RESEARCH ARTICLE

Road Crack Segmentation Algorithm Based on Multiple Attention Fusion and Defect Correction

Rendong Ji[1], **Xiaojun Zhang**[1], **Xiu Tang**[1], **Xiaoyan Wang**[1,*], **Yunlong Xu**[1] **and Jiaxin Shi**[2]

[1] Department of Electronic Information Engineering, Huaiyin Institute of Technology, Huai'an 223001, China
[2] Department of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an 223001, China

Abstract

This paper proposes an enhanced U-Net-based segmentation framework for road crack detection that effectively addresses issues such as incomplete segmentation, detail loss, environmental complexity, and crack-pixel imbalance. The model integrates multiple functional modules to improve segmentation performance across varying crack types and scales. Specifically, an atrous residual convolution (ARC) module is embedded in the encoder to expand the receptive field and capture large-scale features. A multiple attention fusion module (MAFM), combined with an efficient channel attention mechanism, is introduced at the bridge stage to emphasize crack-relevant features. In the decoder, a defect correction module (DCM) with deep supervision and adaptive refinement is designed to restore fine-grained crack boundaries, especially for small or subtle defects. The proposed model achieves F1-scores of 78.11%, 90.02%, and 81.60% on CFD dataset, CRACK500 dataset, and HYCrack dataset, respectively. Compared to existing state-of-the-art segmentation models, the proposed approach achieves superior accuracy and better preservation of crack detail. These results demonstrate its practical value and strong potential for widespread application in road crack detection and infrastructure maintenance.

Keywords: atrous residual convolution, defect correction, multiple attention fusion, road crack, semantic segmentation.

1 Introduction

As one of the most prevalent types of road defects, cracks in pavement exhibit progressive increases in quantity, length, and width over time. The resulting damage causes irreversible impacts on the structural strength, stability, and safety of roads. Multifaceted deterioration issues reduce roadway service life and pose potential risks to pedestrians and vehicles [1]. Consequently, the timely detection and treatment of pavement cracks during routine highway maintenance holds significant practical importance for road rehabilitation, emergency response, and traffic accident prevention.

Traditional crack detection methods primarily rely on manual operation of camera equipment, employing manual photography approaches to record pavement cracks. However, manual detection methods present

Academic Editor:
Jianlei Kong

Submitted: 03 November 2025
Accepted: 22 January 2026
Published: 29 June 2026

Vol. 3, **No.** 2, 2026.
10.62762/TIS.2025.325163

***Corresponding author:**
✉ Xiaoyan Wang
wxygxy@163.com

Citation
Ji, R., Zhang, X., Tang, X., Wang, X., Xu, Y., & Shi, J. (2026). Road Crack Segmentation Algorithm Based on Multiple Attention Fusion and Defect Correction. *ICCK Transactions on Intelligent Systematics*, 3(2), 126–144.

several inherent problems. For instance, inspectors' crack assessment results are often influenced by personal subjective factors, creating difficulties for subsequent crack treatment operations. Consequently, the development of more efficient, accurate, and high-quality detection methods has become an urgent requirement for road crack detection. Currently, image recognition and segmentation algorithms can be broadly classified into traditional image segmentation methods and deep learning-based image segmentation methods. As far as traditional image segmentation methods are concerned, thanks to the development of digital image processing technology, they can be roughly divided into the following three categories: threshold segmentation method, edge detection method and mathematical morphology method. Although these methods can realize different degrees of automatic extraction of road cracks, they still have obvious limitations in dealing with complex crack shapes or significant pavement interference.

In contrast, deep learning methods achieve automated detection without manual intervention through established neural network architectures. The core advantage lies in their ability to automatically learn hierarchical feature representations of cracks through large-scale data training, ranging from low-level edge and texture features to high-level semantic information. Models such as FCN [2], U-Net [3], and DeeplabV3+ [4] have demonstrated excellent performance. Huang et al. [5] systematically analysed the latest advances in machine vision-based crack detection for asphalt pavements, comparing traditional digital image processing methods with deep learning approaches. Their review emphasizes that while traditional methods can achieve varying degrees of automatic crack extraction, they remain limited when handling complex crack shapes and significant pavement interference, whereas deep learning methods demonstrate superior adaptability through automated hierarchical feature learning. More importantly, deep learning methods achieve pixel-level precise crack segmentation without necessitating adjustments for specific features or background interferences, yet accurately extract fine-grained crack characteristics. This automated feature learning capability and adaptability to complex environments enable deep learning methods to significantly outperform traditional approaches in terms of detection accuracy, efficiency, and generalization ability, thereby providing more reliable and efficient technical solutions for pavement crack detection.

Zhang et al. [6] proposed an improved U-Net architecture to solve the problem of insufficient accuracy of crack contour segmentation by U-Net network. Yang et al. [7] proposed MST-Net, a multiscale triple-attention network for end-to-end pixelwise crack segmentation. By integrating triple attention mechanisms with multiscale feature aggregation and deep supervision, this approach achieves accurate detection of microcracks under complex backgrounds, providing a strong foundation for attention-based crack segmentation research. Aiming at road crack detection in complex background, Pan et al. [8] proposed a spatial-channel hierarchical deep learning network for pixel-level automated crack detection, which leverages hierarchical spatial and channel feature extraction to address crack detection challenges in complex background environments. Al-Huda et al. [9] developed a hybrid deep learning method for pavement crack semantic segmentation, leveraging knowledge transfer between class activation maps (KTCAM) and an encoder-decoder segmentation network. In order to improve the segmentation effect of asphalt pavement cracks, Ali et al. [10] contributed to crack detection research by developing the Crack45k dataset and proposing a crack segmentation model integrating vision transformers with tubularity flow field guidance and a sliding-window strategy to handle complex pavement crack structures. Hamishebahar et al. [11] presented a comprehensive review of deep learning-based crack detection methodologies, analysing diverse approaches and their structural health monitoring applications. Xiang et al. [12] proposed a crack-segmentation algorithm that fuses Transformer and convolutional neural network architectures to handle complex detection scenarios, synergistically leveraging global context modeling with local spatial feature extraction in an encoder-decoder framework for more robust crack boundary delineation. Most recently, Sahragard et al. [13] proposed an enhanced UNet architecture that integrates deformable convolution with attention mechanisms to improve boundary delineation and multi-scale feature representation in semantic segmentation tasks, demonstrating the general effectiveness of combining deformable operations with attention-driven feature refinement in encoder-decoder frameworks.

Wang and Su [22] developed an automatic crack segmentation model based on Transformer

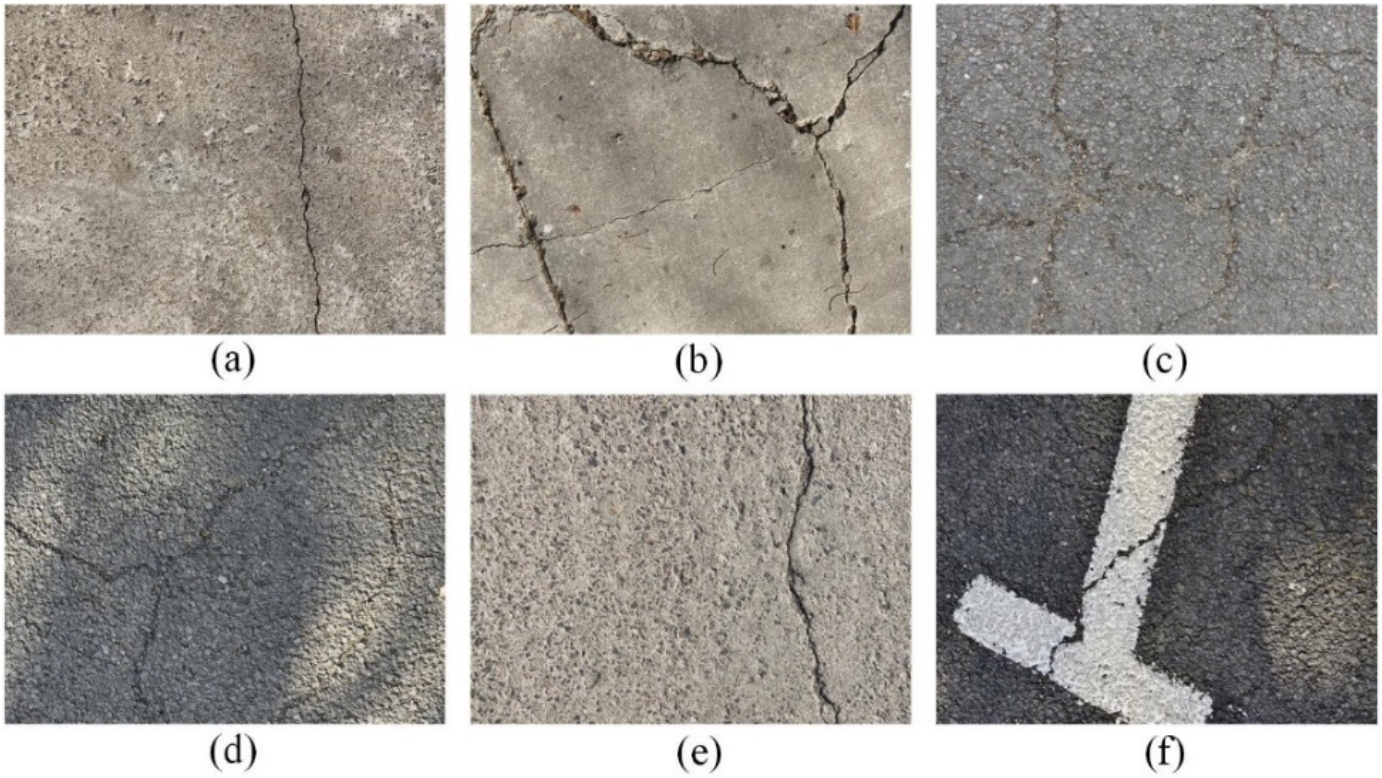

Figure 1. Road crack images under different pavement conditions. (a)Linear. (b)Block. (c)Grid. (d) Shadow occlusion. (e) Adequate lighting. (f) Pavement marking.

