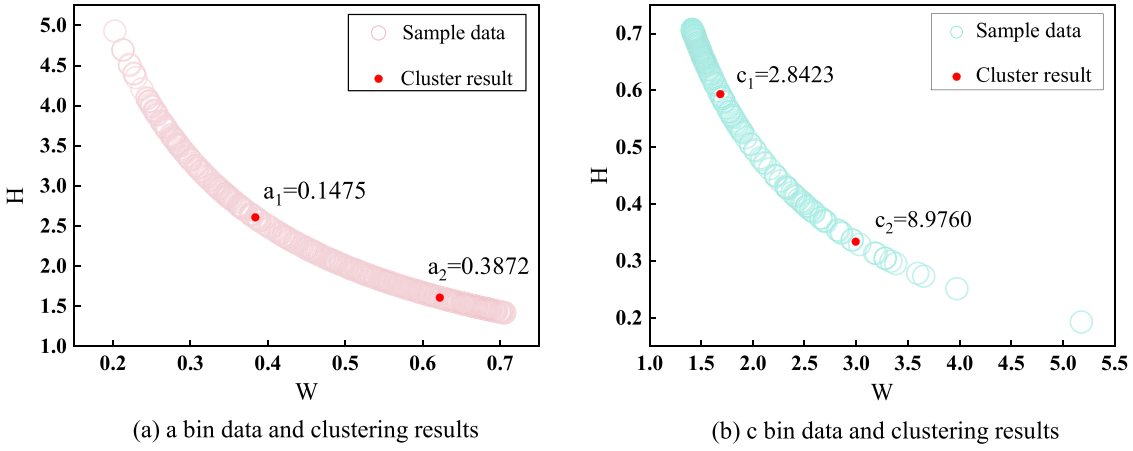


Fig. 13. Examples of clustering results for each bin. The one-dimensional aspect ratio values are transformed into a two-dimensional feature representation to represent the cluster centroids visually. The horizontal axis W represents the defect target width corresponding to the aspect ratio value when the area is 1, while the vertical axis H represents the defect target height corresponding to the aspect ratio value when the area is 1.

2.2.4. Introduction of the Mish activation function

The activation function effectively solves the problem of expression limitation caused by linear superposition by introducing a nonlinear transformation, which enhances the network's ability to learn and model complex nonlinear relationships. In this study, the ReLU activation function used within the Prediction Head module and the FPN of the YOLACT++ model is replaced with the superior performance Mish [46] activation function with the function expression:

$$Mish = x \cdot \tanh(\ln(1 + e^{x}))$$

The comparison of the function curves is shown in Fig. 14. The Mish activation function suppresses saturation in the negative region, allowing gradient conduction with negative inputs and avoiding ReLU's zero issues in the region. The Mish activation function helps to alleviate the problem of gradient vanishing that may occur during the training of neural networks, especially in deep network structures. It is crucial for maintaining the signal strength of the backpropagation.

2.3. Quantitative characterization methods for defect

2.3.1. Cylindrical surface correction algorithm

Cables are cylindrical components, so the defective areas identified by this model exhibit columnar features, which impact the accurate assessment of the defective areas. A set of surface correction algorithms for the defective surface region is designed by combining them with the cable characteristics. The principle is shown in Fig. 15, where point M is the observation position of the camera, BF is the width range of the cable captured in the field of view of the camera, OA is the radius of the cable, and X is an arbitrary pixel point in the region of the captured cable. After unfolding, the pixel position of the pixel point X is calculated from Eq. (1) and Eq. (2) through geometric principles.

$$\begin{cases} \theta = \arccos\left(\dfrac{\mathrm{OF_r}}{\mathrm{OM_r}}\right) \\ OX_p = OF_p = \dfrac{GF_p}{\sin\theta} \\ \alpha = \arcsin\left(\dfrac{DX_p}{OX_p}\right) \end{cases} \tag{1}$$

$$\begin{cases} \widehat{EX}_p = \alpha \cdot OX_p \\ \mathrm{GF_r} = \mathrm{OF_r} \cdot \sin\theta \\ k = \dfrac{\mathrm{GF}_r}{\mathrm{GF_p}} \\ \widehat{EX}_r = \mathrm{k} \cdot \widehat{EX}_p \end{cases} \tag{2}$$

Where $\mathrm{OF_r}$ is the real radius value of the cable in the experiment, $\mathrm{OM_r}$ is the distance from the camera lens to the center of the cable in the experiment, OX_p is the pixel value of the radius of the cable in the image, DX_p is the pixel distance from pixel X to the centerline of the width of the cable C in the image, α is the angular value of DX_p when it is unfolded into radian EX_p, radian EX_p is the pixel length of DX_p after the unfolding, k is the ratio factor of actual length to pixel length, which converts the measured pixel scale to exact size to achieve refined measurement of the defective area.

To achieve planarization of the surface defective region of the cable, the pixels in the image are traversed from left to right and top to bottom. Pixels above the centerline C of the cable width are identified and shifted upward based on their corresponding α values, while pixels below the centerline C are shifted downward using the same principle. This process decreases pixel density and the emergence of gap regions in the new image. Pixel filling through linear interpolation is utilized to maintain integrity and continuity. This method effectively preserves the edges and contour characteristics of the original defective region, creating a more natural appearance in the processed region.

The GrabCut algorithm separates the cable from the background to avoid background disturbances and measure defects accurately. The cable boundary is then smoothed with corrosion-inflated morphological operations. Finally, a straight line is fitted to the smoothed boundary using the least squares method. The principle of the smoothing process is shown in Fig. 16. The black color in the figure represents the background pixels, the white color represents the foreground target pixels, the blue color represents the corroded

Table 1
Statistics on data binning results.

