

Article

A Multi-Scale Attention-Enhanced YOLOv26 Framework for Steel Structure Corrosion Detection and Segmentation

Hongmei Hou ^{1,*}, Zhixin Wang ¹, Jianbo Zheng ¹, Jinzhen Xi ¹ and Libin Tian ²

¹ Intelligent Construction College, Shandong University of Aeronautics, Binzhou 256600, China

² School of Civil Engineering, Qingdao University of Technology, Qingdao 266520, China

* Correspondence: hhm@sdua.edu.cn

Abstract

Steel structures in complex service environments are highly susceptible to corrosion, making accurate detection challenging. This study proposes an improved YOLOv26-based method for corrosion damage segmentation. A diverse dataset is constructed by combining field-collected and public data with varying lighting conditions and multi-scale features. Enhancements to the YOLOv26-seg architecture include integrating Efficient Channel Attention (ECA) in the backbone to strengthen low-contrast feature representation, designing a multi-branch attention mechanism (ECA + CBAM) in the detection head to improve small- and medium-scale target recognition, and introducing Selective Kernel Attention (SKA) in the segmentation branch to refine boundary details. The resulting YOLOv26-ECS model achieves an mAP₅₀ of 0.920 and mAP_{50–95} of 0.851 on the self-constructed dataset, outperforming the baseline by 5.0% and 6.0%, respectively, while maintaining 28.34 FPS. Experiments on public datasets further demonstrate strong generalization. A GUI system is also developed for visualization and practical deployment. Overall, the proposed method delivers accurate and efficient corrosion detection and segmentation, showing strong potential for engineering applications.

Keywords: steel structures; corrosion damage; image segmentation; YOLOv26; attention mechanisms; deep learning

1. Introduction

Large-span steel structures [1], characterized by their light self-weight, long span capacity, and high construction efficiency, have been widely applied in bridge engineering, industrial plants, and large public buildings. However, during long-term service, steel structures are inevitably exposed to environmental factors such as humidity, temperature variations, and corrosive media, which can lead to varying degrees of corrosion damage. Corrosion [2] not only reduces the cross-sectional area of structural members and degrades connection performance, but may also induce local stress concentrations, thereby compromising the overall safety and durability of the structure [3]. Therefore, the development of efficient and accurate methods for corrosion detection and quantitative assessment [4] is of great significance for ensuring structural safety [5].

Traditional corrosion detection methods for steel structures mainly rely on manual inspection or conventional image processing techniques, such as threshold segmentation and edge detection. Although these approaches are simple to implement, they suffer from low efficiency, strong subjectivity, and poor adaptability to complex environments, making them unsuitable for large-scale and rapid inspections. In recent years, with the rapid



Academic Editor: Grzegorz
Ludwik Golewski

Received: 27 April 2026

Revised: 13 May 2026

Accepted: 20 May 2026

Published: 22 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

advancement of deep learning, convolutional neural network (CNN)-based methods [6] for object detection and image segmentation have achieved remarkable progress in defect identification. Among them, the YOLO (You Only Look Once) series [7], owing to its end-to-end training paradigm and excellent real-time performance, has shown great potential in engineering vision tasks. Wang et al. [8] incorporated a simple, parameter-free attention module (SimAM) into YOLOv11 and replaced the original upsampling module with a lightweight and efficient dynamic upsampler (DySample), enabling automatic emphasis on key image features and improved robustness against background shadows, thereby developing the SDS-YOLO algorithm for concrete vibration target recognition. Zhang et al. [9] proposed a YOLOv8-CM framework that integrates YOLOv8 with the Convolutional Block Attention Module (CBAM) to automatically detect and segment various defects and key objects in tunnels. Zhang et al. [10] developed an improved lightweight YOLOv8 network by incorporating Slim-neck and VanillaNet, significantly reducing model complexity; additionally, replacing the loss function with Wise-IoU enhanced feature extraction capability, demonstrating effectiveness in fiber bundle recognition. Du et al. [11] proposed the BLP-YOLOv10 model, which optimizes feature extraction and image processing by adjusting backbone channel parameters, integrating sparse attention mechanisms, and incorporating low-frequency enhancement filters for real-time safety helmet detection. Tu et al. [12] introduced a YOLOv10-UDFishNet model for diseased fish detection, which improves the backbone, neck, and detection head, and employs an innovative backbone with auxiliary branches to capture multi-level image information, thereby enhancing feature representation.

Despite these advances, directly applying existing YOLO models to steel corrosion detection and segmentation [4,5,13] still faces several challenges. On the one hand, corrosion regions often exhibit low contrast, blurred boundaries, and irregular shapes, making it difficult for models to learn stable and discriminative feature representations. On the other hand, corrosion [14–16] in real-world scenarios presents significant multi-scale characteristics, ranging from small pitting to large-area degradation, which complicates accurate detection and segmentation across different scales. Furthermore, background elements such as welds, bolts, and surface textures share visual similarities with corrosion features, increasing the risk of false positives and missed detections.

To address these challenges, this study adopts YOLOv26 [17] as the baseline model and optimizes the network architecture from the perspectives of feature representation and multi-scale modeling. Specifically, an Efficient Channel Attention (ECA) mechanism is introduced into the backbone network to enhance the representation of low-contrast corrosion features. In the detection head, a multi-branch structure combining ECA and the Convolutional Block Attention Module (CBAM) is designed to improve the recognition of small and medium-scale corrosion regions. Meanwhile, a Selective Kernel Attention (SKA) mechanism is incorporated into the segmentation branch to strengthen multi-scale feature fusion and refine boundary delineation. Based on these improvements, an enhanced YOLOv26-ECS model is developed to achieve high-precision detection and segmentation of steel corrosion damage.

To validate the effectiveness of the proposed method, extensive experiments are conducted on a self-constructed steel corrosion dataset as well as multiple public datasets. The model performance is comprehensively evaluated in terms of detection accuracy, segmentation quality, and inference speed. In addition, a visualization system based on PyQt5 is designed and implemented to demonstrate the feasibility and practicality of the proposed model in real-world engineering applications.

2. Overview of Corrosion Characteristics in Steel Structures

Corrosion [13] is a degradation process of steel caused by its interaction with the surrounding environment. According to its causes and mechanisms, corrosion in large-span steel structures can be broadly classified into chemical corrosion and electrochemical corrosion. Chemical corrosion refers to the direct surface reactions between steel and substances such as oxygen, carbon dioxide, sulfur-containing gases in the atmosphere or industrial emissions, as well as non-electrolytic liquids. In contrast, electrochemical corrosion occurs when impurities of different metals exist within the steel, creating differences in electrode potential; when exposed to electrolytes, water, or humid air, galvanic reactions are formed, leading to material degradation. In practical engineering applications, most corrosion processes involve either electrochemical corrosion alone or a combination of chemical and electrochemical corrosion. The corrosion characteristics of large-span steel structures are influenced by multiple factors, including environmental conditions, structural design, construction quality, and maintenance practices.

The color characteristics of corroded surfaces provide the most intuitive visual cues. Different metals and corrosion stages exhibit distinct color variations. Steel corrosion may appear as green, orange-red, or black thin layers, which are often difficult to distinguish from the background in image segmentation tasks. In addition, corrosion regions typically exhibit irregular boundaries, with no clear demarcation between corroded and non-corroded areas, as illustrated in Figure 1. The variability in size and the non-uniform spatial distribution of corrosion further increase the difficulty of accurate feature extraction for detection models.



Figure 1. Corrosion characteristics of steel structures.

3. Segmentation Network Architecture

3.1. Baseline Network

YOLOv26 is a single-stage object detection model proposed by Ultralytics in 2026, specifically optimized for practical deployment. Building upon the classical “Backbone–Neck–Head” architecture of the YOLO series, it achieves efficient detection through structural simplification and improved training strategies. The backbone adopts an enhanced lightweight feature extraction module integrated with attention mechanisms to improve representation capability. The neck employs an optimized multi-scale feature fusion structure to enhance detection performance across objects of different sizes. The head adopts an anchor-free, end-to-end design, eliminating the traditional non-maximum suppression (NMS) post-processing step, thereby reducing inference latency and improving deployment stability.

In addition, YOLOv26 removes the Distribution Focal Loss (DFL) to simplify the regression process, and introduces novel training strategies such as ProgLoss and STAL to accelerate convergence and improve small-object detection performance. The MuSGD optimizer is further employed to enhance training stability. The model supports unified multi-task learning, including detection, segmentation, and pose estimation, and demon-

states strong cross-platform deployment capability (e.g., ONNX and TensorRT). Overall, YOLOv26 significantly improves inference efficiency and engineering applicability while maintaining high detection accuracy, reflecting the transition of object detection models from accuracy-oriented to deployment-oriented design. The network architecture is illustrated in Figure 2.

