Smart Structures and Systems, *Vol. 37, No. 6 (2026) 529-551*
https://doi.org/10.12989/sss.2026.37.6.529

Techno-Press

Building concrete surface crack detection method based on improved YOLOv8

Fangzhen Hu*

School of Architectural Engineering, Wuhan City Polytechnic Wuhan 430068, China

(*Received April 1, 2026, Revised June 8, 2026, Accepted June 9, 2026*)

Abstract. This article proposes a method for detecting surface cracks in building concrete based on improved YOLOv8. By introducing deformable attention mechanism (DAttention) in the backbone network, the crack growth trend can be dynamically focused; Enhance multi-scale feature expression capability in neck design (Cross scale Feature Fusion Module, CCFM); Embedding Efficient Channel Attention (ECA) in the head to enhance the weight of key features; And replace the Complete Intersection over Union (CIoU) loss function with the Scale invariant Intersection over Union (SIoU) loss function to optimize the bounding box regression process. The experimental results show that our method achieved a detection accuracy of 88.4%, a recall rate of 95.2%, and an average precision mean (mAP) of 96.4% on a self built dataset, which is significantly improved compared to the benchmark YOLOv8 model. This method is capable of extracting crack features in a complete and continuous manner, effectively identifying subtle cracks and suppressing background interference, providing reliable technical support for the health assessment and safety risk prevention of building structures.

Keywords: attention mechanism; building concrete; crack detection; feature fusion; improved YOLOv8; loss function

1. Introduction

As the cornerstone of modern engineering construction, the structural integrity and durability of building concrete directly determine the safety and service life of the building. However, during long-term service, concrete structures are inevitably subjected to the coupling effects of multiple factors such as loads, material aging, temperature stress, and chemical erosion, leading to the generation and development of surface cracks [1-3]. These cracks are not just superficial defects, but also "warning signals" for the internal health status of the structure. They will significantly weaken the load-bearing capacity of the structure and provide entry paths for harmful media such as moisture and chloride ions, accelerating the corrosion of internal steel bars and the deterioration of concrete, ultimately posing a serious threat to the overall stability and safety of the structure [4-5]. Therefore, timely and accurate detection of surface cracks in concrete has become an indispensable key link in the health monitoring and safety management of building structures. It is of great significance for preventing catastrophic accidents, formulating scientific maintenance strategies, and extending the service life of structures.

*Corresponding author, M.D., E-mail: hufz1234@163.com

http://www.techno-press.org/sss/5

ISSN: 1738-1584 (Print), 1738-1991 (Online)

Traditional concrete crack detection mainly relies on manual visual inspection, which is not only inefficient and subjective, but also difficult to implement in certain high-risk or concealed environments. In recent years, significant progress has been made in intelligent detection methods based on computer vision and deep learning, providing possibilities for automated detection. However, due to the complex surface texture and diverse background interference of concrete, as well as the characteristics of variable scale and complex topological structure of crack morphology, achieving high-precision and strong robustness in automated detection still faces severe challenges. Although current research methods have their own characteristics, they also have corresponding limitations. Yamaguchi et al. proposed a crack detection method based on convolutional neural networks (CNNs) and self-organizing maps (SOMs) [6]. They used the CNN architecture to extract crack features from images, constructed a SOM network, and conducted cluster analysis on the extracted crack features. By training the SOM network, similar crack features were mapped to adjacent positions in the network, thereby achieving the classification and visualization of crack types. The image to be detected was input into the trained CNN model to extract crack features, and the extracted features were then input into the SOM network. The type and location of the cracks were determined based on the clustering results. The extraction of crack features by CNNs relies on a large amount of high-quality training data. In practical applications, the diversity and complexity of crack images can lead to insufficient feature extraction, thus affecting the detection accuracy. Yadav et al. proposed a crack detection method based on bridged CNNs and transformers [7]. They designed a crack detection model that bridged CNNs and transformers. The CNN architecture employed stacked convolutional and pooling layers for local feature extraction, while the transformer leveraged self-attention to model long-range feature dependencies. The model was trained using a preprocessed image dataset, and data augmentation techniques were employed to increase the diversity of the dataset and improve the generalization ability of the model. The image to be detected was input into the trained model, and the model would output the crack detection results. In real scenarios, cracks may appear in complex backgrounds, such as those with rich textures and uneven illumination, which can cause interference when the model extracts crack features, thus affecting the detection accuracy. Manjunatha et al. proposed a method for crack detection using a deep convolutional neural network based on semantic segmentation [8]. They collected an image dataset containing cracks and annotated the images to generate crack label images for semantic segmentation. They designed and trained a deep convolutional neural network architecture based on semantic segmentation. The image to be detected was input into the trained model, and the model would output the crack detection results. To enhance detection accuracy, post-processing techniques were applied. While semantic segmentation-based DCNNs have demonstrated strong crack detection performance, their generalization capability remains constrained. If the crack features in the image to be detected differ significantly from those in the training data, the model may not be able to accurately identify them. Bhardwaj et al. proposed a crack detection method based on fuzzy c-means clustering and selective edge enhancement [9]. They set the number of clusters and the fuzzy factor and applied the fuzzy c-means clustering algorithm to perform pixel-level clustering on the preprocessed image. Based on the results of the fuzzy C-means clustering, the pixels belonging to the crack region were identified, and edge enhancement processing was performed on the pixels in these crack regions. In the image after edge enhancement, the crack regions were further extracted through methods such as threshold segmentation and morphological operations. The extracted crack regions were identified and annotated to generate the crack detection results. The effectiveness of fuzzy C-means (FCM) clustering is critically influenced by both the cluster count and fuzziness parameter, which directly determine segmentation quality

and subsequent edge enhancement outcomes. Improper parameter settings can lead to the crack regions being incorrectly classified into other regions or poor edge enhancement effects for the cracks. Niu et al. combined artificial intelligence and computer vision technology to propose a pixel level fine automatic recognition and segmentation method for concrete dam cracks in complex backgrounds [10]. To overcome the interference of complex environmental background factors, various digital image processing methods were used to preprocess the crack images of concrete dams. Based on Mask Region-based Convolutional Neural Network (Mask R-CNN), the model backbone network was improved to enhance the crack feature extraction ability. Although this method enhances the ability to extract crack features by improving the backbone network of Mask R-CNN, its average accuracy may still be limited by residual interference from complex backgrounds, which affects the further improvement of overall average accuracy. Su and Wang proposed a road crack recognition method based on YOLO v3 deep learning algorithm [11]. Firstly, the dataset images are scaled to 416 × 416. Then, Labelme is used to label the data for cracks and convert the bounding box position information. Finally, the YOLO v3 algorithm framework is used for model training Although this method utilizes the YOLO v3 algorithm to achieve efficient crack recognition, scaling the input image uniformly to 416 × 416 pixels can result in loss of crack detail information, especially insensitive to subtle cracks, which directly leads to a low recall rate.

In summary, addressing the shortcomings of existing methods such as poor adaptability in complex backgrounds, low sensitivity to fine cracks, and insufficient generalization capabilities, this paper proposes a building concrete surface crack detection method based on an improved YOLOv8. The main contributions and novelty of this work are as follows:

(1) Introducing the deformable attention mechanism (DAttention) into the backbone network enables the convolutional kernel to dynamically adapt to the irregular morphology of cracks, overcoming the limitations of fixed geometric structures and effectively suppressing interference from complex backgrounds.