architecture, leveraging self-attention mechanisms to capture long-range dependencies in crack images, which demonstrates the effectiveness of Transformer-based encoder-decoder designs for pixel-level crack boundary detection. These research advances have inspired our design of the multiple attention fusion module, which integrates efficient channel attention (ECA) and convolutional block module (CBM) to enhance crack feature extraction. The proven effectiveness of multiscale representations and attention-based mechanisms in complex computer vision tasks motivates our integration of these components for adaptive feature refinement in road crack segmentation. Therefore, this paper proposes a road crack segmentation algorithm based on multi-attention fusion and defect correction. The core contributions comprise the following:

- Improve the backbone network structure and reconstruct the convolution layer of the original U-Net network Encoder by integrating atrous convolution and residual networks. The constructed atrous residual convolution (ARC) module captures richer contextual crack information, alleviates the problem of gradient vanishing, and enhances the extraction capability of multiple scale crack features.
- Design a multiple attention fusion module (MAFM) in the Bridge Stage connecting the encoder and decoder. By reinforcing critical information within high-level semantic features, this module enhances feature reconstruction quality during upsampling, effectively addressing issues of crack edge blurring and inaccurate segmentation of crack details.
- A novel defect correction module (DCM) is innovatively designed in the decoder stage. Concurrently, an efficient channel attention (ECA) module is introduced before the output of each decoder layer. This ECA module generates auxiliary losses at each sampling layer. These auxiliary losses are weighted, accumulated, and fused with the backbone model loss to form the total loss function. This approach significantly improves segmentation accuracy for fine crack branches and mitigates discontinuity in fractured regions.

2 Dataset Construction and Characteristics

Deep learning models require extensive image datasets for model training. This study collects the image dataset necessary for road crack detection using campus roads as samples, and achieves real-time and

efficient detection purposes through segmentation models to extract road crack features.

2.1 Collection Method and Implementation

The road crack data includes various types of roads, such as main roads, secondary roads, and pedestrian walkways. Pavement materials primarily consist of asphalt concrete and cement concrete, with road widths ranging from 3 m to 10 m. These cracks typically measure less than 2 mm in width and relatively short in length, predominantly manifesting as single linear cracks in transverse or longitudinal orientations. Long-term service has resulted in substantial degradation of pavement material performance, with cracks exhibiting network distribution characteristics. Multiple cracks interweave to form complex crack networks, with crack widths exceeding 10 mm and correspondingly increased depths.

Our study employs a handheld road crack acquisition device for image collection. The core imaging unit of the device utilizes a 1/1.7-inch CMOS image sensor with a capture resolution of 12 megapixels and an individual pixel size of 1.85 μm. During the capture process, the device strictly maintains a fixed shooting angle and height, with the lens optical axis consistently perpendicular to the ground surface. The vertical distance between the device and the road surface is fixed within 1.2 m. The device captures and saves only a single crack image per shutter trigger, proceeding to capture the next crack image only after confirming successful storage.

To ensure the diversity of road crack images, image acquisition was performed under sunny weather conditions. The acquired crack images encompass both fully sunlit non-shadow regions and shadow regions created by occlusion from trees or other obstructions. The distribution and statistics of collected crack image categories are shown in Table 1. Based on the different road conditions and crack types mentioned above, we constructed the Huaiyin crack (HYCrack) dataset. Among the 633 images in the dataset, each image has a pixel dimension of 4096×3072. The acquired crack images primarily comprise three categories: Crack images under adequate lighting, crack images under shadow occlusion and crack images containing pavement marking interference. The types of crack images collected are shown in Figure 1.

2.2 Image Characteristics of Road Cracks

Comprehending the image characteristics of road cracks constitutes a prerequisite for selecting segmentation models, precisely identifying cracks, and evaluating pavement conditions. Road cracks are broadly categorized into linear cracks, block cracks, and network cracks. Linear cracks can be further subdivided into transverse, longitudinal, and diagonal types. Crack images collected in campus scenarios typically exhibit the following characteristics:

- Crack regions display darker coloration and lower pixel values compared to the overall pavement background. All crack types are accompanied by varying degrees of branch extension. Due to external factors such as pavement materials, collection locations, and collection times, crack images exhibit brightness variations.
- From the perspective of pavement texture and contamination levels, campus roads demonstrate significant heterogeneity. Certain road segments maintain clean surfaces, with asphalt or cement surfaces preserving their original coloration and texture, resulting in pronounced grayscale contrast between cracks and background. Conversely, other pavement surfaces are distributed with contaminants including oil stains, soil, fallen leaves, and tire marks, whose coloration approximates that of cracks, readily causing misidentification.
- Various traffic markings are distributed across campus roads, presenting high-intensity features in images. When cracks intersect or run parallel to these markings, mutual interference between the two elements complicates accurate crack boundary identification. Additionally, manholes and their peripheral regions constitute key objects for data collection.

Due to the above reasons, crack feature extraction presents considerable challenges, necessitating the design of more precise and efficient segmentation models. Therefore, this study proposes a road crack segmentation algorithm incorporating multi-attention fusion and defect correction to meet the requirements for accurate and efficient crack features extraction.

2.3 Image Annotation Method

Road crack segmentation models require both the original pre-annotation images and their corresponding binary annotated images to complete the training task. Therefore, the study employs

Table 1. Distribution and statistics of collected crack image categories.

Categories	Groups			
	Adequate lighting	Shadow occlusion	Pavement marking	Total number of images
Linear	252	70	29	351
Block	113	51	33	197
Grid	53	21	11	85
Number of images	418	142	73	633

Adobe Photoshop as the annotation tool to label cracks. Addressing the complex characteristics of irregular crack morphology, variable orientations, and pavement marking interference, this study adopts a pixel-level annotation method for road crack annotation tasks, ensuring that annotation results accurately reflect the true boundaries and morphological features of cracks.

Pixel-level annotation is a precise method that operates at the individual pixel level, generally depicting crack contours through the point-by-point annotation approach to ultimately form a complete closed region. Initially, the main trajectory of the crack is identified and annotated, namely the principal crack line. During annotation, tracing is performed along the crack centerline to ensure that the annotated region completely covers the crack width. For cracks exhibiting obvious width variations, the path width must be adjusted in real-time during annotation to maintain consistency between the annotated region and the actual crack width. After completing the main crack annotation, branch cracks are sequentially processed. Branch cracks must be completely annotated even when they are minute branches measuring only a few pixels in length. At the junctions between branches and the main trunk, particular attention must be paid to ensuring the continuity of the annotated region, avoiding any fractures or gaps.

Upon completion of annotation, binary images in PNG format are generated with image dimensions consistent with the original dimensions. The specific annotation process is illustrated in Figure 2. In the generated crack binary images, black pixels with a value of 0 represent road background information, namely non-crack regions, while white pixels with a value of 255 represent road target crack information, namely the pixel positions where cracks are located. To ensure annotation consistency and quality, all crack images were independently reviewed by two trained annotators after initial annotation, with disagreements resolved through discussion and consensus. Adobe Photoshop was selected as the annotation tool due to its precise pen tool supporting sub-pixel boundary tracing, which is particularly suited for the irregular and narrow morphology of pavement cracks.

2.4 Image Enhancement Method

Crack image enhancement aims to reduce noise and surrounding environmental interference, improve image quality, while maximally preserving the authenticity and significance of crack information. Addressing the issues of crack discontinuity and edge blurring during image transformation, this study optimizes the original captured images through enhancement from different perspectives.

On the one hand, targeting the crack discontinuity and edge blurring problems occurring during image transformation, this study optimizes the original captured images. In terms of geometric transformation, considering that rotating transverse cracks, longitudinal cracks, and diagonal cracks by 90° or 270° would alter the classification of road defect types, this study only employs horizontal flip, vertical flip, and small-angle rotation to modify crack morphology and distribution characteristics. The small-angle rotation transformation rotates crack images by 15° to achieve slight angular adjustment, ensuring that small branches in crack regions do not blur due to rotation, thereby effectively maintaining crack detail information. In addition, to adapt to illumination intensity variations across different time periods, this study employs a dynamic brightness enhancement strategy. Specifically, brightness is enhanced by a factor of 1.3 to simulate crack states under strong illumination conditions, while brightness is reduced to 0.7 times to simulate crack states in low-light environments. This bidirectional brightness adjustment approach enables the model to adapt to various illumination conditions ranging from strong to weak light, enhancing the model's robustness to different environments. Gaussian Blur techniques are employed to suppress noise and reduce other interference factors in images, assisting the model

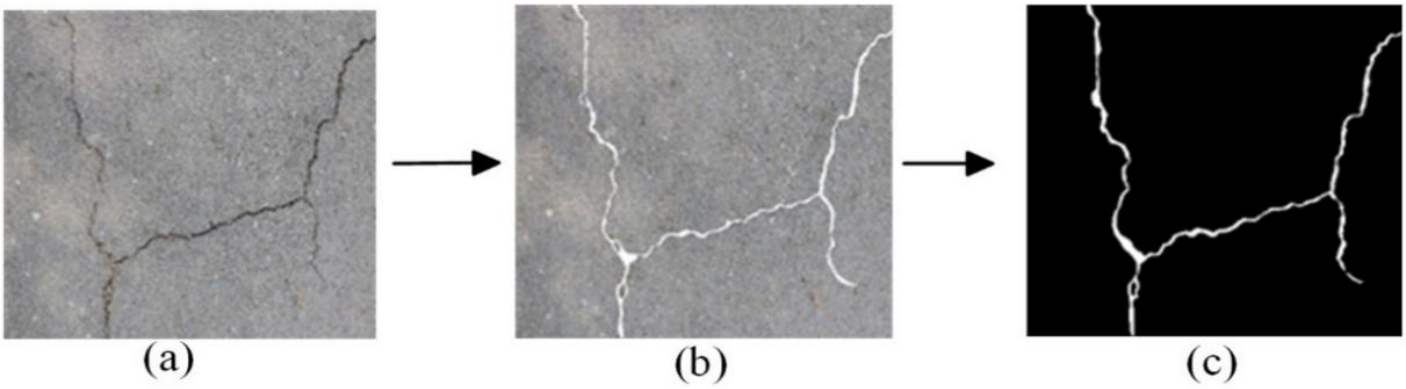

Figure 2. Road crack annotation process. (a)Original image. (b)Pixel-level annotation. (c)Binary image.

in focusing on crack regions, thereby improving detection accuracy. The image enhancement results are shown in Figure 3.

