Article citation info:
Wang X. Deep learning-based fatigue crack detection method for weld-free joints of circular steel tube trusses . Diagnostyka. 2026;27(4):2026405. https://doi.org/10.29354/diag/231241.

DIAGNOSTYKA, 2026, Vol. 27, No. 4

e-ISSN 2449-5220
DOI: 10.29354/diag/231241

DEEP LEARNING-BASED FATIGUE CRACK DETECTION METHOD FOR WELD-FREE JOINTS OF CIRCULAR STEEL TUBE TRUSSES

Xinyi WANG

Faculty of Art and Design, Guang Xi Vocational & Technical College, Nanning 530226, China
Corresponding author, e-mail: Sddzwxy1977@163.com

Abstract

To improve the detection accuracy and efficiency of fatigue cracks in weld-free joints of circular steel tube trusses, and address crack omission, information loss, and secondary damage issues of traditional sensor-based methods in complex joint structures, this paper proposes a method based on improved YOLO11 and SegFormer. Aiming at YOLO11's large parameters, limited inference speed, and poor adaptability to narrow, stress-concealed cracks, the YOLO-CD model is proposed: StarNet reduces computational costs for mobile scenarios; HSANet enhances capture of crack edge details and weak features to resist corrosion/weld interference; and the OSCD head optimizes multi-scale crack detection. The improved SegFormer-HF integrates the FFM and HLFDB modules to suppress downsampling information loss and improve segmentation consistency and completeness under complex backgrounds. A "detection-localization first, segmentation-refinement later" scheme combined with skeletonization algorithm realizes automated accurate calculation of crack length and width. Traditional fatigue crack detection methods mainly rely on manual inspection, ultrasonic testing, magnetic particle inspection, and sensor-based monitoring approaches. However, these methods often suffer from low efficiency, limited coverage, sensitivity to environmental interference, and difficulty in detecting narrow early-stage cracks under complex structural backgrounds. Recent deep learning-based methods have improved crack localization and segmentation accuracy, but challenges related to lightweight deployment, narrow crack recognition, and robustness under complex backgrounds still remain. Experiments on the constructed dataset (cracks under different fatigue stages/stress loads) show: YOLO-CD achieves 67.8% F1, 71.5% mAP50, 46.4% mAP50-95, with GFLOPs reduced by 47.6% vs. YOLO11; SegFormer-HF reaches 91.50% F1, 90.51% mIoU, 85.15% mPA, outperforming DeepLabv3+ and UNet. This lightweight method balances speed and accuracy, adapts to narrow/concealed cracks and complex backgrounds, and can be deployed on mobile devices (e.g., inspection drones), supporting dynamic monitoring and safety assessment of weld-free joint fatigue cracks.

Keywords: fatigue crack detection; deep learning; weld-free joints； object detection; semantic segmentation

1. INTRODUCTION

As core load-bearing components in long-span buildings and transportation infrastructure, weld-free joints of circular steel tube trusses are widely used in engineering scenarios such as bridges, factories, and stadiums, thanks to advantages including easy assembly, uniform stress distribution, and excellent seismic performance. Their structural integrity directly determines the bearing safety and service life of the entire truss system. However, relying on bolted connections or interlocking assembly, weld-free joints are prone to initiating fatigue cracks in stress concentration areas (e.g., around bolt holes and at the intersection interfaces of tube trusses) when subjected to long-term cyclic loads, vibrational impacts, and environmental corrosion (such as moisture and salt erosion). Such fatigue cracks feature narrowness, concealment, directional propagation along stress directions, and low pixel ratio in the early stage, making them difficult to detect initially. Once entering the rapid propagation phase, they may lead to joint connection failure, thereby triggering overall truss collapse and causing severe engineering accidents. Additionally, the complex structure of weld-free joints—with interfering factors like bolt shadows, assembly gaps, and rust traces on the surface—easily confuses fatigue cracks. Meanwhile, detection must meet non-contact and non-destructive requirements to avoid node disassembly or damage to the connection structure, imposing stringent demands on the detection technology's anti-interference capability, narrow crack recognition accuracy, and dynamic tracking ability. Therefore, accurately identifying the characteristics of fatigue cracks in weld-free joints and real-time monitoring their propagation

Received 2026-03-20; Accepted 2026-08-18; Available online 2026-11-19

© 2026 by the Authors. Licensee Polish Society of Technical Diagnostics (Warsaw. Poland). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

trends are crucial to ensuring the safe service of circular steel tube truss structures. The detection effect directly relates to the scientific nature of engineering operation and maintenance decisions and the durability of structural service.

Currently, fatigue crack detection for weld-free joints of circular steel tube trusses still relies mainly on traditional methods: Contact sensors (e.g., strain gauges, ultrasonic probes) need to be installed close to the joints. Limited by the complex structure of the joints, they are difficult to cover all potential crack areas, prone to missed detections, and the installation process may interfere with the stress state of the joints. Manual inspection relies on visual observation and non-destructive testing equipment (e.g., magnetic particle inspection), which is inefficient and highly subjective, failing to capture early narrow fatigue cracks and achieve full-coverage detection of large-scale truss structures. With the development of computer vision technology [1-2], digital image technology has been gradually applied in the field of steel structure crack detection due to its advantages of non-contact, high efficiency, and low cost. Some scholars have realized the identification and measurement of steel structure surface cracks through image acquisition and processing technology [3], and other studies have improved the detection accuracy of cracks under complex backgrounds with the help of deep learning algorithms [4-6]. In the field of object detection, the Faster R-CNN algorithm proposed by Ren et al. [7] realized end-to-end detection, providing ideas for the rapid localization of steel structure cracks; Zhou et al. [8] proposed a real-time tunnel crack detection method based on the YOLOX-G algorithm, which addresses the problems of low accuracy and slow speed of traditional detection methods for tunnel lining cracks. In the field of semantic segmentation, the improved DeepLabv3 model by Chen et al. [9] enhanced segmentation accuracy, providing technical support for crack contour extraction. However, existing methods still have obvious limitations for fatigue crack detection of weld-free joints of circular steel tube trusses: They are prone to misclassifying linear noise as cracks under complex backgrounds (e.g., bolt edges, tube truss textures, rust traces); the recognition sensitivity for narrow and early fatigue cracks is insufficient; and they lack the ability to track the dynamic propagation of fatigue cracks, failing to meet the core engineering demand of "early warning and continuous monitoring" [10].

1.1. Categorization of Existing Fatigue Crack Detection Methods

(1) Traditional Manual Inspection Methods
Visual inspection
Magnetic particle testing
Ultrasonic testing
Limitations:
Low efficiency
Human subjectivity
Poor early crack sensitivity
(2) Sensor-Based Monitoring Methods
Strain gauges
Acoustic emission sensors
Fiber optic sensing
Limitations:
Expensive
Limited coverage
Difficult installation in complex joints
(3) Image Processing-Based Methods
Thresholding
Edge detection
Morphological processing
Limitations:
Sensitive to noise
Weak robustness
(4) Deep Learning-Based Methods
Faster R-CNN
YOLO
U-Net
DeepLabV3+
Transformer-based segmentation
Advantages:
Automated detection
High accuracy
Real-time capability
Remaining challenges:
Narrow crack detection
Lightweight deployment
Complex background interference

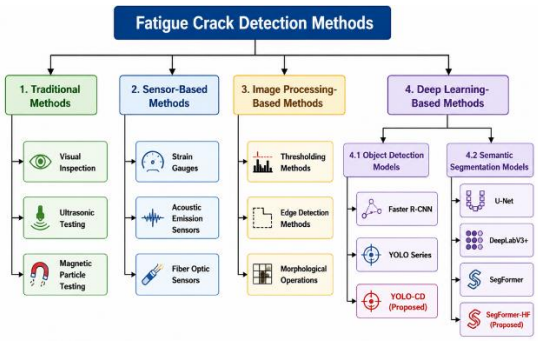


Fig. 1. Categorization of Existing Fatigue Crack Detection Methods

Figure 1 categorizes the major fatigue crack detection approaches used in structural health monitoring applications. Traditional methods mainly rely on manual inspection and non-destructive testing techniques such as ultrasonic and magnetic particle testing. Sensor-based methods utilize physical sensing devices to monitor crack propagation and structural response. Image processing-based methods perform crack extraction using handcrafted algorithms such as thresholding and edge detection. In recent years, deep learning-based methods have become dominant due to their superior capability in automated feature extraction and complex crack recognition. These approaches are mainly divided into object detection and

semantic segmentation frameworks. The proposed YOLO-CD and SegFormer-HF models belong to these two categories respectively and are specifically designed for fatigue crack detection under complex weld-free joint backgrounds.