(2) Design a cross-scale feature fusion module (CCFM) in the neck network, which integrates deep semantic and shallow detail information through lightweight upsampling and weighted concatenation, addressing the issue of fine cracks being easily lost in the feature pyramid.

(3) An Efficient Channel Attention (ECA) mechanism is embedded in the head network to adaptively recalibrate channel weights with minimal computational overhead, enhancing crack-related features and improving the model's discriminative capability.

(4) Replace the CIoU loss function with the SIoU loss function, introducing angular cost and redefined distance cost to optimize the convergence stability and localization accuracy of bounding box regression.

The synergistic effects of the aforementioned four contributions provide reliable technical support for the health assessment of building concrete structures.

2. Building concrete surface crack detection method

To clearly demonstrate the overall architecture and core improvements of the concrete crack detection method proposed in this article, a flowchart of the construction concrete surface crack detection method based on the improved YOLOv8 is provided, as shown in Fig. 1.

2.1 Construction of the building concrete surface crack detection model based on YOLOv8

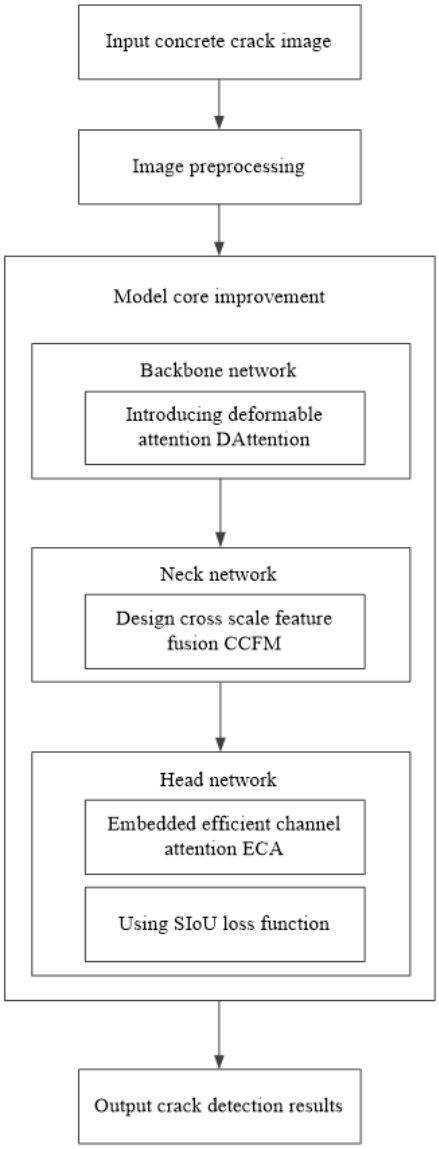


Figure 1. Overall process of surface crack detection method for building concrete based on improved YOLOv8

Developed by Ultralytics in 2023, YOLOv8 represents a significant advancement in real-time object detection and image segmentation technology. This latest iteration introduces architectural refinements that enhance both computational efficiency and recognition performance compared to its predecessors, achieving superior balance between speed and precision. Therefore, when conducting building concrete surface crack detection, the YOLOv8 network model is selected. However, the core innovation of this study is not simply the use of YOLOv8, but rather the systematic improvement of YOLOv8 to address specific issues such as irregular crack morphology, variable scales, strong background interference, and easy omission of subtle cracks in the detection of surface cracks in building concrete. Specifically, this article proposes targeted optimization

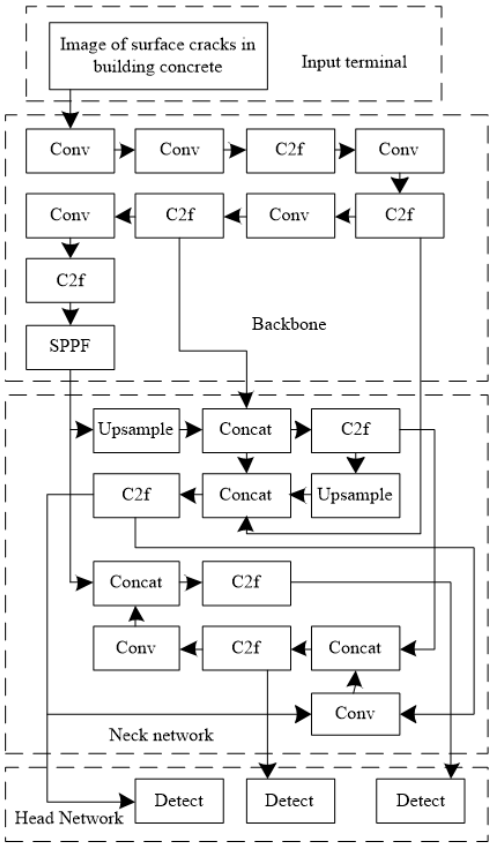


Figure 2. Structural diagram of the building concrete surface crack detection model based on YOLOv8

strategies in the following four aspects: (1) introducing deformable attention mechanism (DAttention) in the backbone network to enhance the model's spatial adaptability to irregular cracks; (2) Design a Cross Scale Feature Fusion Module (CCFM) in the neck network to enhance the fusion effect of multi-scale crack features; (3) Embedding efficient channel attention mechanism (ECA) in the head network to enhance the feature expression of crack related channels; (4) Using SIoU loss function instead of CIoU to optimize the convergence stability and positioning accuracy of bounding box regression. The detailed principle of the above improvements can be found in section 2.2. Fig. 2 shows the structure of the building concrete surface crack detection model based on YOLOv8 [12], which is mainly composed of four parts: the input end (Input), the backbone network (Backbone), the neck (Neck), and the head (Head).

As the starting point for the model to perform the building concrete surface crack detection task, the input end mainly pre-processes the input building concrete surface crack images, such as adaptive input size, mosaic data enhancement, and adaptive grayscale filling, to improve the robustness of the model to the size changes of building concrete surface cracks.

The backbone network extracts crack features from concrete surfaces, where the C2f module replaces YOLOv5's C3 structure with fewer parameters while enhancing feature extraction efficiency. This optimized architecture maintains detection accuracy while significantly reducing computational complexity.

As a key component connecting the backbone network and the head, the neck part is designed for multi - scale feature fusion and generating multi - scale building concrete surface crack feature maps, which can perform high - precision detection of building concrete surface cracks of different sizes under the requirements of low computational cost and low latency.

The head is the last layer of the YOLOv8 detection model, which is responsible for processing the extracted building concrete surface crack features and outputting the final building concrete surface crack detection results.

By leveraging multi-layer collaboration, YOLOv8 significantly improves its processing capability for concrete surface crack images in buildings. This architectural enhancement boosts crack detection accuracy while maintaining computational efficiency, simultaneously strengthening the model's robustness and generalization capacity across diverse structural scenarios.

2.2 Improvement of the YOLOv8 network model

To systematically enhance YOLOv8's ability to detect irregular and subtle cracks in complex backgrounds, this paper improves the benchmark model from four dimensions: feature extraction, multi-scale fusion, channel recalibration, and bounding box regression. Each module forms a progressive optimization chain, and the specific integration method is as follows.

2.2.1 Improvement of the backbone network based on the deformable attention mechanism

Building concrete typically dominates the background in crack detection images, where cracks predominantly exhibit elongated, narrow characteristics. To enhance detection precision, we integrate the deformable attention mechanism (DAttention) into YOLOv8's backbone network. This spatial adaptation module enables dynamic focus adjustment during image processing [13-14], particularly optimizing performance for thin linear features characteristic of concrete surface cracks. It can better capture the growth trend and direction of surface cracks in building concrete, to minimize background interference while enhancing both recognition precision and detection efficiency for concrete surface cracks, Fig. 3 illustrates the modified YOLOv8's backbone architecture.

