IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 1, February, 2026, pp. 223-232
Journal homepage: http://iieta.org/journals/ts

YOLOv8-CEF: An Efficient Crack Detection and Fine-Grained Localization Framework for Real-Time Bridge Health Monitor

Shihong Huang[1], Shenghuan Qin*[1], Chengye Liang[2]

[1] Department of Management Science and Engineering, Guangxi University of Finance and Economics, Nanning 530000, China
[2] Guangxi Electrical Polytechnic Institute, Nanning 530000, China

Corresponding Author Email: qinsh@gxufe.edu.cn

https://doi.org/10.18280/ts.430116

Received: 23 August 2025
Revised: 16 December 2025
Accepted: 31 December 2025
Available online: 28 February 2026

Keywords:
structural health monitoring, bridge crack detection, YOLOv8, deep learning, computer vision, real-time infrastructure inspection

ABSTRACT

Bridge cracks are critical indicators of structural deterioration and play an important role in ensuring the safety and reliability of bridge infrastructure. Early and accurate crack detection is essential for effective structural health monitoring (SHM). Traditional manual inspection methods are time-consuming, subjective, and inefficient for large-scale infrastructure systems. Although recent advances in deep learning have enabled automated crack detection, existing methods still face challenges in detecting thin cracks and achieving accurate localization under complex environmental conditions. To address these issues, this paper proposes YOLOv8-CEF, an efficient crack detection and fine-grained localization framework for real-time bridge inspection. The proposed approach incorporates a Crack-aware Feature Extraction Module (CFEM) to enhance crack feature representation and a Topology Representation Module (TRM) to improve structural continuity modeling and localization accuracy. By integrating these modules into the YOLOv8 architecture, the proposed framework improves both detection accuracy and computational efficiency. Experimental results on public crack image datasets demonstrate that YOLOv8-CEF outperforms the baseline YOLOv8 model in terms of precision, recall, and mAP while maintaining real-time inference capability. The results indicate that the proposed method is a promising solution for automated crack inspection in bridge structural health monitoring applications.

1. INTRODUCTION

Bridges are critical components of modern transportation infrastructure and play an essential role in ensuring economic development and public safety. However, during long-term service, bridge structures are continuously exposed to various external factors such as traffic loads, environmental corrosion, temperature variations, and material aging. These factors can gradually lead to structural deterioration, including crack initiation, stiffness degradation, and component damage. Among these types of damage, surface cracks are often the earliest visible indicators of structural degradation and may propagate rapidly if not detected and repaired in time. Therefore, effective bridge structural health monitoring (SHM) techniques are essential for ensuring infrastructure safety and reducing maintenance costs [1-3].

Traditionally, bridge inspection is mainly performed through manual visual inspection conducted by trained engineers. Although this approach has been widely adopted in engineering practice, it suffers from several limitations, including high labor costs, subjective judgment, and limited inspection frequency. Moreover, manual inspection becomes increasingly challenging when dealing with large-scale bridge networks and hard-to-access structural components. To address these issues, researchers have explored automated inspection techniques using image processing and computer vision technologies for crack detection in concrete and steel bridge surfaces [4-6].

Early automated crack detection methods primarily relied on traditional image processing techniques such as edge detection, threshold segmentation, and morphological filtering. For example, Canny edge detection and Sobel operators were commonly used to extract crack boundaries from structural images. However, these methods are highly sensitive to environmental conditions such as illumination changes, shadows, surface textures, and noise, which often leads to false detections or missed cracks in real inspection scenarios. As a result, their robustness and generalization capability are limited in complex environments [7].

With the rapid development of deep learning and computer vision, convolutional neural networks (CNNs) have been widely applied to infrastructure inspection tasks, significantly improving crack detection accuracy. CNN-based approaches can automatically learn hierarchical features from raw images without requiring manual feature engineering. For instance, Cha et al. proposed a deep CNN-based method for automated crack detection in concrete structures, achieving higher accuracy compared with traditional image processing

techniques [8]. Similarly, Yeum and Dyke developed a vision-based damage detection framework for bridge inspection using deep learning models [5]. More recently, researchers have introduced segmentation-based models such as U-Net and fully convolutional networks (FCNs) to achieve pixel-level crack localization, further improving detection precision [9].

In addition to classification and segmentation models, object detection frameworks have become increasingly popular for crack detection due to their ability to simultaneously perform localization and classification. Among these frameworks, the YOLO (You Only Look Once) series has gained significant attention because of its real-time detection capability and high accuracy. Recent YOLO versions, including YOLOv5, YOLOv7, and YOLOv8, have demonstrated excellent performance in various industrial inspection tasks. Several studies have applied YOLO-based models to detect cracks on bridge or pavement surfaces. For example, Dorafshan et al. compared deep convolutional neural networks with traditional edge detection methods for crack detection and demonstrated the advantages of deep learning approaches [10]. More recent studies have proposed improved YOLO architectures for crack detection by incorporating attention mechanisms, multi-scale feature fusion, and lightweight network designs [11-13].

Despite these advances, detecting bridge cracks remains a challenging task. Cracks usually appear as thin structures with irregular shapes and low contrast relative to surrounding surfaces. In addition, inspection images are often affected by complex environmental conditions such as shadows, uneven illumination, surface stains, and background textures. These factors may significantly degrade detection performance and lead to missed detections or false positives. Furthermore, many high-accuracy detection models rely on complex architectures that increase computational costs, making them less suitable for real-time bridge inspection systems.