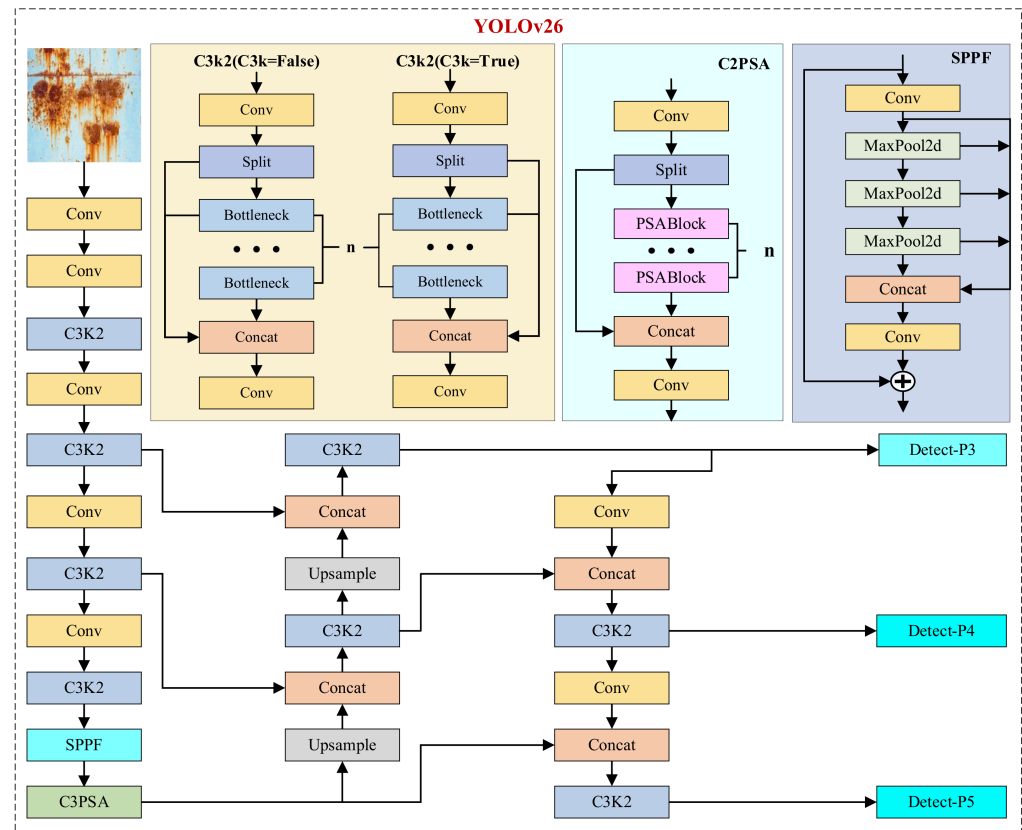


Figure 2. YOLOv26 network architecture.

3.2. Network Improvements

Steel corrosion damage segmentation is a core task in structural health monitoring and safety assessment, requiring both object detection (localizing corrosion regions) and pixel-level segmentation (precisely delineating corrosion areas). This task presents unique challenges: corrosion regions exhibit irregular shapes and uneven intensity distributions; multi-scale characteristics are prominent, ranging from tiny corrosion spots (small targets) to medium-scale patches and large-area corrosion (large targets); and interference from welds, bolts, and surface textures often leads to misclassification.

Although the original YOLOv26-seg model demonstrates strong multi-scale feature fusion capability, it still suffers from missed detections of small corrosion spots, localization deviations in medium-scale regions, and blurred segmentation boundaries. Moreover, the backbone network lacks sufficient focus on low-contrast and irregular corrosion features.

To address these issues, this study improves the YOLOv26-seg architecture from three perspectives: backbone feature enhancement, detection head attention reinforcement, and segmentation refinement. Specifically, three modules—ECA, CBAMHead, and SKA—are introduced to construct an improved YOLOv26-ECS model, which integrates multi-scale attention and refined segmentation mechanisms while maintaining model efficiency and lightweight characteristics.

3.2.1. Backbone Feature Enhancement

The original YOLOv26 backbone follows a “convolution–residual–feature fusion” paradigm, utilizing C3k2 modules, SPPF, and C2PSA attention modules for multi-scale feature extraction. However, corrosion features are typically low-contrast, blurred, and irregular, and repeated downsampling leads to the loss of fine-grained information, particularly for small corrosion spots. Meanwhile, background noise (e.g., welds and bolts) reduces feature discriminability.

To overcome these limitations, Efficient Channel Attention (ECA) modules are embedded into the high-level feature outputs (P4 and P5). ECA [18] is a lightweight channel attention mechanism that adaptively recalibrates channel weights via 1D convolution without introducing significant computational overhead. It enhances corrosion-related features while suppressing background interference.

The ECA module operates as follows:

- (1) global average pooling compresses spatial information into channel descriptors;
- (2) 1D convolution learns channel-wise dependencies with adaptive kernel size;
- (3) a sigmoid activation generates normalized channel weights, which are applied to the original feature map.

By embedding ECA into P4 (512 channels) and P5 (1024 channels), the model can better capture fine-grained corrosion features such as edges, textures, and intensity variations, mitigating information loss and improving downstream detection and segmentation accuracy (Figure 3).

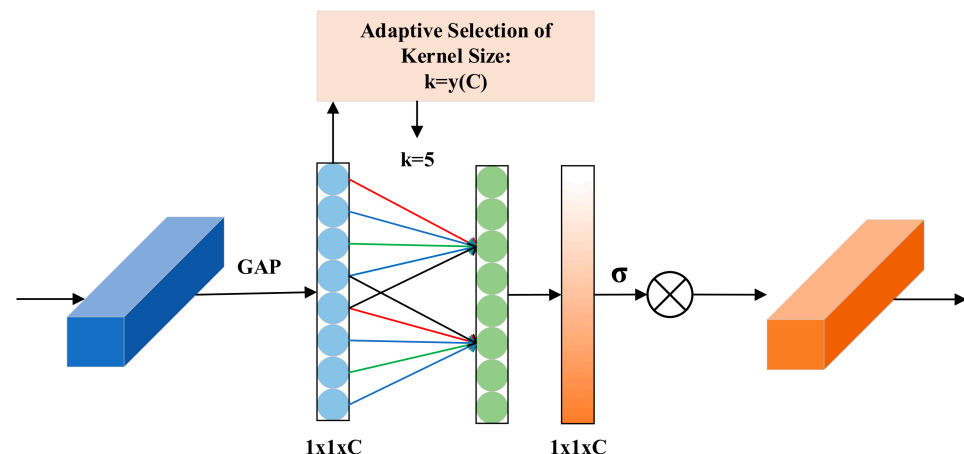


Figure 3. ECA network architecture.

3.2.2. Multi-Branch Attention Mechanism in Detection Head

The original YOLOv26-seg detection head adopts an FPN + PAN bidirectional feature fusion architecture, producing three feature maps at different scales: P3 ($8 \times$ downsampling), P4 ($16 \times$ downsampling), and P5 ($32 \times$ downsampling), corresponding to small-, medium-, and large-scale object detection and segmentation, respectively. In the task of steel corrosion damage segmentation, corrosion regions exhibit pronounced multi-scale characteristics, including tiny corrosion spots (small targets, e.g., point-like corrosion on steel surfaces), medium-scale corrosion patches (e.g., localized flake-like corrosion), and large-area corrosion regions (large targets, e.g., overall corrosion of structural components). Moreover, the features of corrosion regions vary significantly across scales.

However, the original detection head lacks scale-specific enhancement, leading to several limitations: small corrosion spots are easily missed due to weak feature responses; medium-scale corrosion regions are prone to confusion with interference such as welds;

and large-scale corrosion regions often suffer from blurred segmentation boundaries. These issues hinder the precise segmentation required for steel corrosion assessment.

To address these challenges, this study introduces differentiated enhancement modules into the three feature branches, constructing a multi-branch attention mechanism to strengthen corrosion feature representation across scales. The specific improvements are as follows:

- (1) P3 branch (small targets): Dual-attention enhancement with ECAHead and CBAMHead

The P3 feature map ($8 \times$ downsampling) is responsible for detecting and segmenting tiny corrosion spots. Although it has high spatial resolution and can capture fine details, it contains relatively weak semantic information. Due to the small pixel proportion and low contrast of corrosion spots, feature responses are often weak and susceptible to background noise (e.g., welds and bolts), resulting in missed detections and localization errors.

To mitigate this issue, ECAHead and CBAMHead modules are sequentially introduced after the P3 feature layer, forming a dual-attention mechanism that significantly enhances feature response strength and discriminability.

CBAMHead (Convolutional Block Attention Module Head) [19] integrates both channel and spatial attention. The channel attention submodule combines global average pooling and global max pooling to extract channel-wise global information and learns channel weights via fully connected layers, emphasizing features related to corrosion intensity and texture. The spatial attention submodule applies average and max pooling along the channel dimension to generate two spatial maps, which are concatenated and processed by a convolution layer to learn spatial attention weights. This enables precise localization of corrosion regions while suppressing background interference. Its architecture is shown in Figure 4.

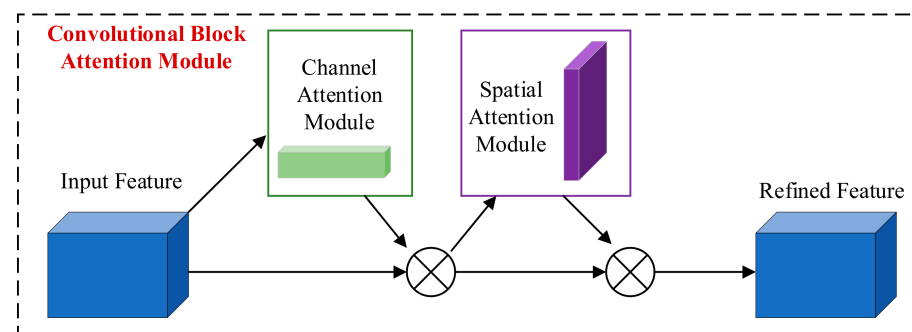


Figure 4. CBAMHead network architecture.

Within the P3 branch, the ECAHead module first enhances channel-wise feature representation, improving the contrast between corrosion and background. The CBAMHead module then refines spatial attention, enabling accurate localization and discrimination of corrosion from interference such as welds and bolts. The combination of these two mechanisms effectively improves the detection and localization of small corrosion spots, ensuring comprehensive identification of corrosion regions and providing reliable data for quantitative assessment.

- (2) P4 branch (medium targets): Lightweight enhancement with single ECAHead

The P4 feature map ($16 \times$ downsampling) is responsible for detecting and segmenting medium-scale corrosion regions, such as flake-like or strip-like corrosion. This layer balances spatial resolution and semantic richness, making it critical for segmentation performance. However, medium-scale corrosion features often resemble background

elements such as welds and oxide layers, resulting in insufficient discriminability and segmentation errors.