The DAttention mechanism enhances crack detection through deformable convolution and attention integration. Its key innovation involves:

(1) Augmenting standard convolution with learnable offsets, enabling dynamic sampling point adjustment to match crack morphology and orientation.

Let the input feature map be $X \in R^{H\times W\times C}$, where H, W and C represent the height, width, and number of channels of the output features of surface cracks in building concrete. An additional convolutional layer is used to generate the offset Δp_k, where k represents the index of the sampling point of the convolution kernel

$$\Delta p_k = Conv(X) \tag{1}$$

Among them, p represents the position of the output feature map, p_k is the sampling point of the standard convolution kernel. Sample the input feature map according to the offset Δp_k to obtain the deformable feature X_d

$$X_d(p) = \sum_{k=1}^{K} w_k \cdot X(p + p_k + \Delta p_k) \tag{2}$$

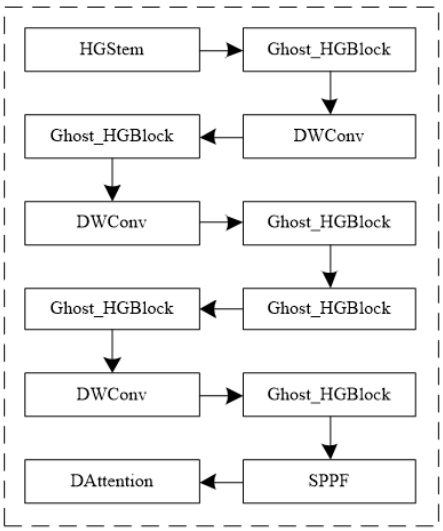


Figure 3. Improved YOLOv8 Model Backbone Network

Among them, w_k is the weight of the convolution kernel.

(2) The attention mechanism assigns adaptive weights to sampling points, emphasizing crack features while suppressing background interference.

$$\alpha_k = Soft\,max\big(Conv(X_d)\big) \tag{3}$$

(3) Combine the features extracted by deformable convolution with the attention weights to generate a more accurate representation of crack features

$$y_p = \sum_{k=1}^{K} \alpha_k \cdot X_d(p) \tag{4}$$

Embed the DAttention module into the backbone network of YOLOv8, replacing the original standard convolutional layer. In this way, the model can better capture the local details and global structural information of cracks while reducing the interference of background noise.

To gain a deeper understanding of how the DAttention module guides the model to focus on the crack area, the decision logic is explained by extracting and analyzing the spatial attention map output by the module. This graph is directly generated by the spatial attention weight A in the DAttention module, and its value $A(x, y)$ represents the importance of position (x, y) on the feature map.

Assuming the input image is I, after forward propagation through the backbone network, the spatial attention matrix A can be obtained by embedding the DAttention layer. The generation process of this matrix can be expressed as

$$A = \varphi_s(F) \tag{5}$$

Among them, F is the input feature map of this layer; φ_s is the spatial attention calculation

function.

By aggregating the L2 norm of the spatial attention matrix A along the channel dimension and upsampling it to the input image resolution using bilinear interpolation, a heatmap H is generated that is consistent with the input image size

$$H = U_p\left(\sqrt{\sum_c A_c^2}\right) \tag{6}$$

Among them, the areas with higher values in H (warm tones) represent the key areas that the model focuses on when making crack detection decisions. If H shows high response in the crack pixel area and forms a significant contrast with the background area, it proves that the DAttention mechanism successfully guides the model's attention to crack features, thus intuitively verifying its ability to reduce background interference and capture crack growth trends.

The DAttention module is embedded in the last three C2f modules of the backbone network, replacing the original standard 3×3 convolutional layer. Its core function is to enable the model to have spatial deformation adaptability during the feature extraction stage, dynamically fit the crack direction, and suppress background texture interference such as aggregate and bubble holes on the concrete surface. This module provides purer initial features for subsequent CCFM and ECA.

The DAttention module proposed in this article is an original design that differs fundamentally from existing deformable convolution or attention mechanism methods in the following three aspects:

(1) Different from standard deformable convolution: Standard DCN only adapts to geometric deformation by learning the offset of sampling points, but the feature importance corresponding to the offset sampling points is treated equally, lacking the ability to distinguish differences. This can make it difficult for the model to distinguish the true crack features from the background like aggregates, bubble holes, etc. in concrete crack detection. The DAttention module in this article introduces a sampling point level attention weight learning mechanism within the framework of deformable convolution, enabling the model to synchronously optimize the sampling position and feature importance, and achieve selective focusing on the crack area.

(2) Different from existing attention mechanisms: Existing attention mechanisms usually calculate the weights of channels or spatial dimensions on fixed grid feature maps, and their receptive fields and sampling positions are fixed in advance, which cannot dynamically adapt to the irregular geometric shape of cracks. This article presents the first end-to-end joint optimization of deformable sampling and attention weight calculation in the DAttention module. The allocation of attention weights no longer relies on fixed position convolution features, but is based on a dynamically adjusted set of sampling points, enabling the model to more accurately capture the growth trend and direction of cracks.

(3) Different from other fields of deformable attention applications: In natural language processing or general object detection, deformable attention is often used to handle global contextual relationships. This article addresses the special needs of concrete crack detection by designing lightweight offset prediction branches and attention weight generation branches, and embedding the DAttention module into the last three C2f modules of the YOLOv8 backbone network to replace the original standard convolutional layer. This embedding aims to enable the model to have spatial deformation adaptability during the high-level semantic feature extraction stage, providing purer crack features for subsequent cross scale fusion and channel recalibration.

In summary, the DAttention module is an improved module proposed in this article for the task of detecting cracks in building concrete. Its core innovation lies in the organic integration of deformable sampling mechanism and sampling point level attention weight learning.

2.2.2 Neck network improvement based on the CCFM network model

In the task of detecting cracks on the surface of building concrete, the feature maps of building concrete surface cracks at different layers have different spatial resolutions and semantic information. Lower-layer features capture detailed edge and texture information, whereas higher-layer features encode abstract semantics. The CCFM integrates these multi-level representations of concrete surface cracks, significantly improving performance in complex backgrounds and for fine crack detection. Consequently, CCFM was adopted to enhance the neck network architecture.