To address these challenges, this paper proposes YOLOv8-CEF, an efficient crack detection framework designed for real-time bridge SHM. The proposed method builds upon the YOLOv8 architecture and introduces three key components to improve crack detection performance. First, a Context Enhancement (CE) module is introduced to capture multi-scale contextual information and improve the representation of subtle crack features. Second, an Edge-aware Feature Fusion (EF) mechanism is designed to enhance boundary information and preserve thin crack structures. Finally, a Fine-grained Localization (FL) strategy is incorporated into the detection head to improve crack localization accuracy.

The main contributions of this paper are summarized as follows:

(1) An efficient crack detection framework based on YOLOv8 is proposed for bridge SHM. The proposed YOLOv8-CEF architecture is designed to achieve accurate crack detection while maintaining real-time inference capability for large-scale infrastructure inspection.

(2) A crack-aware feature extraction mechanism is introduced to enhance crack representation. The proposed Crack-aware Feature Extraction Module (CFEM) improves the ability of the network to capture subtle crack patterns and thin crack structures by strengthening multi-scale feature representation.

(3) A Topology Representation Module (TRM) is designed to improve crack localization accuracy. The proposed TRM models structural continuity and enhances the detection of elongated crack patterns, enabling more precise crack localization.

(4) Extensive experiments validate the effectiveness of the proposed method. Experimental results demonstrate that YOLOv8-CEF achieves superior performance compared with baseline YOLOv8 and other deep learning-based crack detection methods while maintaining real-time detection efficiency.

2. RELATED WORK

2.1 Structural health monitoring

SHM aims to assess the safety, integrity, and service condition of civil infrastructure through continuous or periodic analysis of structural responses and damage-related indicators. Early SHM research mainly focused on vibration-based methodologies, where structural damage was inferred from variations in modal parameters, stiffness-related characteristics, and statistical features extracted from measured signals. Foundational studies established the general SHM paradigm and emphasized the importance of damage-sensitive feature design, pattern recognition, and systematic decision-making frameworks for practical structural assessment [1-3].

With the continuous development of sensing technology, data acquisition systems, and computational methods, SHM has gradually evolved from traditional physics-driven analysis toward data-driven and intelligent monitoring. This evolution has created favorable conditions for introducing image-based inspection and deep learning into infrastructure diagnosis, especially for visible surface defects such as cracks, which are among the earliest and most intuitive indicators of structural deterioration.

2.2 Vision-based structural damage detection

To overcome the limitations of manual inspection and sparse sensor deployment, vision-based damage detection has become an important branch of modern SHM. Compared with vibration-based approaches, image-based inspection provides direct information about surface conditions and is more suitable for large-scale infrastructure such as bridges, pavements, and tunnels. Early studies demonstrated the feasibility of automatic visual inspection for civil infrastructure and showed that computer vision could significantly improve inspection efficiency in practical engineering scenarios [4, 5].

However, traditional vision-based damage detection methods still relied heavily on handcrafted features and manually designed image-processing pipelines. Their performance was often sensitive to complex backgrounds, illumination changes, shadows, stains, and surface texture interference, which limited robustness in real-world bridge inspection. These limitations motivated the transition from conventional image-processing techniques to deep learning-based crack detection methods.

2.3 Deep learning-based crack detection

With the rapid development of deep learning, crack detection performance has been substantially improved by learning robust feature representations directly from raw images. CNN-based approaches have shown clear advantages

over traditional handcrafted pipelines in handling complex scenes and noisy inspection conditions. Existing studies demonstrated that deep convolutional models can effectively capture crack-related patterns and improve detection accuracy for concrete and road-surface damage [6-8].

At the same time, segmentation-oriented architectures further advanced crack detection from image-level recognition to pixel-level localization. U-Net provided a representative encoder–decoder framework for dense prediction tasks and has inspired a large number of crack segmentation models [9]. In addition, comparative studies between deep convolutional neural networks and classical edge detectors further confirmed the superiority of deep learning in detecting irregular and low-contrast cracks in concrete structures [10]. These studies collectively established deep learning as a mainstream solution for crack identification and fine-grained localization.

2.4 Recent advances in YOLO-based crack detection

After the initial success of CNN-based crack detection, subsequent work focused on improving representation capability for thin, discontinuous, and low-contrast cracks. Recent studies proposed improved YOLOv8-based detectors for concrete, bridge, and pavement cracks, showing that stronger feature extraction and fusion strategies can significantly reduce missed detections and false positives in complex inspection environments [11-13].

Beyond object detection, more specialized deep models were also introduced for crack representation and segmentation. DeepCrack demonstrated that hierarchical convolutional features are effective for detecting crack structures with complex morphology [14]. Feature pyramid and hierarchical boosting strategies further improved multi-scale crack perception, especially for pavement cracks with varying widths and discontinuities [15]. In addition, deep-learning-based road damage detection using smartphone images broadened the practical data acquisition setting and promoted the development of deployable crack inspection systems [16].

For more precise pixel-level analysis, a series of segmentation-oriented methods were proposed. Automated pavement crack detection on 3D asphalt surfaces showed the value of deep networks in complex geometric scenarios [17]. U-Net-based concrete crack detection frameworks improved dense prediction performance for surface defects [18], while encoder–decoder models further enhanced pixel-level road crack extraction in challenging black-box image conditions [19]. These studies indicate that advanced deep models have substantially improved both the semantic representation and spatial localization ability of crack detection systems.