Bin number	a bin	b bin	c bin
R=width/height	R<0.5	0.5≤R≤2	R>2
Sample number	608	1499	176
Percentage (%)	26.63	65.66	7.71
Number of optimal clusters	2	-	2
Optimized anchor box parameters	0.1475,0.3872	1	2.8423,8.9760

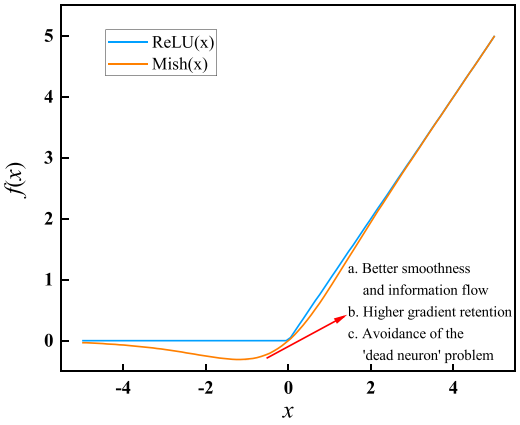
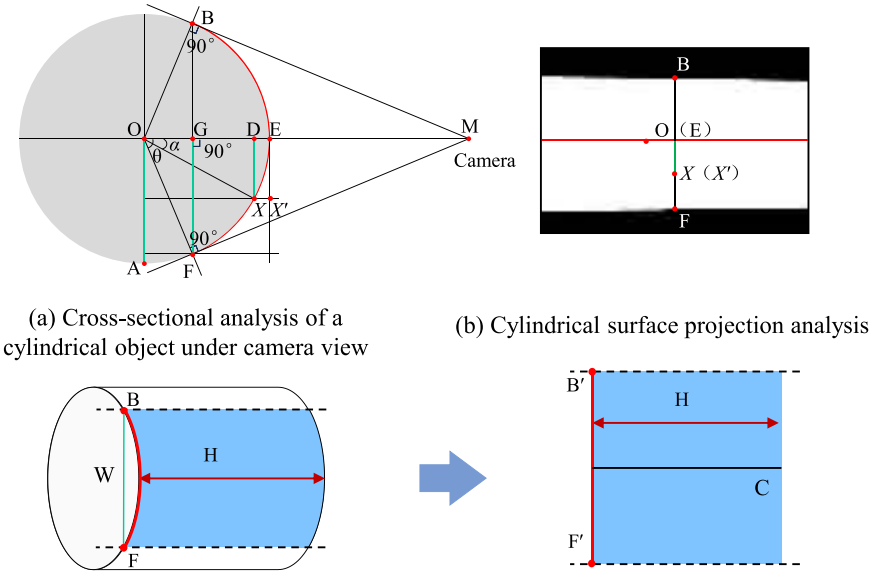


Fig. 14. Comparison of ReLU and Mish activation function curves.



(a) Cross-sectional analysis of a cylindrical object under camera view

(b) Cylindrical surface projection analysis

(c) Simulated unfolding analysis of a cylindrical surface

Fig. 15. Principle of surface unwrapping.

pixel area, and the red color represents the inflated pixel area. After completing the above process, the surface correction of the cable and defects are then performed, and the whole process is shown in Fig. 17. To demonstrate the correction effect, the identified defective images were binarized using the global thresholding method.

2.3.2. *Quantification defect*

Accurate measurement of the defective area is achieved by pixel measurement of the planarized processed defective area. The true area, length, and width of the defective area were obtained based on the scale factor k between the actual size and the pixel size. The estimation of the area of the defective region can be divided into three steps: counting the number of pixel points in the defective region, calculating the pixel area of the defective region, and calculating the proper area of the defective region. The calculation process is shown in Eq. (3).

$$\begin{aligned} S_i &= \sum_{i \in \Omega} w_i \cdot h_i \\ S_r &= k^2 \cdot S_i \end{aligned} \tag{3}$$

Where w_i is the width of the i_{th} pixel point in the planarized defective region, h_i is the height of the i_{th} pixel point in the planarized defective region, Ω is all the pixel points in the planarized defective region, S_i is the pixel area of the defective region, and S_r is the true area of the defective region.

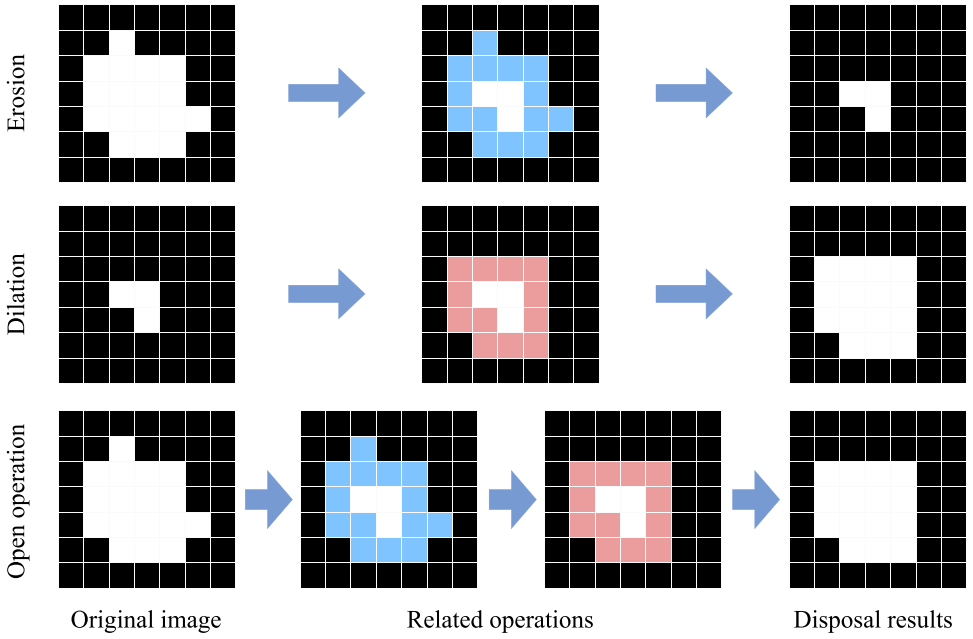


Fig. 16. Split boundary smoothing effect schematic.