CCFM first extracts information from the input feature map of building concrete surface cracks through a series of lightweight convolutional layers. Subsequently, it horizontally fuses the feature maps of building concrete surface cracks at different scales through the 1×1 convolution kernel [15-16], realizing the interaction and supplementation between the features of building concrete surface cracks and enhancing the expression ability of these features. During the fusion process of building concrete surface cracks, CCFM can utilize the features of building concrete surface cracks from multiple scales to obtain comprehensive information, extract the features of building concrete surface cracks from convolutional layers at different depths, and select the feature maps of building concrete surface cracks to be fused. First, the low-resolution feature map of building concrete surface cracks is adjusted to the same size as the high-resolution feature map through upsampling operations such as bilinear interpolation, obtaining the high-resolution feature map of building concrete surface cracks y_{up}

$$y_{up} = U(y_{low}) \tag{7}$$

Subsequently, y_{up} and the high-resolution feature map y_{high} are weighted and spliced to achieve multi-scale fusion of the features of building concrete surface cracks. The fused feature map of building concrete surface cracks y_{fused} is

$$y_{fused} = \alpha \cdot y_{high} + (1 - \alpha) y_{up} \tag{8}$$

Subsequently, CCFM applies the Rectified Linear Unit (RELU) activation function and convolution operation to deeply process the feature map of building concrete surface cracks to enhance the model's expressiveness. The formula is described as follows

$$y_p = RELU\left(\partial\left(y_{fused}\right)\right) \tag{9}$$

The CCFM module is integrated between the FPN and PAN of the neck network. Specifically, the P3, P4, and P5 feature maps output by the backbone network are convolved by 1 × 1 to align the number of channels and input into CCFM; Within CCFM, low resolution feature maps are upsampled to high-resolution sizes through bilinear interpolation, and then weighted and concatenated with high-resolution feature maps to generate fused feature maps [17]; After convolution and ReLU activation, the fused image is fed back to the corresponding level of PAN. CCFM complements DAttention: DAttention ensures the purity of single scale features, while CCFM ensures the fusion of multi-scale features, solving the problem of subtle cracks being easily diluted in the feature pyramid.

2.2.3 Improvement of the head network based on the ECA module

To enhance detection of subtle concrete surface cracks, we integrate the Efficient Channel Attention (ECA) mechanism [18] into YOLOv8's head network. The ECA module adaptively recalibrates channel-wise feature weights, prioritizing crack-relevant information while suppressing background interference. Compared with the commonly used SE attention mechanism and CBAM attention mechanism, ECA has lower computational complexity and fewer parameters, which can effectively enhance the feature expression ability of building concrete surface cracks while avoiding excessive consumption of computational resources.

This article provides a detailed design of the integration method of ECA module in YOLOv8 detection head. The ECA module is inserted after the 3×3 convolutional layer of each detection branch and before the classification/regression output, and is embedded in both the classification and regression branches, forming a structure of "convolution → ECA → output". In terms of application scope, the ECA module is applied to the three detection branches P3, P4, and P5 in a scale independent configuration. The parameters of each branch learn independently to adapt to the differences in the number of channels in feature maps of different scales (P3 has 256 channels, P4 has 512 channels, and P5 has 1024 channels). Taking the P3 scale (256 × 80 × 80) as an example, the data stream is as follows: the feature map is convolved by 3 × 3, compressed into channel vectors through global average pooling, and adaptively convolved to generate channel weights. After Sigmoid activation, it is multiplied back to the original feature map channel by channel and finally sent to the output layer. The processing flow for P4 and P5 scales is the same, with only differences in feature map size and channel count.

In the ECA module, global average pooling is first performed on the input building concrete surface crack feature map to compress the original building concrete surface crack feature map and achieve the interaction of global context information. Then, a one-dimensional convolution kernel is used to process the compressed building concrete surface crack feature map to ensure that the coverage of cross-channel interaction is proportional to the channel dimension. Among them, the one-dimensional convolution kernel affects the range of the receptive field and the ability to capture inter-channel dependencies. A smaller one-dimensional convolution kernel captures local information with high computational efficiency; a larger one-dimensional convolution kernel captures global information but incurs greater computational overhead. The value of the one-dimensional convolution kernel is

$$M = \sigma\left(\frac{log(C)}{\xi} + \frac{b}{\xi}\right) \tag{10}$$

Where: ξ and b are constants, with values of 2 and 1 respectively.

The weights of each channel are calculated using one-dimensional convolution, and the weights are mapped between 0 and 1 through an activation function. Finally, the obtained weight values are multiplied and weighted to the original input building concrete surface crack feature map to enhance the feature expression ability.

To further demonstrate how the ECA module influences the final detection decision of the model through channel weighting strategy, the gradient weighted channel activation mapping (Grad CAM's channel attention variant) technique is introduced. This technology calculates the gradient information of the target category (i.e., cracks) and flows it back to the convolutional layer before the ECA module, and weights the feature map of this layer to generate a heatmap that can explain the decision-making basis of the model.

Let the convolutional layer before the ECA module in the head network output a feature map A^k, where k is the feature map index. The score of the model for crack categories is denoted as y^c. The importance weight α_k^c of each feature map A^k for crack detection decisions is obtained by calculating the gradient $\frac{\partial y^c}{\partial A^k}$

$$\alpha_k^c = \frac{1}{B}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{11}$$

Among them, B is the spatial size of the feature map; i, j is the spatial position index.

Subsequently, by weighting the importance weight α_k^c with the original feature map A^k and filtering out negative response pixels through the ReLU activation function, the final explanatory heatmap L_G^c is generated

$$L_G^c = Re\,L\,U\left(\sum_k \alpha_k^c A^k\right) \tag{12}$$

The ECA module is inserted into each feature map branch of the detection head, located between the output of the convolutional layer and the classification/regression branch. The workflow is as follows: the input feature map is first compressed by global average pooling of channel dimensions, then channel weights are generated through adaptive kernel size one-dimensional convolution, and finally the weights are multiplied back to the original feature map channel by channel. ECA and CCFM work closely together: CCFM is responsible for the fusion of cross scale spatial features, and ECA recalibrates the semantic information carried by channels based on this, enhancing discriminative channels with minimal parameter count and suppressing redundant channels, thereby directly improving the ability to distinguish between subtle cracks and background noise.

2.2.4 Improvement of the loss function

In the task of detecting surface cracks in building concrete, cracks exhibit diversity and complexity in angles and shapes. Conventional loss functions poorly accommodate building concrete crack characteristics. Excessive penalization of the model's flexibility in identifying crack geometry can destabilize training and induce overfitting. The CIoU loss enhances crack detection by jointly optimizing bounding box regression through three geometric factors: aspect ratio consistency, centroid alignment, and intersection-over-union coverage. However, when the aspect ratios match, CIoU's penalty term becomes zero, allowing poor-quality samples to disproportionately affect regression loss and destabilize convergence. In contrast, the SIoU loss function enhances crack detection accuracy and stability by emphasizing crack position and orientation. It consists of four key components.

Aspect 1: Angle cost. Cracks in images usually exhibit obvious directionality (such as longitudinal cracks, transverse cracks, or oblique cracks), and the angle deviation between the predicted box and the real box directly affects the convergence path of subsequent regression. The angle cost is reduced by introducing a trigonometric penalty term to prioritize the reduction of angle deviation during the training process of the model. Specifically, when the horizontal offset angle between the predicted box and the real box is greater than 45°, the model tends to prioritize optimizing the vertical alignment; On the contrary, priority should be given to optimizing the horizontal direction. This design enables the predicted box to rotate faster in the same direction as the real box, thereby accelerating convergence and improving the accuracy of locating directional

cracks.

The calculation formula for defining the angle cost as Ψ, Ψ is

$$\Psi = 1 - 2 \cdot \sin^2\left(\arcsin x - \frac{\pi}{4}\right) \tag{13}$$

$$x = \frac{\delta_h}{\sigma} = sin\,\eta \tag{14}$$

$$\sigma = \sqrt{\left(e_{c_x}^{gt} - e_{c_x}\right)^2 + \left(e_{c_y}^{gt} - e_{c_y}\right)^2} \tag{15}$$

$$\delta_h = \max\left(e_{c_y}^{gt} - e_{c_y}\right) - \min\left(e_{c_y}^{gt} - e_{c_y}\right) \tag{16}$$

In the formula: η is the horizontal offset angle between the predicted and real crack boxes on the building concrete surface; σ is the distance between the center points of the real and predicted crack boxes on the building concrete surface; δ_h is the height difference between the center points of the real and predicted crack boxes on the building concrete surface; $e_{c_x}^{gt}$ is the abscissa of the center of the real crack box on the building concrete surface; e_{c_x} is the abscissa of the center of the predicted crack box on the building concrete surface; $e_{c_y}^{gt}$ is the ordinate of the center of the predicted crack box on the building concrete surface; e_{c_y} is the ordinate of the center of the predicted crack box on the building concrete surface.