Similarly, accurate crack length calculation methods are also crucial components in crack detection [11]. Nishiyama et al. [12] proposed a method to calculate crack width by placing targets on both sides of the crack and monitoring changes in their positions. Zhe et al. [13] developed an approach based on the orthogonal boundary algorithm, which utilizes crack boundaries and skeleton directions to determine crack propagation and measure crack width. Zhou Ying et al. [14] put forward a crack length calculation method based on Euclidean distance. Tian et al. [15] proposed a crack detection method based on computer vision techniques for structural surface analysis. Wang Zhengdong et al. [16] extracted single-pixel skeletons and contours along crack boundaries based on the orthogonal projection method, enabling continuous crack width measurement. The above research results indicate that crack segmentation and size measurement based on computer vision technology can achieve pixel-level accuracy. Compared with traditional manual detection methods, these approaches not only significantly improve the accuracy of detection results but also greatly enhance the automation level of the detection process.

The YOLOv11 [17] object detection algorithm has optimized detection speed and accuracy based on YOLOv8 [18], making it more suitable for the high-efficiency detection requirements of engineering sites. Considering that weld-free joints of circular steel tube trusses are mostly deployed in long-span engineering scenarios, detection needs to be integrated into mobile devices such as drones and inspection robots, and fatigue cracks are characterized by narrowness, concealment, and low pixel ratio, this paper selects the YOLO11-n model from YOLOv11 for targeted improvement and proposes the YOLO-CD (You Only Look Once-Crack Detect) object detection model. While ensuring lightweight performance and meeting the deployment requirements of mobile devices, it enhances the recognition capability of narrow features of fatigue cracks in stress concentration areas of weld-free joints (e.g., around bolt holes), avoiding missed detections and false detections under complex backgrounds.

Meanwhile, given the complex structure of weld-free joints (including bolt shadows, assembly gaps, and rust traces) and the core demand of balancing accuracy and computational efficiency for high-resolution image segmentation, the problems of large parameter volume and high memory usage in previous semantic segmentation models are particularly prominent. Based on the lightweight semantic segmentation model SegFormer [20] under the Transformer [19] framework, this paper proposes the improved SegFormer-HF model. By optimizing the feature extraction and fusion mechanisms, it reduces model latency while enhancing the semantic consistency and edge integrity of fatigue crack segmentation under complex backgrounds, adapting to the high-resolution image processing requirements of weld-free joints.

To further improve detection efficiency and accuracy, this paper proposes a joint scheme of "detection-localization first, then segmentation-refinement": the YOLO-CD algorithm is used to quickly lock in fatigue crack regions in weld-free joints, followed by the SegFormer-HF algorithm to achieve accurate segmentation of crack contours. Combined with geometric parameter calculation methods for fatigue cracks (length and width measurement), this scheme ultimately realizes the automatic identification, localization, and quantitative analysis of fatigue cracks in weld-free joints of circular steel tube trusses. Experimental verification shows that this joint scheme adapts to the complex detection scenarios of weld-free joints and is feasible in practical engineering applications, providing data support for subsequent fatigue crack propagation monitoring.

The main contributions of this study are summarized as follows:

(1) A lightweight YOLO-CD object detection model is proposed for fatigue crack localization in weld-free joints of circular steel tube trusses. The proposed model integrates StarNet, HSANet, and OSCD modules to improve narrow crack detection accuracy while reducing computational complexity.
(2) An improved SegFormer-HF semantic segmentation model is developed to enhance crack edge continuity and segmentation robustness under complex structural backgrounds such as bolt shadows, assembly gaps, and rust interference.
(3) A joint “detection-localization first and segmentation-refinement later” framework is constructed to achieve automated fatigue crack localization, contour extraction, and quantitative geometric parameter measurement.
(4) A crack skeleton-based geometric measurement strategy is introduced to calculate crack length and width with engineering-level accuracy under complex weld-free joint conditions.
(5) Extensive experiments including ablation studies, comparative analysis, and cross-dataset generalization experiments demonstrate that the proposed method achieves superior lightweight performance and detection accuracy compared with existing mainstream methods.

The remainder of this paper is organized as follows. Section 2 introduces the proposed YOLO-CD object detection model and SegFormer-HF semantic segmentation model. Section 3 presents the

dataset construction, experimental settings, and performance evaluation results. Section 4 describes the fatigue crack geometric parameter calculation process, including crack length and width measurement. Finally, Section 5 summarizes the conclusions and discusses future research directions.

2. FATIGUE CRACK DETECTION MODEL FOR STRUCTURAL SURFACES

2.1. YOLO-CD Object Detection Model

Targeting the characteristics of "narrow and concealed cracks, complex backgrounds (including bolt holes, assembly gaps, and rust traces)" in fatigue crack detection of weld-free joints of circular steel tube trusses, as well as the requirements for model lightweighting and real-time inference from on-site mobile inspection equipment (drones, inspection robots), this paper addresses the inference speed bottleneck caused by the large parameter volume of the YOLO11 network and proposes the YOLO-CD (YOLO Crack Detection) object detection model adapted for weld-free joint fatigue crack detection. Its network structure is shown in Fig. 2. Based on YOLO11, the YOLO-CD model adopts the StarNet [21] module to replace the original backbone network, reducing computational costs while enhancing the ability to capture narrow crack features. A lightweight Hybrid Selective Attention Network (HSANet) module is designed to replace the neck network, which recovers the detailed features of fatigue cracks obscured by background interference through an adaptive nonlinear interpolation strategy (Dysample module). Meanwhile, the Optimized Spatial Context Detection Head (OSCD) is employed, which adapts to fatigue cracks in different propagation stages of weld-free joints (narrow early-stage cracks and expanded late-stage cracks) through shared convolutional layers and dynamic scaling technology, further improving the detection efficiency of multi-scale cracks.

Due to the large parameter volume and slow inference speed of the C3k2 module in YOLO11's backbone feature extraction network, it cannot meet the real-time requirements of mobile inspection for weld-free joints and struggles to efficiently capture the weak features of narrow fatigue cracks. To address this, this paper adopts the lightweight StarNet network to replace YOLO11's backbone network, aiming to reduce model complexity, improve inference speed, and enhance the feature expression capability for narrow cracks. StarNet is a lightweight network based on the Star Operation, whose structure is shown in Fig. 2. Here, H and W represent the height and width of the input image, respectively, and d denotes the number of channels of the input feature map. The Star Operation realizes implicit high-dimensional mapping of input features through element-wise multiplication, which can strengthen nonlinear interactions between features, thereby more accurately capturing the differential features between "narrow fatigue cracks and complex backgrounds". Its mathematical expression is as follows:

$$Y = (W_1 X) * (W_2 X) \tag{1}$$

Where W1 and W2 are weight matrices, X is the input feature, and * denotes element-wise multiplication.

Unlike the addition operation in traditional convolution, the Star Operation enhances nonlinear interactions between features through element-wise multiplication, enabling more effective distinction of the complex feature relationships between weld-free joint fatigue cracks (narrow linear features) and background interference (bolt shadows, rust textures). Traditional convolution typically improves expressive capability by stacking layers or increasing kernel size, but this significantly increases computational costs (e.g., MobileNetv3's depthwise separable convolution needs to balance accuracy and speed). In contrast, the Star Operation generates a high-dimensional feature space similar to polynomial kernels through element-wise multiplication alone.