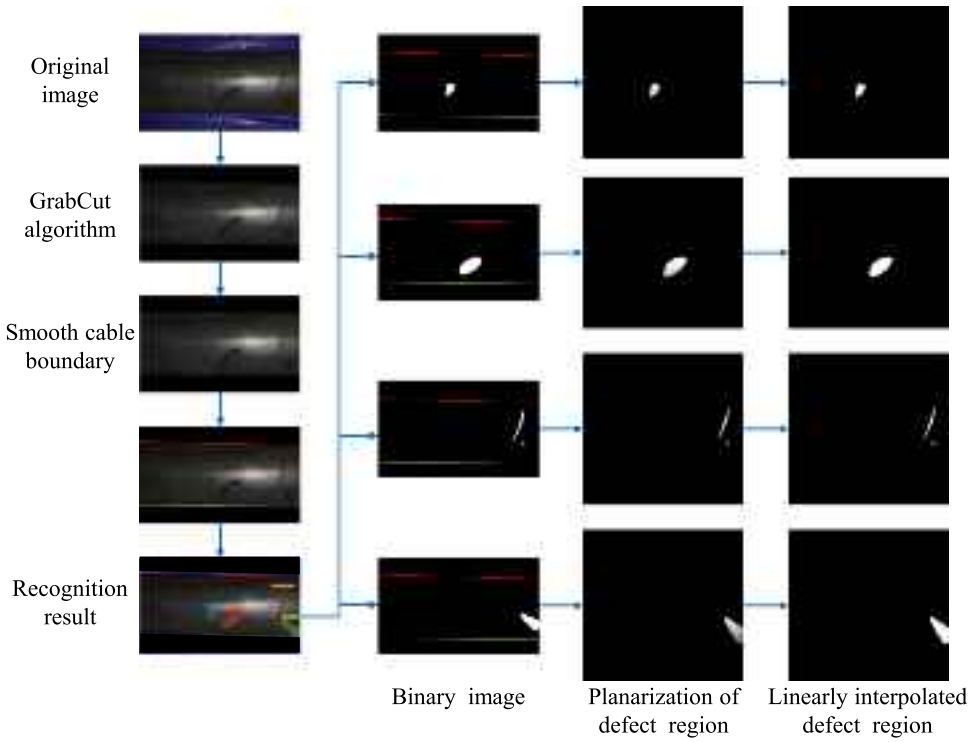


Fig. 17. Flowchart for planarization of defect regions.

To objectively assess the safety of cables, the length and maximum width of crack-type defects are being further estimated. Therefore, by standardizing the defect features, training the SVM model, and performing secondary classification on the defects identified by this model, the final classification accuracy exceeds 91 %. The estimation is categorized based on the length and maximum width of the crack defect, which can be divided into three steps: crack skeleton extraction, calculation of crack pixel length and actual length, and calculation of crack maximum pixel and actual width.

Step 1: skeleton extraction. The Zhang-Suen algorithm is used to iterate the crack region continuously, deleting pixel points in the non-skeleton region according to specific conditions. The effect of single-pixel skeleton extraction of the crack is shown in Fig. 18.

Step 2: pixel length and actual length estimation. The skeleton line pixel length is counted by traversing the image pixel by pixel from top to bottom and from left to right. Then, the actual length is estimated based on the scale factor k, as shown in Eq. (4), where L_p is the pixel length, L_r is the actual length, W_p is the maximum pixel width, and W_r is the maximum actual width.

Step 3: maximum pixel and actual width estimation. This is done by traversing each pixel on the skeleton and calculating the crack pixel value of each skeleton pixel point at 6 angles (0°, 30°, 60°, 90°, 120°, 150°) along with its corresponding opposite direction. The crack pixel width of the skeleton pixel point is determined by taking the minimum value, and the maximum value of all the crack pixel widths of skeleton pixel points is considered the maximum pixel width of the crack. The whole process is shown in Fig. 18, and the maximum actual width is estimated based on the scale factor k between the actual size and the pixel size.

$$L_r = \mathrm{k} \times L_p;\ W_r = \mathrm{k} \times W_p \tag{4}$$

Fig. 19 shows in detail the training and defect quantitative characterization steps of this model. Regarding defect segmentation, the YOLACT++ model was first integrated with the CBAM, enhancing the network's perception of key features and improving the recognition accuracy of defect contours. Secondly, the anchor box generation mechanism was optimized to solve the problem of identifying defects with different shapes. Finally, a smooth Mish activation function was introduced to promote the information transfer of the model to the defect area. Regarding defect quantitative characterization, the cylindrical surface correction algorithm was first used to planarize the defect area. Secondly, the defect pixel area is obtained through the pixel counting method. Finally, the crack pixel length and maximum pixel width are obtained through skeleton extraction and angle search, improving the precision of defect quantitative characterization. In summary, the refined segmentation and quantitative characterization of surface defects in cables have been achieved.

2.4. Evaluation metrics

The fundamental objective of the instance segmentation algorithm is accurate segmentation of each object of interest in the scene. The essence involves addressing challenges in pixel-level classification and category recognition. In this study, mAP is selected as the criterion to evaluate the performance of the prediction box and instance segmentation mask. The calculation formula for mAP is shown in Eq. (5). For example, mAP@0.5 indicates the mAP value calculated with an IoU threshold of 0.5, and mAP@0.5:0.95 represents the IoU threshold ranging from 0.5 to 0.95, with a step size of 0.05. mAP@0.5 focuses on a specific IoU threshold of 0.5 and is suitable for essential performance evaluation. mAP@0.5:0.95, by calculating average precision across multiple IoU thresholds from 0.5 to 0.95, provides a more detailed performance assessment suitable for understanding the model performance across varying precision requirements. In addition, FPS is used to evaluate the model inference speed.

$$\begin{aligned} AP_c &= \sum_n (R_n - R_{n-1}) P_n \\ mAP &= \frac{1}{\mathrm{N}} \sum_{\mathrm{c}=1}^{\mathrm{N}} AP_c \end{aligned} \tag{5}$$

Where N represents all defect classes, c represents the defect category, P_n denotes the precision at the n_{th} threshold, R_n is the recall at the n_{th} threshold, AP_{c} is the average precision for category c.

3. Experiment tests

The experiments in this study use the operating system Ubuntu 18.04.5, CPU model Intel Xeon Platinum 8255 C, GPU model NVIDIA GeForce RTX 2080 Ti with 40 G of video memory, hard disk size of 75 GB, and deployed on Pytorch 1.8.1. Deep learning framework, configured with parallel architecture CUDA10.2.

The improved YOLACT++ model training loss is shown in Fig. 20. Epoch represents the number of training rounds. Box_loss indicates the model's loss in identifying the bounding boxes of targets. Class_loss represents the model's loss in classifying categories, and Mask_loss denotes the mask loss calculated from an instance segmentation perspective. As training progresses, the model understanding and fitting ability gradually improve, eventually reaching a relatively stable state.

The improved YOLACT++ model's mAP@0.5 and mAP@0.5:0.95 variation curves for Box and Mask on the validation set are shown in Fig. 21. As the number of training epochs increases, the model's recognition accuracy on the validation set continuously improves, with the mAP eventually stabilizing at a high value. This indicates that the model maintains good generalization ability on unseen sample data and can effectively identify defect targets in unknown images.