On the other hand, to address the issues of cracks, discontinuities, and edge blurring that occur during image transformation, we employed an adaptive bilinear interpolation method to achieve image downscaling and upscaling. This approach preserves the overall crack structure while maintaining key features, significantly enhancing crack edge definition and textural details. Adaptive bilinear interpolation particularly emphasizes the capability to retain crack details, effectively maintaining crack continuity and reducing the loss of crack detail information. Rotation Transformation operations rotate crack images counter-clockwise by 10° to achieve small-angle image rotation while ensuring that fine branches in crack regions do not blur. Unsharp masking techniques deepen edge contrast along cracks by enhancing texture representation, making crack contours more distinctly discernible. Perspective Transformation methods simulate diverse viewing angles, improving the model's adaptability to viewpoint variations, enabling the model to accurately identify crack features under different shooting angles. The image enhancement results of the aforementioned methods are shown in Figure 4.

2.5 Datasets

The crack dataset in our research comprises three parts: CFD dataset [14], CRACK500 dataset [15], and HYCrack dataset. To enhance network generalizability and mitigate overfitting risks, we employed the aforementioned image augmentation techniques to expand all three datasets, generating augmented images for each respectively. CFD dataset is a publicly available datasets containing 118 annotated crack images with a resolution of approximately 480×320 pixels. After augmentation, it yielded 1602 crack images, with 1122 images allocated to the training data set and 480 images to the test data set. The CRACK500 dataset is sourced from the road pavement crack conditions of the main campus of Temple University. The original dataset consists of a total of 700 images with a resolution of about 2560×1440 pixels. After image enhancement, 3000 images for training and 962 images for testing were finally obtained.

Our custom HYCrack dataset images required manual annotation after collection. The annotation dataset was then expanded using image enhancement methods, resulting in 6878 crack images with a resolution of 448×448 pixels. We obtain 5697 training images and 1181 test images. Detailed statistics of all crack datasets are shown in Table 2.

Table 2. Comparison of crack dataset before and after augmentation.

Dataset	CFD		CRACK500		HYCrack	
Original Numbers	118		700		633	
Original Size	480×320		2560×1440		4096×3072	
Enhanced Numbers	Train	Test	Train	Test	Train	Test
	1122	480	3000	962	5697	1181
Enhanced Size	448×448					

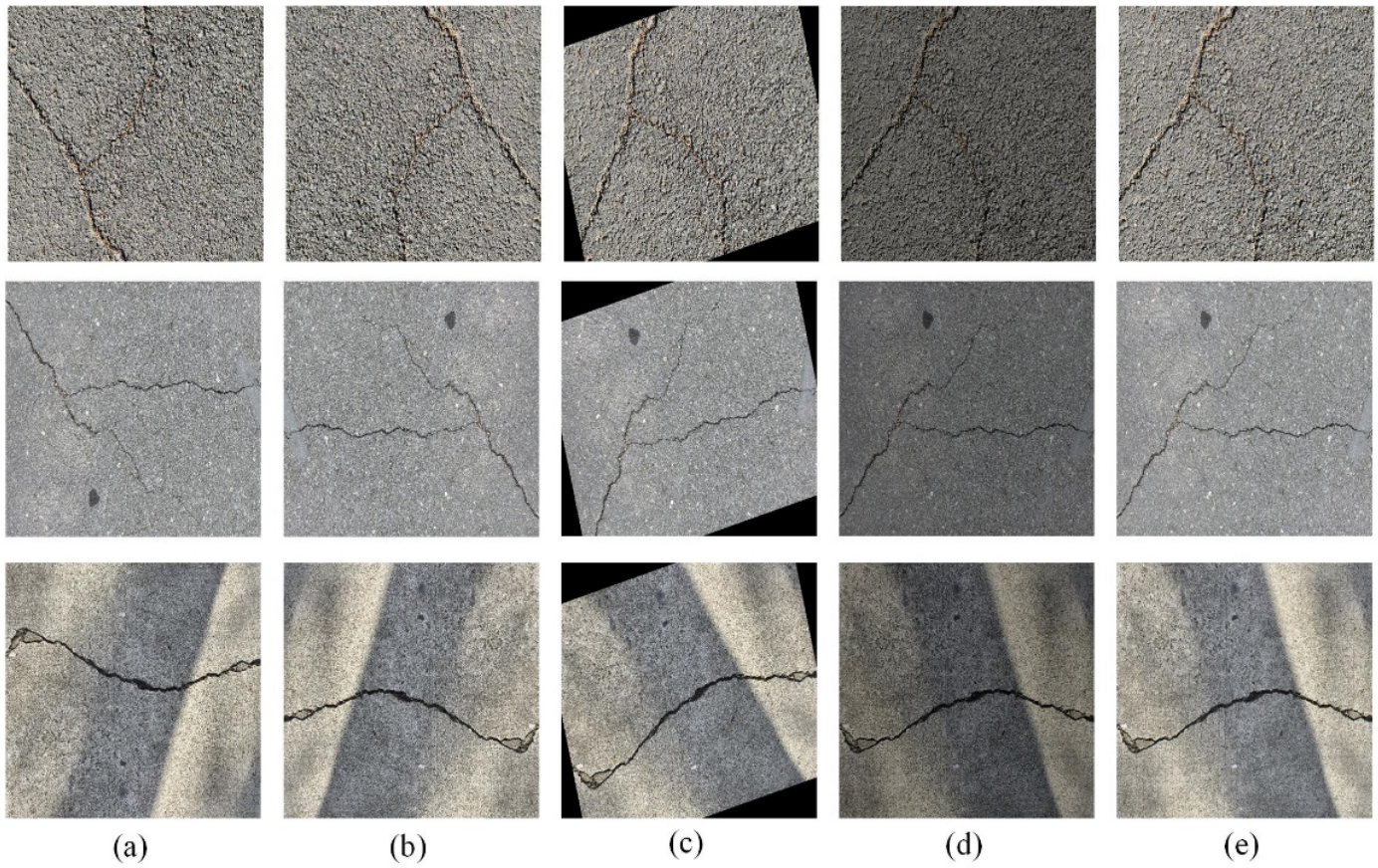

Figure 3. Geometric transformation enhanced image. (a)Vertical. (b)Horizontal. (c)Rotation 15°. (d)Darkness. (e)Blur.

3 Methodology

This section briefly outlines the improved network model and its components. The model structure design is based on U-Net. The proposed improved model consists of three main parts: Encoder Stage, Bridge Stage, and Decoder Stage. The improved modules are atrous residual convolution module (ARC), multiple attention fusion module (MAFM) and defect correction module (DCM) respectively.

3.1 Improved Network Model

Road crack segmentation targets cracks, which typically occupy a small proportion of pixels within an entire image. For such small-target segmentation tasks, the U-Net network serves as a primary baseline. Our approach improves upon this original structure to achieve road crack image segmentation. Compared to other convolutional neural network architectures for semantic segmentation [16], U-Net achieves more precise pixel-level segmentation with lower training data requirements through its symmetric encoder-decoder structure and skip connections, making it particularly well-suited for small-target detection tasks such as road crack segmentation where annotated data is scarce.

This paper presents a model improvement based on the U-Net network, primarily focusing on feature extraction, attention mechanisms, and deep supervision capabilities for road crack image segmentation. We construct a deep learning model integrating multiple attention fusion and defect correction. Figure 5 shows the improved network architecture, which is mainly divided into three parts:

- Encoder Stage: The original convolutional layers are replaced with ARC modules. Each residual module incorporates atrous convolution and skip connections, forming atrous residual convolutional layers that serve as the backbone for feature extraction.

- Bridge Stage: The MAFM is embedded in the transition path connecting the encoder's output to the decoder within the U-shaped network. By combining the advantages of spatial attention and efficient channel attention, this module significantly enhances crack segmentation accuracy.

- Decoder Stage: A four-stage upsampling path restores the spatial resolution of the image. Each