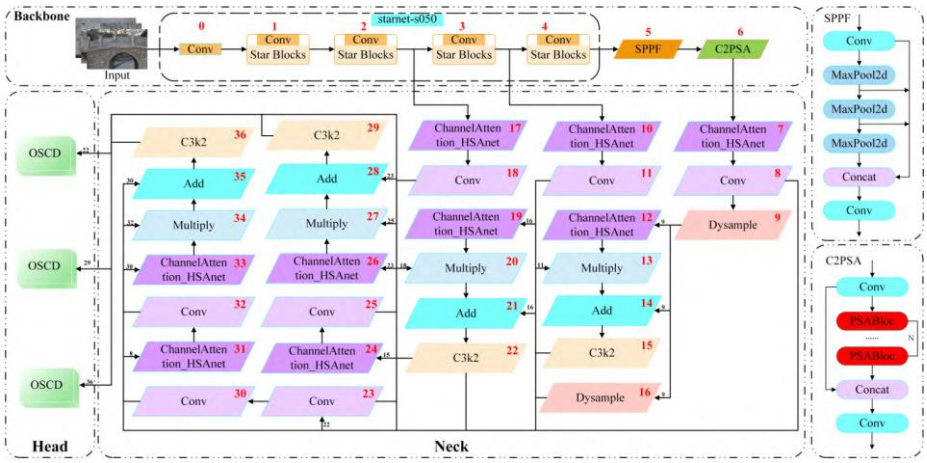


Fig. 2. YOLO-CD network structure diagram

While maintaining a narrow network structure to adapt to the computational resources of mobile inspection equipment, it enhances the nonlinear capture capability for narrow fatigue cracks and avoids the computational overhead caused by increasing the number of channels in traditional convolution.

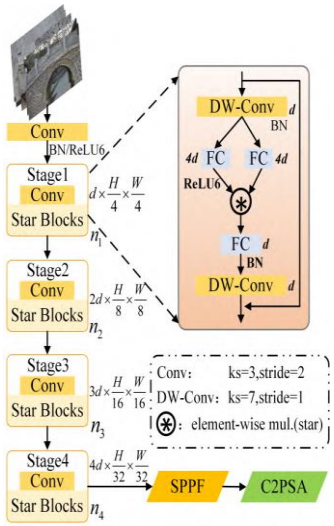


Fig. 3. StarNet architecture overview

StarNet adopts a phased design. After each stage performs downsampling and increases the number of channels through convolutional layers, it concatenates multiple Star Block modules, as shown in Fig. 3. Each Star Block performs depthwise convolution (DWConv), two fully connected (FC) transformations, and the Star Operation in sequence to gradually extract and fuse features. A single-layer Star Operation module generates nonlinear interactions of features through two linear transformations (fully connected layers) and one element-wise multiplication operation, and its calculation formula is as follows:

$$Y_l = (W_{l1}Y_{l-1}) * (W_{l2}Y_{l-1}) \quad (2)$$

Where Y_{l-1} denotes the output feature of the previous layer, and W_{l1} and W_{l2} are the weight matrices of the current layer. In multi-layer stacking, the Star Operation module of each layer takes the implicit high-dimensional features from the previous layer as input for further mapping and fusion. As the number of layers increases, the dimension of implicit features grows exponentially, but this expansion is implicit and does not significantly affect the explicit number of channels or computational complexity. Through such layer-wise stacking, StarNet can efficiently express high-dimensional nonlinear features and realize complex interactions under limited computational resources, thereby improving model performance while maintaining lightweight properties.

To address the detection challenges of fatigue cracks in weld-free joints of circular steel tube trusses-characterized by narrowness, concealment, and the susceptibility of details in low-resolution feature maps to being obscured by background interference such as bolt shadows and rust textures. This paper proposes the Hybrid Selective Attention Network (HSANet), whose structure is shown in Fig. 4. The network utilizes a Channel Attention (CA) mechanism to focus on effective features of fatigue cracks and filter out background interference. It integrates a Selective Feature Fusion (SFF) module to aggregate multi-scale features of fatigue cracks in different propagation stages. Meanwhile, a Dysample module is introduced to preserve the narrow spatial structure information of fatigue cracks by dynamically adjusting interpolation weights, avoiding the edge blurring problem caused by traditional bilinear interpolation. This enhances the YOLO-CD model's ability to retain edge details of fatigue cracks in low-resolution feature maps.

HSANet consists of two stages: feature selection and feature fusion. In the feature selection stage, the CA module performs channel-wise weighting on the input feature maps to filter out irrelevant and inefficient information such as bolts and rust on the surface of weld-free joints, directing the model to focus on the core features of fatigue cracks. The calculation formula of the CA module is as follows:

$$C(F) = \sigma(W \cdot GAP(F)) \quad (3)$$

Where F denotes the input feature map containing the background of weld-free joints, W is a trainable weight matrix, σ represents the sigmoid activation function, and GAP (global average pooling) denotes the global average pooling operation.

In the feature fusion stage, the filtered low-scale (early-stage narrow fatigue cracks) features are fused with high-scale (late-stage expanded fatigue cracks) semantic features. The SFF mechanism enhances the expressiveness of low-scale narrow crack features while preserving the edge detail information of fatigue cracks. The fusion process of the SFF module can be expressed as:

$$S(F_h, F_1) = F_h + \sigma(W \cdot F_1) \quad (4)$$

Where F_h and F_l denote the high-scale (expanded cracks) and low-scale (narrow cracks) feature maps, respectively, W is a trainable weight matrix, and σ represents the sigmoid activation function.

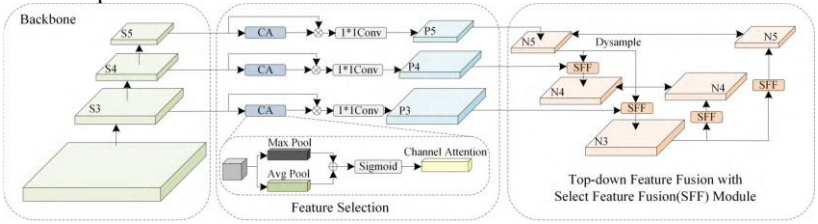


Fig. 4. HSANet network structure diagram

Meanwhile, to further enhance the detail recovery capability of narrow fatigue cracks in low-resolution feature maps, the Dysample upsampling module is utilized in the feature fusion stage. It effectively preserves the linear spatial structure of fatigue cracks and reduces common issues such as edge blurring or distortion during upsampling. This is crucial for identifying early fatigue cracks with an extremely low pixel ratio in weld-free joints.

By integrating Group Normalization (GN) [22] with multi-layer convolutional modules, the OSCD detection head enhances the detection capability for fatigue cracks of different scales (from initial narrow cracks to expanded cracks) in weld-free joints while reducing model complexity. Its modular structure consists of alternately stacked 1×1 and 3×3 convolutional kernels: first, 1×1 convolution is used to compress the channel dimension to eliminate redundant features in the weld-free joint background; subsequently, 3×3 convolution extracts the spatial context information of fatigue cracks, and the GN layer normalizes the feature distribution to alleviate the gradient instability issue during training. To further balance the requirements of detection accuracy for weld-free joint fatigue cracks and model lightweighting (adapting to mobile inspection equipment), the OSCD employs shared convolutional layers to significantly reduce the number of model parameters. Meanwhile, a Scale layer is introduced after the shared layers. By learning the size distribution parameters of fatigue cracks in weld-free joints, this layer dynamically scales the feature resolution to adaptively match the size differences of cracks in different propagation stages, thereby enhancing the robustness of multi-scale detection.

At the end of the detection head, there are independent classification (Conv.Cls) and bounding box regression (Conv.Reg) modules: Conv.Cls outputs the class probabilities of "fatigue cracks/background interference", while Conv.Reg predicts the position and angle parameters of rotated bounding boxes adapted to the linear physical characteristics of fatigue cracks. The latter is constrained to the range of [-45°, 135°] through the Sigmoid function, ensuring that the bounding box matches the actual morphology of fatigue cracks propagating along stress directions in weld-free joints.