3.1. Ablation experiments

Ablation experiments were conducted to verify the effectiveness of each improvement on model performance enhancement. The effect of these optimizations on the model performance was explored by removing or replacing the optimization strategies in the model, and the results of the experiments are shown in Table 2.

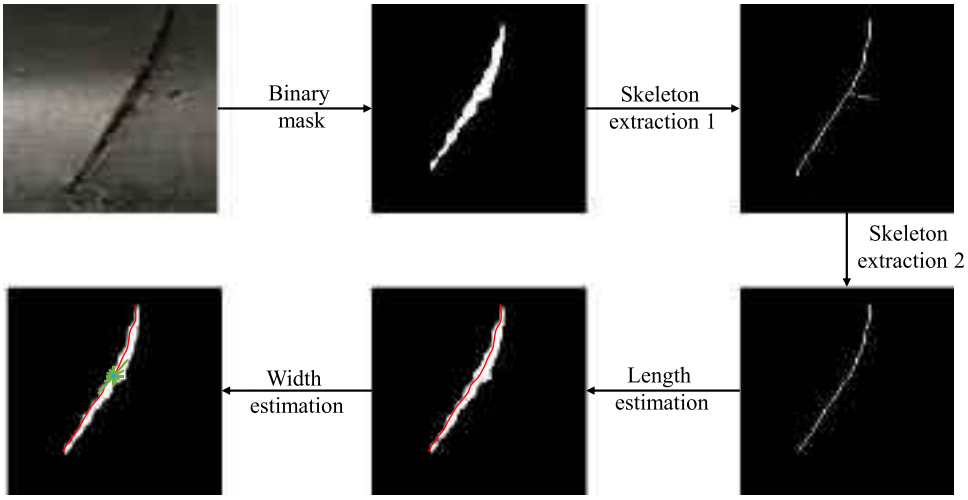


Fig. 18. Flowchart for crack parameter calculation.

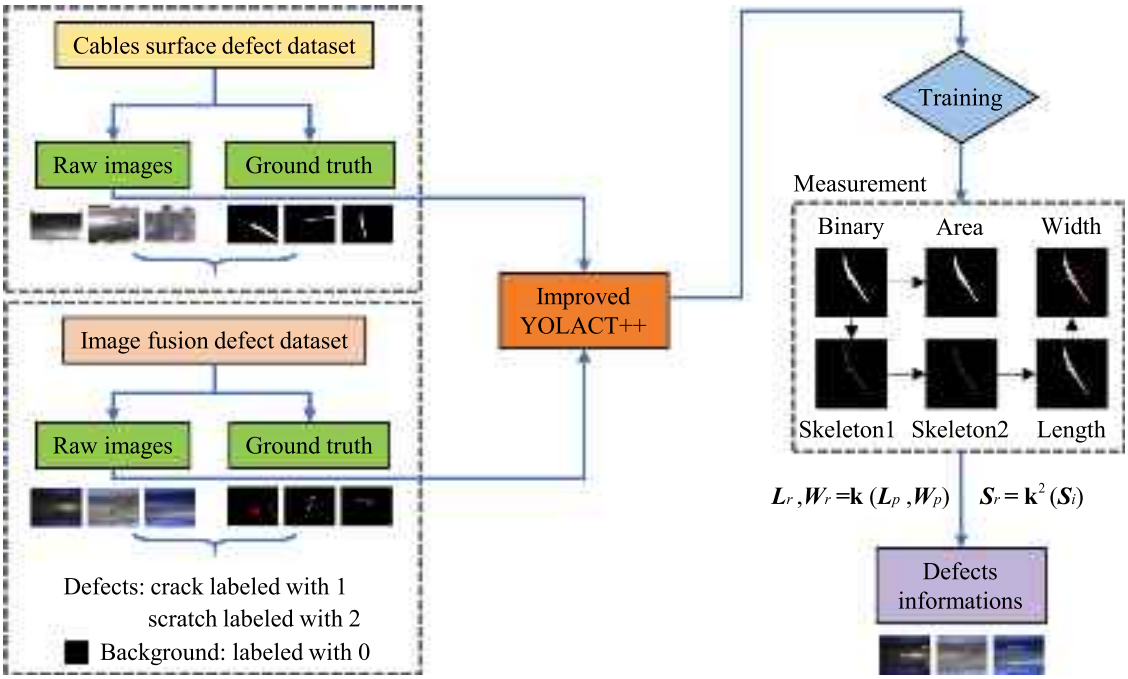


Fig. 19. Flowchart of the cable surface defect identification method.

The comparison between experiments 1 and 2 shows that incorporating the CBAM before feature fusion, the mask mAP@0.5 and mAP@0.5:0.95 increased by 2.89 % and 1.45 %, respectively. It demonstrates that the CBAM can effectively capture the deep semantic features of cable surface defects, suppress irrelevant background noise interference, and significantly enhance model detection accuracy.

The comparison between experiments 1 and 3 shows that the model inference speed is reduced by 0.5 FPS after optimizing the anchor box generation mechanism for the Prediction Head. However, the mask mAP@0.5 and mAP@0.5:0.95 are improved by 4.5 % and 2.96 %, respectively. The generated candidate boxes are more suitable for the defect target, which enhances the model robustness and generalization.

Comparing the results of experiments 1 and 4, replacing the ReLU activation function in Prediction Head and FPN with the Mish activation function led to improvements in various aspects. Specifically, the inference speed increased by 0.61 FPS, while the mask mAP@0.5 and mAP@0.5:0.95 saw improvements of 3.28 % and 1.19 % respectively. This can be attributed to the Mish activation function's ability to preserve valuable information when dealing with noisy or outlier-filled input data, thereby enhancing the model's

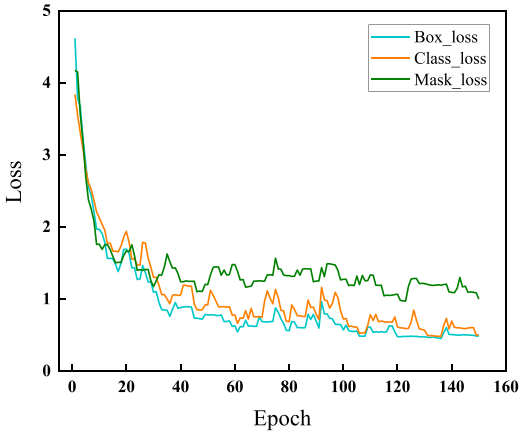


Fig. 20. Change curve of loss function.