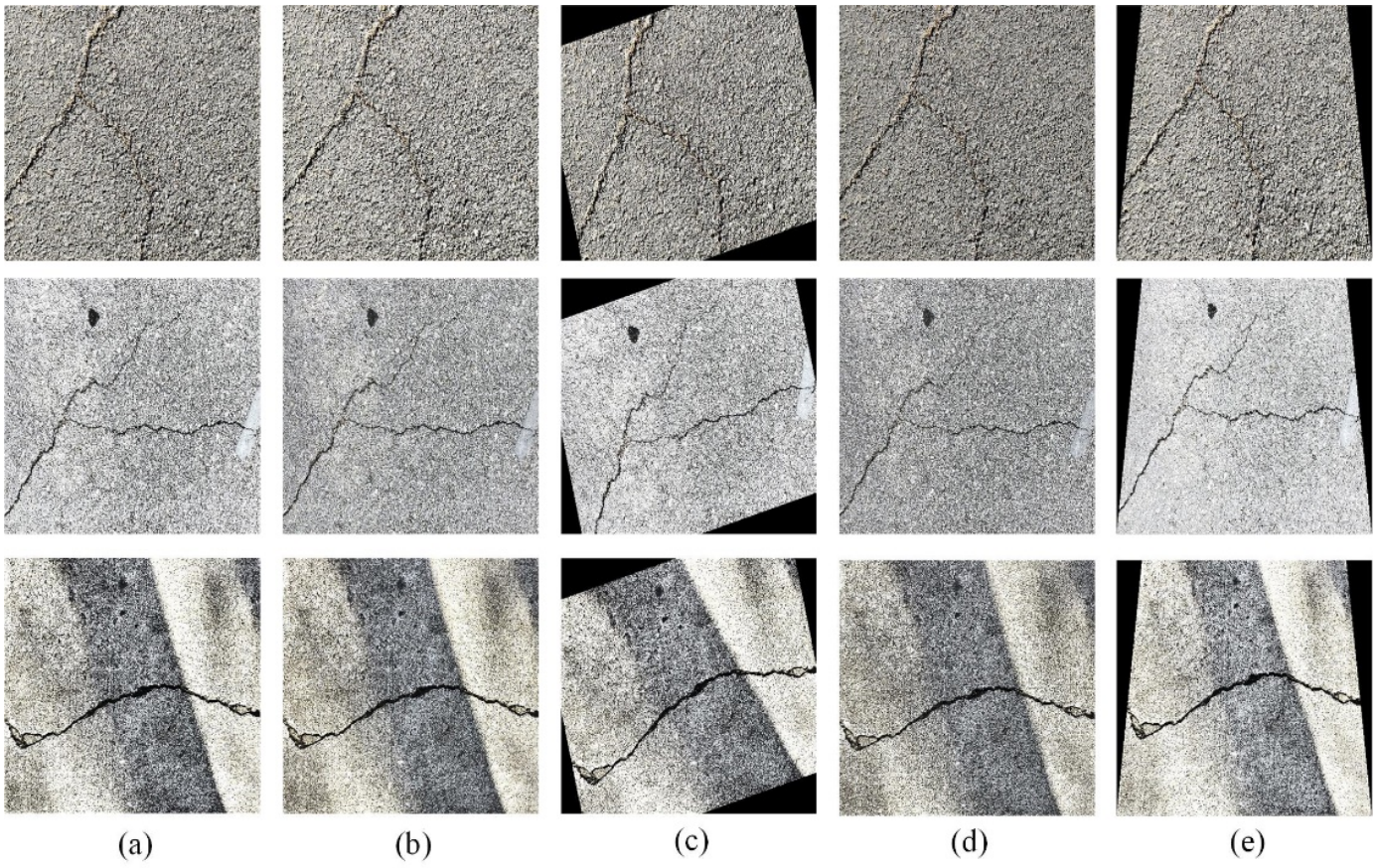

Figure 4. Bilinear interpolation enhanced image. (a)Downscale. (b)Upscale. (c)Rotation10°. (d)Texture. (e)Perspective.

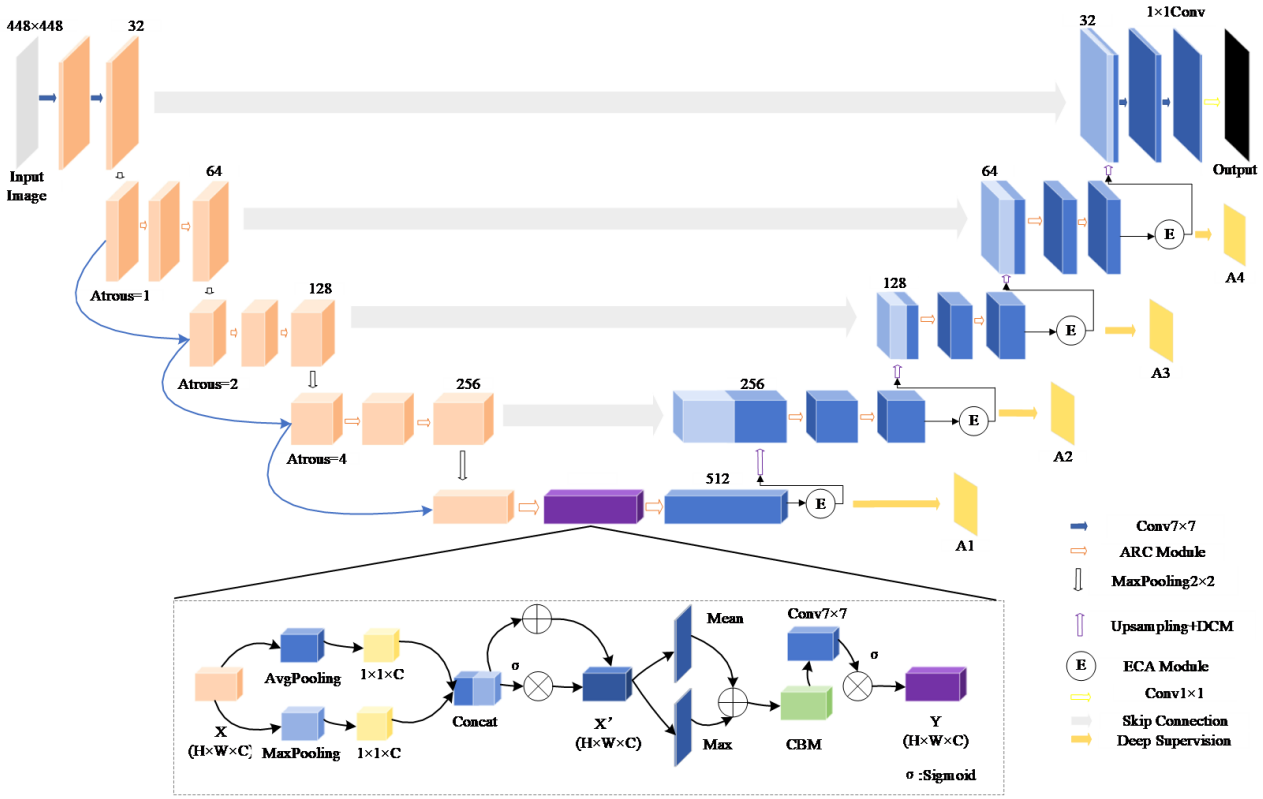

Figure 5. Overall architecture of the enhanced network.

upsampling stage incorporates the innovatively designed DCM, followed by concatenation with an ECA module. Feature maps from the preceding stage are fused with feature maps of the corresponding scale from the subsequent stage via upsampling operations.

When the model is trained, the original road crack image with a size of 448×448 is first input into the improved "U" network, and the low-level feature information of the image is extracted in the double convolution layer, and the number of channels and feature information of the input image are gradually extracted by the atrous residual convolution module with atrous rates of 1, 2 and 4 respectively. The multi-attention fusion module is introduced in the Bridge Stage between the encoder and the decoder, which not only improves the focusing ability of the network model on the crack region, but also reduces the large-scale interference of the crack image background. Then, in the decoder stage, the model uses the skip connection to stitch the up-sampled feature map and the feature map of the encoder stage. This will fuse the shallow feature information and deep feature information of the crack image, and the fused feature map will be output to the next upsampling stage under the further processing of the ECA module and the defect correction module. The improved decoder part improves the recovery ability of the edge details of the output feature map after stitching, accurately detects and segments the small branches of the cracks. Finally, the feature map of the up-sampling output of each layer of the decoder is calculated by the defect correction module to obtain loss values.

3.2 Atrous Residual Convolution Module

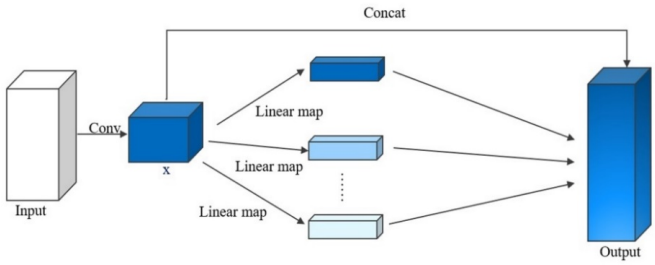

Figure 6. Atrous convolution module structure.

The Atrous Residual Convolution module (ARC) integrates the structure characteristics of atrous convolution and residual network [17]. It mainly enhances feature extraction capabilities while maintaining network efficiency. Figure 6 is a diagram of an atrous convolution module. This module takes feature maps containing multiple scale crack information from the preceding layer as input. It extracts multi-scale features through three parallel atrous convolutional layers with dilation rates of 1, 2, and 4, respectively. Each convolutional layer is immediately followed by Batch Normalization (BN) layer and ReLU to increase the model's nonlinear processing capability. The enhanced feature maps are then fused with the original input feature maps via residual connections. This fusion integrates shallow crack information and reinforces crack features. The final output is an enhanced crack feature map incorporating shallow-layer information. This module not only strengthens the model 's ability to capture different scale features, but also mitigates the vanishing gradient problem in deep network training through its residual connections. Consequently, it improves the robustness and accuracy of road crack detection. The detailed structure is illustrated in Figure 7.

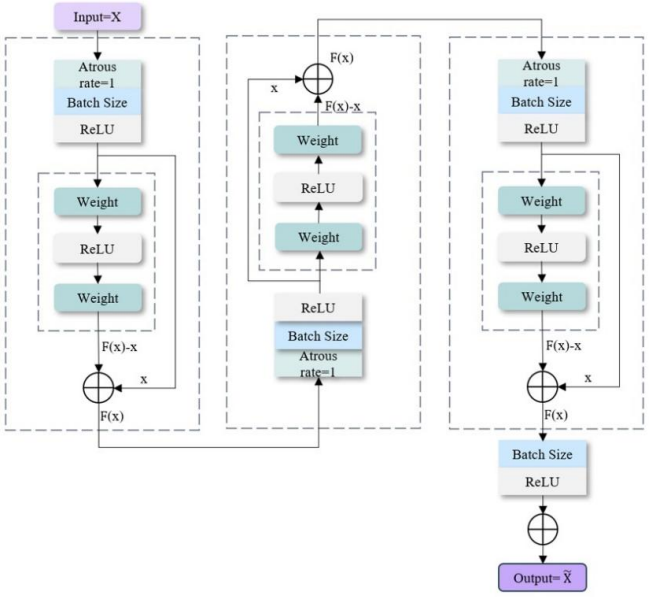

Figure 7. Atrous residual convolution structure.

3.3 Multiple Attention Fusion Module

The Multiple Attention Fusion Module (MAFM) integrates the core advantages of the Convolutional Block Attention Module (CBAM) [18] and the Efficient Channel Attention (ECA) module [25]. It replaces the channel attention mechanism in CBAM with the ECA module while retaining CBAM's spatial attention component. The part of attention retains the Convolutional Block Module (CBM) in CBAM. MAFM not only has the efficient channel feature selection ability of ECA, but includes the spatial information

optimization ability of CBM. When applied in the Bridge Stage of the network, MAFM significantly enhances crack detection performance in complex backgrounds.