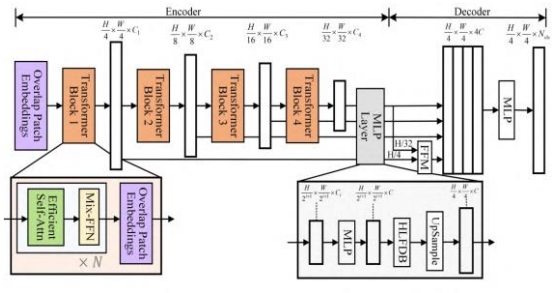


Fig. 5. SegFormer-HF network structure diagram

2.2. SegFormer-HF Semantic Segmentation Model

To address the core challenges in fatigue crack segmentation of weld-free joints of circular steel tube trusses—including complex background interference (bolt holes, assembly gaps, rust traces), low pixel ratio of narrow cracks, and adaptation to edge devices for mobile inspection. This paper proposes the SegFormer-HF semantic segmentation model. In its structure, H and W denote the height and width of the input image, respectively; (C1, C2, C3, C4) represent the number of channels of feature maps at each stage; and Ncls is the total number of classes ("fatigue cracks/background interference"). To balance segmentation accuracy and lightweight requirements for edge deployment, the model selects MiT B0 as the encoder. As an efficient model based on the Transformer architecture, the MiT encoder of SegFormer-HF consists of three core modules: Overlapped Patch Embedding (OPE), Efficient Multi-Head Self-Attention (EMSA), and Mixed Feed-Forward Network (MixFFN). OPE enhances the local feature continuity of narrow fatigue cracks on the surface of weld-free joints through overlapping convolution operations, preventing their linear contours from breaking during block division. EMSA extracts global context information via the attention mechanism, enabling it to focus on the core features of fatigue cracks in complex backgrounds and reduce the risk of missegmentation. MixFFN replaces traditional positional encoding with 3×3 convolution to implicitly model the spatial positional relationship of fatigue cracks propagating along stress directions. The decoder unifies the channels of the four-stage features output by the encoder to 256 dimensions through 1×1 convolution, restores them to 1/4 resolution via bilinear interpolation, performs feature concatenation, and finally generates segmentation results through two 1×1 convolutions. This design enhances the feature expression of early narrow fatigue cracks and accurately extracts their complete contours.

In the SegFormer-HF model adapted for fatigue crack segmentation of weld-free joints of circular steel tube trusses, a Features Fusion Module (FFM) is introduced to efficiently fuse multi-scale features: this module fuses the features output by the CNN encoder and the Transformer encoder, enabling simultaneous capture of the narrow edge details (low-level features) and semantic category information (high-level features) of fatigue cracks in weld-free joints. It effectively improves the semantic consistency of fatigue crack segmentation under complex backgrounds (bolt holes, assembly gaps, rust traces) and enhances the model's ability to distinguish between "cracks and background interference". Meanwhile, addressing the problem that fatigue cracks in weld-free joints (especially early narrow ones) have a low pixel ratio and are prone to losing key details after downsampling, a High-Low Frequency Decomposition Block (HLFDB) is added after the downsampling process

of the MiT encoder. By separating and preserving the high-frequency detail information of cracks, it solves the problem of information loss of the narrow structure of fatigue cracks during downsampling.

The SegFormer-HF model processes the multi-scale features extracted by the MiT encoder through the SegFormerHead decoder. However, there are significant differences between low-level and high-level features in the weld-free joint scenario: low-level features can capture detailed spatial information such as the narrow edges and linear structures of fatigue cracks, but are mixed with noise like bolt shadows and rust textures; high-level features possess the ability to recognize abstract semantics of fatigue cracks, yet lack the spatial details of narrow cracks. Due to the imbalance between the two in terms of semantic levels and spatial resolution, simple feature fusion is prone to causing deterioration of segmentation performance-either low-level noise interferes with crack recognition accuracy, or insufficient high-level semantics fail to provide effective guidance for narrow cracks. This is also the core reason why targeted optimization of feature fusion logic is required for fatigue crack segmentation in weld-free joints.

To address the feature fusion imbalance problem in fatigue crack segmentation of weld-free joints of circular steel tube trusses—where low-level features are mixed with noise such as bolt shadows and rust textures, and high-level features lack the spatial details of narrow cracks. The SegFormer-HF model introduces a Features Fusion Module (FFM), whose structure is shown in Fig. 6. Here, Ci and Ti denote the input features of the CNN branch (capturing narrow edges of fatigue cracks) and the Transformer branch (extracting semantic categories of cracks), respectively, while C'I and T'I represent the processed features. This module enhances the ability of low-level features to distinguish between "fatigue cracks and background interference" by injecting the semantic information of high-level features into low-level features. The fused low-level features not only retain the linear spatial details of narrow cracks but also possess clear semantic guidance, which facilitates the semantic alignment of low-level and high-level features in the decoder. This improves the semantic consistency of fatigue crack segmentation under complex backgrounds and ultimately achieves more accurate extraction of crack contours.

In the FFM module, linear interpolation is first performed on high-level features to match their spatial dimensions with low-level features. This is to align the linear structures of narrow fatigue cracks in weld-free joints. Then, 1×1 convolution is used for dimensionality reduction of high-level features to reduce their channel proportion, preventing high-level semantic information from obscuring the key details of low-level narrow cracks. Next, the dimension-reduced high-level features are concatenated with low-level features, fusing the spatial details of narrow cracks (low-level) and the semantic category information of cracks (high-level). Finally, 3×3 convolution further enhances the expressive capability of the fused features, fully combining the CNN's advantage in capturing local details and the Transformer's advantage in understanding global semantics to improve the accuracy of fatigue crack segmentation for weld-free joints.

To address the problem that early narrow fatigue cracks (high-frequency details) of weld-free joints are prone to being lost during the downsampling process of the MiT encoder, the SegFormer-HF model introduces a High-Low Frequency Decomposition Block (HLFDB). Based on the spatial-frequency feature decoupling principle (as shown in Fig. 7), this module achieves detail recovery by separating and preserving the high-frequency details of cracks and the low-frequency information of the background: First, it extracts low-frequency components from the features via average pooling (Avgpool), which correspond to background information such as bolts and rust of weld-free joints. Then, it separates high-frequency components (corresponding to the narrow edge details of fatigue cracks) through residual calculation between the original features and the upsampled low-frequency components. The specific calculations are shown in Eqs. (5)-(7):

$$F_{\text{low}} = Avgpool(F) \tag{5}$$

$$F' = Upsampling(F_{\text{low}}) \tag{6}$$

$$F_{\text{high}} = F - F' \tag{7}$$

Where Avgpool (·) denotes average pooling, and Upsampling (·) denotes bilinear interpolation upsampling. F_{low} represents the separated low-frequency information, and F_{high} represents the separated high-frequency information [25].