Considering the need for computational efficiency in real-world applications, a single ECAHead module is introduced after the P4 feature layer to achieve lightweight enhancement. This module adaptively optimizes channel features, emphasizing corrosion-related texture and intensity channels while suppressing interference from welds and oxide layers. Compared with the dual-attention design in the P3 branch, the single ECAHead module prioritizes efficiency and practicality, improving segmentation accuracy for medium-scale corrosion without significantly increasing computational cost.

(3) P5 branch (large targets): Feature enhancement with SKA

The P5 feature map ($32 \times$ downsampling) is mainly used to generate prototype features for large-area corrosion regions. Although it contains rich semantic information and captures overall structural contours, its low resolution and fixed receptive field limit its ability to model irregular boundaries, texture variations, and intensity gradients, resulting in coarse segmentation masks and loss of detail.

To address this issue, a Selective Kernel Attention (SKA) module [20] is introduced after the P5 feature layer. SKA enables adaptive multi-scale feature fusion through dynamic selection of convolution kernels, enhancing both spatial and channel representations. This significantly improves boundary accuracy and detail preservation in segmentation masks.

SKA is a hybrid attention mechanism with adaptive receptive field selection capability. It dynamically integrates features extracted by convolution kernels of different sizes, allowing the network to select the most suitable receptive field for a given target. Its workflow consists of three stages: Multi-scale feature decomposition: convolution kernels (e.g., 3×3 and 5×5 or dilated convolutions) extract local details and global structural features; Adaptive gated fusion: a squeeze-and-excitation mechanism learns weights for different scales, enabling dynamic selection of appropriate receptive fields; Feature reconstruction and enhancement: weighted multi-scale features are fused to generate more discriminative representations.

Compared with traditional single-scale convolution, SKA is better suited for irregular, multi-scale, and weak-boundary corrosion regions. It enhances boundary continuity, suppresses background noise, and improves mask smoothness, making it particularly effective for fine-grained segmentation of large-area corrosion. Its architecture is shown in Figure 5.

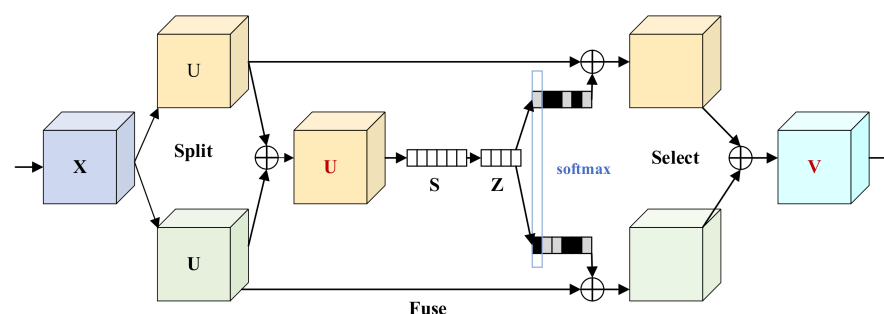


Figure 5. SKA module architecture.

By incorporating the SKA module into the P5 segmentation branch, the model can adaptively focus on key texture and boundary features of corrosion regions, producing segmentation masks that more closely match the true corrosion contours. This significantly improves the accuracy of corrosion area quantification and provides a reliable basis for subsequent damage assessment.

To facilitate implementation and ensure experimental reproducibility, the detailed configurations, insertion positions, and computational overhead of the three attention modules are quantitatively summarized in Table 1.

Table 1. Detailed implementation specifications and computational overhead of the integrated attention modules.

Module	Insertion Position	Kernel Size	Layer Configuration	Additional Parameters	GFLOP Increment
ECA	Backbone: P4, P5 feature layers	1D convolution: 3, 5	Global average pooling + 1D conv + Sigmoid activation	0.008 M	+0.8%
CBAMHead	Detection head: P3 feature layer	Spatial convolution: 7×7	Channel attention (MLP) + Spatial attention	0.42 M	+1.5%
SKA	Segmentation branch: P5 feature layer	3×3 conv, 5×5 dilated conv	Dual-branch multi-scale fusion + squeeze-excitation mechanism	1.39 M	+2.3%
Total	—	—	—	1.82 M	+4.6%

Note: The GFLOPs increment is computed relative to the baseline YOLOv26-seg model.

3.2.3. Overall Network Structure Preservation and Compatibility Optimization

Steel corrosion damage segmentation must be adaptable to real-world engineering scenarios, requiring high training stability, weight compatibility, and practical deployability for on-site inspection systems. To this end, the proposed improvement strategy strictly follows a “plug-and-play without altering the original architecture” principle. Modular enhancements are introduced based on the original YOLOv26 network without modifying its core structure or key parameters, ensuring seamless adaptation to both training and deployment requirements for corrosion segmentation tasks.

The improved network retains the original FPN + PAN bidirectional feature fusion architecture of YOLOv26, preserving the scale relationships, feature concatenation logic, and downsampling ratios of the P3, P4, and P5 feature layers. This design guarantees effective multi-scale feature fusion, enabling the model to simultaneously handle segmentation of small, medium, and large corrosion regions. In addition, the Segment26 segmentation head is maintained, which accepts multi-scale feature inputs and outputs bounding boxes (for localization), class probabilities (to distinguish corrosion from background), and pixel-level segmentation masks (to delineate corrosion regions). This enables joint inference for detection and segmentation, meeting the comprehensive evaluation requirements of engineering applications.

All newly introduced modules (ECA, CBAMHead, and SKA) adopt adaptive channel designs, eliminating the need for manual parameter tuning. These modules automatically adjust to different feature layer dimensions and integrate seamlessly with the original network, while effectively enhancing corrosion feature extraction and segmentation performance. Furthermore, the improved network does not require modifications to hyper-parameters, loss functions, or training strategies, and can be trained end-to-end using the original YOLOv26 training framework. This reduces training cost and deployment complexity while balancing academic innovation and engineering practicality, allowing rapid deployment in real-world steel structure inspection systems. The architecture of the improved YOLOv26-ECS model is illustrated in Figure 6.

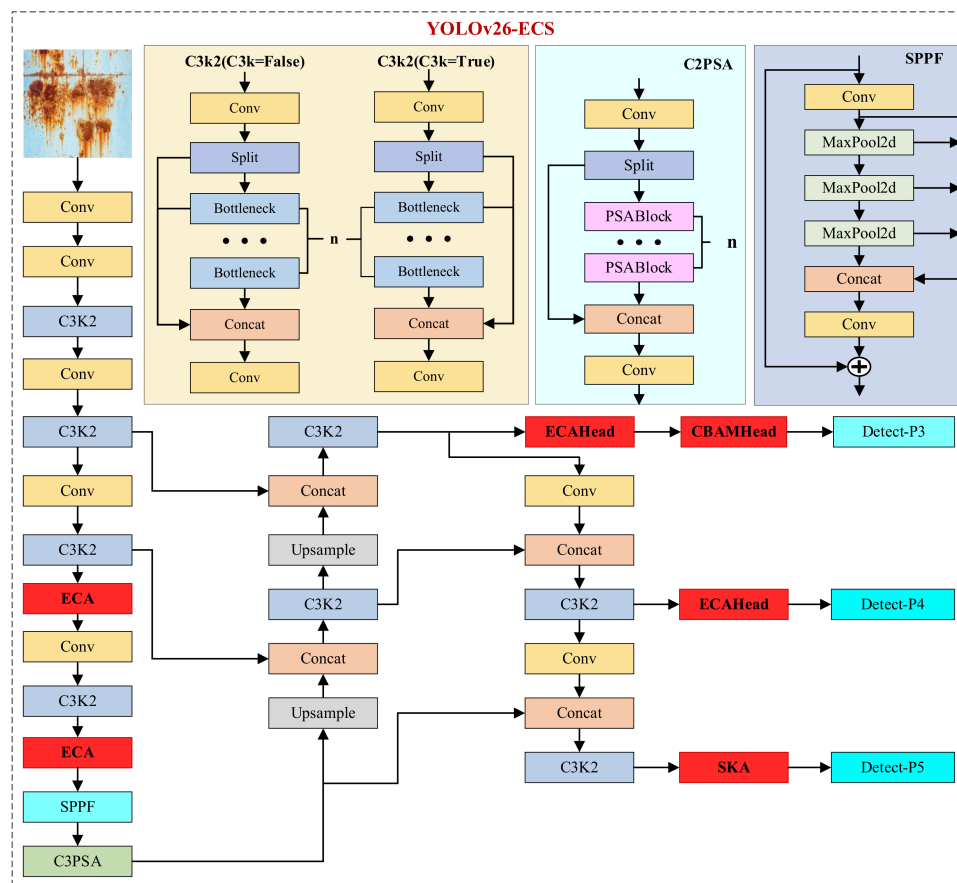


Figure 6. YOLOv26-ECS network architecture.

4. Network Training

4.1. Dataset Preparation and Preprocessing

In this study, a diversified steel structure corrosion dataset characterized by rich scene variations and fine annotations is established through the integration of field-collected images and publicly available dataset images.

A total of 2000 field images covering steel structural components (e.g., beams, columns, joints, bolts, and welds) are collected from industrial buildings, bridges, and large-span public structures. Image acquisition is performed using a digital single-lens reflex camera under natural daylight, indoor artificial illumination, backlight, low-illumination, and uneven-illumination environments. Shooting distances are controlled within the range of 0.5 m to 3.0 m, and top-view, front-view, and side-view angles are adopted. All images are captured at a uniform resolution of 4032×3024 pixels and resized to 640×640 pixels in the training phase.

Corrosion severity is classified into three graded levels:

Level 1 (Slight corrosion): Localized pitting or shallow rust, covering less than 10% of the structural surface;

Level 2 (Moderate corrosion): Flaky corrosion with distinct color variation, covering 10–30% of the structural surface;

Level 3 (Severe corrosion): Large-area continuous rust with rough surface texture and evident peeling, covering more than 30% of the structural surface.

The dataset presents favorable class balance, with 682 images for slight corrosion, 745 images for moderate corrosion, and 573 images for severe corrosion. The total number of labeled corrosion instances is 3246, with a class distribution ratio of approximately 1:1.09:0.84, eliminating significant bias toward specific corrosion grades.