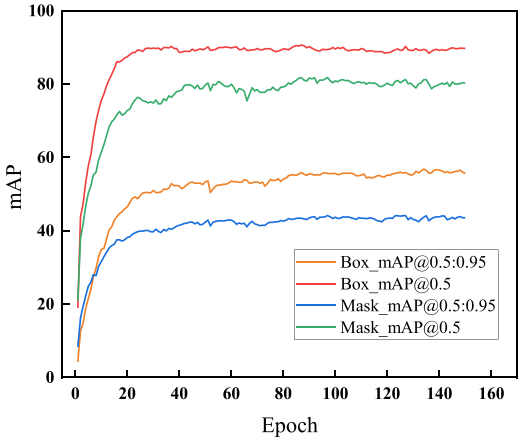


Fig. 21. Change curve of mAP on the validation set.

Table 2
Performance comparison of ablation experiments.

Test no.	Model	CBAM	Anchor	Mish	mAP@0.5(%)		mAP@0.5:0.95(%)		FPS
					Box	Mask	Box	Mask	
1	YOLACT++	×	×	×	81.75	76.60	52.77	40.58	26.55
2	A	√	×	×	84.88	79.49	53.49	42.03	25.12
3	B	×	√	×	87.14	81.10	56.34	43.54	26.05
4	C	×	×	√	84.31	79.88	54.92	41.77	27.16
5	D	√	√	×	87.14	80.52	55.24	43.29	25.02
6	E	×	√	√	87.11	79.68	53.91	43.27	26.69
7	F	√	×	√	84.86	78.28	54.44	42.40	27.11
8	Ours	√	√	√	90.28	81.24	55.85	44.16	25.74

detection accuracy. Additionally, the non-monotonic nature of the Mish function allows critical points within positive and negative intervals to transition smoothly through a segment of moderate negative gradient, ensuring the coherence of information flow and ultimately benefiting the model's inference speed.

The comparison between experiments 1 and 8 shows that the inference speed of the optimized model is only reduced by 0.81 FPS compared to the original model. However, the mask mAP@0.5 and mAP@0.5:0.95 are improved by 4.64 % and 3.58 % respectively. This shows that the optimized model is more advantageous for the cable surface defect segmentation. To visualize the effect of the present model on the perception of the target features, the outputs of the intermediate feature layers and critical structures of the present model are visualized, as shown in Fig. 22.

Fig. 22 shows that ResNet focuses on a specific region of the body while incorporating the CBAM allows for the extraction of more

global features. Protonet's feature layer can effectively focus on multiple defect targets in the image in a balanced manner, demonstrating the model's ability to identify surface defects on cables accurately.

3.2. Comparative experiment

To verify its superiority, the model was compared with Mask R-CNN [47], SOLO v1- v2 [48,49], YOLACT [50], YOLACT-Edge [51], YOLACT++, and SparseInst [52]. All of these models were tested in the same experimental environment and training strategy as the method in this study. The segmentation results are shown in Table 3 and Fig. 23.

Mask R-CNN extends Faster R-CNN by adding a branch, which generates pixel-level masks for each detected instance. The introduced region of interest align (ROI Align) operation precisely aligns the region feature maps.SOLO v1-v2 utilizes object center location information for instance segmentation and belongs to the basic network in the domain of instance segmentation. YOLACT integrates detection and segmentation into a neural network using FPN to extract multi-scale features, the Prediction Head to generate bounding boxes, masks, and category information, and Protonet to generate image-level prototype masks. YOLACT-Edge is a lightweight version of YOLACT that optimizes TensorRT and exploits temporal redundancy in videos for instance segmentation. YOLACT++ introduces DCN, fast non-maximum suppression, and optimized Prediction Head on top of YOLACT. SparseInst introduces the sparse mask head, which predicts instance masks by learning a set of sparse, deformable convolutional kernels, thereby enhancing the computational speed and segmentation accuracy. Additionally, SparseInst employs the query interaction module to strengthen interactions between features.

As seen in Table 3, the proposed method achieved the mask mAP@0.5:0.95 of 44.16 %. Based on the contribution of the real-time detection models, the mask mAP@0.5:0.95 of YOLACT, YOLACT-Edge, and YOLACT++ were 38.24 %, 36.89 %, and 40.58 %, respectively. SparseInst achieved the mask mAP@0.5:0.95 of 37.41 %. Mask R-CNN and SOLO v1-v2 obtained lower scores. For Mask R-CNN, the significant variation in the size of cracks and scratches likely contributes to poorer detection performance. SOLO v1-v2, using single-stage architectures, are prone to missed detections or false positives. Additionally, the proposed method had an inference speed of 25.74 FPS. YOLACT had the fastest inference speed at 28.89 FPS, but its mask mAP@0.5:0.95 is 5.92 % lower than the proposed method. Mask R-CNN had the slowest inference time, primarily due to the need to perform ROI Align for each generated bounding box.

As seen in Table 3 and Fig. 23, the proposed method achieved the best score in the mask mAP@0.5 with 81.24 %. YOLACT, YOLACT++, and SparseInst were 76.43 %, 76.60 %, and 76.30 %, respectively. In summary, the proposed model significantly improves detection accuracy while maintaining a good speed performance.

3.3. Segmentation and quantitative characterization of defect

The most common defects on the surface of cables are scratches and cracks. As shown in Fig. 24, three images with typical features