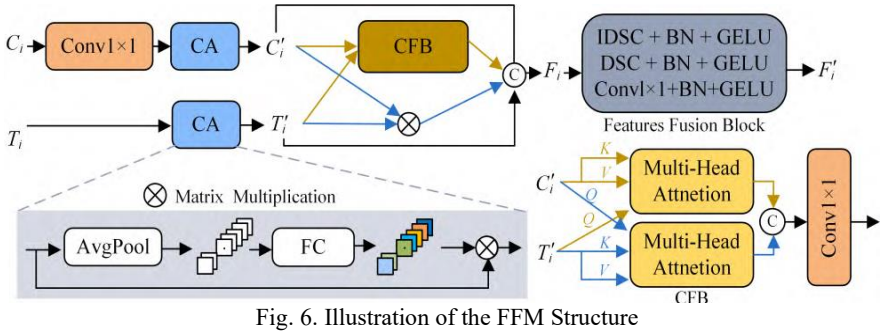


Fig. 6. Illustration of the FFM Structure

Targeting the frequency-domain characteristic differences of fatigue cracks in weld-free joints of circular steel tube trusses (where high-frequency components correspond to the edge details of narrow cracks, and low-frequency components correspond to background interference such as bolt shadows and rust textures), the HLFDB module employs dual-branch targeted processing paths: The high-frequency branch adopts a Local Feature Extraction Block. By stacking 3×3 small convolutional kernels and a dense connection mechanism, it accurately enhances the modeling of local edges and linear structures of narrow fatigue cracks in weld-free joints. This avoids large receptive fields confusing background interference with crack details, ensuring that the high-frequency features of narrow cracks (especially early cracks with a low pixel ratio) are not obscured. The low-frequency branch efficiently captures the global structural dependencies of weld-free joints through three-level downsampling operations and Channel-wise Self-Attention. Meanwhile, it introduces a multi-scale fusion module based on Discrete Wavelet Transform (DWT), which decomposes low-frequency features into {LL, LH, HL, HH} subbands. By fusing low-scale features with the global structural information of the LL subband, and then reconstructing background structural information through Inverse Discrete Wavelet Transform (IDWT), it effectively suppresses the loss of key details of narrow fatigue cracks during downsampling. Finally, the outputs of the two branches are integrated through a cross-scale feature fusion strategy, achieving the synergistic optimization of high-frequency crack details and low-frequency background information. This not only restores the narrow structural information of fatigue cracks lost during downsampling but also enhances the anti-interference capability of features against the complex background of weld-free joints, significantly improving the robustness and accuracy of model segmentation.

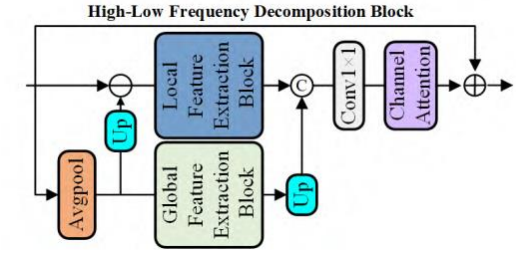


Fig. 7. Illustration of the HLFDB Structure

3. EXPERIMENTS ON STRUCTURAL SURFACE CRACK DETECTION

3.1. Selection of Fatigue Crack Data

This study constructs the initial dataset based on fatigue crack images of weld-free joints of circular steel tube trusses field-collected by the research team, combined with adapted and modified open-source crack datasets ("Crack-seg [26]" and "Crack500 [27]" created by the Roboflow team).

The images collected by the research team are from weld-free joint areas of circular steel tube trusses in engineering scenarios such as bridges and large-span workshops. These images feature unique complex background interference inherent to weld-free joints (e.g., bolt hole shadows, tube truss assembly gaps, surface rust traces). The fatigue cracks exhibit typical characteristics including narrowness, concealment, directional propagation along stress directions, and low pixel ratio in the early stage, posing significant detection challenges. The open-source datasets Crack-seg and Crack500 are screened and modified: samples that do not match the characteristics of fatigue cracks in weld-free joints (such as non-steel structure backgrounds and wide cracks) are eliminated, while crack samples with narrow morphology and low pixel ratio are retained. The screened samples from Crack500 are dominated by narrow linear cracks with relatively simple backgrounds. The screened samples from Crack-seg cover linear crack morphologies similar to those of weld-free joints, with partial samples suffering from blur or occlusion, which can enhance the model's adaptability to complex working conditions. The two modified open-source datasets are highly similar to weld-free joint fatigue cracks in terms of narrow morphology and linear structural characteristics, which can help improve the model's generalization ability.

To improve data quality and adaptability to weld-free joint scenarios, we filtered out images with non-steel structure backgrounds and unclear crack features in accordance with unified standards. We also adopted data augmentation techniques such as flipping, cropping, and adding rust-simulated noise to simulate the states of weld-free joints under different operation and maintenance conditions, thereby expanding sample diversity. For the two tasks of object detection and semantic segmentation, X-annotabeling software and EISeg software were used respectively to perform unified annotation of fatigue cracks in weld-free joints. Finally, the crack-detect detection dataset (6,070 images) and the crack-mask segmentation dataset (7,460 images) were constructed. The datasets were divided into training set, validation set, and test set in an 8:1:1 ratio. The proportion of each dataset is shown in Table 1, and the original image examples are presented in Fig. 8.

Table 1. The percentage distribution of datasets

Dataset	Self-built	Crack500	Crack-seg
crack-detect	72%	10%	18%
crack-mask	59%	16%	25%



Fig. 8. Some typical crack images

3.2. Experimental Environment and Hyperparameter Settings

The experimental environment supports the training and validation of the fatigue crack detection models for weld-free joints of circular steel tube trusses (including the YOLO-CD object detection model and SegFormer-HF semantic segmentation model). It adopts the Ubuntu 22.04 server operating system, with the development environment built on Python 3.10.8, PyTorch 2.4.1, and CUDA 12.1. For computing resources, the setup includes an NVIDIA GeForce RTX 4090D graphics card (24GB video memory) and a CPU configuration of 18 vCPUs (AMD EPYC 9754 128-Core Processor). The training process relies on the Graphics Processing Unit (GPU) to achieve efficient computing.

For the detection characteristics of fatigue cracks in weld-free joints-*narrowness, concealment, and complex backgrounds*-targeted hyperparameter settings are adopted for model training: the object detection (YOLO-CD) training is based on the Ultralytics 8.3.9 framework, with 800 training epochs (stopping if no improvement is achieved within 50 consecutive epochs), an initial learning rate of 0.01, a batch size of 16, an input image size of 640×640 pixels, and the SGD optimizer selected, while the semantic segmentation (SegFormer-HF) training uses 300 epochs, an initial learning rate of 0.0001, a batch size of 8, an input image size of 512×512 pixels, and the AdamW optimizer. Meanwhile, F1-score, mAP (mean Average Precision), and GFLOPs (Giga Floating-point Operations Per Second) are employed as the core evaluation metrics, with the specific calculation formulas shown in Eqs. (8)-(12).

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{8}$$

$$\mathrm{AP} = \int_0^1 P\,(R)\,dR \tag{9}$$

$$\mathrm{mAP} = \frac{1}{N}\sum A\,P \tag{10}$$

$$\begin{cases} \mathrm{FLOPs}_{\mathrm{depth}} = H_{\mathrm{in}} \cdot W_{\mathrm{in}} \cdot K_h \cdot K_w \cdot C_{\mathrm{in}} \\ \mathrm{FLOPs}_{\mathrm{point}} = H_{\mathrm{out}} \cdot W_{\mathrm{out}} \cdot C_{\mathrm{in}} \cdot C_{\mathrm{out}} \\ \mathrm{FLOPs}_{\mathrm{total}} = \mathrm{FLOPs}_{\mathrm{depth}} + \mathrm{FLOPs}_{\mathrm{point}} \end{cases} \tag{11}$$

$$\mathrm{GFLOPs} = \frac{\sum \mathrm{FLOPs}}{10^9} \tag{12}$$

In Eqs. (8)-(12), P denotes Precision, and R denotes Recall; N is the number of defect categories, and AP represents the area enclosed in the PR (Precision-Recall) curve; Hin is the input height of the convolutional layer, Hout is the output height of the convolutional layer, Win is the input width of the convolutional layer, Wout is the output width of the convolutional layer, Kh is the kernel height, Kw is the kernel width, Cin is the number of input channels, and Cout is the number of output channels.