Additionally, the public Corrosion Instance Segmentation dataset is introduced from the Roboflow Universe platform, which provides 734 images with fine instance-level annotations. This dataset incorporates diverse illumination, background, and steel surface conditions, thereby improving the diversity and generalization capacity of the fused training dataset.

The complete integrated dataset is strictly partitioned into training, validation, and test subsets at a ratio of 7:2:1 to guarantee independent and objective model evaluation.

Pixel-level polygonal annotation is implemented using the LabelmeV5.3 tool. Standardized annotation protocols are followed as listed below:

- (1) Closed contours are applied to fit the actual boundaries of corrosion regions precisely;
- (2) Ambiguous or negligible corrosion areas are excluded from labeling;
- (3) Overlapped corrosion regions are labeled individually to satisfy the requirements of instance segmentation;
- (4) Double-checking is conducted by two experienced annotators to ensure labeling consistency and accuracy.

Annotation results are stored in JSON format and converted into binary mask labels compatible with neural network training. Data augmentation strategies are adopted to enhance model robustness and generalization, including random rotation ($\pm 15^\circ$), horizontal flipping, vertical flipping, brightness adjustment (0.6–1.4), contrast adjustment, and Gaussian blurring. The training subset is expanded to eight times the original scale without introducing additional labeling noise.

4.2. Training Environment

Model training was conducted on a high-performance workstation equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB memory) and dual Intel Xeon (R) E5-2699 v3 @ 2.30 GHz CPUs with 128 GB RAM. The training environment was configured using the Anaconda platform, with the PyTorch 2.2.2 deep learning framework installed. Code execution was carried out in the PyCharm 2024.3.5 development environment.

4.3. Training Strategy and Evaluation Metrics

In this study, the model was trained for 250 epochs. Input images were uniformly resized to 640×640 pixels, and the batch size was set to 20 to balance computational efficiency and training effectiveness. Through iterative optimization, the model achieved good convergence and generalization performance for corrosion detection tasks.

A multi-task joint optimization strategy was adopted, incorporating several loss functions: bounding box regression loss (box_loss), classification loss (cls_loss), distribution focal loss (dfl_loss), segmentation loss (seg_loss), and semantic segmentation loss (sem_loss). Specifically, box_loss constrains the positional and scale differences between predicted and ground-truth bounding boxes; cls_loss measures classification discrepancies; dfl_loss improves localization accuracy via distribution modeling; and seg_loss and sem_loss optimize pixel-level predictions for instance and semantic segmentation, respectively, enhancing the model's ability to capture fine-grained corrosion features.

Model performance was comprehensively evaluated using commonly adopted metrics in object detection, including Precision, Recall, mean Average Precision (mAP), and Frames Per Second (FPS).

Precision measures the proportion of correctly predicted positive samples among all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

where TP (True Positive) denotes correctly detected corrosion targets, and FP (False Positive) denotes incorrectly detected targets. Higher precision indicates a lower false detection rate.

Recall evaluates the model's ability to detect all true targets:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

where FN (False Negative) represents missed detections. Higher recall indicates stronger detection capability.

For average precision, mAP@0.5 (mAP50) and mAP@0.5:0.95 (mAP50–95) were adopted. mAP50 represents the average precision at an IoU threshold of 0.5:

$$\text{mAP}_{50} = \frac{1}{N} \sum_{i=1}^N AP_i(\text{IoU} = 0.5) \quad (3)$$

mAP50–95 averages AP over multiple IoU thresholds from 0.5 to 0.95 (step size 0.05), providing a more comprehensive evaluation:

$$\text{mAP}_{50-95} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T AP_i^{(\text{IoU}_t)} \right) \quad (4)$$

where N is the number of categories and T is the number of IoU thresholds.

Average Precision (AP) is defined as the area under the Precision–Recall curve:

$$AP = \int_0^1 P(R) dR \quad (5)$$

FPS measures real-time performance and is defined as:

$$\text{FPS} = \frac{1}{T} \quad (6)$$

where T is the average processing time per image (in seconds). A higher FPS indicates faster inference speed. In practice, a balance between detection accuracy and FPS is required to meet engineering demands.

Overall, the combination of multi-loss optimization and multi-dimensional evaluation metrics enables a comprehensive assessment of detection accuracy, localization performance, and real-time capability.

4.4. Training Results

The training results are illustrated in Figure 7. YOLOv26-ECS model exhibits stable convergence and consistent optimization throughout the training process. All loss components (box_loss, seg_loss, cls_loss, dfl_loss, and sem_loss) show a clear downward trend and gradually stabilize. Specifically, the significant reduction in box_loss indicates improved localization accuracy, while the decrease in seg_loss reflects enhanced segmentation precision. The convergence of cls_loss and dfl_loss demonstrates effective optimization in classification and bounding box regression. Meanwhile, sem_loss supervises pixel-level classification by measuring the discrepancy between predicted and ground-truth labels for each pixel, contributing to improved semantic consistency.

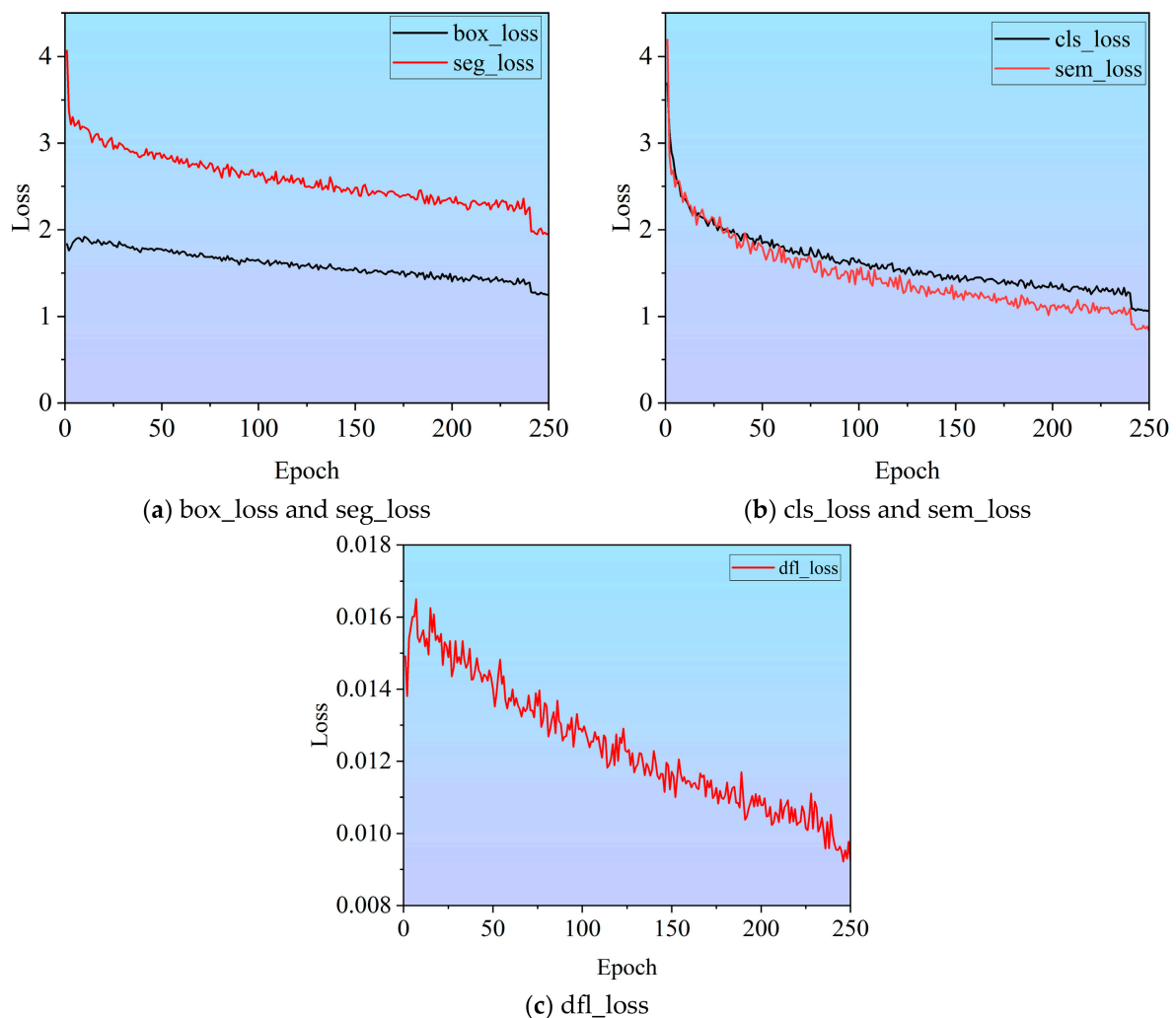


Figure 7. Loss function curves.

From Figure 8, performance metrics show a consistent upward trend. Both Precision and Recall improve steadily, indicating reduced false positives and false negatives. The mAP50 rapidly increases during early training and stabilizes at a high level, reaching a maximum of 0.920, demonstrating strong detection accuracy at IoU = 0.5. Meanwhile, mAP50–95 also increases steadily, achieving a maximum of 0.851, indicating robust performance under stricter localization criteria and superior boundary prediction capability.

The training process exhibits a typical staged pattern. In the early stage, the model focuses on feature learning, resulting in lower and fluctuating performance metrics. As training progresses, the model captures key corrosion features, leading to rapid performance improvement. In the later stage, metrics stabilize, indicating convergence and sustained high performance.

Regarding real-time performance, inference speed was evaluated for the improved YOLOv26-ECS model. Under an input size of (1, 3, 640, 640), the processing time per image includes 3.61 ms for preprocessing, 25.03 ms for inference, and 6.64 ms for postprocessing, with a total latency of 35.28 ms. This corresponds to an FPS of approximately 28.34. These results demonstrate that the model achieves a good balance between accuracy and efficiency, meeting practical engineering requirements for real-time detection.