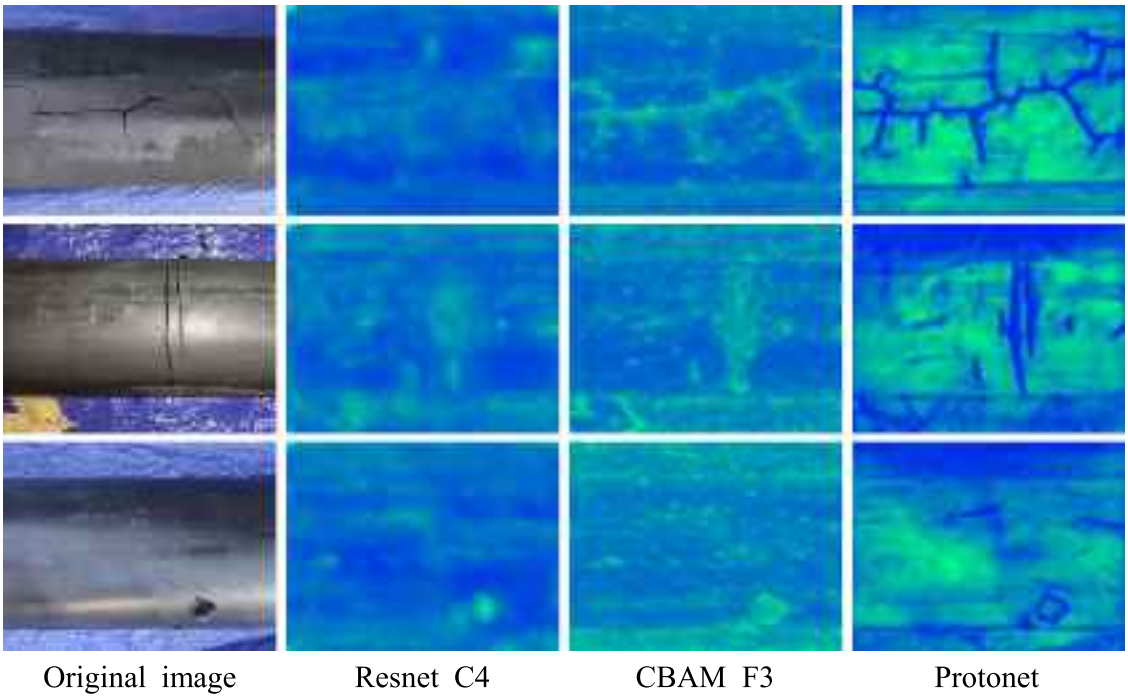


Fig. 22. Visualization results of features from various feature extraction networks.

Table 3
Comparisons of comprehensive performance on our dataset.

Model	mAP@0.5(%)		mAP@0.5:0.95(%)		FPS
	Box	Mask	Box	Mask	
Mask R-CNN	88.80	63.80	56.10	31.80	8.55
SOLO v1	-	65.68	-	33.20	15.80
SOLO v2	-	72.88	-	36.29	17.30
YOLACT	81.30	76.43	49.16	38.24	28.89
YOLACT-Edge	79.87	75.87	46.24	36.89	26.15
YOLACT++	81.75	76.60	52.77	40.58	26.55
SparseInst	-	76.30	-	37.41	25.13
Ours	90.28	81.24	55.85	44.16	25.74

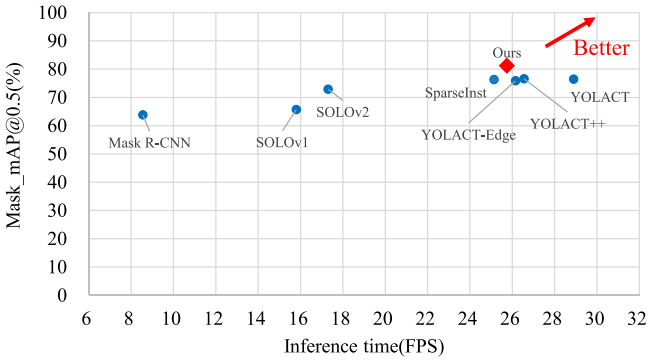


Fig. 23. Scatterchart comparison of mAP and inference time.

are selected (from top to bottom, the first column is numbered 1081, 1082, and 1083, respectively) to fully demonstrate the application of the proposed algorithm in defect segmentation and quantitative characterization. The "Detect result" is the defect segmentation result, "Evaluated area" is the estimation of the defect area, "Skeleton" is the extraction of the crack skeleton, and "Evaluated length and width" is the evaluation of crack length and width, where L is the evaluated crack length and W is the evaluated crack width.

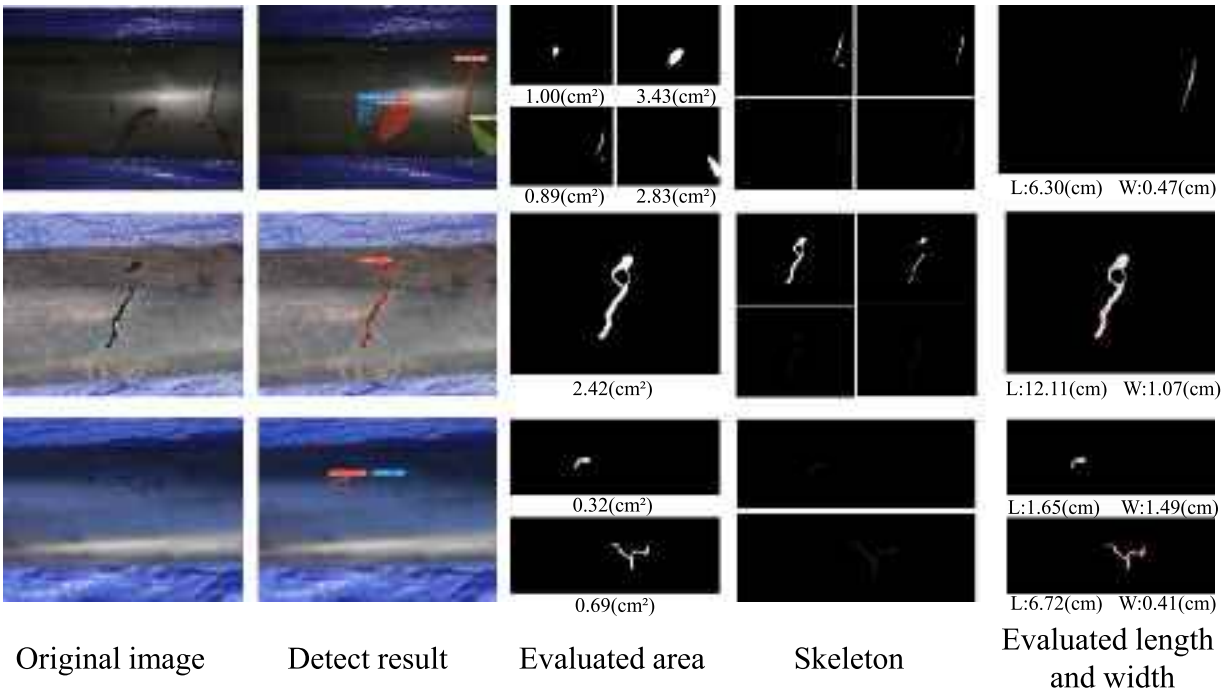


Fig. 24. Recognition results.

To verify the effectiveness of this method for defect quantitative characterization, we analyzed the defect characterization information of 270 images and compared it with manual measurement results. The analysis shows that the error is kept within 10 %.