Aspect 2: Distance cost. After the angle cost guides the predicted box to the direction channel of the real box, the distance cost is responsible for reducing the center point distance between the two. Distance cost is weighted on the basis of angle cost, and when the angle cost is small (i.e., the predicted box is aligned with the true box direction), distance cost plays a major role; When the angle cost is high, the contribution of distance cost is suppressed, avoiding blind optimization of the model in inappropriate directions. This collaborative mechanism significantly improves regression efficiency.

The calculation formula for defining the distance cost as Ω, Ω is

$$\Omega = \sum_{t=x,y} (1 - \Sigma^{-\gamma\rho_t}) \tag{17}$$

In the formula: Σ represents the exponential function. When κ is close to 0, the contribution cost of the distance is greatly reduced. When κ is close to 45°, the distance cost Ω contributes more. As the angle increases, γ is defined as the distance value with time priority.

Aspect 3: Shape cost. Shape cost is used to constrain the consistency between the predicted box and the real box in terms of width and height ratios. This article fixes it at 1 in order to avoid excessive interference of shape constraints on the optimization of predicted box positions and angles during the early stages of training. For concrete crack detection, the width of cracks is usually much smaller than their length, and the aspect ratio of different cracks varies greatly. Applying shape constraints too early or too strongly may cause the model to fall into local optima and be difficult to adapt to the changing shapes of cracks. After fixing the shape cost to 1, the model can more flexibly explore the position and direction of cracks, and fine adjust the shape through IoU cost after the

predicted box and the real box are basically overlapped.

The calculation formula for defining the shape cost as Z, Z is

$$Z = \sum_{t=w,h} (1 - \Sigma^{-N_t})^{\theta} \tag{18}$$

The value of Z defines the importance of the shape cost. When Z is 1, it will limit the shape of the predicted volume and its free movement. To avoid the excessive loss of the shape from affecting the movement of the prediction box for cracks on the surface of building concrete, this value is set to 1.

Aspect 4: IoU cost. IoU cost is the fundamental term of the loss function, which directly measures the degree of overlap between the predicted box and the true box. During the training process, the four cost components interact synergistically according to the following logic: firstly, angle cost guides the prediction box to quickly align the direction of cracks; Subsequently, the distance cost is reduced based on directional alignment to minimize the distance from the center point; Meanwhile, the shape cost constrains the aspect ratio with fixed strength, but does not dominate the optimization direction; Ultimately, IoU cost is used as the core positioning indicator to finely optimize the degree of overlap. The four costs are fused into the final SIoU loss function through a combination of multiplication and addition, enabling the model to achieve efficient and stable bounding box regression for crack detection tasks.

The calculation formula for defining the ratio of the intersection to the union of the real box and the prediction box as I_{IoU}, I_{IoU} is

$$I_{IoU} = \frac{|Q \cap Q^{GT}|}{|Q \cup Q^{GT}|} \tag{19}$$

So far, the calculation formula for defining the value of the SIoU loss function as L_{SIoU}, L_{SIoU} is

$$L_{SIoU} = 1 - I_{IoU} + \frac{\Omega + \Psi}{2} \tag{20}$$

On the basis of the original IoU loss function, the angle cost Ψ and the distance cost Ω are added, which reduces the probability that the penalty term is 0, weakens the excessive penalty of geometric factors on the model, makes the convergence process more stable, and improves the prediction accuracy of cracks on the surface of building concrete.

The SIoU loss function directly replaces the CIoU loss function in the benchmark YOLOv8 for optimizing the boundary box regression branch. DAttention, CCFM, and ECA jointly improve the initial quality of crack prediction boxes, while the regression loss function determines the path and stability of convergence from the predicted box to the true box. By introducing angle cost and redefined distance cost, SIoU reduces the probability of zero penalty term, weakens the excessive penalty of geometric factors on the model, and achieves more accurate and stable crack localization.

3. Experimental results

3.1 Experimental environment

This paper studies a method for detecting cracks on the surface of building concrete based on the

Table 1. Experimental parameter settings

Parameter	Parameter values
Input size	640×640
Initial learning rate	0.01
Minimum learning rate	0.0001
Number of iterations	300
Batch size	32
Optimizer	Stochastic Gradient Descent (SGD)
Ratio	0.8

improved YOLOv8. To verify the practical application performance of the method in this paper, the following experimental environment is set:

The operating system used in the experiment is Windows 11, the Graphics Processing Unit (GPU) is NVIDIA GeForce RTX 3090, the GPU memory is 24 GB, the Compute Unified Device Architecture (CUDA) version is 12.1, the Torch version is 1.12.0. The model training uses YOLOv8s as the benchmark network, with 300 iterations during training, Batchsize of 32, initial learning rate of 0.01, and minimum learning rate of 0.0001. The optimizer uses SGD (momentum of 0.937, weight decay of 0.0005), and the input image size is uniformly 640 × 640 pixels. The specific experimental parameter settings are shown in Table 1.

3.2 Dataset collection and processing

At present, there is a lack of publicly available datasets specifically for detecting surface cracks in building concrete. To construct a dataset suitable for this study, crack images from multiple publicly available datasets were integrated, and some self collected images were supplemented. Specifically, the data mainly comes from the following three channels: (1) concrete structure crack samples in the public dataset SDNET2018; (2) Select images that meet the characteristics of building concrete from the Crack Detection project on the GitHub open-source platform; (3) Concrete surface crack images captured using high-definition cameras at actual construction sites (acquisition equipment: Sony α 7R IV, resolution: 7952 × 5304). The distribution of data from various sources is as follows: SDNET2018 contributed 1562 pieces (accounting for 46.3%), including 492 cracks in perforated concrete slabs, 538 cracks in plain concrete slabs, and 532 cracks in reinforced concrete beams; The GitHub open-source platform contributed 1024 files (accounting for 30.4%), mainly from publicly available datasets of road and bridge cracks; We collected and contributed 785 images (accounting for 23.3%), covering crack images under different lighting conditions (strong light, shadow, cloudy) and backgrounds (plain concrete, formwork joints, rust stains) at the construction site. By integrating multiple sources of data, a total of 3371 representative images of surface cracks in building concrete were collected, ensuring the diversity and representativeness of the dataset.

In terms of crack type distribution, the dataset covers 1124 longitudinal cracks (33.3%), 986 transverse cracks (29.3%), 703 network cracks (20.9%), and 558 micro cracks (with a maximum width less than 2 mm) (16.5%). In terms of lighting conditions, normal lighting images account for 65%, strong light (including local overexposure) images account for 18%, shadow area images account for 10%, and low light images account for 7%. In terms of background environment, it includes clean concrete surfaces (52%), surfaces with exposed aggregates (23%), surfaces with

template joints or holes (15%), and surfaces covered with rust stains or pollutants (10%). The above distribution reflects the complexity and diversity of concrete surface cracks in actual engineering, while also revealing possible biases in the dataset: the proportion of fine crack samples is relatively low (16.5%), and there are fewer images under low light conditions (7%), which may result in relatively weak detection performance of the model in these scenarios.