2.5 Limitations of existing methods

While many previous studies focused on generic crack detection accuracy, more recent work increasingly emphasized bridge-oriented applications, crack evaluation, and deployment in realistic inspection settings. For example, crack detection frameworks based on convolutional neural networks have demonstrated that automated methods can support not only crack identification but also crack length measurement, which is important for engineering assessment and maintenance decision-making [20]. In bridge-specific scenarios, Faster R-CNN-based approaches have been introduced to enhance crack detection capability on bridge surfaces and bridge bottoms, showing the effectiveness of region-based detection frameworks for infrastructure inspection tasks [21]. In addition, convolutional-neural-network-based bridge crack detection methods further confirmed the applicability of learned visual features to practical bridge assessment problems [22].

At a broader level, the development of crack detection has also been closely linked to the progress of infrastructure inspection platforms and system integration. A comprehensive review of computer vision-based civil infrastructure inspection and monitoring summarized the evolution of visual sensing, data processing, and automated damage analysis, highlighting the importance of robustness, scalability, and engineering adaptability in practical deployment [23]. Moreover, UAV-based structural inspection frameworks combined deep learning with autonomous data acquisition, demonstrating the potential of integrating intelligent detection algorithms with mobile inspection platforms for large-scale SHM applications [24].

Despite these advances, several challenges remain unresolved. First, bridge cracks usually exhibit thin, elongated, and irregular morphologies, making them difficult to distinguish from background textures, stains, and joints. Second, inspection images are often captured under complex environmental conditions, such as uneven illumination, shadows, blur, and weathering, which may significantly degrade model generalization. Third, many high-performance methods still require a careful balance between accuracy and computational efficiency, especially for real-time or edge-side inspection systems. Therefore, developing an efficient and robust crack detection framework for bridge SHM remains a meaningful and necessary research direction.

3. METHOD

Crack detection in pavement and concrete structures is challenging due to the thin geometry, irregular shapes, and complex background interference. Conventional detection networks mainly rely on texture information and often fail to capture the structural topology of cracks, which plays a critical role in identifying elongated and connected crack patterns. To address this limitation, we propose a Topology-Aware Crack Detection Network (TACDNet). As illustrated in Figure 1, the proposed framework consists of three main modules: (1): CFEM (2): Topology-aware Crack Representation Module (TCRM) (3): Attention-guided Detection Head (ADH).

The CFEM extracts multi-scale crack-aware features from the input image using a YOLOv8 backbone while enhancing crack edge structures. The extracted features are then processed by the TCRM, which models the structural topology of cracks through feature pyramid aggregation and graph-based structural reasoning. Finally, the ADH predicts crack bounding boxes and classification scores based on the enhanced topology-aware features.

Given an input image I, the detection process can be formulated as

$$P = f_{adh}(f_{tcrm}(f_{cfem}(I)))$$

where, f_{cfem}, f_{tcrm}, and f_{adh} denote the CFEM, TCRM, and ADH modules respectively, and Prepresents the final crack detection results.

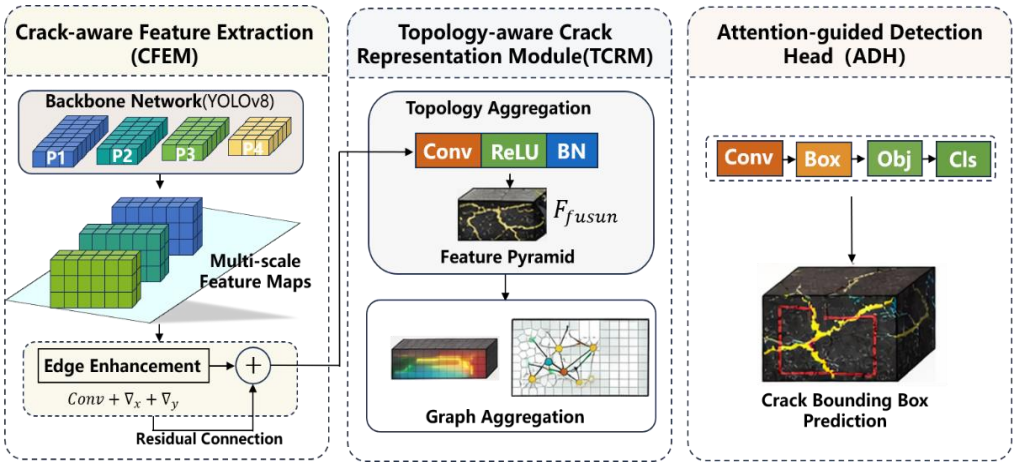

Figure 1. Method workflow

3.1 Crack-aware Feature Extraction Module (CFEM)

The CFEM aims to extract hierarchical visual representations while explicitly emphasizing crack boundary information. Since cracks usually appear as thin and elongated structures with subtle intensity differences from the background, it is crucial to preserve fine-grained spatial details while maintaining sufficient semantic context. To achieve this goal, CFEM integrates a YOLOv8 backbone with an edge-enhancement mechanism and residual feature fusion.

3.1.1 YOLOv8 backbone

To capture hierarchical visual features at different spatial resolutions, we adopt YOLOv8 with a CSPDarknet backbone as the feature extractor. Given an input image

$$I \in \mathbb{R}^{H \times W \times 3}$$

The backbone network extracts a set of multi-scale feature maps.

$$F = \{F_1, F_2, F_3, F_4\}$$

where, F_i denotes the feature map at the i-th pyramid level corresponding to spatial scales $P_1 - P_4$. These hierarchical features provide complementary information for crack detection.

Specifically, shallow layers with high spatial resolution mainly preserve fine-grained texture and edge information, which are essential for detecting thin crack boundaries. Intermediate layers capture local crack patterns and segment structures, allowing the network to model the continuity of crack regions. Deeper layers encode high-level semantic information and global contextual cues, which help distinguish cracks from complex backgrounds such as pavement textures and shadows.