To optimize MAFM's performance, a dynamic weighting strategy is incorporated during attention computation. This enables the module to automatically adjust parameters based on specific feature layers. The spatial information of the shallow feature map is important, so it gives the CBM higher weight. However, we found the semantic information of deep feature map is richer, and the relationship between channels is more critical, so the parameters of ECA module are improved. This adaptive strategy dynamically adjusts the distribution of attention weights according to input characteristics, thereby strengthening the model's ability to extract crack features in complex backgrounds.

This paper introduces the MAFM into the Bridge Stage of the U-shaped network. It enhances the model's overall feature expression and crack location capabilities while retaining more spatial information in feature map. This design offers certain advantages in extracting the context information of cracks, and focuses on improving the perception of crack branches and details. The detailed structure is illustrated in Figure 8.

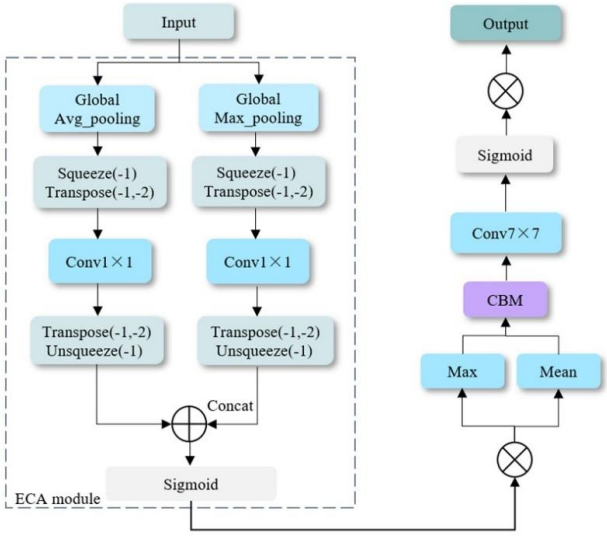

Figure 8. Multiple attention fusion module structure.

3.4 Defect Correction Module

For the decoder part of the network, in order to enable the model to learn multiple level details of crack area across feature maps at different scales, thereby improving and correcting potential missed detections or false positives in the decoder output, this paper innovatively designs a Defect Correction Module. Our module primarily consists of a deep supervision mechanism and an adaptive refinement mechanism, with its structure illustrated in Figure 9.

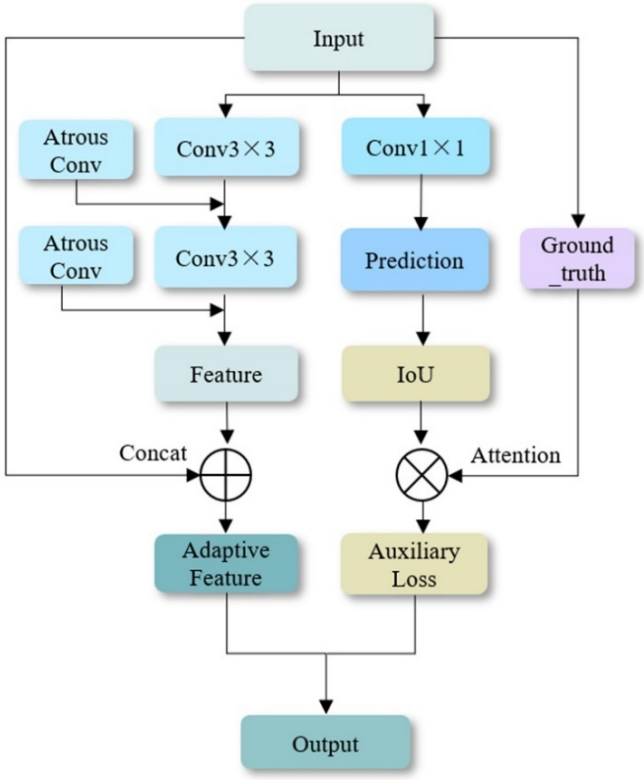

Figure 9. Defect correction module structure.

The improved defect correction module presented in this paper, compared to traditional defect correction methods, which typically employ single-layer supervision at the final decoder output and rely on post-processing operations after segmentation, cannot adaptively learn from the data. This not only leads to gradient vanishing in deep networks but also fails to guide the learning of intermediate features. In contrast, DCM implements multi-layer deep supervision at each decoder stage and uses a weighted auxiliary loss to ensure strong gradient signals throughout the entire network depth. The adaptive refinement mechanism in DCM is integrated into the decoder, learning optimal edge enhancement features during training using parallel dilated convolutions with dilation rates of 2 and 4.

The synergistic combination of deep supervision and adaptive refinement enables progressive defect correction across multiple scales. Each decoder layer benefits from direct supervision, preventing

the accumulation of segmentation errors during upsampling, while adaptive refinement continuously improves edge accuracy. This dual-mechanism design addresses both false negative and false positive issues, resulting in more accurate and complete crack segmentation results than traditional single-mechanism methods.

Through the deep supervision mechanism, DCM generates output feature maps F_i after each of the four decoder upsampling layers. These feature maps incorporate both semantic and detailed information, enabling multiple level learning of crack regions and reducing missed detections. The enhanced feature maps from each upsampling layer are processed through a 1×1 convolutional layer D_i to produce intermediate feature prediction results P_i. Thus, the prediction results function is defined as:

$$P_i = D_i(F_i), \tag{1}$$

where F_i is the output feature map. These predictions are subsequently compared against ground truth maps at the corresponding pixel locations to obtain auxiliary losses A_i. Equation (2) represents the auxiliary loss function.

$$\begin{aligned} A_i &= \mathcal{L}_{\text{BCE-Dice}}(P_i, Y) \\ &= -\frac{1}{N}\sum_{j=1}^{N}[Y_j \log P_{ij} + (1 - Y_j)\log(1 - P_{ij})] \\ &\quad - \frac{2\sum_j P_{ij} Y_j}{\sum_j P_{ij} + \sum_j Y_j} \end{aligned} \tag{2}$$

where Y denotes the binary ground truth mask; i indexes the decoder layer (1, 2, 3, 4). The BCE-Dice loss combines binary cross-entropy with the Dice coefficient to jointly optimize pixel-wise classification accuracy and region overlap, which is particularly effective for the class-imbalanced crack segmentation task. The model performs weighted summation on the auxiliary losses of all upsampling layers to obtain the total loss L for the deep supervision mechanism, which can be defined as:

$$L = \sum_{i=1}^{4} \gamma_i A_i, \tag{3}$$

where γ_i are the weight coefficients, which are 0.1, 0.2, 0.3, 0.4, respectively.

The DCM module uses an adaptive refinement mechanism to optimize the upsampled output feature maps, thereby enhancing the features in crack edges and complex regions. The refinement convolution layer utilizes two parallel dilated convolutions to extract crack edge details. Each upsampled output feature map F_i undergoes secondary processing through this refinement layer. The first and second layers use 3×3 dilated convolution with dilated rates of 2 and 4, respectively. Followed by batch normalization layers and ReLU activation functions to obtain refined feature maps.

This refined feature map F_{aci} is fused with the original output feature map to generate an adaptively refined feature map F_{rei}. We use the adaptive refinement feature map into the next upsampling layer and repeat the above operation. This mechanism of progressive refinement from shallow to deep layers enhances the precision of extracted crack edge features and complex details for the segmentation operation. The calculation process is shown in Equations (4) and (5).

$$F_{aci} = C_{aci}(F_i), \tag{4}$$

$$F_{rei} = F_i + C_{aci}(F_{aci}), \tag{5}$$

where F_{rei} is the output feature map after adaptive refinement; C_{aci} denotes the adaptive refinement convolution layer.

4 Experiments

This section describes the experimental parameters and evaluation metrics. It further analyses the results through ablation studies and comparative experiments, providing comprehensive details regarding implementation specifics and relevant outcome metrics.

4.1 Experimental Parameters

All experiments were implemented using PyTorch 1.13.1 and Python-3.10. The operating system was Windows 11, the running memory was 16GB, the graphics card was NVIDIA GeForce RTX 4060, the processor was AMD Ryzen 9 7945HX, the main frequency was 2.5 GHz, and the specific parameters were 16 cores and 32 threads. In order to ensure the fairness of the experimental process, all models are trained using the Adam optimizer. The size of the training batch was set to 4, and the learning rate parameter was $1{\times}10^{-4}$. To ensure statistical reliability of the reported results, all experiments were conducted with three independent runs using different random

Table 3. Effect of different auxiliary loss functions on the DCM module, evaluated on the CRACK500 dataset during the loss function selection stage (before full model optimization).

Loss functions	Precision/%	Recall/%	Dice/%	IoU/%	F1/%
BCE	92.16	80.67	85.66	75.55	86.24
BCE_DICE	90.16	86.30	87.52	78.11	89.07
Weighted BCE_DICE	81.14	90.34	85.13	74.75	85.24
Tversky	89.30	80.35	84.59	73.96	85.69

seeds (42, 123, 456). The mean and standard deviation of evaluation metrics are reported in Tables 5, 6, and 7.

4.2 Model Evaluation Indicators

In road crack image segmentation tasks, accurately identifying fundamental characteristics such as crack location, width, and length is essential. This study uses the evaluation indicators used in segmentation tasks, including Precision, Recall, Dice Similarity Coefficient (DSC), Intersection over Union (IoU) and F1-score.

Precision represents the proportion of correctly detected crack pixels to the original correct pixels; Recall reflects the percentage of actual crack pixels successfully detected against all ground truth crack pixels; The Dice Similarity Coefficient calculates similarity between two samples, reaching its maximum value of 1 under perfect segmentation. As the standard metric for semantic segmentation, IoU represents the ratio of overlap area between ground truth and prediction to total combined area. The F1-score harmonizes both Precision and Recall, yielding higher values when fewer crack pixels are missed while maintaining minimal false positives. Model evaluation indicators parameter formulas are shown in Equations (6) to Equations (10).