In the data preprocessing stage, a script is used to randomly divide the dataset into a training set (2359 images) and a testing set (1012 images) in a 7:3 ratio. To ensure the rigor of the evaluation, it is necessary to ensure that images from the same source are evenly distributed in the training and testing sets during partitioning to avoid data leakage. Create PASCAL VOC format annotation files for all images using the open-source data annotation tool LabelImg. During the annotation process, use the smallest bounding box to accurately label crack areas, ensuring that each bounding box contains only one complete crack and strictly avoiding overlap between bounding boxes. All crack samples are uniformly labeled as the "Crack" category. Through the systematic data collection and annotation work mentioned above, the quality and applicability of the dataset have been effectively ensured.

During the model training phase, the method of transfer learning was adopted, and the same training process as that for dataset division was used for model training. To accelerate the convergence speed of the model and improve the training effect, the weight parameters pre-trained on the large-scale Common Objects in Context (COCO) dataset were selected as the initial weights of the model. Subsequently, training was carried out on the surface crack dataset of building concrete, with a total of 300 iterations.

3.3 Evaluation plan

To objectively evaluate the detection performance of the method proposed in this article, the following evaluation plan is formulated:

(1) Evaluation indicators: Quantitative evaluation is conducted using four indicators: Precision, Recall, Mean Precision, and F1 score. Among them, accuracy measures the correctness of the detection results, recall measures the ability to detect cracks, average precision reflects the positioning and classification performance of the model under different confidence thresholds, and F1 value is the harmonic average of the first two.

(2) Data partitioning and validation method: The dataset is randomly divided into a training set (2359 images) and a testing set (1012 images) in a 7:3 ratio. All models are trained and tested under the same partition, and the test set is only used for final performance evaluation and does not participate in any model tuning process. To eliminate the bias caused by random partitioning, five repeated experiments with different random seeds were conducted, and the average and standard deviation of each indicator were taken as the final result.

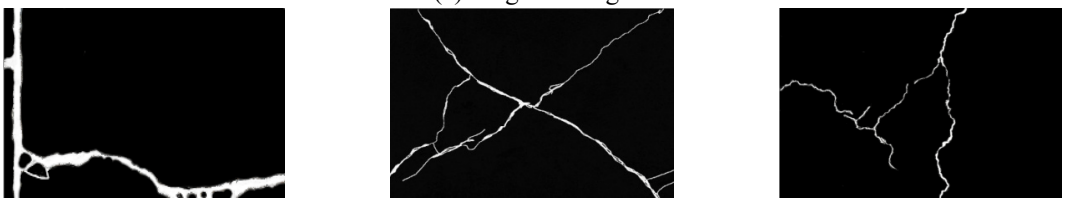
(3) Testing process: The model performs forward inference on the test set, and after non maximum suppression (NMS) of the predicted results, calculates the IoU with the annotation box. When the IoU between the predicted box and the real box is greater than 0.5 and the category is consistent, it is judged as correct detection (TP); Otherwise, it will be judged as FP or FN based on the confidence threshold. Draw a P-R curve, calculate the area under the curve to obtain the AP value, and then take the average of the crack categories to obtain mAP.

(4) Comparison experiment setup: In order to verify the progressiveness of the method in this paper, three representative methods are selected for comparison: the method based on CNN and self-organizing mapping in reference [6], the method based on bridging CNN and Transformer

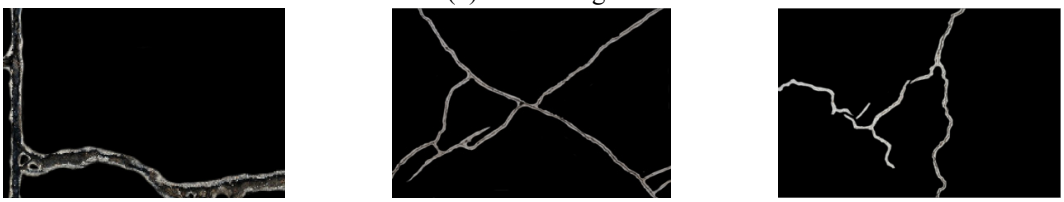
in reference [7], and the method based on semantic segmentation depth convolution neural network in reference [8]. All comparison methods were retrained and tested on the same dataset and evaluation protocol, with hyperparameters set according to the recommended settings in the original reference.

(5) Experimental setup for ablation: (1) Clearly redefine the original model I as the "baseline model"; (2) Independently construct four comparative models, namely "Baseline+Attention", "Baseline+CFM", "Baseline+ECA", and "Baseline+SIOU", so that the independent utility of each module can be quantified separately; (3) Add a complete model (Ours) to verify the synergistic effects between modules.

(a) Original image



(b) Label image



(c) The detection results of this article's method

(d) Results of method detection in reference [6]

(e) Results of method detection in reference [7]

Figure 4. Crack detection results

(f) Results of method detection in reference [8]

Figure 4. Continued

Table 2. Quantitative comparison results

Method	Accuracy/%	Recall rate/%	mAP	F1 value
Method in reference [6]	76.2	80.5	81.3	78.29
Reference [7] method	79.8	83.1	84.6	81.42
Reference [8] method	81.5	86.4	87.2	83.88
Proposed method	88.4	95.2	96.4	91.66

3.4 Experimental results

3.4.1 Analysis of crack detection performance

Three original surface crack images of building concrete were randomly selected from the surface crack test set of building concrete (as shown in Fig. 4(a)). The open-source data annotation tool Labeling was used to create label images for the selected original images (as shown in Fig. 4(b)). The trained model was used to perform crack detection on the selected images. To further validate the detection performance of the proposed method, four advanced methods were selected for comparison: the method based on CNN and self-organizing mapping in reference [6], the method based on bridging CNN and Transformer in reference [7], and the method based on semantic segmentation deep convolutional neural network in reference [8]. The method in this paper and the three comparison methods were respectively used to perform crack detection on the three selected initial images, and the results are shown in Fig. 4.

By analyzing Fig. 4, for the first crack image, the methods in references [6], [7], and [8] misidentified a noise point in the background as a crack feature, while the method proposed in this paper can extract the crack feature information more accurately without being much interfered by noise points. For the second crack image, the methods in references [6], [7], and [8] have the problems of incomplete crack feature recognition and discontinuous recognition, while the method in this paper can recognize the cracks more completely and continuously. For the third crack image, the methods in references [6], [7], and [8] cannot recognize the fine cracks, while the method in this paper can recognize finer cracks. This indicates that the performance of the method in this paper has been significantly improved in the task of identifying cracks on the surface of building concrete.

3.4.2 Quantitative comparison results

In order to further verify the progressiveness of the method in this paper, quantitative comparison is made with the above four methods on the complete test set (1012 images), and the results are shown in Table 2.