Table 4 lists seven typical examples, showing an average error of 6.1 % for defect area, 6.2 % for crack length, and 5.9 % for crack width. The total average defect area error for 270 test images is 6.7 %, the average crack length error is 7.5 %, and the average crack width error is 6.5 %. This demonstrates the reliability of the proposed method in defect information statistics and provides a more objective basis for the safety evaluation of cables.

4. Conclusion

This study focuses on achieving high-precision, real-time recognition and refined characterization of surface defects on cables. To this end, the YOLACT++ model has been improved by incorporating the CBAM, anchor generation mechanism, and Mish activation function, significantly enhancing the accuracy of defect area detection. Additionally, a surface correction algorithm has been applied to flatten the defect regions, further improving the accuracy of defect quantification. The results show that this method offers high accuracy and reliability in detecting defects under complex backgrounds, making it effective for evaluating cable damage states. Future research may extend to the detection of more minor defects, and by incorporating more intelligent algorithms, the model adaptability can be further improved, enhancing the efficiency and accuracy of defect detection on cables. Two primary benefits of this study are outlined below:

(1) This model effectively balances the speed and accuracy of detection. Digital image processing and image fusion techniques are used to effectively expand the crack samples, which solves the problem of insufficient surface crack samples of cables. The incorporation of the CBAM enhances the network's ability to perceive defect features, reduces noise interference, and improves the model accuracy for defect detection. The proposed method of an optimized anchor box generation mechanism enhances the model's accuracy in detecting defects. The Mish activation function, which provides better smoothing, is used to improve the detection speed and accuracy of the model.

(2) The mask mAP@0.5:0.95 of this model is 44.16 %, and the inference speed is 25.74 FPS. Compared to the YOLACT++, the inference speed is reduced by only 0.81 FPS, while the detection accuracy is improved by 3.58 %. Based on the defective area identified by this model, the surface correction algorithm is used to planarize the defective area. The defective area is measured using the pixel measurement method, and the crack skeleton is extracted for length and width measurement. The experimental results show that the error rate of the statistics on the defective scale is within 10 %.

Further optimization of the backbone network could be carried out to improve the model performance further. Much real-scale defect data could be collected to enhance the estimation of real-scale defects.

CRediT authorship contribution statement

Jianting Zhou: Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Tengjiao Jiang:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Tianyu Hu:** Writing – review & editing, Methodology, Investigation, Data curation. **Yanfeng Gong:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Conceptualization. **Xiaogang Jiang:** Validation, Methodology, Investigation, Data curation. **Jiangxia He:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Hong Zhang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

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.

Table 4
Measurement error of the proposed method.

Defect no.	Defect type	E_a (cm^2)	M_a (cm^2)	Error (%)	E_l (cm)	M_l (cm)	Error (%)	E_w (cm)	M_w (cm)	Error (%)
1081_1	scratch	1.00	0.95	5.3	-	-	-	-	-	-
1081_2	scratch	3.43	3.56	3.7	-	-	-	-	-	-
1081_3	crack	0.89	0.83	7.2	6.30	6.65	5.3	0.47	0.45	4.4
1081_4	scratch	2.83	2.71	4.4	-	-	-	-	-	-
1082_1	crack	2.42	2.64	8.3	12.11	11.25	7.6	1.07	0.99	8.0
1083_1	crack	0.32	0.34	5.9	1.65	1.74	5.2	1.49	1.56	4.5
1083_2	crack	0.69	0.64	7.8	6.72	6.31	6.5	0.41	0.44	6.8
Mean	-	-	-	6.1	-	-	6.2	-	-	5.9
Test image	-	-	-	6.7	-	-	7.5	-	-	6.5

E_a is the Evaluated defect area, E_l is the Evaluated crack length, E_w is the Evaluated maximum crack width, M_a is the Measured defect area, M_l is the Measured crack length, M_w is the Measured maximum crack width.

Acknowledgements

The support from the National Natural Science Foundation of China (52278291, U20A20314), the Chongqing Natural Science Foundation of China (CSTB2022NSCQLZX0006, CSTB2022TIAD-KPX0205, cstc2022ycjh-bgzxm0086), and the Chongqing Jiaotong University Graduate Student Research Innovation Program (2023S0083).

Data availability

Data will be made available on request.