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}}, \tag{6}$$

$$R = \frac{T_{TP}}{T_{TP} + F_{FN}}, \tag{7}$$

$$\text{Dice} = \frac{2 \times T_{TP}}{2 \times T_{TP} + F_{FN} + F_{FP}}, \tag{8}$$

$$\text{IoU} = \frac{T_{TP}}{T_{TP} + F_{FP} + F_{FN}}, \tag{9}$$

$$F1 = \frac{2 \times P \times R}{P + R}, \tag{10}$$

where P denotes Precision; R denotes Recall; T_{TP} represents the number of pixels correctly classified as crack pixels; F_{FP} represents the number of pixels incorrectly classified as crack pixels; F_{FN} represents the number of crack pixels incorrectly classified as background pixels.

4.3 Comparative Analysis of Loss Function

To objectively evaluate the performance of the novel DCM module proposed in this paper, comparative experiments were conducted using four different auxiliary loss functions: Binary Cross-Entropy Loss, Tversky Loss, BCE_DICE Loss, and Weighted BCE_DICE Loss. All experiments were performed on the CRACK500 dataset using the same hyperparameters, including model architecture, optimizer, and batch size. The training period was fixed at 60 epochs. Furthermore, these auxiliary loss functions only participated in gradient calculation and did not modify the structure of the backbone network or the DCM module. Their computational complexity is the same as the main loss function. Therefore, introducing these auxiliary loss functions does not significantly increase the total training time. Table 3 shows the results of the impact of different loss functions on model.

Experimental results show that different loss functions exhibit different performance characteristics. Using BCE Loss as the baseline method, the BCE_DICE loss function achieved significant performance improvements through its fusion mechanism, with metrics reaching 90.16%, 86.30%, 87.52%, 78.11%, and 89.07%, respectively. Conversely, the weighted BCE_DICE loss function showed a different optimization trend. Recall reached its highest value of 90.34%, but precision correspondingly decreased to 81.14%. These experimental results demonstrate the adaptability and reliability of the DCM module combined with different auxiliary loss functions. After comprehensive evaluation, this model selects the BCE_DICE loss function as the auxiliary loss function for the DCM module, ensuring its compatibility with the loss function of the backbone network. Figures 10 and 11 show the convergence curve of the loss function and the F1 score change curve,

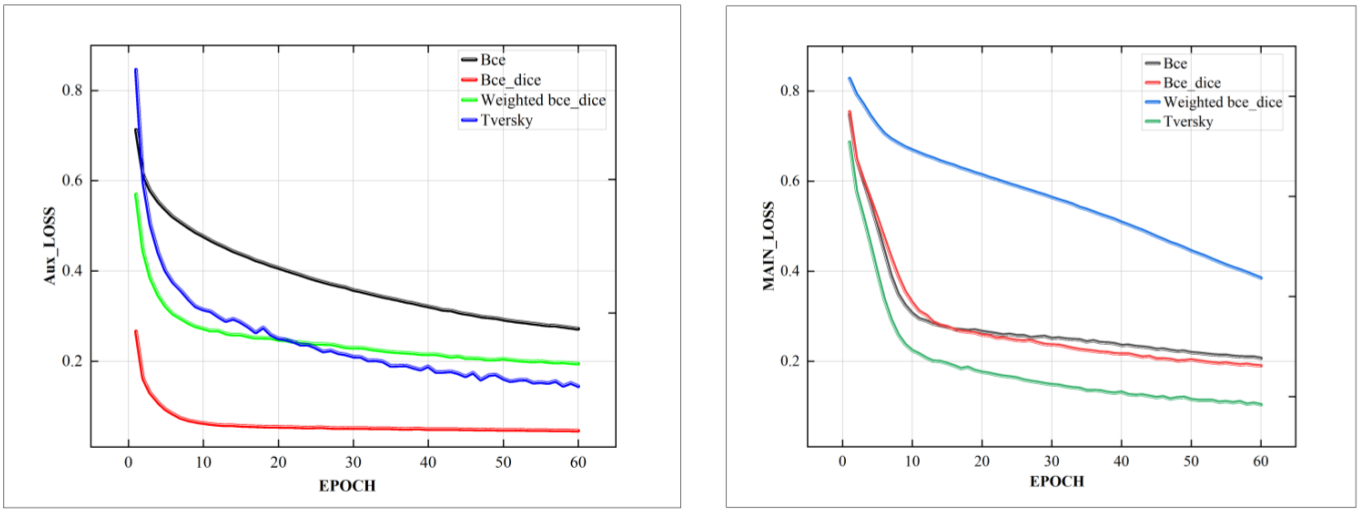

Figure 10. The variation of model loss function with epochs: (a) Auxiliary loss function; (b) Main loss function.

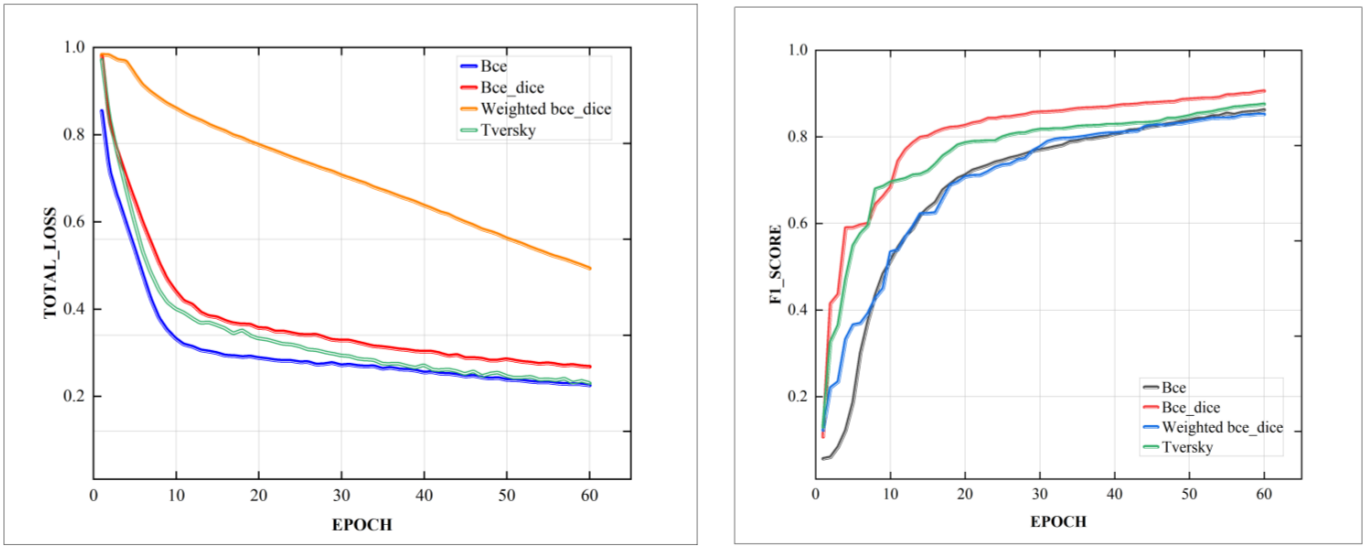

Figure 11. The variation of model loss function with epochs: (a) Total loss function; (b) F1-score function.

respectively. When using the BCE_DICE loss function, this module achieves the best road crack segmentation performance in the model. Note that these results reflect an intermediate training stage for loss function selection; the final model performance after complete training is reported in Table 6.

4.4 Ablation and Comparative Experiment

To validate the effectiveness of the proposed road crack segmentation algorithm, we conducted ablation experiments by progressively integrating the ARC, MAFM, ECA, and DCM modules into the U-Net architecture. Each module was added sequentially, and performance was evaluated on the CFD dataset under identical experimental conditions. The results are presented in Table 4. Specifically, the introduction of the ARC module to the baseline U-Net network improved the F1-score from 67.25% to 69.60%, achieving an absolute gain of 2.35 percentage points. Subsequently, incorporating the MAFM module into the ARC-enhanced architecture increased precision and recall by 2.67% and 3.02% respectively, resulting in a 3.20% improvement in F1-score. This validates the effectiveness of multi-scale feature aggregation in capturing crack patterns at varying scales.

The integration of the ECA module at the decoder stage further enhanced performance by 1.22%, achieving an F1-score of 74.02%. The ECA module provides a lightweight channel attention mechanism with minimal computational overhead, contributing to refined feature representations at the decoder stage. Finally, implementing the DCM module after the upsampling layers yielded the most substantial single-module improvement, boosting the F1-score

Table 4. CFD dataset ablation experimental results.

U-Net	ARC	MAFM	ECA	DCM	Precision /%	Recall/%	Dice/%	IoU/%	F1/%
✓					67.63	65.45	66.51	49.84	67.25
✓	✓				71.34	66.72	68.95	52.62	69.60
✓	✓	✓			74.01	69.74	71.81	56.02	72.80
✓	✓	✓	✓		75.17	70.99	73.10	57.51	74.02
✓	✓	✓	✓	✓	80.59	74.23	76.65	62.95	78.11

Table 5. Training results of different segmentation models on the CFD dataset.

Method	Precision /%	Recall /%	Dice /%	IoU /%	F1 /%
U-Net	67.63	65.45	66.51	49.84	67.25
PSPNet	64.14	62.46	63.29	47.88	63.72
FPN	65.83	61.64	63.67	47.04	64.50
DeeplabV3+	69.77	66.90	68.03	51.17	69.48
DeepCrack	68.34	67.44	67.69	51.73	68.01
CrackSAM	72.22	75.25	73.65	58.48	74.82
SegCrack	70.61	65.30	67.01	51.99	68.83
SCCDNet	72.02	79.96	75.41	61.02	77.63
Ours	80.59	74.23	76.65	62.95	78.11

Table 6. Training results of different segmentation models on the CRACK500 dataset.