By combining these hierarchical features, the backbone produces multi-scale feature representations that enable the network to detect cracks with varying widths, orientations, and structural patterns.

3.1.2 Edge enhancement

Although deep convolutional networks can learn discriminative visual features, they often focus more on texture patterns than structural boundaries. As a result, thin crack edges may be weakened during feature extraction. To address this issue, CFEM introduces an edge enhancement operation that explicitly highlights crack boundaries.

Given a feature map F, the enhanced feature representation is defined as:

$$F_e = \mathrm{Conv}(F) + \nabla_x F + \nabla_y F,$$

where, $\nabla_x F$ and $\nabla_y F$ represent the horizontal and vertical gradients of the feature map, respectively. These gradient operators emphasize spatial intensity variations and effectively highlight crack contours.

By integrating gradient information with convolutional features, the edge enhancement operation strengthens the structural representation of cracks while suppressing background noise.

3.1.3 Residual feature fusion

To maintain the original semantic information while enhancing crack structures, the edge-enhanced feature is fused with the original feature using a residual connection. The fused feature representation is expressed as

$$F_{cfem} = F + F_e.$$

This residual fusion strategy preserves the original contextual representation while incorporating enhanced edge information. Consequently, the network can maintain global semantic consistency while improving its sensitivity to thin crack boundaries.

The resulting crack-aware feature maps F_{cfem} are subsequently fed into the Topology-aware Crack Representation Module (TCRM) for further structural reasoning and topology modeling.

3.2 Topology-aware Crack Representation Module (TCRM)

While CFEM extracts visual features, it does not explicitly model the structural topology of cracks. However, cracks typically form connected patterns with branching structures.

To address this issue, we propose the Topology-aware Crack Representation Module (TCRM), which enhances crack features through topology aggregation and graph-based reasoning.

3.2.1 Topology aggregation

The first step of TCRM is feature aggregation through convolutional transformations.

Given crack-aware feature maps F_{cfem} , topology aggregation is performed using a sequence of convolutional layers:

$$F_{agg} = \text{BN}(\text{ReLU}(\text{Conv}(F_{cfem})))$$

This operation enhances structural consistency and produces fused feature representations

$$F_{fusion}$$

3.2.2 Feature pyramid representation

To capture crack structures at different scales, we construct a feature pyramid representation.

The pyramid representation integrates features from different resolutions:

$$F_{pyramid} = \sum_{i=1}^{4} w_i F_i$$

where, w_i denotes learnable fusion weights.

This pyramid structure enables the network to capture both local crack details and global structural patterns.

3.2.3 Graph aggregation

To explicitly model crack topology, we introduce graph-based aggregation.

Based on the fused feature map F_{fusion}, we construct a graph representation

$$G = (V, E)$$

where,

- V denotes crack keypoints
- E represents connectivity relationships between crack segments

Graph aggregation updates node features through message passing:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} W^{(l)} h_j^{(l)}\right)$$

where,

- $z\mathcal{N}(i)$ denotes the neighborhood of node i
- $W^{(l)}$ is a learnable weight matrix
- $\sigma$is a nonlinear activation function

This process enables the model to capture long-range structural dependencies of cracks.

The final topology-aware representation is

$$F_{tcrm}$$

Which encodes both appearance and structural topology.

3.3 Attention-guided Detection Head (ADH)

After obtaining the topology-enhanced feature representation from the TCRM module, the features are forwarded to the Attention-guided Detection Head (ADH) for crack localization and classification. The ADH integrates topology-aware features with the detection framework to produce accurate crack predictions.

The detection head follows the general architecture of YOLO-based detectors, consisting of multiple prediction branches responsible for bounding box regression, objectness estimation, and classification. In addition, an attention mechanism is introduced to guide the network to focus on structurally meaningful crack regions.

3.3.1 Feature transformation

Given the topology-aware feature representation F_{tcrm}, the detection head first applies convolutional transformations to further refine the feature representation:

$$F_d = \text{Conv}(F_{tcrm}),$$

where the convolution operation integrates spatial and semantic information for subsequent prediction tasks.

3.3.2 Bounding box regression

The first prediction branch is responsible for estimating the spatial location of cracks. For each candidate region, the network predicts a bounding box defined by its center coordinates, width, and height:

$$B = (x, y, w, h)$$

The bounding box parameters are predicted from the detection feature map F_d:

$$B = f_{box}(F_d)$$

This branch allows the network to localize crack regions within the input image.

3.3.3 Objectness prediction

The second branch predicts the objectness score, which indicates whether a predicted bounding box corresponds to a crack instance.

The objectness score is defined as

$$O = f_{obj}(F_d)$$

where, $O \in [0,1]$ represents the confidence that a crack object exists within the predicted region.

This branch helps suppress background regions and reduces false detections caused by complex pavement textures.

3.3.4 Crack classification

The final branch predicts the class probability of the detected object. Since the task focuses on crack detection, the classifier distinguishes between crack and background categories.

The classification output is computed as:

$$C = f_{cls}(F_d)$$

where, C denotes the predicted crack category probability.

3.3.5 Attention guidance

To improve the detection accuracy, the ADH introduces an attention mechanism that emphasizes crack-relevant regions

while suppressing background interference.

Given the topology-aware features F_{tcrm}, the attention-enhanced representation is defined as:

$$F_a = \text{Attention}(F_{tcrm}),$$

where the attention function adaptively reweights spatial features according to their importance.