References

[1] Jianhua Hu, et al., Field monitoring and response characteristics of longitudinal movements of expansion joints in long-span suspension bridges, Measurement 162 (2020) 107933.
[2] Wei Huang, et al., Design and construction of super-long span bridges in China: review and future perspectives, Front. Struct. Civ. Eng. 14 (2020) 803–838.
[3] Sinno Jialin Pan, et al., A survey on transfer learning. IEEE Trans. Knowl. data Eng. 22.10 (2009) 1345–1359.
[4] Dongsheng Li, et al., Monitoring and failure analysis of corroded bridge cables under fatigue loading using acoustic emission sensors, Sensors 12 (4) (2012) 3901–3915.
[5] Matthew Jake Deeble Sloane, et al., Experimental analysis of a nondestructive corrosion monitoring system for main cables of suspension bridges, J. Bridge Eng. 18 (7) (2013) 653–662.
[6] K.M. Mahmoud, et al., Fracture strength for a high strength steel bridge cable wire with a surface crack, Theor. Appl. Fract. Mech. 48 (2) (2007) 152–160.
[7] C.X. Li, et al., Fatigue crack growth of cable steel wires in a suspension bridge: multiscaling and mesoscopic fracture mechanics, Theor. Appl. Fract. Mech. 53 (2) (2010) 113–126.
[8] Fadi Althoey, et al., Machine learning based computational approach for crack width detection of self-healing concrete, Case Stud. Constr. Mater. 17 (2022) e01610.
[9] Jun Luo, et al., Development of cable maintenance robot for cable-stayed bridges, Ind. Robot.: Int. J. 34 (4) (2007) 303–309.
[10] Fengyu Xu, et al., Cable inspection robot for cable-stayed bridges: design, analysis, and application, J. Field Robot. 28 (3) (2011) 441–459.
[11] Obrien E.J. Mcgetrick PJ GonzalezA, et al., Adrive-by inspection system via vehicle moving force identification, Smart Struct. Syst. 13 (5) (2014) 821–848.
[12] Brodie Chan, et al., Towards UAV-based bridge inspection systems: a review and an application perspective, Struct. Monit. Maint. 2 (3) (2015) 283–300.
[13] Junwon Seo, et al., Drone-enabled bridge inspection methodology and application, Autom. Constr. 94 (2018) 112–126.
[14] Kazuma Shibano, et al., Improvement of crack detectivity for noisy concrete surface by machine learning methods and infrared images, Case Stud. Constr. Mater. 20 (2024) e02984.
[15] Christoph Schaal, et al., Damage detection in multi-wire cables using guided ultrasonic waves, Struct. Health Monit. 15 (3) (2016) 279–288.
[16] Hasnae Zejli, et al., Detection of the presence of broken wires in cables by acoustic emission inspection, J. Bridge Eng. 17 (6) (2012) 921–927.
[17] Fengyu Xu, et al., Inspection method of cable-stayed bridge using magnetic flux leakage detection: principle, sensor design, and signal processing, J. Mech. Sci. Technol. 26 (2012) 661–669.
[18] T. Jiang, G.T. Frøseth, S. Wang, Ø.W. Petersen, A. Rønnquist, A 6-DOF camera motion correction method using IMU sensors for photogrammetry and optical measurements, Mech Syst Signal, Process 210 (2024) 111148, https://doi.org/10.1016/j.ymssp.2024.111148.
[19] Tengjiao Jiang, et al., A visual inspection and diagnosis system for bridge rivets based on a convolutional neural network, Comput. Aided Civ. Infrastruct. Eng. (2024).
[20] Tengjiao Jiang, et al., A robust bridge rivet identification method using deep learning and computer vision, Eng. Struct. 283 (2023) 115809.
[21] Chen, Jianuo, et al. Surface defect detection of cable based on threshold image difference. 2021 IEEE far east NDT new technology & application forum (FENDT). IEEE, 2021.
[22] Tengjiao Jiang, et al., A robust line-tracking photogrammetry method for uplift measurements of railway catenary systems in noisy backgrounds, Mech. Syst. Signal Process. 144 (2020) 106888.
[23] Tengjiao Jiang, et al., A detailed investigation of uplift and dam** of a railway catenary span in traffic using a vision-based line-tracking system, J. Sound Vib. 527 (2022) 116875.
[24] Hoai-Nam Ho, et al., An efficient image-based damage detection for cable surface in cable-stayed bridges, Ndt E Int. 58 (2013) 18–23.
[25] Xinke Li, et al., Particle swarm optimization-based SVM for classification of cable surface defects of the cable-stayed bridges, IEEE Access 8 (2019) 44485–44492.
[26] Zhiqiang Li, et al., Gabor wavelet transform combined with area CNN in appearance intelligent detection of stayed cables, J. Vibroeng. 25 (8) (2023) 1465–1479.
[27] Rana Ehtisham, et al., Classification of defects in wooden structures using pre-trained models of convolutional neural network, Case Stud. Constr. Mater. 19 (2023) e02530.
[28] Zengsheng He, et al., A novel MO-YOLOv4 for segmentation of multi-class bridge damages, Adv. Eng. Inform. 62 (2024) 102586.
[29] Jiaxiu Dong, et al., MFAFNet: An innovative crack intelligent segmentation method based on multi-layer feature association fusion network, Adv. Eng. Inform. 62 (2024) 102584.
[30] Fengyu Xu, et al., Nondestructive testing of bridge stay cable surface defects based on computer vision. CMC-Comput. Mater. Contin. 75. 1 (2023) 2209–2226.
[31] Jiaxiu Dong, et al., Innovative method for pavement multiple damages segmentation and measurement by the Road-Seg-CapsNet of feature fusion, Constr. Build. Mater. 324 (2022) 126719.
[32] Li Wang, et al., Effective small crack detection based on tunnel crack characteristics and an anchor-free convolutional neural network, Sci. Rep. 14 (1) (2024) 10355.
[33] Song Chen, et al., Pavement crack detection based on the improved Swin-Unet model, Buildings 14 (5) (2024) 1442.
[34] Qiong Zhang, et al., Improved U-net network asphalt pavement crack detection method, Plos One 19 (5) (2024) e0300679.
[35] Huang Huang, et al., A three-stage detection algorithm for automatic crack-width identification of fine concrete cracks, J. Civ. Struct. Health Monit. (2024) 1–10.
[36] Shitong Hou, et al., Inspection of surface defects on stay cables using a robot and transfer learning, Autom. Constr. 119 (2020) 103382.
[37] Ashish Gaur, et al., A novel approach for industrial concrete defect identification based on image processing and deep convolutional neural networks, Case Stud. Constr. Mater. 19 (2023) e02392.
[38] Yeping Peng, et al., Non-uniform illumination image enhancement for surface damage detection of wind turbine blades, Mech. Syst. Signal Process. 170 (2022) 108797.
[39] Prahar M. Bhatt, et al., Image-based surface defect detection using deep learning: A review, J. Comput. Inf. Sci. Eng. 21 (4) (2021) 040801.
[40] R. Hussin, et al., Digital image processing techniques for object detection from complex background image, Procedia Eng. 41 (2012) 340–344.
[41] Antonio Torralba, et al., Labelme: online image annotation and applications. Proc. IEEE 98.8 (2010) 1467–1484.
[42] Lin, Tsung-Yi, et al. Microsoft coco: Common objects in context. Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6-12, 2014, proceedings, part V 13. Springer international publishing, 2014.

[43] Bolya, Daniel, et al. Yolact++ better real-time instance segmentation. University of California, Davis, 2020.
[44] He, Kaiming, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[45] Vaswani, Ashish, et al. Attention is all you need. Advances in neural information processing systems 30 (2017).
[46] Misra, Diganta. Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681 (2019).
[47] He, Kaiming, et al. Mask R-CNN [C]|| Proceedings of 2017 IEEE international conference on computer vision. Venice, Italy: IEEE (2017): 2980-2988.
[48] Wang, Xinlong, et al. Solo: Segmenting objects by locations. Computer vision–ECCV 2020: 16th european conference, glasgow, UK, August 23–28, 2020, proceedings, part XVIII 16. Springer international publishing, 2020.
[49] Wang, Xinlong, et al. Solov2: Dynamic and fast instance segmentation. Advances in neural information processing systems 33 (2020): 17721-17732.
[50] Bolya, Daniel, et al. YOLACT: Real-time instance segmentation. Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[51] Liu, Haotian, et al. YolactEdge: Real-time instance segmentation on the edge.2021 IEEE international conference on robotics and automation (ICRA). IEEE, 2021.
[52] Cheng, Tianheng, et al. Sparse instance activation for real-time instance segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.