Method	Precision /%	Recall /%	Dice /%	IoU /%	F1 /%
U-Net	78.45	72.89	75.26	60.38	76.88
PSPNet	77.16	71.60	74.28	59.02	75.91
FPN	77.30	72.25	74.70	59.06	76.39
DeeplabV3+	80.17	72.47	76.13	61.83	78.20
DeepCrack	75.57	70.04	72.70	58.42	73.15
CrackSAM	85.38	88.76	87.98	76.18	89.92
SegCrack	80.92	74.17	76.56	64.92	78.26
SCCDNet	89.90	82.67	86.20	75.63	87.24
Ours	94.34	86.54	90.19	81.43	90.02

Table 7. Training results of different segmentation models on the HYCrack dataset.

Method	Precision /%	Recall /%	Dice /%	IoU /%	F1 /%
U-Net	74.77	67.43	70.91	54.19	73.03
PSPNet	71.41	64.27	67.65	51.79	69.10
FPN	72.08	65.36	68.57	52.34	70.62
DeeplabV3+	78.61	73.13	75.78	60.87	77.06
DeepCrack	76.11	69.80	73.89	58.41	75.36
CrackSAM	76.45	80.70	77.05	64.22	78.82
SegCrack	78.15	71.63	74.01	60.33	76.05
SCCDNet	80.27	75.41	76.29	64.40	77.23
Ours	82.65	78.73	80.91	69.22	81.60

from 74.02% to 78.11%, an increase of 4.09%. This significant gain demonstrates the effectiveness of the defect correction mechanism in addressing segmentation errors and refining crack boundaries. Overall, the proposed model architecture achieves a 10.86% improvement in F1-score compared to the baseline U-Net network, validating the synergistic effect of all proposed modules.

To demonstrate the superiority of the proposed road crack segmentation algorithm, we conducted comprehensive comparative experiments against several state-of-the-art methods: U-Net, DeeplabV3+, PSPNet [19], FPN [20] (adapted with a semantic

segmentation head for pixel-level crack prediction), DeepCrack [21], SegCrack [22], CrackSAM [23] and SCCDNet [24]. All comparative methods were re-implemented and trained from scratch under identical experimental settings on the three datasets. For the Transformer-based crack segmentation model of Wang and Su [22], we adopt the abbreviation SegCrack to refer to this method in subsequent tables for brevity. In the comparative experiments, FPN [20] refers to the general feature pyramid architecture of Lin et al., which is distinct from the pavement-crack-specific feature pyramid network of Yang et al. [15]. The CrackSAM results reported in Tables 5–7 were obtained by re-implementing the SAM-based crack detection framework of Rakshitha et al. [23] under identical experimental settings, trained from scratch on each dataset independently. The reported values reflect our re-implementation performance and may differ from the original publication due to differences in training data, augmentation strategy, and optimization configuration.

The quantitative results presented in Tables 5, 6 and 7 demonstrate the consistent superiority of the proposed method across all three datasets. On the HYCrack dataset, evaluation metrics indicate suboptimal performance across all comparative methods, with CrackSAM yielding the best results at 76.45%, 80.70%, 77.05%, 64.22%, and 78.82%. While the corresponding evaluation indexes of our method are 82.65%, 78.73%, 80.91%, 69.22% and 81.60% respectively. Compared with several state-of-the-art crack segmentation models, the proposed approach exhibits superior performance in detecting edge details.

Model performance was evaluated through both quantitative metrics and visual segmentation results. As shown in Figures 12 and 13, the models show significant disparities in crack segmentation results. Compared with other comparative methods, the proposed method has better recognition rate, higher accuracy, and more complete crack segmentation for all datasets. Comparative methods exhibit blurring, discontinuities, and missed detections. Consequently, our approach not only comprehensively captures crack morphology with exceptional background robustness but also effectively reduces false positives, enhancing segmentation accuracy and reliability. Furthermore, the model predicted binarized fracture images with distinct boundary transitions between crack and non-crack regions, demonstrating exceptional performance in preserving and extracting fine crack structures.

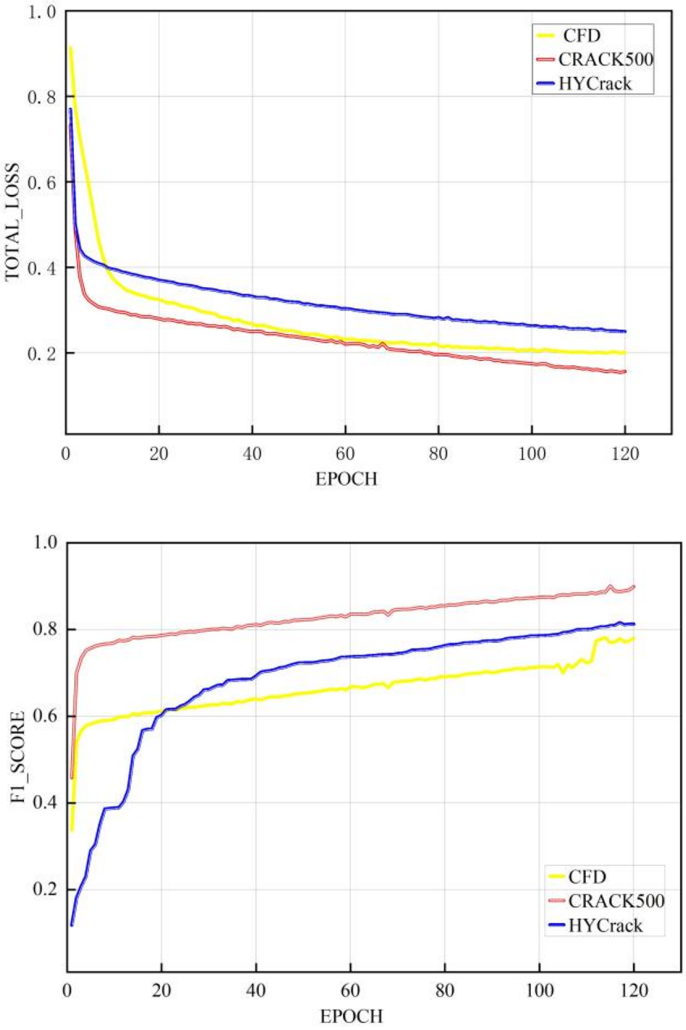

Figure 12. Convergence curves of total loss and F1-scores across training epochs on the CFD, CRACK500, and HYCrack datasets.

The proposed model achieves optimal performance on the CRACK500 dataset, followed by the HYCrack dataset in second place. This verifies the model has strong generalization capability. It can ensure the structural integrity of road cracks, adapt to various crack types, and maintain stable performance in various complex scenarios. Whether processing fractures in complex background or capturing characteristics of varied crack types, the model shows obvious advantages, it is ahead of other segmentation models in evaluation indexes.

4.5 Robustness and Generalization Analysis

To evaluate the robustness and generalization capability of the proposed model across diverse scenarios, we conducted a comprehensive

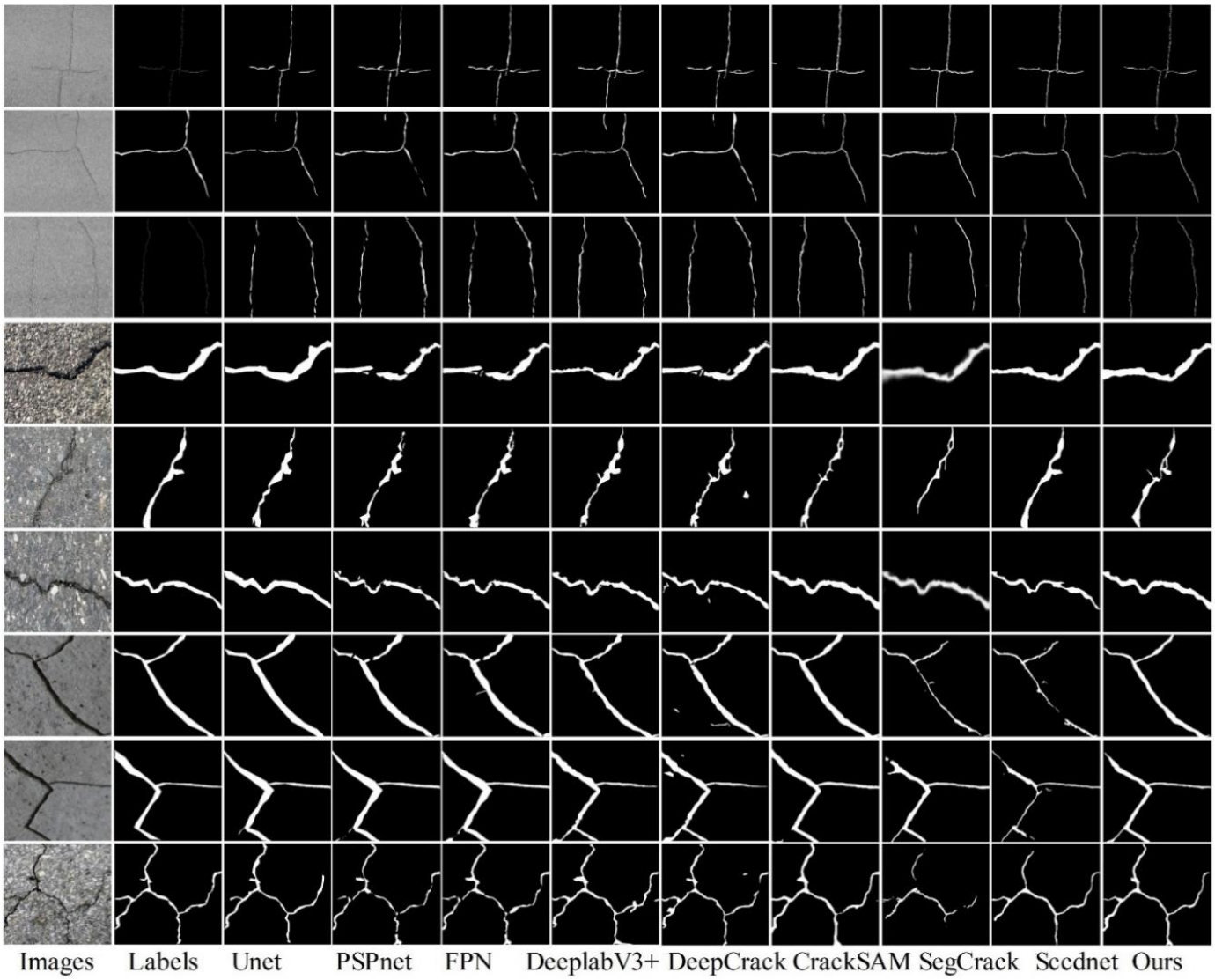

Figure 13. The segmentation results of the proposed model and the comparative model on three datasets.