In summary, the improved YOLOv26-ECS model achieves high detection accuracy and high-quality segmentation, demonstrating strong synergy between detection and

segmentation tasks. The high mAP50 and mAP50–95 values indicate excellent localization precision and generalization capability.

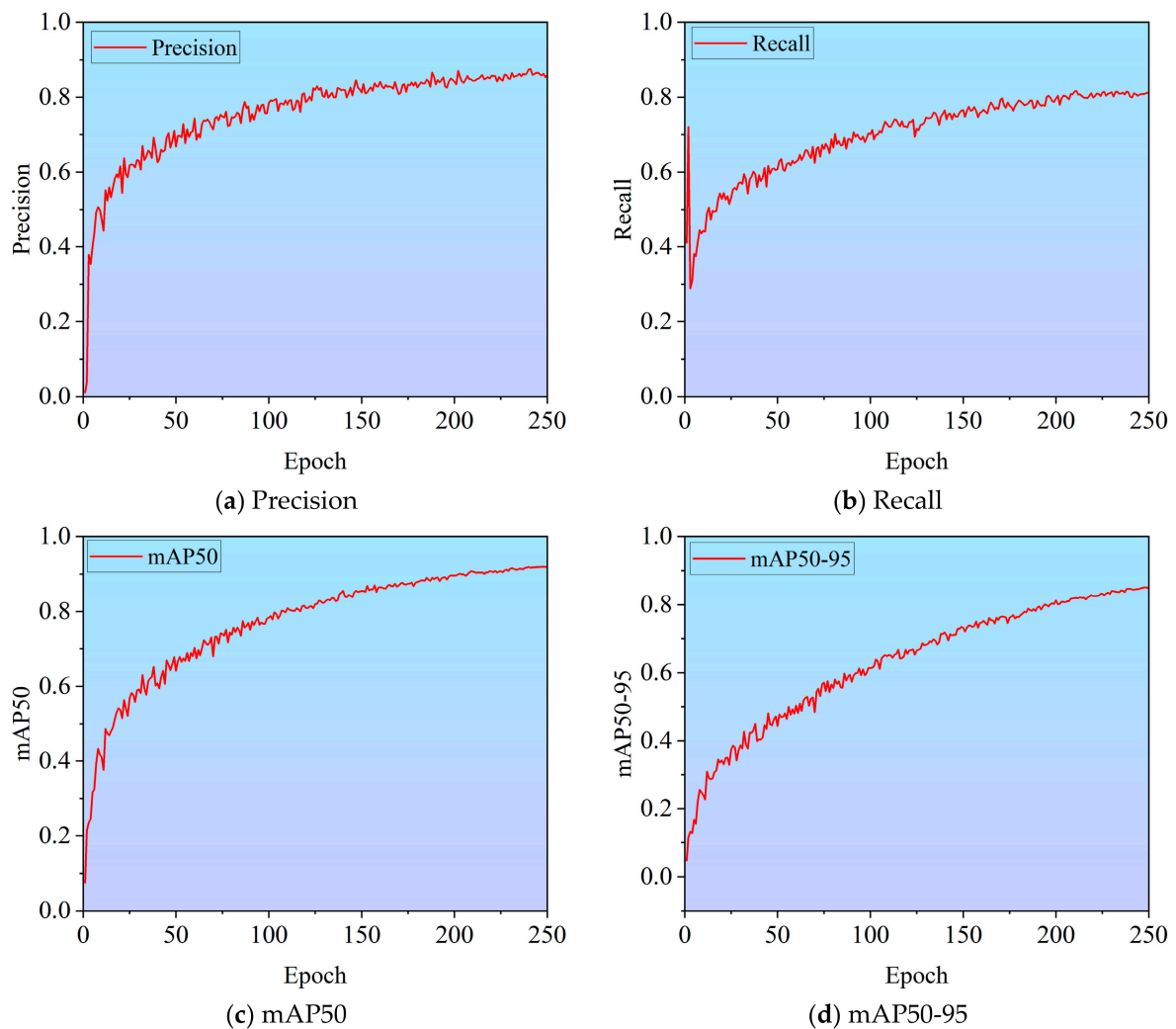


Figure 8. Performance metrics of YOLOv26-ECS.

5. Experimental Analysis

5.1. Ablation Experiments

To evaluate the contribution of each attention module to the overall model performance, systematic ablation experiments were conducted by introducing ECA, CBAM, and SKA modules into the baseline YOLOv26 model. The results are summarized in Table 2.

Table 2. Results of the ablation Experiments.

Model	ECA	CBAM	SKA	mAP50	FPS
1				0.870	29.94
2	✓			0.882	29.71
3		✓		0.888	29.36
4			✓	0.895	29.18
5	✓	✓		0.900	28.92
6	✓		✓	0.906	28.75
7		✓	✓	0.909	28.51
8	✓	✓	✓	0.920	28.34

From the single-module experiments, it can be observed that all three attention mechanisms effectively improve detection performance. Specifically, incorporating the ECA module increases mAP50 from 0.870 to 0.882; adding the CBAM further improves it to 0.888; and introducing the SKA module yields the most significant improvement, achieving an mAP50 of 0.895. These results indicate that all attention mechanisms enhance feature representation, with SKA demonstrating superior capability in multi-scale feature modeling.

In the dual-module experiments, the performance is further improved. The combination of ECA and CBAM achieves an mAP50 of 0.900; ECA combined with SKA reaches 0.906; and CBAM combined with SKA achieves 0.909. Notably, all combinations that include the SKA module outperform those without it, indicating that SKA exhibits strong complementary effects with other attention mechanisms and significantly enhances feature representation.

When all three modules (ECA, CBAM, and SKA) are integrated, the model achieves the best performance, with an mAP50 of 0.920, representing a 5.0% improvement over the baseline. Compared with dual-module configurations, the triple-module fusion still provides additional performance gains, demonstrating that joint modeling across both channel and spatial dimensions can further improve feature representation capability.

In terms of inference speed, as the number of attention modules increases, the computational complexity also rises, resulting in a decrease in FPS from 29.94 to 28.34, corresponding to an overall reduction of approximately 5.3%. Despite this slight decline, the impact on real-time performance is limited, indicating that the introduced modules effectively enhance accuracy while maintaining practical efficiency.

In summary, the progressive integration of ECA, CBAM, and SKA modules leads to consistent performance improvements, with SKA contributing the most significant gains. The multi-module fusion strategy further exploits feature information and achieves a favorable balance between accuracy and efficiency, thereby validating the effectiveness and rationality of the proposed approach.

5.2. Comparative Experiments

To further validate the performance advantages of the proposed model, comparative experiments were conducted on the same steel corrosion dataset by evaluating the improved YOLOv26-ECS model against YOLOv11 [21], YOLOv13 [22], and the baseline YOLOv26. To ensure a reasonable and task-consistent comparison, only state-of-the-art real-time YOLO-series instance segmentation models are selected as benchmarks, since this study focuses on real-time, deployment-oriented steel corrosion instance segmentation. YOLOv26 represents the latest and strongest version of the YOLO family; thus, comparison with YOLOv11, YOLOv13, and the original YOLOv26 baseline is sufficiently comprehensive and persuasive.

All models were trained under identical training strategies and experimental settings to ensure fairness and comparability. Quantitative results in terms of Precision, Recall, mAP50, mAP50–95, and FPS are presented in Table 2, while a visual comparison of model performance is shown in Figure 9.

As indicated in Table 3, the proposed YOLOv26-ECS model demonstrates superior performance across all evaluation metrics. In terms of detection accuracy, YOLOv26-ECS achieves a Precision of 0.86 and a Recall of 0.81, both of which are improvements over the baseline YOLOv26 (0.83 and 0.80, respectively), indicating enhanced capability in reducing both false positives and false negatives.