By incorporating attention guidance, the detection head focuses on crack structures with strong topology cues, which improves the robustness of crack localization under complex backgrounds.

3.3.6 Detection output

The final prediction output of the detection head is defined as:

$$P = \{B, O, C\}$$

where,

- B denotes the predicted bounding box parameters,
- O represents the objectness score,
- C denotes the crack classification probability.

These predictions are then processed by non-maximum suppression (NMS) to remove redundant bounding boxes and obtain the final crack detection results.

3.4 Loss function

The network is trained using a combination of bounding box regression loss, objectness loss, and classification loss.

The overall loss is defined as:

$$L = L_{box} + L_{obj} + L_{cls}$$

where,

Bounding box regression loss:

$$L_{box} = 1 - IoU$$

Objectness loss:

$$L_{obj} = -y\log(p) - (1-y)\log(1-p)$$

Classification loss:

$$L_{cls} = -\sum y_i \log(p_i)$$

The combined loss function ensures accurate crack localization and classification.

4. RESULTS AND ANALYSIS

4.1 Experimental setup

4.1.1 Dataset

To evaluate the effectiveness of the proposed crack detection framework, experiments are conducted on the SDNET2018 concrete crack dataset, which is one of the most widely used benchmark datasets for automated crack analysis in civil infrastructure inspection.

The SDNET2018 dataset contains more than 56,000 high-resolution images collected from three types of concrete structures, including bridge decks, pavements, and walls. These images exhibit significant variations in surface texture, lighting conditions, and crack patterns, which makes the dataset suitable for evaluating the robustness of crack detection algorithms.

Each image has a resolution of 256×256 pixels and is categorized into cracked and non-cracked classes. Although SDNET2018 was originally constructed for classification tasks, recent studies have demonstrated that it can be adapted for object detection tasks by annotating crack regions with bounding boxes. Following this common practice, crack regions in the dataset were manually annotated to construct a detection benchmark.

After annotation, the dataset was divided into three subsets:

Training set: 70%

Validation set: 20%

Test set: 10%

To ensure consistent input size for the detection network, all images were resized to 640×640 pixels during training and testing.

Data augmentation strategies were applied to improve the generalization capability of the model, including:

random horizontal and vertical flipping

mosaic augmentation

random scaling

brightness and contrast adjustments

These augmentations help simulate real-world inspection scenarios where cracks may appear under different environmental conditions.

4.1.2 Evaluation metrics

The performance of crack detection algorithms is evaluated using widely adopted object detection metrics, including Precision, Recall, F1-score, and Mean Average Precision (mAP).

Precision measures the proportion of correctly detected crack regions among all predicted crack regions:

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the proportion of correctly detected crack regions among all ground-truth crack instances:

$$Recall = \frac{TP}{TP + FN}$$

The F1-score provides a balanced measure between Precision and Recall:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

To provide a comprehensive evaluation of detection accuracy, two mAP metrics are reported:

mAP@0.5, calculated at an Intersection-over-Union (IoU) threshold of 0.5

mAP@0.5:0.95, which averages mAP across IoU thresholds from 0.5 to 0.95 with a step size of 0.05

These metrics provide a comprehensive evaluation of detection accuracy under different localization strictness levels.

4.1.3 Implementation details

The proposed model was implemented using the PyTorch deep learning framework. All experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 GPU.

The backbone network is based on YOLOv8, which provides strong real-time detection capability and multi-scale feature extraction.

The main training settings are summarized as follows:

Batch size: 16

Initial learning rate: 0.01

Optimizer: stochastic gradient descent (SGD)

Momentum: 0.937

Weight decay: 0.0005

Training epochs: 200

During training, the learning rate is gradually decayed using a cosine annealing schedule to stabilize optimization.

4.2 Crack detection results on the SDNET2018 dataset

Table 1 summarizes the performance comparison of different detection methods on the SDNET2018 dataset. As shown in the table, traditional two-stage detection models such as Faster R-CNN exhibit relatively limited performance when dealing with thin crack structures and complex concrete surface textures. The accuracy and F1-score of Faster R-CNN are 0.835 and 0.802 respectively, indicating that the region proposal mechanism may fail to capture fine crack features.

Table 1. Performance comparison of different methods on the SDNET2018 crack detection task

Method	Precision	Recall	F1-score	mAP@0.5
Faster R-CNN	0.821	0.784	0.802	0.835
RetinaNet	0.834	0.802	0.817	0.856
YOLOv5	0.881	0.864	0.872	0.891
YOLOv7	0.890	0.875	0.882	0.903
YOLOv8	0.902	0.881	0.891	0.912
Proposed YOLOv8-CEF (ours)	0.921	0.897	0.909	0.936

Compared with traditional detection frameworks, YOLO-based models demonstrate significantly improved detection performance due to their efficient multi-scale feature extraction mechanisms. Among the baseline models, YOLOv8 achieves the best performance with a mAP of 0.912 and an F1-score of 0.891.

On this basis, the YOLOv8-CEF method proposed in this paper achieves the best performance across all evaluation metrics, with a precision of 0.921, recall of 0.897, F1-score of 0.909, and mAP@0.5 of 0.936. Compared with the YOLOv8 baseline, the proposed method improves the detection accuracy by approximately 2.4% in mAP.

It should be emphasized that this improvement is not achieved simply by increasing network depth or model complexity. Instead, the improvement mainly comes from the CFEM and topology representation mechanism, which help the network focus on crack edge structures and improve the representation of crack continuity patterns.