Table 8. Model performance analysis under varying conditions on HYCrack dataset.

	Groups	Training Samples	Precision /%	Recall /%	IoU /%	F1/%
1	Adequate lighting cracks	3762	90.25	87.35	79.51	89.71
	Shadow occlusion cracks	1278	84.27	83.24	72.22	83.51
	Pavement marking cracks	657	81.09	77.31	65.34	78.37
2	Linear cracks	3159	91.61	89.06	82.19	90.44
	Block cracks	1773	88.25	81.54	73.29	84.30
	Grid cracks	765	78.34	74.32	61.31	76.85
3	Asphalt pavement cracks	3609	81.41	82.12	69.29	81.73
	Concrete pavement cracks	2088	85.52	87.14	75.68	86.29

performance analysis on the HYCrack dataset under varying environmental and structural conditions. We divide the test samples into three groups based on different characteristics: environmental conditions, crack morphology, and pavement type. Due to natural overlap in real-world scenarios, some samples may belong to multiple categories simultaneously.

4.5.1 Robustness Analysis

Due to significant environmental interference in outdoor data acquisition, the dataset naturally includes images captured under different environmental conditions. We divided the test samples into three categories based on average pixel intensity and contrast features: adequate lighting images,

shadow occlusion cracks, and pavement marking cracks, as shown in Table 8. The F1 score reached 89.71% on Adequate lighting cracks and 78.37% on pavement marking cracks, showing a performance decrease of approximately 11.3% from optimal to challenging environmental conditions. This variation indicates that the model has reasonable robustness to environmental changes, but the performance degradation in complex environments suggests that dataset could be further improved through data augmentation or specialized pre-processing techniques to enhance model performance.

4.5.2 Pavement material generalization

Dataset contains cracks on both asphalt and concrete pavements, which exhibit significant visual differences in texture and reflective properties. Our method evaluates the model's performance on these two pavement types separately. The model achieved an F1 score of 81.73% on asphalt pavement cracks and 86.29% on concrete pavement cracks, with corresponding IoU values of 69.29% and 75.68%. The difference in F1 scores between the two types of cracks is 4.56%, and the difference in IoU values is 6.39%, indicating that despite the varying surface textures and optical properties of different pavement materials, the proposed model can effectively generalize to different pavement materials, validating the effectiveness of the multi-scale mechanism in capturing material-independent crack features.

5 Conclusion

To address challenges including blurred crack boundaries, loss of detail information, and low segmentation accuracy in road crack segmentation tasks, this paper proposes a crack segmentation algorithm integrating multi-attention fusion and defect correction mechanisms. Building upon the U-Net network, we incorporate ARC and ECA within the encoder-decoder framework, while innovatively designing MAFM and DCM. These modules improve the network feature extraction capability and enhances segmentation precision at crack boundaries. The proposed segmentation algorithm not only strengthens multi-scale crack feature capture and mitigates potential gradient vanishing during training, but also improves the completeness and continuity of segmentation results.

The experimental results demonstrate that the proposed algorithm can accurately identify crack details in complex backgrounds. This capability significantly mitigates issues such as crack edge blurring and the omission of minor branches. Future research directions include two aspects. Firstly, on roads with severe weathering and surface damage, crack boundaries become blurred, leading to performance degradation; future work should introduce uncertainty estimation to flag unreliable predictions. Secondly, the model requires an input resolution of 448×448. Although the current algorithm performs well on dataset containing nearly ten thousand images, applying it to larger road crack datasets and more complex road environments may present difficulties.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province, China under Grant SJCX25_2194.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study involved image collection on campus roads using a handheld device. No human subjects were recruited and no personal data were collected. All images were captured at ground level targeting pavement surfaces; any incidentally captured pedestrians or vehicles were anonymized through automatic blurring prior to dataset construction. Formal ethical approval was not required for this type of infrastructure monitoring study under the institutional guidelines of Huaiyin Institute of Technology.

References

[1] Wang, W., Wang, M., Li, H., Zhao, H., Wang, K., He, C., ... & Chen, J. (2019). Pavement crack image acquisition methods and crack extraction algorithms: A review.

Journal of Traffic and Transportation Engineering (English Edition), 6(6), 535-556. [CrossRef]

[2] Yang, X., Li, H., Yu, Y., Luo, X., Huang, T., & Yang, X. (2018). Automatic pixel-level crack detection and measurement using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering, 33*(12), 1090-1109. [CrossRef]

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing. [CrossRef]

[4] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

[5] Huang, S., Chen, H., Yan, L., Zou, X., Li, B., & Bi, Y. (2025). A review of the progress in machine vision-based crack detection and identification technology for asphalt pavements. *Digital Transportation and Safety, 4*(1), 65-79. [CrossRef]

[6] Zhang, Q., Chen, S., Wu, Y., Ji, Z., Yan, F., Huang, S., & Liu, Y. (2024). Improved U-net network asphalt pavement crack detection method. *Plos one, 19*(5), e0300679. [CrossRef]

[7] Yang, L., Bai, S., Liu, Y., & Yu, H. (2023). Multi-scale triple-attention network for pixelwise crack segmentation. *Automation in Construction, 150*, 104853. [CrossRef]

[8] Pan, Y., Zhang, G., & Zhang, L. (2020). A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. *Automation in Construction, 119*, 103357. [CrossRef]

[9] Al-Huda, Z., Peng, B., Algburi, R. N. A., Al-antari, M. A., Al-Jarazi, R., & Zhai, D. (2023). A hybrid deep learning pavement crack semantic segmentation. *Engineering Applications of Artificial Intelligence, 122*, 106142. [CrossRef]

[10] Ali, L., Jassmi, H. A., Khan, W., & Alnajjar, F. (2023). Crack45K: integration of vision transformer with tubularity flow field (TuFF) and sliding-window approach for crack-segmentation in pavement structures. *Buildings, 13*(1), 55. [CrossRef]

[11] Hamishebahar, Y., Guan, H., So, S., & Jo, J. (2022). A comprehensive review of deep learning-based crack detection approaches. *Applied Sciences, 12*(3), 1374. [CrossRef]

[12] Xiang, C., Guo, J., Cao, R., & Deng, L. (2023). A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenarios. *Automation in Construction, 152*, 104894. [CrossRef]

[13] Sahragard, E., Farsi, H., & Mohamadzadeh, S. (2025). Advancing semantic segmentation: Enhanced UNet algorithm with attention mechanism and deformable convolution. *PloS one, 20*(1), e0305561. [CrossRef]

[14] Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic road crack detection using random structured forests. *IEEE transactions on intelligent transportation systems, 17*(12), 3434-3445. [CrossRef]

[15] Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., & Ling, H. (2019). Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE transactions on intelligent transportation systems, 21*(4), 1525-1535. [CrossRef]

[16] Yan, Y., Deng, C., Li, L., Zhu, L., & Ye, B. (2023). Survey of image semantic segmentation methods in the deep learning era. *Journal of Image and Graphics, 28*(11), 3342-3362. [CrossRef]

[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. [CrossRef]

[18] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018, September). CBAM: Convolutional Block Attention Module. In European Conference on Computer Vision (pp. 3-19). Cham: Springer International Publishing. [CrossRef]

[19] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017, July). Pyramid Scene Parsing Network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6230-6239). IEEE. [CrossRef]

[20] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017, July). Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 936-944). IEEE. [CrossRef]

[21] Zou, Q., Zhang, Z., Li, Q., Qi, X., Wang, Q., & Wang, S. (2018). Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE transactions on image processing, 28*(3), 1498-1512. [CrossRef]

[22] Wang, W., & Su, C. (2022). Automatic concrete crack segmentation model based on transformer. *Automation in Construction, 139*, 104275. [CrossRef]

[23] Rakshitha, R., Srinath, S., Vinay Kumar, N., Rashmi, S., & Poornima, B. V. (2024). Crack SAM: enhancing crack detection utilizing foundation models and Detectron2 architecture. *Journal of Infrastructure Preservation and Resilience, 5*(1), 11. [CrossRef]

[24] Li, H., Yue, Z., Liu, J., Wang, Y., Cai, H., Cui, K., & Chen, X. (2021). Sccdnet: A pixel-level crack segmentation network. *Applied Sciences, 11*(11), 5074. [CrossRef]

[25] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11534-11542). IEEE. [CrossRef]

Rendong Ji received the Ph.D. degree from Nanjing University of Aeronautics and Astronautics, in 2015. He has been engaged in research work in the field of photoelectric detection for a long time, and has carried out research on intelligent photoelectric detection and information processing. He is currently conducting research on the application of deep learning in related fields. (Email: jrdgxy@163.com)

Xiaojun Zhang is currently studying at Huaiyin Institute of Technology, pursuing the M.S. degree in transportation. His main research interests are deep learning and road crack image detection technology. (Email: zxxxxj8800@163.com)

Xiu Tang is currently pursuing the M.S. degree in Transportation Engineering at Huaiyin Institute of Technology. Her research focuses on image detection technology, with applications in electric vehicle (EV) safety. (Email: 2474855223@qq.com)

Xiaoyan Wang received the Ph.D. degree from Nanjing University of Aeronautics and Astronautics in 2020. She has been engaged in research work in the field of photoelectric detection for a long time and has carried out research in spectral detection and analysis. She is currently conducting research on the application of deep learning in related fields. (Email: wxygxy@163.com)

Yunlong Xu is currently pursuing the M.S. degree in Transportation Engineering at Huaiyin Institute of Technology. He is currently engaged in research in the field of deep learning. (Email: yunlongxv293@gmail.com)

Jiaxin Shi is currently studying at Huaiyin Institute of Technology, pursuing a M.S. degree in Computer and Software Engineering. His main research interest is hyperspectral target detection. (Email: shijiaxin32@163.com)