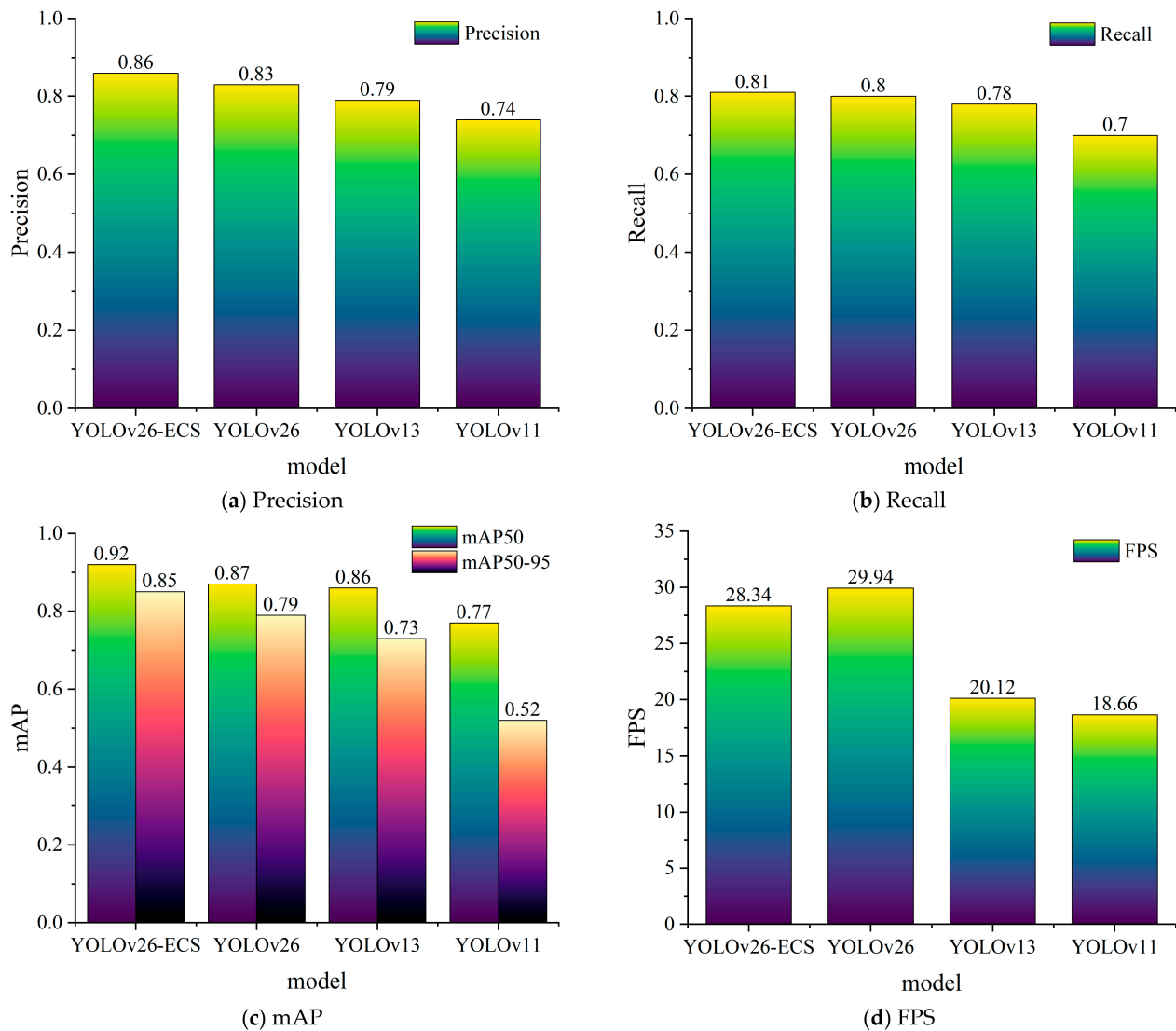


Figure 9. Comparison of different network performance indicators (Precision, Recall, mAP, FPS).

Table 3. Performance comparison of different networks.

Model	Precision	Recall	mAP50	mAP50-95	FPS
YOLOv26-ECS	0.86	0.81	0.92	0.85	28.34
YOLOv26	0.83	0.8	0.87	0.79	29.94
YOLOv13	0.79	0.78	0.86	0.73	20.12
YOLOv11	0.74	0.70	0.77	0.52	18.66

Regarding comprehensive evaluation metrics, YOLOv26-ECS attains an mAP50 of 0.92, representing a 5.0% improvement over YOLOv26, and an mAP50–95 of 0.85, which is 6.0% higher than the baseline value of 0.79. These results demonstrate that the proposed method not only achieves higher detection accuracy under relatively loose IoU thresholds but also maintains strong performance under stricter evaluation criteria, reflecting improved localization accuracy and robustness.

Compared with other models, the superiority of YOLOv26-ECS is even more pronounced. Relative to YOLOv13, mAP50 and mAP50–95 are improved by 6.0% and 12.0%, respectively; compared to YOLOv11, the improvements reach 15.0% and 33.0%. These findings indicate that with structural optimization and the integration of attention mechanisms, the model's ability to represent complex corrosion features is significantly enhanced.

In terms of inference speed, YOLOv26-ECS achieves 28.34 FPS, which is slightly lower than the baseline YOLOv26 (29.94 FPS) but still significantly faster than YOLOv13 (20.12 FPS) and YOLOv11 (18.66 FPS). Overall, the proposed model achieves substantial accuracy improvements with only a minor decrease in speed (approximately 5.3%), demonstrating a favorable balance between accuracy and efficiency.

In summary, the YOLOv26-ECS model exhibits superior overall performance in steel corrosion detection tasks. It significantly improves detection accuracy while maintaining high inference speed, thereby validating the effectiveness and practical value of the proposed method.

The predicted segmentation results of different models are illustrated in Figure 10. By comparing segmentation outputs in representative steel corrosion scenarios, the differences in corrosion region recognition and boundary delineation among various models can be intuitively analyzed.

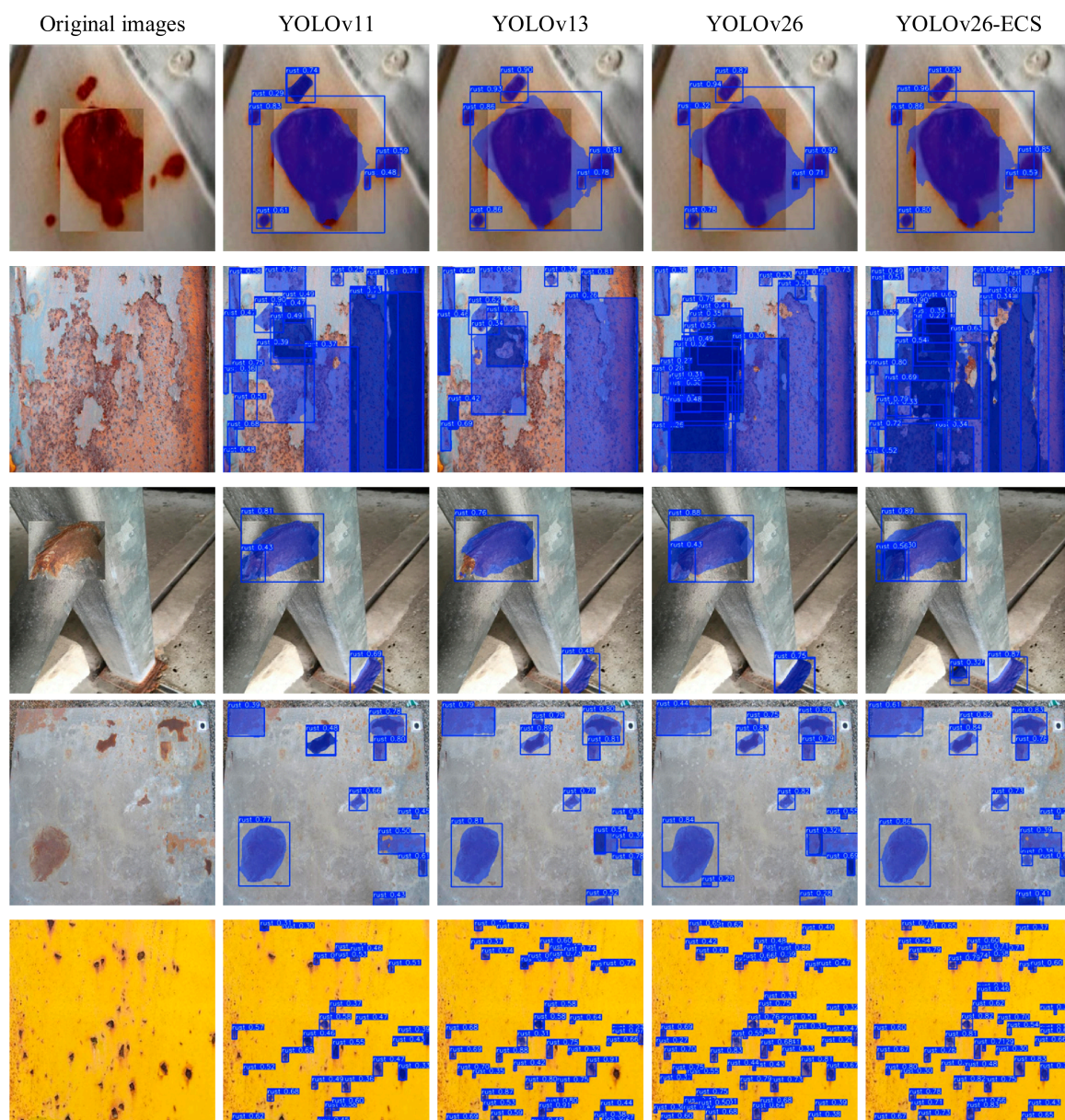


Figure 10. Segmentation prediction results of different networks.

From the overall segmentation performance, the baseline YOLOv26 model is capable of accurately identifying large-area corrosion regions; however, it still exhibits limitations

in fine-grained detail representation. Specifically, it tends to miss small corrosion spots, produces incomplete segmentation for some medium-scale corrosion regions, and generates blurred or overly smoothed boundaries. Moreover, in the presence of background interference such as welds and bolts, the model is prone to false segmentation, incorrectly classifying non-corroded regions as corrosion, indicating insufficient discriminative capability in complex environments.

In comparison, YOLOv11 and YOLOv13 show overall inferior segmentation performance relative to YOLOv26. YOLOv11 demonstrates weak responsiveness under complex lighting and low-contrast conditions, often resulting in large-scale missed detections. Additionally, its segmentation outputs are fragmented and fail to form coherent corrosion regions. Although YOLOv13 improves detection performance to some extent, it still suffers from unclear boundaries and insufficient detail representation in segmentation tasks. In particular, its ability to detect small targets and weak-texture regions remains limited when corrosion of multiple scales coexists.

The proposed YOLOv26-ECS model exhibits clear advantages in segmentation performance. As shown in Figure 10, it not only effectively captures large-area corrosion regions but also accurately detects small corrosion spots, significantly reducing missed detections. For medium-scale corrosion patches, the predicted segmentation is more continuous and complete, closely aligning with the true corrosion distribution. Furthermore, the segmentation masks generated by the improved model present sharper and smoother boundaries, enabling more accurate delineation of transitions between corrosion and background, clearly outperforming the comparison models.

Further analysis indicates that YOLOv26-ECS demonstrates stronger robustness in complex backgrounds. For confusing regions such as welds, bolts, and surface textures, the model effectively suppresses false positives and distinguishes them from actual corrosion areas. This improvement can be attributed to the introduced multi-scale attention mechanisms, which enable the model to focus on critical features in both channel and spatial dimensions, thereby enhancing feature discriminability and robustness.

In summary, the comparative analysis of segmentation results demonstrates that the proposed YOLOv26-ECS model outperforms other models in terms of corrosion region completeness, boundary precision, and adaptability to complex scenarios. It enables more accurate and stable segmentation of steel corrosion damage, providing reliable technical support for subsequent corrosion quantification and structural safety assessment.