For the semantic segmentation task, in addition to the F1-score, the mean Intersection over Union (mIoU) and mean Pixel Accuracy (mPA) are further introduced as key evaluation metrics. mIoU measures the pixel-level matching accuracy by calculating the ratio of the intersection to the union between predicted segmentation regions and ground truth annotations, the closer its value is to 1, the more consistent the segmentation result is with the ground truth. mPA, on the other hand, statistics the proportion of correctly classified pixels across the entire image, reflecting the global segmentation accuracy. The specific calculations are shown in Eqs. (13)-(14).

$$\mathrm{mIou} = \frac{1}{k+1}\sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{13}$$

$$\mathrm{mPA} = \frac{1}{N}\sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{14}$$

In Eqs. (13)-(14), pii denotes the number of pixels correctly predicted as class i from class i; pij and pji represent the number of pixels incorrectly predicted as class j from class i and as class i from class j, respectively; and k is the total number of classes. N denotes the total number of classes (note: k and N are consistently defined as the total number of classes); TPi is the True Positive (TP) count for class i, FPi is the False Positive (FP) count for class i, and FNi is the False Negative (FN) count for class i (correcting a potential typo in the original text: FNi corresponds to False Negative, not False Positive, as per standard evaluation metrics in computer vision).

3.3. Test Results and Analysis of the YOLO-CD Model

This study takes YOLO11-n as the base model and conducts ablation experiments on the StarNet, HSANet, and OSCD modules respectively. The performance of different modules is shown in Table 2, with F1-score, mAP50, mAP50-95, and GFLOPs used as evaluation metrics.

Table 2. Results of ablation experiments

Model	F1	mAP50	mAP 50-95	GFLOPs
YOLO11	0.644	0.686	0.446	6.3
Starnet	0.651	0.685	0.445	5.0
HSANet	0.660	0.687	0.444	5.3
OSCD	0.680	0.724	0.488	5.7
Starnet+HSANet	0.658	0.679	0.435	3.9
HSANet+OSCD	0.670	0.713	0.472	4.6
YOLO-CD	0.678	0.715	0.464	3.3

The data in Table 2 shows that when the backbone network and neck network constructed with the StarNet module and HSANet module are used to replace those of the base model, compared with the base model, the computational complexity of the model decreases by 20.63% and 15.87% respectively, the F1-score increases by 1.087% and 2.484%, and the mAP remains basically consistent. When the OSCD module is used alone, compared with the base model, there is a further improvement in accuracy and a reduction in computational complexity. When combining all three modules simultaneously, the F1-score, mAP50, and mAP50-95 values of the YOLO-CD model increase by 5.28%, 4.23%, and 4.04% compared with the base model, and the computational complexity is reduced by a factor of 1.909, significantly lowering the computational cost. These results indicate that the YOLO-CD model maintains lightweight properties while achieving high detection accuracy, and its overall performance is superior to that of the YOLO11 model, fully verifying the effectiveness of the model improvements. The detection results are shown in Fig. 9.



Fig. 9. Schematic Diagram of YOLO-CD Detection Results

To further verify the performance advantages of the YOLO-CD object detection model proposed in this study, mainstream YOLO models at the current stage are selected as comparison models, including YOLOv5, YOLOv8, and YOLO11. These models represent the development levels of different versions in the YOLO series, covering a variety of structures ranging from medium-sized to large-scale models. By comparing with the aforementioned YOLO models, the comprehensive performance of the YOLO-CD model in terms of detection accuracy, computational efficiency, and real-time performance can be evaluated more comprehensively. The detection comparison results between the YOLO-CD model and these mainstream YOLO models are shown in Table 3, detailing the differences between each model in terms of various performance metrics.

Table 3. Comparative experimental results

Model	F1	mAP50	mAP50-95	GFLOPs
YOLO11	0.644	0.686	0.446	6.3
YOLOv5	0.652	0.661	0.401	4.1
YOLOv8	0.638	0.667	0.434	8.1
YOLO-CD	0.678	0.715	0.464	3.3

The data in Table 3 shows that the YOLO-CD model achieves the highest F1-score, mAP50, and mAP50-95, reaching 67.8%, 71.5%, and 46.4% respectively: 4.00%, 8.17%, and 15.71% higher than those of YOLOv5, 6.26%, 7.19%, and 6.91% higher than those of YOLOv8, and 5.28%, 4.23%, and 4.04% higher than those of YOLO11. In terms of computational complexity and model lightweighting, the YOLO-CD model also demonstrates significant advantages, with reductions of 19.51%, 59.26%, and 47.62% compared to YOLOv5, YOLOv8, and YOLO11 respectively. In summary, the analysis indicates that the YOLO-CD model outperforms YOLOv5, YOLOv8, and YOLO11 in both accuracy and lightweighting.

The proposed YOLO-CD model achieved an average inference speed of approximately 78 FPS on an NVIDIA RTX 4090D GPU, demonstrating its suitability for real-time mobile inspection applications. In addition, the reduced GFLOPs and lightweight architecture facilitate deployment on edge devices such as drones and inspection robots.

3.4. Test Results and Analysis of the SegFormer-HF Model

This study takes SegFormer as the base model and conducts ablation experiments on the FFM and HLFDB modules respectively. The performance of different modules is shown in Table 4, with mIoU (mean Intersection over Union), mPA (mean Pixel Accuracy), and F1-score used as evaluation metrics.

Table. 4 Results of ablation experiments

Model	mIoU	mPA	F1
SegFormer_B0	84.08%	89.52%	90.49%
FFM	84.80%	89.99%	91.20%
HLFDB	84.61%	89.81%	91.12%
SegFormer-HF	85.15%	90.51%	91.50%

The data in Table 4 shows that when incorporating the FFM module and HLFDB module respectively, compared with the base model, the mIoU, mPA instead of mAP (mean Average Precision)) and F1-score increase by 0.86%, 0.53%, 0.78% and 0.45%, 0.32%, 0.69% respectively. When both modules are introduced simultaneously, the mIoU, mPA, and F1-score of the SegFormer-HF model are 1.27%, 1.10%, and 1.12% higher than those of the base model. These results indicate that the SegFormer-HF model achieves high segmentation accuracy, and its overall performance is superior to that of the SegFormer_B0 model, fully verifying the effectiveness of the model improvements. The segmentation results are shown in Fig. 10.

To further verify the performance advantages of the improved SegFormer semantic segmentation model, this study selects mainstream semantic segmentation models at the current stage for comparison, including Deeplabv3+ [28], Unet [29], and PSPnet [30], as shown in Table 5.

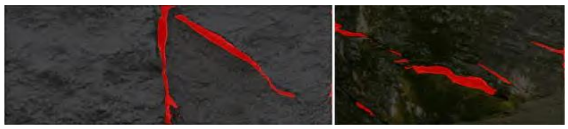


Fig. 10. Schematic Diagram of SegFormer-HF Detection Results

Table 5. Comparative experimental results

Model	mIoU	mPA	F1
Segformer_B0	84.08%	89.52%	90.49%
DeepLabv3+	83.52%	88.75%	90.05%
Unet	83.78%	88.86%	90.45%
PSPnet	83.93%	89.24%	89.85%
Segformer-HF	85.15%	90.51%	91.50%

The data in Table 5 shows that the SegFormer-HF model achieves the highest mIoU, mPA, and F1-score, reaching 85.15%, 90.51%, and 91.5% respectively. It is 1.95%, 1.98%, and 1.61% higher than those of Deeplabv3+, 1.64%, 1.86%, and 1.16% higher than those of Unet, and 1.46%, 1.45%, and 1.84% higher than those of PSPnet. In summary, the results indicate that the improved SegFormer-HF model outperforms PSPnet, Deeplabv3+, Unet, and the original SegFormer model.

3.4.1. Comparison with Existing Studies

Compared with existing crack detection studies, the proposed method demonstrates improved adaptability to narrow fatigue cracks and complex structural backgrounds. Traditional CNN-based approaches such as U-Net and DeepLabV3+ mainly focus on general surface crack segmentation and often suffer from information loss when detecting narrow crack regions under interference conditions. Transformer-based methods improve global semantic representation but usually require higher computational resources.

In contrast, the proposed YOLO-CD and SegFormer-HF framework combines lightweight design, multi-scale feature fusion, and high-frequency detail preservation mechanisms to achieve both computational efficiency and detection robustness. Experimental results indicate that the proposed method achieves higher F1-score and mIoU while maintaining lower computational complexity, making it more suitable for real-time mobile inspection applications in weld-free joint scenarios.