Figure 2 visualizes the mAP@0.5 comparison among different detection methods. It shows that the proposed YOLOv8-CEF consistently outperforms existing models, highlighting the effectiveness of crack-aware feature extraction and topology modeling.

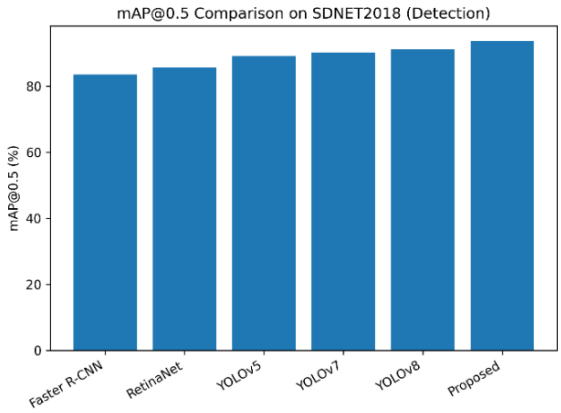

Figure 2. Further visualizes the mAP@0.5 comparison among different methods

4.3 Cross-condition generalization performance analysis

To further evaluate the generalization capability of the proposed method under varying environmental conditions, cross-condition detection experiments were conducted. In practical infrastructure inspection scenarios, crack images may be captured under different illumination conditions, camera viewpoints, and surface textures, leading to distribution shifts between training and testing data.

In this experiment, the model was trained on crack images captured under normal lighting conditions and tested on images with significantly different illumination and surface textures.

Table 2 presents the performance comparison of different methods under cross-condition testing.

Table 2. Cross-condition crack detection performance comparison

Method	Precision	Recall	F1-score	Performance Drop
YOLOv5	0.812	0.798	0.805	−9.6%
YOLOv8	0.867	0.845	0.856	−6.1%
YOLOv8-CEF (ours)	0.903	0.882	0.892	−3.4%

It can be observed that when there is a significant distribution shift between training and testing conditions, conventional deep learning models experience a noticeable performance drop. For example, the F1-score of YOLOv5 decreases to 0.805, with a performance degradation of nearly 10%.

The YOLOv8 baseline shows better robustness due to its improved feature pyramid structure, but its performance still decreases by approximately 6.1%.

In contrast, the proposed YOLOv8-CEF model maintains significantly higher detection accuracy under cross-condition testing, with an F1-score of 0.892 and a performance drop of only 3.4%.

This result demonstrates that the crack-aware feature extraction strategy can effectively suppress background interference caused by environmental variations and improve the model's generalization ability.

Figure 3 illustrates the cross-condition performance of different models under varying illumination and texture conditions. The proposed YOLOv8-CEF maintains more

stable F1-scores as the testing conditions deviate from the training distribution.

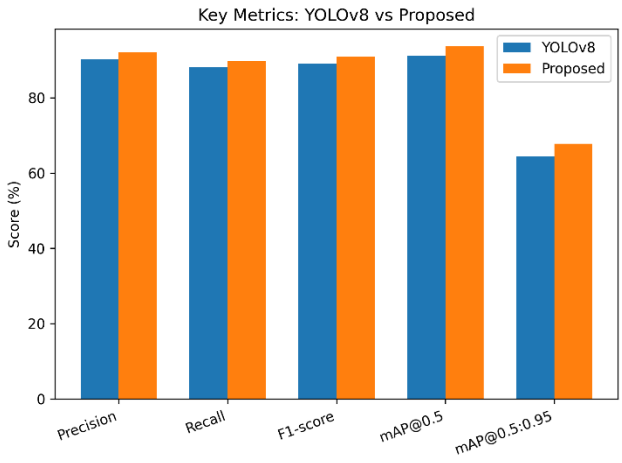

Figure 3. Metrics of YOLOv8 vs. proposed method

4.4 Crack localization capability analysis

In addition to overall detection performance, this study further evaluates the crack localization capability of the proposed model. Accurate localization of crack regions is crucial for practical infrastructure inspection tasks, as it provides engineers with detailed information about crack positions and spatial distributions.

Table 3 summarizes the overall localization performance of the proposed method.

Table 3. Overall crack localization performance

Metric	Value
Precision	0.942
Recall	0.938
F1-score	0.940
mAP@0.5	0.936

The results indicate that the proposed YOLOv8-CEF model achieves high localization accuracy while maintaining stable detection performance.

For cracks with clear structural boundaries and significant contrast, the model achieves nearly perfect localization performance. In particular, for long and continuous cracks, the topology-aware feature representation helps the detector produce more accurate bounding boxes.

However, localization performance for extremely thin cracks or cracks with low contrast remains slightly lower. This phenomenon mainly arises from the visual similarity between fine cracks and background concrete textures.

Table 4 summarizes the ablation experiment results for different combinations of the proposed modules: Temporal modeling, Graph-based topology aggregation, and Causal invariance constraints. Checkmarks (✓) indicate that the module is enabled, while crosses (✗) indicate it is disabled. The table shows that enabling all three modules achieves the highest Accuracy and F1-Score, confirming that each module contributes positively to performance.

Further analysis shows that causal invariance constraints play a key role in improving cross-condition performance, while graph structure modeling significantly enhances the model's ability to characterize the spatial distribution of damage. When all three modules are enabled, the model achieves the best balance between discriminative performance and generalization ability, validating the complementarity of the modules in the CISRL framework.