5.3. Dataset Experiments

To further validate the generalization capability and cross-domain adaptability of the proposed YOLOv26-ECS model, two representative public datasets were selected for evaluation: the crack-bphdr dataset and the road_segmentationv1.1 dataset, both sourced from the Roboflow Universe platform. These datasets differ significantly in task characteristics and scene complexity, providing a comprehensive benchmark for assessing the robustness and versatility of the model.

The crack-bphdr dataset comprises approximately 4000 images of concrete cracks, encompassing a wide range of background environments, illumination conditions, shooting perspectives, and crack morphologies. Such diversity introduces substantial variability and complexity, making it well-suited for evaluating fine-grained feature extraction and structural preservation capabilities. In contrast, the road_segmentationv1.1 dataset contains 2766 road images captured under different real-world scenarios, including variations in road types, lighting conditions, occlusions, and surface textures. This dataset poses higher demands on semantic segmentation accuracy and boundary delineation.

To ensure experimental consistency and fairness, the same training strategies and hyperparameter settings described in the previous section were applied to both datasets. The qualitative segmentation results of YOLOv26-ECS on these datasets are illustrated in Figures 11 and 12.

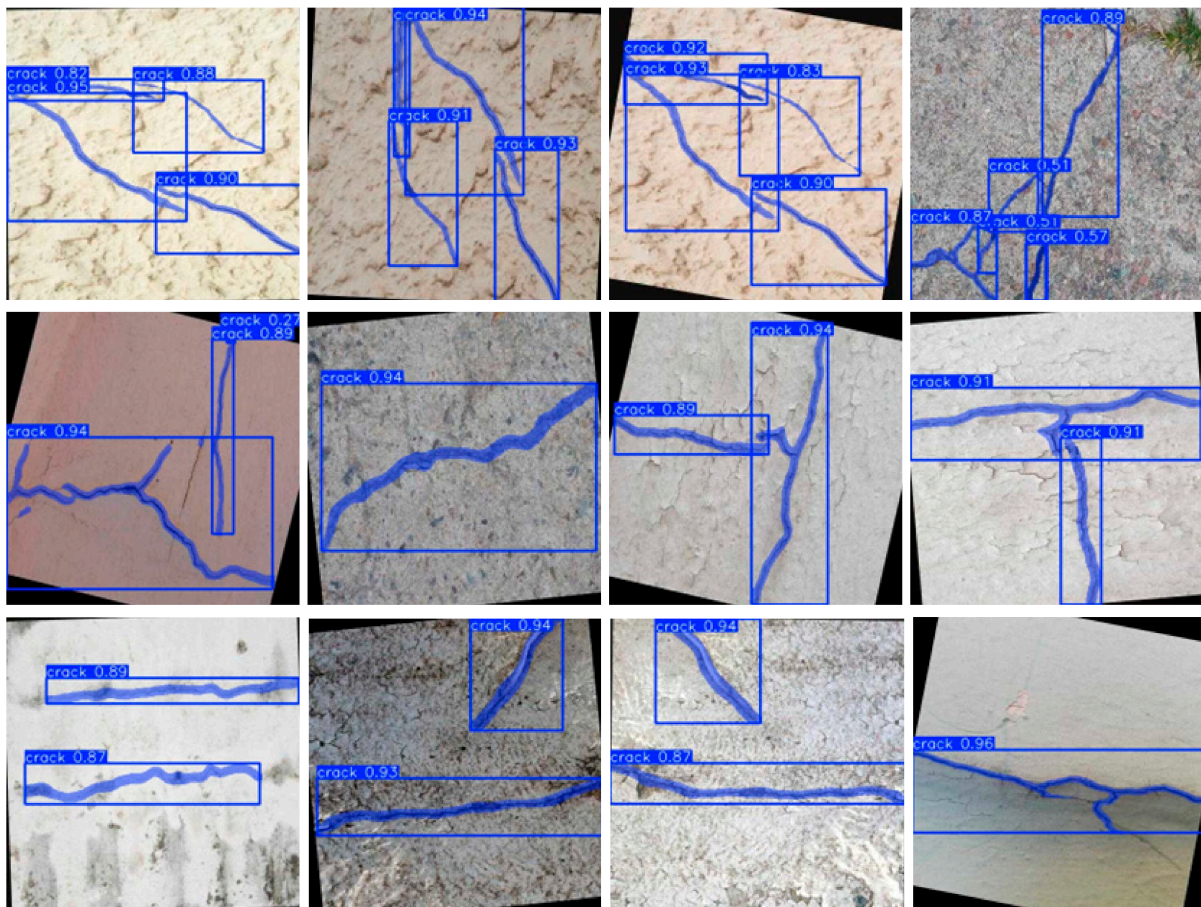


Figure 11. Segmentation results on the crack-bphdr dataset.

On the crack-bphdr dataset, the proposed model demonstrates strong capability in accurately extracting crack regions with diverse shapes and structures. For slender and elongated cracks, YOLOv26-ECS maintains excellent continuity with minimal fragmentation, indicating effective preservation of fine structural details. In cases of branched or network-like cracks, the model successfully captures complex topological patterns, reflecting its strong multi-scale representation ability. Furthermore, even under challenging conditions such as cluttered backgrounds and uneven illumination, the model is able to highlight crack regions while effectively suppressing irrelevant noise, demonstrating robust feature discrimination. Nevertheless, slight missed detections are observed in extremely fine crack regions, suggesting that further enhancement is still needed for ultra-small target perception.

On the road_segmentationv1.1 dataset, YOLOv26-ECS also achieves competitive performance. The model accurately distinguishes between road and non-road regions, producing segmentation outputs with high continuity and well-defined boundaries. Under difficult conditions, including shadows, vehicle occlusions, and variations in road surface textures, the model maintains stable and reliable performance, indicating strong robustness against environmental interference. Additionally, the model exhibits consistent segmentation accuracy across roads of varying widths, shapes, and structural layouts, without noticeable over-segmentation or under-segmentation issues.

Overall, the experimental results on both datasets demonstrate that YOLOv26-ECS not only achieves strong performance on the original steel corrosion task but also generalizes effectively to other structurally different segmentation tasks. Its ability to produce high-quality segmentation results in both crack detection and road scene understanding validates the effectiveness of the proposed architecture in feature extraction, multi-scale modeling, and complex scene adaptation. These findings highlight the model's strong versatility and provide a solid foundation for its application in a wide range of engineering inspection and real-world vision tasks.

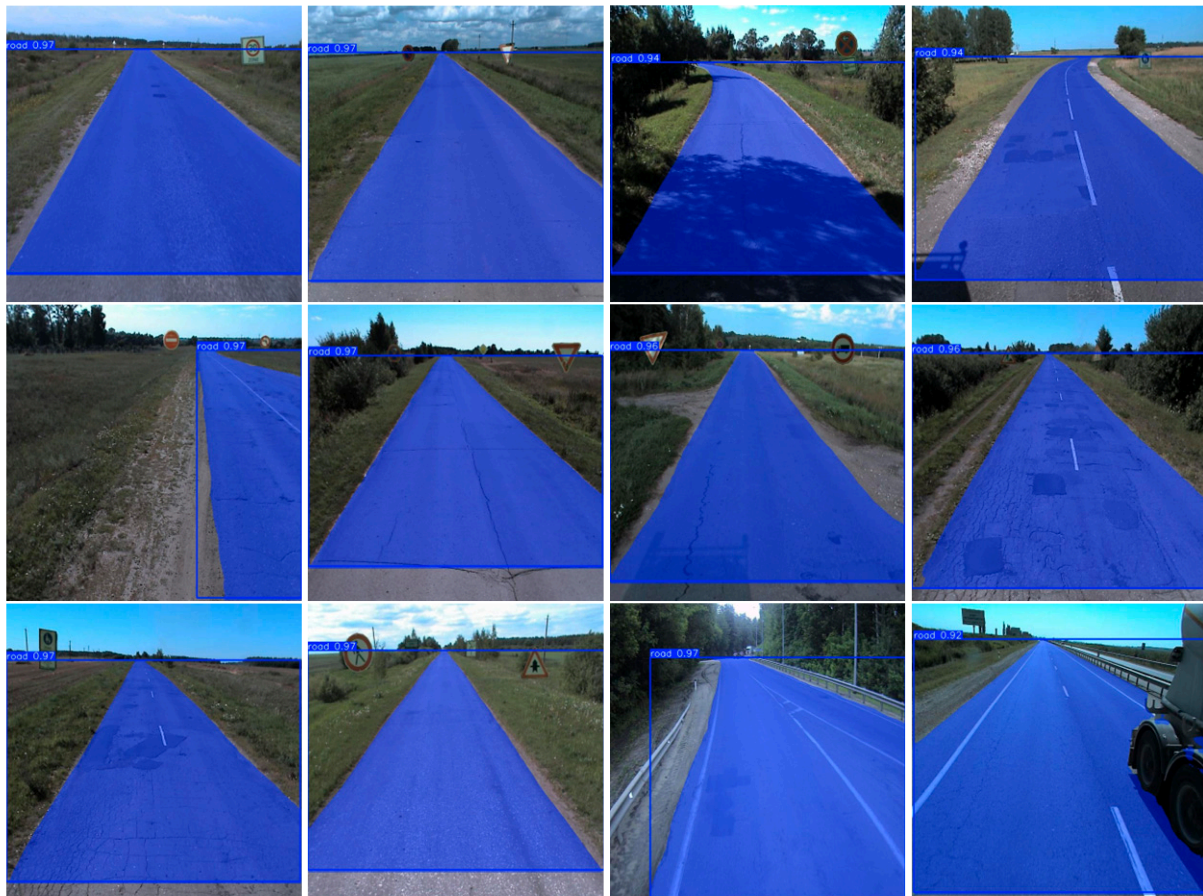


Figure 12. Segmentation results on the road_segmentationv1.1 dataset.