3.5. Model Generalization Experiments

To further verify the model generalization ability, cross-scenario tests were conducted on the CrackDetect and DeepCrack datasets using the improved YOLO-CD object detection model and SegFormer-HF semantic segmentation model respectively. The results are shown in Table 6. The above results show that without retraining, although the test results of each model on the CrackDetect and DeepCrack datasets are slightly lower than their performance on the original training set, they outperform or are comparable to the unimproved original model. This further verifies the robustness and generalization ability of the proposed models across different data scenarios.

4. CALCULATION OF FATIGUE CRACK LENGTH AND WIDTH

The images used in this experiment were selected from the crack-mask weld-free joint fatigue crack dataset constructed in Section 3.1. These images are consistent with engineering inspection scenarios, featuring complex backgrounds such as bolt hole shadows, tube truss assembly gaps, and surface rust traces. The fatigue cracks exhibit typical characteristics of narrowness, concealment, and directional extension along stress directions, making them suitable as test objects for geometric parameter calculation. Image acquisition was performed using high-definition cameras mounted on drones (or handheld inspection cameras), which are commonly used in engineering inspections. The shooting distance for local areas of weld-free joints is approximately 0.5-1m to meet the requirements of node detail collection. After processing, the image resolution is 512×512, ensuring the detail clarity of the narrow edges of fatigue cracks and providing support for the accurate measurement of geometric features.

Table 6. Generalization Experiment Results of YOLO-CD & SegFormer-HF in Weld-Free Joint Fatigue Crack Detection of Circular Steel Tube Trusses

Dataset	YOLO-CD (Object Detection)				SegFormer-HF (Semantic Segmentation)		
	F1 Score	mAP50	mAP50-95	GFLOPs (Lightweight Metric)	mIoU (Segmentation Accuracy)	mPA (Pixel Accuracy)	F1 Score (Segmentation Balance)
Weld-Free Joint Fatigue Crack Dataset (CrackDetect)	0.661	0.705	0.455	3.3	84.22%	89.78%	90.67%
General Steel Structure Crack Dataset (DeepCrack)	0.654	0.696	0.448	3.3	84.03%	89.35%	90.12%

After the identification and segmentation of weld-free joint fatigue crack images by the YOLO-CD and SegFormer-HF models, it is necessary to extract geometric features through targeted digital image processing and then calculate the length and width, considering the narrow morphology of the cracks and complex background interference. The specific steps are as follows:

(1) Fatigue crack region localization: Use the YOLO-CD algorithm to locate fatigue crack regions in weld-free joints, preliminarily filtering out background interference such as bolts and rust;

(2) Refined segmentation of crack regions: Input the localized regions into the SegFormer-HF model to complete the contour segmentation of fatigue cracks and retain the details of narrow edges;

(3) Targeted binarization and denoising: Combined with the background characteristics of weld-free joints, an adaptive thresholding method and non-local means filtering are adopted to achieve binarization. Meanwhile, noise caused by rust textures and bolt shadows is suppressed to simplify subsequent calculations;

(4) Fatigue crack skeleton extraction: Capture the "skeleton points" of each segment of the crack through a skeletonization algorithm and connect them into a continuous skeleton line. This line conforms to the directional propagation morphology of weld-free joint fatigue cracks, providing a basis for length calculation (as shown in Fig. 11).

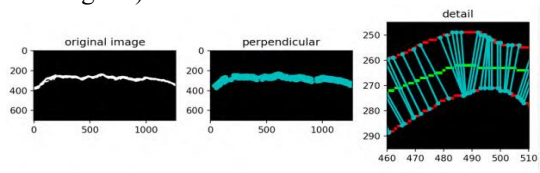


Fig. 11. Schematic diagram of crack feature calculation based on crack skeleton line

(5) Crack length calculation: The crack length is obtained by traversing each point on the skeleton line, calculating the Euclidean distance between adjacent points, and then summing these distances to get the total crack length L. The specific calculation formula is as follows:

$$L = \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (15)$$

Where: n is the number of points on the skeleton line; (xi,yi) and (xi+1,yi+1) are the coordinates of adjacent points respectively.

(6) Average crack width calculation: The average crack width W is obtained by averaging the widths calculated at all skeleton points:

$$W = \frac{1}{n}\sum_{i=1}^{n}(w_{i1} - w_{i2}) \quad (16)$$

Where: n is the number of points on the skeleton line; *wi*1 and *wi*2 are the upper width and lower width of the crack at the i-th skeleton point, respectively.

(7) Calculation of the scale factor μ: To achieve physical size conversion, this study calculates the scale factor μ (unit: mm/pixel) based on the lens imaging principle. The calculation process is as follows:

First, according to the pinhole imaging principle, the object distance, focal length, and image distance satisfy the equal relationship shown in Equation (17):

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \quad (17)$$

Through identity transformation, the image distance v can be obtained as:

$$v = \frac{fu}{u-f} \quad (18)$$

Where: f is the camera focal length; u is the distance from the camera to the crack surface, i.e., the object distance; v is the image distance. Then, according to the mathematical relationship of the camera imaging model, it can be obtained that:

$$L_1 = \frac{Mu}{v} \quad (19)$$

Where: L1 is the actual crack size; M is the imaging size. Substituting Equation (18) into Equation (19) yields:

$$L_1 = \frac{u-f}{f}M \quad (20)$$

Thus, the scale factor μ corresponding to a single pixel is:

$$\mu = \frac{(u-f)\times s}{S\times f} \quad (21)$$

Where: s is the physical size of the photosensitive element; S is the number of pixels.

(8) Calculation of actual crack length and width:

$$L_{\text{real}} = L \times \mu, \quad W_{\text{real}} = W \times \mu \quad (22)$$

The calculation results of crack length and width are shown in Table 7.

Table 7. Crack Length and Width Calculation Results

Serial No.	Length Measurement					Width Measurement				
	Pixel Length/px	Actual Length/mm	Manual Measurement/mm	Absolute Error/mm	Relative Error/%	Pixel Width/px	Actual Width/mm	Manual Measurement/mm	Absolute Error/mm	Relative Error/%
①	1892.73	327.82	317.50	10.32	3.25	129.10	22.36	21.50	0.86	4.00
②	4322.11	748.59	766.75	18.16	2.37	14.03	2.43	2.25	0.18	8.00
③	3484.76	603.56	612.80	9.24	1.51	155.49	26.93	26.35	0.58	2.20
④	3658.84	633.71	653.40	19.69	3.01	82.91	14.36	13.75	0.21	4.44

Note: The scale factor is 0.1732

During the experiment, we conducted targeted tests on the influence of fatigue cracks of different sizes in weld-free joints of circular steel tube trusses on calculation errors. This test aligns with the core engineering requirements of "early warning of narrow initial cracks and crack monitoring during the propagation stage." The experimental results show that as the size of fatigue cracks in weld-free joints decreases (corresponding to the initial cracks that require key early warning in engineering), the measured relative error gradually increases; in particular, when the width or length of the fatigue crack is less than 2mm (a typical size of initial fatigue cracks in weld-free joints), the relative error will exceed 10%; while when the crack size is larger than this threshold (entering the propagation stage), the error is basically controlled within 10%. This indicates that the proposed method has good error stability for the measurement of fatigue cracks in weld-free joints during the propagation stage.

In addition, the pixel precision of fatigue cracks in weld-free joints in images is strongly correlated with the shooting distance, image resolution, and focal length settings in engineering inspection scenarios. All images used in the calculations of this study were captured at a fixed distance (approximately 1m) suitable for local inspection of weld-free joints, with a uniform resolution of 512×512. These parameters not only match the on-site collection conditions of drones and handheld inspection cameras but also ensure that the edge details of narrow fatigue cracks are clearly distinguishable. Under this premise, when the length and width of fatigue cracks in weld-free joints reach a sufficient pixel ratio in the image, error fluctuations can be effectively reduced, achieving stable engineering-level measurement.