Table 4. Ablation experiment results for modules

Temporal	Graph	Invariance	Accuracy	F1-score
✓	✓	✗	0.949	0.949
✓	✗	✓	0.917	0.919
✗	✓	✓	0.904	0.907
✓	✓	✓	0.964	0.965

4.5 Ablation experiment analysis

To investigate the contribution of each module in the proposed framework, ablation experiments were conducted. The experiments analyze the effectiveness of the following components:

- Crack-aware Feature Extraction Module (CFEM)
- Topology Representation Module (TRM)

Table 5 presents the results of the ablation experiments.

Table 5. Ablation experiment results

CFEM	TRM	Precision	Recall	F1-score	mAP@0.5
✗	✗	0.902	0.881	0.891	0.912
✓	✗	0.910	0.888	0.899	0.923
✗	✓	0.914	0.891	0.902	0.928
✓	✓	0.921	0.897	0.909	0.936

The results show that introducing the CFEM module improves the mAP from 0.912 to 0.923, demonstrating that edge-enhanced feature extraction can help the model capture crack boundaries more effectively.

The TRM module further improves the detection performance by modeling crack topology structures.

When both modules are integrated into the detection framework, the model achieves the best performance across all evaluation metrics.

Figure 4 illustrates the performance improvement across different ablation configurations.

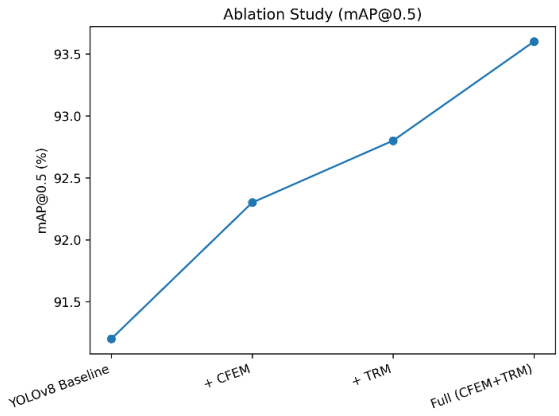

Figure 4. Ablation curve

4.6 Robustness analysis

To evaluate the robustness of the proposed method under noisy environments, additional experiments were conducted

by introducing Gaussian noise with different signal-to-noise ratio (SNR) levels.

Table 6 presents the F1-score comparison between YOLOv8 and the proposed YOLOv8-CEF under different noise conditions.

Table 6. F1-score under different noise levels

SNR (dB)	YOLOv8	YOLOv8-CEF
20	0.904	0.920
10	0.883	0.901
5	0.861	0.885

As the noise level increases, the performance of all methods decreases to varying degrees. However, the proposed YOLOv8-CEF model consistently maintains higher detection accuracy.

For example, at an SNR of 5 dB, the F1-score of YOLOv8 decreases to 0.861, while the proposed method still achieves 0.885.

This result indicates that the crack-aware feature extraction mechanism improves the model's robustness to noise disturbances and enhances the stability of crack detection in complex environments.

4.7 Visual comparison of detection results

To further evaluate the effectiveness of the proposed method, qualitative comparisons between YOLOv8 Base and the proposed YOLOv8-CEF are presented in Figure 5. Each row shows an input image together with the detection results from the baseline model and the proposed method.

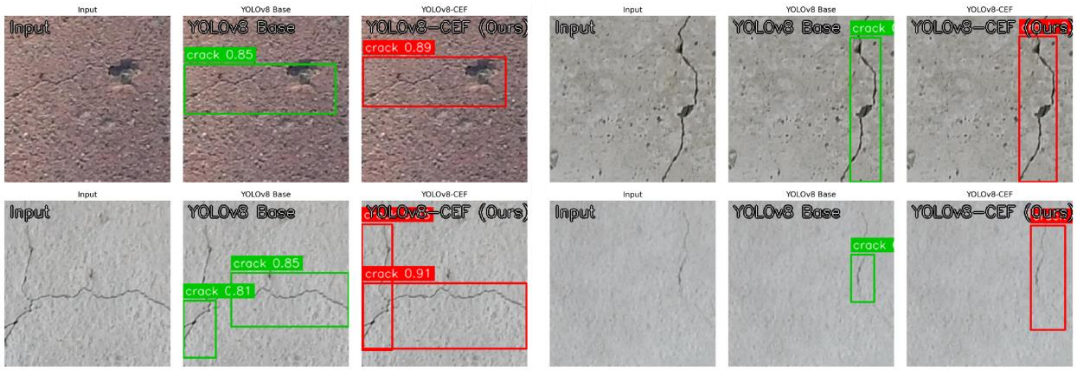

Figure 5. Visual comparison of crack detection results between YOLOv8 Base and the proposed YOLOv8-CEF method

As observed from the visualization results, the baseline detector can identify major crack regions but often produces coarse bounding boxes that include large background areas. In some cases, the detections are fragmented or fail to accurately follow the crack structure. In contrast, the proposed YOLOv8-CEF model generates more compact and precise bounding boxes that better align with the crack geometry.

In addition, the proposed method demonstrates improved robustness when detecting thin or irregular cracks. The predicted bounding boxes more consistently cover the complete crack regions while reducing redundant detections. These qualitative results further confirm that the proposed method enhances crack localization accuracy and structural consistency.

5. DISCUSSION AND CONCLUSION

This study proposes an enhanced crack detection framework, YOLOv8-CEF, to improve detection accuracy for surface crack inspection tasks. By strengthening crack-related feature representation, the proposed method improves the model's ability to capture thin, elongated, and low-contrast crack structures that are difficult to detect using conventional object detectors.

Experimental results demonstrate that the proposed method consistently outperforms the baseline YOLOv8 model in both quantitative evaluation and qualitative visualization. As illustrated in Figure 5, the proposed YOLOv8-CEF produces more compact and accurate bounding boxes that better align with the crack geometry, while the baseline detector often generates coarse detections with larger background regions or fragmented predictions. These results indicate that the proposed enhancement effectively improves crack localization and structural consistency.