Analysis of Table 2 shows that the method proposed in this paper significantly outperforms all

Table 3. Calculation Results of Crack Length and Width

Crack image number	Fracture orientation	Actual length/ pixel points	Based on the method proposed in this article, the length/pixel point detection results are obtained	Absolute error	Relative error (%)
1	Longitudinal	327.829	326.035	1.794	0.55
	Horizontal	290.725	289.954	0.771	0.27
2	Longitudinal	305.600	304.997	0.603	0.20
	Horizontal	268.401	269.003	0.602	0.22
3	Longitudinal	69.844	70.264	0.420	0.60
	Horizontal	13.686	14.691	1.005	7.34
4	Longitudinal	194.268	193.558	0.710	0.37
	Horizontal	113.581	114.690	1.109	0.98

comparison methods in terms of accuracy, recall, mAP, and F1 score. Specifically, compared to the best performing method in reference [8], the accuracy, recall, mAP, and F1 values of our method have been improved by 6.9, 8.8, 9.2, and 7.78 percentage points, respectively. Mainly due to the enhanced detection of subtle cracks by the CCFM cross scale fusion module; The improvement in accuracy is attributed to the synergistic suppression of background false positives by DAttention and ECA; The mAP improvement reflects the optimization of the SIoU loss function on the quality of bounding box regression. The above results validate the effectiveness and synergy of the four improvement strategies proposed in this paper.

3.4.3 Calculation of crack length and width

The detected crack dimensions were compared with actual measurements, with results presented in Table 3.

Analysis of Table 3 shows that the detection results of crack length by the method proposed in this article are basically consistent with the actual length. Except for the transverse length of the fine crack (No. 3), which has a relative error of 7.34% due to the small number of pixels, the relative error of the other cracks is controlled within 1%. The method used in this article for crack detection has a positive impact on the calculation of crack length and width, which can help engineers more accurately evaluate the integrity and safety of building structures.

3.4.4 Comparison of loss functions

The improved model's loss convergence was verified, with training curves for both loss functions shown in Fig. 5.

Analysis of Fig. 5 shows that the SIoU loss function exhibits better convergence characteristics during the training process. Specifically, after 300 training cycles, the SIoU loss value remained stable at 0.115, while the CIoU loss value remained at 0.130; SIoU reaches the convergence platform within 150 cycles, which is about 37.5% faster than the 240 cycles required by CIoU. These data demonstrate that the SIoU loss function effectively reduces the bounding box regression loss, significantly accelerates the model convergence process, and improves generalization ability.

3.4.5 Ablation experiment

To verify the independent effectiveness of each module in the proposed improved model,

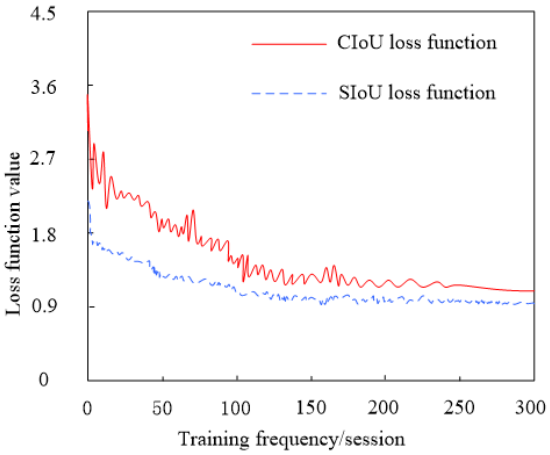


Figure 5. Comparison of loss functions

Table 4. Results of Ablation Experiment

Model	Accuracy/%	Recall rate/%	mAP/%	F1 value/%
Baseline	83.8	91.2	92.8	87.36
Baseline+DAttention	85.9	92.8	94.7	89.20
Baseline+CCFM	86.1	93.1	94.9	89.44
Baseline+ECA	87.8	94.8	95.8	91.18
Baseline+SIoU	84.2	91.5	93.1	87.67
Ours	88.4	95.2	96.4	91.66

ablation experiments were conducted on the YOLOv8 baseline model. The standard YOLOv8 network is referred to as the baseline model (Baseline). The deformable attention mechanism (DAttention), cross-scale feature fusion module (CCFM), efficient channel attention mechanism (ECA), and SIoU loss function were independently introduced into the baseline model, resulting in four comparative models, sequentially denoted as Baseline+DAttention, Baseline+CCFM, Baseline+ECA, and Baseline+SIoU. Additionally, all modules were jointly introduced into the baseline model, referred to as the complete model proposed in this paper (Ours). All models were trained and tested under identical dataset splits, training parameters, and evaluation protocols, with experimental results shown in Table 4.

Analysis of Table 4 shows that each independent module can bring performance improvements compared to the baseline model. Among them, after independently introducing the ECA module, the accuracy increased from 83.8% to 87.8% (an increase of 4.0 percentage points), and the F1 value increased from 87.36% to 91.18%, which is the most significant improvement effect in a single module, indicating that the efficient channel attention mechanism can effectively recalibrate channel feature weights; After independently introducing the CCFM module, the recall rate increased from 91.2% to 93.1%, reflecting the enhanced effect of cross scale feature fusion on detecting subtle cracks; After independently introducing the DAttention module, mAP increased from 92.8% to 94.7%, demonstrating that the deformable attention mechanism helps reduce background interference; After independently introducing the SIoU loss function, the improvement in various indicators was relatively small (accuracy increased by 0.4 percentage points, mAP increased by 0.3

percentage points), but the contribution was significant when combined with other modules in the complete model. After integrating the four modules in a coordinated manner, the complete model achieved optimal performance in all indicators, verifying the good complementarity and collaborative enhancement effect between the modules.

4. Discussion

4.1 Advantages of the proposed method

The improved YOLOv8 model proposed in this paper achieves significant performance enhancements in the task of building concrete surface crack detection, with its advantages primarily reflected in the following four aspects:

First, high robustness under complex backgrounds. As shown in Fig. 4, for the first crack image containing background noise, the methods in references [6-8] all exhibit false detection of noise as cracks, whereas the proposed method accurately extracts crack features with virtually no interference. This advantage is mainly attributed to the DAttention deformable attention mechanism, which learns offsets for the sampling points of convolution kernels, enabling the model to dynamically adapt to the irregular morphology of cracks and effectively avoid interference from background textures such as aggregate and air bubbles on the concrete surface.

Second, high crack detection capability. The comparison methods suffer from incomplete or discontinuous crack identification, while the proposed method achieves complete and continuous detection. This advantage stems from the CCFM cross-scale feature fusion module, which deeply integrates deep semantic information with shallow high-resolution detail features, mitigating the dilution of fine cracks in traditional feature pyramids. As shown in Table 3, the recall of the proposed method reaches 95.2%, significantly outperforming that of reference [8] at 86.4%, further validating the enhanced detection capability of CCFM.

Thirdly, the efficient enhancement of feature discrimination ability. The ablation experiment (Table 4) showed that after independently introducing the ECA module, the model accuracy increased from 83.8% to 87.8% (an increase of 4.0 percentage points), and the F1 value increased from 87.36% to 91.18%, which is the most significant improvement effect among individual modules. This indicates that the ECA efficient channel attention mechanism achieves adaptive recalibration of channel features with low computational complexity, effectively enhancing crack related features and suppressing redundant channel information.

Fourth, precise and stable bounding box regression. Loss function comparisons demonstrate that the SIoU loss function reaches a convergence plateau within 150 epochs, whereas CIoU requires 240 epochs, and the final loss value of SIoU (0.115) is lower than that of CIoU (0.130). Additionally, crack length and width calculations (Table 3) show that, except for extremely small cracks with very few pixels, the relative error between the detected length and the actual length is controlled within 1%. This indicates that the SIoU loss function effectively optimizes the convergence path and localization accuracy of prediction boxes by incorporating angle cost and redefined distance cost.