4.1 Implementation Challenges and Solutions

During the implementation of the proposed fatigue crack detection framework, several practical challenges were encountered. First, the narrow morphology and low pixel ratio of early-stage fatigue cracks caused feature information loss during downsampling operations. To address this issue, the HLFDB module was introduced to preserve high-frequency crack edge information and enhance narrow crack representation.

Second, complex background interference such as bolt shadows, assembly gaps, and rust textures frequently caused false detections during object localization and semantic segmentation. To improve robustness, the HSANet module and channel attention mechanisms were adopted to suppress irrelevant background features and strengthen crack-related semantic information.

Third, deployment on mobile inspection devices required balancing detection accuracy and computational complexity. Therefore, the lightweight StarNet backbone and shared convolution strategy in the OSCD head were employed to reduce GFLOPs and improve inference efficiency.

Finally, fatigue crack sizes and propagation patterns varied significantly under different stress conditions. Multi-scale feature fusion and adaptive dynamic scaling strategies were integrated to improve the model’s adaptability to cracks at different propagation stages.

5. CONCLUSIONS

This study constructed an integrated fatigue crack detection algorithm for weld-free joints of circular steel tube trusses based on the YOLO-CD and SegFormer-HF models, covering core links such as image preprocessing, localization and recognition, segmentation refinement, geometric parameter calculation, and visualization. Combined with a skeleton line algorithm adapted to narrow cracks, it achieves non-contact, automated, and engineering-level high-precision detection. Ablation and comparison experiments verify the effectiveness of the Starnet, HSANet, and OSCD modules in YOLO-CD, as well as the FFM and HLFDB modules in SegFormer-HF. The results show that YOLO-CD outperforms YOLOv5/8/11 in terms of precision and lightweight, with 3.3 GFLOPs making it suitable for mobile inspection devices; SegFormer-HF surpasses mainstream models such as Deeplabv3+ and UNet in indicators like mIoU and mPA, exhibiting strong robustness and high accuracy in crack recognition under complex backgrounds such as bolt shadows and rust, and achieving efficient detection with a small number of parameters. This study focuses on fatigue crack detection of weld-free joints. In the future, steel structure damage datasets under different working conditions can be incorporated to extend to apparent damage recognition of multiple types of components; additionally, lidar and 3D point cloud modeling can be combined to realize 3D visualization of cracks and propagation trend tracking, providing support for steel structure operation and maintenance decisions. Although the proposed method achieved strong detection performance, the current study mainly focuses on two-dimensional crack image analysis under controlled acquisition conditions. Performance under extreme illumination variations and highly occluded structural environments still requires further investigation.

Conflicts of Interest Declaration

Funding: *2023 Guangxi Vocational & Technical College Research Project《Fatigue Performance Study of Overlapped CHS (Circular Hollow Section) Truss Joints with Hidden Weld Left Unwelded》 project number：20230211119.*

Competing Interests: *The authors have declared that no competing interests exist*

Data availability Statement: *The data used to support the findings of this study are available from the corresponding author upon request.*

References

1. Han XJ, Zhao ZC. Structural surface crack detection method based on computer vision technology. Journal of Building Structures. 2018;39(S1):418-427.
2. Yuan Q, Shi Y, Li M. A review of computer vision-based crack detection methods in civil infrastructure: Progress and challenges. Remote Sensing. 2024; 16(16):2910. https://doi.org/10.3390/rs16162910.
3. Hussain H, Kim DK. Advancements and future prospects of buckling restrained braces for corrosive-environments: A comprehensive literature review. Buildings. 2023;13(9):2156. https://doi.org/10.3390/buildings13092156.
4. Tang Qian-long, Tan Yuan, Peng Li-min, et al. On crack identification method for tunnel linings based on digital image technology. Journal of Railway Science and Engineering. 2019,16(12):3041-3049.
5. Liu YF, Fan JS, Nie JG, et al. Review and prospect of digital-image-based crack detection of structure surface. China Civil Engineering Journal. 2021;54(6): 79-98.
6. Vanian V, Zamanakos G, Pratikakis I. Improving performance of deep learning models for 3D point cloud semantic segmentation via attention mechanisms. Computers & Graphics. 2022;106:277-287. https://doi.org/10.1016/j.cag.2022.06.010.
7. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence. 2016;39(6):1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031.
8. Zhou Zhong, YAN Longbin, Zhang Junjie, et al. Realtime detection of tunnel cracks based on YOLOX-G algorithm. Journal of Railway Science and Engineering. 2023, 20(7): 2751-2762.
9. Chen LC, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV). 2018:801-818. https://doi.org/10.1007/978-3-030-01234-2_49.
10. WANG Qingyuan, XU Ying, QIAN Sheng. Study on Damage Evolution of Concrete Specimens Based on Machine Vision and Digital Image Correlation Technology. Journal of Hunan University. 2023, 50(11):169-180.
11. Protopapadakis E, Voulodimos A, Doulamis A, Doulamis N, Stathaki T. Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing. Applied intelligence. 2019; 49(7):2793-2806. https://doi.org/10.1007/s10489-018-01396-y.
12. Nishiyama S, Minakata N, Kikuchi T, et al. Improved digital photogrammetry technique for crack monitoring. Advanced Engineering Informatics. 2015;29(4):851-858. https://doi.org/10.1016/j.aei.2015.05.005.
13. Li Z, Zhang T, Miao Y, et al. Automated quantification of crack length and width in asphalt pavements]. Computer - Aided Civil and Infrastructure Engineering. 2024;39(21):3317-3336. https://doi.org/10.1111/mice.13344.
14. ZHOU Ying, LIU Tong. Computer Vision-Based Crack Detection and Measurement on Concrete Structure. Journal of Tongji University. 2019;47(9): 1277-1285.http://dx.doi.org/10.11908/j.issn.0253-374x.2019.09.007.
15. Tian Y, Wang Y. Crack detection method of highway side slope based on computer vision. In 2021 International Conference on Aviation Safety and Information Technology. 2021;742-746. https://doi.org/10.1145/3510858.3511375.
16. WANG Weidong, WU Zheng, QIU Shi, et al. Multi-objective cascade deep learning based crack measurement for ballastless track slab. Journal of Railway Science and Engineering. 2023;20(9):3592-3603.
17. Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2024;2410.17725. https://doi.org/10.48550/arXiv.2410.17725.
18. Varghese R, Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In 2024 International conference on advances in data engineering and intelligent computing systems (ADICS). 2024:1-6. https://doi.org/10.1109/ADICS58448.2024.10533619
19. Child R, Gray S, Radford A, et al. Generating long sequences with sparse transformers. arXiv preprint arXiv:2019;1904.10509. https://doi.org/10.48550/arXiv.1904.10509.
20. Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems. 2021;34:12077-12090.
21. Ma X, Dai X, Bai Y, Wang Y, Fu Y. Rewrite the stars. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024:5694-5703. https://doi.org/10.1109/CVPR52733.2024.00544.
22. Tian Z, Shen C, Chen H, He T. FCOS: A simple and strong anchor-free object detector. IEEE transactions on pattern analysis and machine intelligence. 2020;44(4):1922-1933. https://doi.org/10.1109/TPAMI.2020.3032166.
23. Zhang J, Zeng Z, Sharma PK, et al. A dual encoder crack segmentation network with Haar wavelet-based high–low frequency attention. Expert Systems with Applications. 2024;256:124950. https://doi.org/10.1016/j.eswa.2024.124950.
24. Chen G, Dai K, Yang K, et al. Bracketing image restoration and enhancement with high-low frequency decomposition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024:6097-6107. https://doi.org/10.1109/CVPRW63382.2024.00616.



Wang XINYI, Associate Professor, Postdoctoral Fellow, Ph.D. Main research interests include: static and fatigue performance of steel structures, durability of steel-concrete composite structures, seismic behavior of prefabricated underground structures, intelligent inspection and reinforcement of structures.
e-mail: Sddzwxy1977@163.com