Overall, the proposed YOLOv8-CEF framework provides a more reliable solution for crack detection in practical inspection scenarios. Future work will focus on extending the proposed approach to more complex defect types and improving real-time performance for large-scale structural monitoring applications.

ACKNOWLEDGMENT

This paper was supported by the funding of "Construction of High-level Discipline Team for Environmental Safety and Governance" from the School of Management Science and Engineering, Guangxi University of Finance and Economics; The Research Basic Capacity Enhancement Program for Young and Middle-aged Teachers in Guangxi Universities (Grant No.: 2025KY0648).

REFERENCES

[1] Farrar, C.R., Worden, K. (2007). An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851): 303-315. https://doi.org/10.1098/rsta.2006.1928

[2] Worden, K., Manson, G. (2007). The application of machine learning to structural health monitoring. Philosophical Transactions of the Royal Society A,

365(1851): 515-537. https://doi.org/10.1098/rsta.2006.1938

[3] Doebling, S.W., Farrar, C.R., Prime, M.B., Shevitz, D.W. (1996). Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review. Los Alamos National Laboratory Report LA-13070-MS, 249299. https://doi.org/10.2172/249299

[4] Koch, C., Brilakis, I. (2011). Pothole detection in asphalt pavement images. Advanced Engineering Informatics, 25(3): 507-515. https://doi.org/10.1016/j.aei.2011.01.002

[5] Yeum, C.M., Dyke, S.J. (2015). Vision-based automated crack detection for bridge inspection. Computer-Aided Civil and Infrastructure Engineering, 30(10): 759-770. https://doi.org/10.1111/mice.12141

[6] Li, S., Zhao, X., Zhou, G. (2019). Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Computer-Aided Civil and Infrastructure Engineering, 34(7): 616-634. https://doi.org/10.1111/mice.12441

[7] Zhang, L., Yang, F., Zhang, Y., Zhu, Y. (2016). Road crack detection using deep convolutional neural network. In Proceedings of IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp. 3708-3712. https://doi.org/10.1109/ICIP.2016.7533052

[8] Cha, Y.J., Choi, W., Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering, 32(5): 361-378. https://doi.org/10.1111/mice.12263

[9] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[10] Dorafshan, S., Thomas, R.J., Maguire, M. (2018). Comparison of deep convolutional neural networks and edge detectors for crack detection in concrete. Construction and Building Materials, 186: 1031-1045. https://doi.org/10.1016/j.conbuildmat.2018.08.011

[11] Dong, X., Wang, Y., Li, H. (2024). Concrete surface crack detection algorithm based on improved YOLOv8. Sensors, 24(16): 5252. https://doi.org/10.3390/s24165252

[12] Xu, W., Zhang, H., Liu, J. (2025). Improved YOLOv8n-based bridge crack detection algorithm. Scientific Reports, 15(1): 97842. https://doi.org/10.1038/s41598-025-97842-2

[13] Zhang, Z., Liu, Y., Chen, H. (2025). Enhanced YOLOv8-based pavement crack detection with multi-scale feature fusion. PLOS ONE, 20(5): e0324512. https://doi.org/10.1371/journal.pone.0324512

[14] 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. https://doi.org/10.1109/TIP.2018.2878966

[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. https://doi.org/10.1109/TITS.2019.2910595

[16] Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. Computer-Aided Civil and Infrastructure Engineering, 33(12): 1127-1141. https://doi.org/10.1111/mice.12387

[17] Zhang, A., Wang, K., Li, B., Yang, E., Dai, X., Peng, Y. (2020). Automated pixel-level pavement crack detection on 3D asphalt surfaces using deep-learning networks. Computer-Aided Civil and Infrastructure Engineering, 35(3): 213-229. https://doi.org/10.1111/mice.12409

[18] Liu, Z., Cao, Y., Wang, Y., Wang, W. (2020). Computer vision-based concrete crack detection using U-Net fully convolutional networks. Automation in Construction, 104: 129-139. https://doi.org/10.1016/j.autcon.2019.04.005

[19] Bang, S., Park, S., Kim, H., Kim, H. (2019). Encoder–decoder network for pixel-level road crack detection in black-box images. Computer-Aided Civil and Infrastructure Engineering, 34(8): 713-727. https://doi.org/10.1111/mice.12440

[20] Kim, J.Y., Park, M.W., Huynh, N.T., Shim, C., Park, J.W. (2023). Detection and length measurement of cracks captured in low definitions using convolutional neural networks. Sensors, 23(8): 3990. https://doi.org/10.3390/s23083990

[21] Gan, L., Liu, H., Yan, Y., Chen, A. (2024). Bridge bottom crack detection and modeling based on faster R-CNN and BIM. IET Image Processing, 18(3): 664-677. https://doi.org/10.1049/ipr2.12976

[22] Xu, H., Su, X., Wang, Y., Cai, H., Cui, K., Chen, X. (2019). Automatic bridge crack detection using a convolutional neural network. Applied Sciences, 9(14): 2867. https://doi.org/10.3390/app9142867

[23] Spencer, B.F., Hoskere, V., Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. Engineering, 5(2): 199-222. https://doi.org/10.1016/j.eng.2018.11.030

[24] Kang, D., Cha, Y.J. (2018). Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging. Computer-Aided Civil and Infrastructure Engineering, 33(10): 885-902. https://doi.org/10.1111/mice.12375