4.2 Limitations

Although the proposed method achieves good detection performance, the following limitations remain:

(1) Insufficient detection accuracy for extremely fine cracks. In Table 3, the actual transverse length of Crack No. 3 is only 13.686 pixels, with a detection relative error of 7.34%, significantly higher than that of other cracks. This indicates that the model's discriminative capability is still limited when crack dimensions are extremely small, likely due to irreversible loss of fine features caused by multiple downsampling operations.
(2) Strong dependence on data diversity. The training data of this model mainly come from public datasets such as SDNET2018 and some self-collected images. In practical engineering, concrete surfaces may exhibit more complex appearance characteristics due to factors such as formwork material, curing conditions, carbonation degree, and lighting variations, leading to potential performance degradation of the model in such insufficiently covered scenarios.
(3) The positive-negative sample imbalance remains to be optimized. Crack pixels typically account for a very low proportion in concrete surface images, causing negative samples (background) to dominate during model training. Although the introduction of attention mechanisms has alleviated this issue, the fact that recall (95.2%) is slightly higher than precision (88.4%) in Table 4 indicates that some false detections still exist, suggesting room for further improvement in the model's ability to distinguish hard examples.

4.3 Future work

Based on the above limitation analysis, future research will focus on the following directions:
(1) Detection of extremely fine cracks: Explore deep integration of super-resolution reconstruction techniques with feature pyramids, or introduce auxiliary branches with pixel-level supervision into the network to recover detailed information potentially lost during downsampling.
(2) Data diversity: Study domain adaptation techniques and combine generative adversarial networks to synthesize diverse crack samples, improving model robustness in unseen scenarios.
(3) Model lightweighting: Employ knowledge distillation or neural architecture search to compress model volume while maintaining accuracy, meeting the deployment requirements of edge computing devices.
(4) Positive-negative sample imbalance: Introduce strategies such as Focal Loss or online hard example mining to make the model focus more on hard-to-classify crack pixels during training, thereby improving training efficiency and detection accuracy.

5. Conclusions

This article proposes a detection method based on improved YOLOv8 to address the problems of difficult identification under complex backgrounds and easy omission of subtle cracks in the detection of surface cracks in building concrete. DAttention, CCFM, ECA, and SIoU are introduced to construct a progressive optimization chain. The results show that our method achieved an accuracy of 88.4%, a recall of 95.2%, and a mAP of 96.4% on a self built dataset; In the calculation of crack length and width, except for extremely fine cracks, the relative error is controlled within 1%. In terms of potential applications, the method proposed in this article can be deployed in key parts of concrete structures such as bridges and dams to achieve automated crack inspection. It can also be combined with drones for remote detection of high-risk areas and provide data support for structural health assessment and maintenance decision-making. Future work will focus on improving the accuracy of fine crack detection and deploying lightweight models.

Acknowledgements

No funding. The author declared no conflict of interest.

References

1. Parizat, S., Kamai, R., Shaked, Y., Shmerling, A. (2024). Seismic vulnerability of a pre-code, reinforced concrete, apartment-block building. Bulletin of Earthquake Engineering, 22(15), 7547-7587. https://doi.org/10.1007/s10518-024-02054-0.
2. Mostafa, M., Hogan, L., Stephens, M.T., Olsen, M.J., Elwood, K.J. (2024). A detailed damage investigation of an instrumented ductile reinforced concrete building following the m-7.8 kaikoura earthquake. Earthquake Spectra, 40(3), 2179-2209. https://doi.org/10.1177/87552930231218750.
3. Kumar, C., Sinha, A.K. (2023). Automated crack detection and a web tool using image processing techniques in concrete structures. Russian Journal of Nondestructive Testing, 59(11), 1119-1135. https://doi.org/10.1134/S1061830923600569.
4. Inoue, Y., Nagayoshi, H. (2023). Weakly-supervised crack detection. IEEE Transactions on Intelligent Transportation Systems, 24(11), 12050-12061.
5. Whiteman, M.L., Marin-Artieda, C.C., Tezcan, J. (2024). Convolutional neural network approach for vibration-based damage state prediction in a reinforced concrete building. Journal of Computing in Civil Engineering, 38(2), 0402003. https://doi.org/10.1061/JCCEE5.CPENG-5511.
6. Yamaguchi, T., Mizutani, T. (2024). Road crack detection interpreting background images by convolutional neural networks and a self-organizing map. Computer-aided Civil and Infrastructure Engineering, 39(11), 1616-1640. https://doi.org/10.1111/mice.13132.
7. Yadav, D.P., Sharma, B., Chauhan, S., Dhaou, I.B. (2024). Bridging convolutional neural networks and transformers for efficient crack detection in concrete building structures. Sensors, 24(13), 4257. https://doi.org/10.3390/s24134257.
8. Manjunatha, P., Masri, S.F., Nakano, A., Wellford, L.C. (2024). Crackdenselinknet: a deep convolutional neural network for semantic segmentation of cracks on concrete surface images. Structural Health Monitoring, 23(2), 796-817. https://doi.org/10.1177/14759217231173305.
9. Bhardwaj, M., Khan, N.U., Baghel, V. (2024). Fuzzy c-means clustering based selective edge enhancement scheme for improved road crack detection. Engineering Applications of Artificial Intelligence: 136(A), 108955. https://doi.org/10.1016/j.engappai.2024.108955.
10. Niu, H.Y., Bao, T.F., Li, Y.T., Huang, S.W. (2023). Pixel-level crack detection method of concrete dam based on improved Mask R-CNN. Advances in Science and Technology of Water Resources, 43(01), 87-92.
11. Su, W.G., Wang, J.X. (2023). Research on road crack recognition model based on YOLO v3 deep learning algorithm. Journal of China, Foreign Highway, 43(02), 58-63. https://doi.org/10.14048/j.issn.1671-2579.2023.02.010.
12. Elhanashi, A., Dini, P., Saponara, S., Zheng, Q. (2024). Telestroke: real-time stroke detection with federated learning and yolov8 on edge devices. Journal of Real-Time Image Processing, 21(4), 1-16. https://doi.org/10.1007/s11554-024-01500-1.
13. Dagar, D., Vishwakarma, D.K. (2024). Tex-net: texture-based parallel branch cross-attention generalized robust deepfake detector. Multimedia Systems, 30(4), 233. https://doi.org/10.1007/s00530-024-01424-7.

14. Sarma, D., Dutta, H.P.J., Yadav, K.S., Bhuyan, M.K., Laskar, R.H. (2024). Attention-based hand semantic segmentation and gesture recognition using deep networks. Evolving Systems, 15(1), 185-201. https://doi.org/10.1007/s12530-023-09512-1.
15. Vorugunti, C.S., Pulabaigari, V., Mukherjee, P., Sharma, A. (2020). Deepfuseosv: online signature verification using hybrid feature fusion and depthwise separable convolution neural network architecture. IET Biometrics, 9(6), 259-268. https://doi.org/10.1049/iet-bmt.2020.0032.
16. Vo, T., My, L.N.T. (2025). A novel approach of multi-channel attention mechanism for long-sequential multivariate time-series prediction problem. Soft Computing, 29(2), 629-644. https://doi.org/10.1007/s00500-025-10501-6.
17. Setiawan, F., Yahya, B.N., Lee, S.L. (2023). Normalized attention inter-channel pooling (naip) for deep convolutional neural network regularization. Neural Processing Letters, 55(7), 9315-9333. https://doi.org/10.1007/s11063-023-11203-6.
18. Yan, H., Shen, S.L., Liu, L.K. (2024). Saliency Object Detection Combining Multi-Level Supervision and Boundary Loss. Computer Simulation, 41(06), 293-298.