5.4. Statistical Significance Analysis

To verify that the performance improvements of the proposed YOLOv26-ECS model are statistically significant rather than random fluctuations, an independent-sample t-test was conducted on the core metric mAP@50. Each model was trained and evaluated 10 independent times with identical dataset splits, training strategies, and hyperparameter settings. The significance level was set at $\alpha = 0.05$.

The statistical results demonstrate that the mAP@50 improvement of YOLOv26-ECS relative to the baseline YOLOv26 is statistically significant ($p < 0.05$). Compared with YOLOv13 and YOLOv11, the performance advantages of the proposed model are even more significant ($p < 0.01$). In the ablation study, the mAP@50 gain from the joint integration of ECA, CBAM, and SKA is also statistically significant relative to single-module and dual-module combinations ($p < 0.05$).

These statistical analyses confirm that the performance gains of YOLOv26-ECS arise from the designed multi-scale attention-enhanced strategy, ensuring the reliability and repeatability of the experimental results.

6. GUI

To facilitate engineering application and visual interaction of the steel corrosion detection model, a concise and efficient graphical user interface (GUI) system [23] was developed based on the PyQt5 framework. The system integrates functions such as image loading, model inference, result visualization, and output saving, enabling rapid detection and segmentation of corrosion regions. The interface is shown in Figure 13.

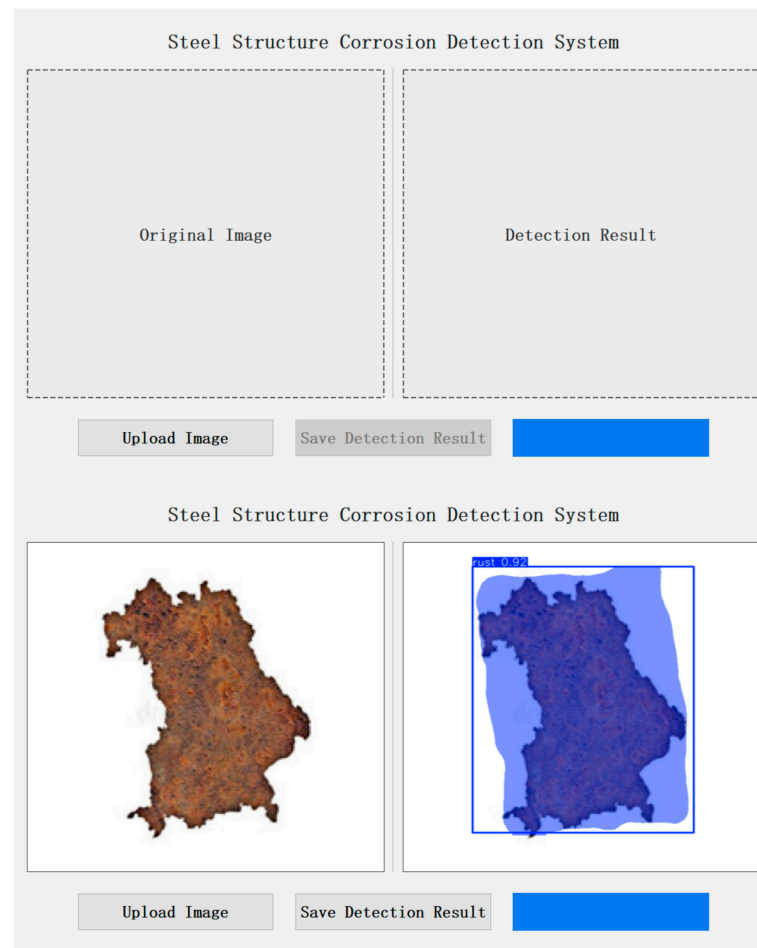


Figure 13. GUI.

From a structural perspective, the GUI adopts a typical top–bottom layout. The upper section is dedicated to image display, while the lower section contains functional controls. The display area is arranged in a side-by-side format: the left panel shows the input image, and the right panel displays the segmentation result. A separator is used between the two panels, allowing users to intuitively compare the original and processed images. This “input–output” comparison facilitates clear evaluation of the model’s performance in corrosion extraction, boundary delineation, and noise suppression.

In terms of functionality, the system provides two main operations: “Upload Image” and “Save Results.” When the user clicks the “Upload Image” button, a file selection dialog is opened, allowing the import of images in common formats (e.g., JPG, PNG, BMP). Once the image is loaded, it is displayed on the left panel, and the system automatically invokes the preloaded YOLOv26-ECS model to perform inference without requiring additional user interaction, thereby improving usability and efficiency.

During inference, the system utilizes the Ultralytics framework to load trained model weights and perform end-to-end segmentation prediction. The output is visualized using built-in rendering functions, including corrosion masks and contour information, and

displayed as an image in the right panel. The results demonstrate that the model effectively highlights corrosion regions and accurately delineates boundaries, confirming its feasibility in real-world applications.

For result management, the system provides a “Save Results” button, allowing users to export the segmentation output to a specified local path with customizable file names. The results are saved in standard image formats for further analysis and engineering use. Notably, the button remains disabled before detection is performed, preventing unintended operations and reflecting a user-friendly design.

From an implementation perspective, the GUI utilizes OpenCV for image reading and processing, while QImage and QPixmap are employed for image format conversion and display. To ensure visual consistency, adaptive image scaling is applied, preserving the aspect ratio while enabling smooth rendering, thereby enhancing user experience.

In summary, the proposed GUI system features a clear structure and simple operation, enabling efficient detection and intuitive visualization of steel corrosion images. By integrating deep learning models with a user-friendly interface, the system enhances practical applicability and provides strong technical support for deployment in real-world inspection scenarios.

7. Summary

7.1. Conclusions

- (1) Considering the characteristics of steel corrosion, such as low contrast, multi-scale distribution, and blurred boundaries, the YOLOv26 model is systematically enhanced at the backbone, detection head, and segmentation branch. By incorporating multiple attention mechanisms, including ECA, CBAM, and SKA, the model’s ability to represent corrosion features is significantly improved.
- (2) Experimental results demonstrate that the proposed model achieves mAP50 and mAP50–95 values of 0.920 and 0.851, respectively, on the self-constructed dataset, showing substantial improvements over the baseline model while maintaining high inference speed. This indicates a favorable balance between accuracy and efficiency.
- (3) Ablation studies confirm the effectiveness of each attention module. Among them, the SKA module contributes most significantly to multi-scale feature modeling, while the combination of multiple modules further enhances overall performance.
- (4) Tests on public datasets show that the model maintains strong segmentation performance in different tasks, such as crack detection and road segmentation, demonstrating robust generalization ability.
- (5) A GUI system developed based on PyQt5 enables visualization and engineering application of the model, improving its practicality and supporting on-site deployment for steel structure corrosion detection.

7.2. Limitations

- (1) Limited dataset scale and diversity. Although the dataset integrates both self-collected and public data, its overall size is still limited and mainly focused on typical steel structure scenarios. Insufficient coverage of extreme conditions (e.g., strong reflections, severe occlusions, complex industrial backgrounds) may restrict the model’s generalization capability.
- (2) Insufficient detection of extremely small targets and early-stage corrosion. Although multi-scale attention mechanisms improve the detection of small corrosion regions, slight missed detections still occur for extremely fine pitting or early-stage corrosion due to weak and indistinct features.

- (3) Trade-off between model complexity and real-time performance. While multiple attention mechanisms significantly enhance detection accuracy, they also increase computational complexity, leading to a slight decrease in inference speed. This may pose challenges for deployment on resource-constrained or embedded devices.
- (4) Gap between experimental setup and real-world applications. Experiments are primarily conducted on static image datasets, without fully considering real-world factors such as continuous video streams, multi-view fusion, and dynamic illumination changes. Long-term stability in practical engineering environments remains to be further validated.

7.3. Future Works

- (1) Construction of large-scale, high-quality datasets with multi-source fusion. Future work can expand the dataset to include more complex conditions and extreme environments. Integration of UAV inspection images, video data, and multimodal information (e.g., infrared and LiDAR point clouds) can further enhance model generalization and environmental adaptability.
- (2) Improvement in small defect detection and early-stage damage recognition. For extremely small targets, techniques such as super-resolution reconstruction, optimized feature pyramids, or Transformer-based architectures can be introduced to strengthen global modeling capability and improve early-stage corrosion detection.
- (3) Model lightweighting and efficient deployment. Techniques such as model pruning, knowledge distillation, and quantization can be explored to reduce model complexity while maintaining accuracy, thereby improving inference speed and enabling deployment on edge devices.
- (4) Extension to practical engineering applications and intelligent system integration. In practical scenarios, the model can be integrated with UAV inspection systems and structural health monitoring platforms to form a comprehensive system combining data acquisition, intelligent analysis, and early warning. Meanwhile, the GUI system can be further enhanced with functions such as batch processing, video detection, and cloud deployment to improve usability and intelligence.

Author Contributions: Conceptualization, H.H.; methodology, H.H. and L.T.; Validation, J.X.; Formal analysis, Z.W., J.Z. and L.T.; Investigation, Z.W., J.Z. and J.X.; Resources, J.X.; Data curation, Z.W. and J.Z.; Writing—review and editing, L.T.; Supervision, H.H.; Funding acquisition, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Development of asphalt concrete for impermeable layers considering the maximum aggregate size (Grant No. SUA2026KJ016).

Data Availability Statement: The data will be provided as needed.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fan, S.; Ding, R.; Xu, T.; Liu, M. Experimental Study on Progressive Collapse Resistance of Steel Frame Structures under Local Real Fire. *J. Build. Eng.* **2026**, *117*, 114785. [[CrossRef](#)]
2. Liu, C.; Zhang, S.; Yan, L.; Liu, L.; Miao, J. Fire-Induced Bonding Degradation in Reinforced Concrete: Effects of Corrosion and Exposure Time on Residual Performance. *J. Build. Eng.* **2026**, *119*, 115205. [[CrossRef](#)]
3. Simoncelli, M.; Aloisio, A.; Zucca, M.; Venturi, G.; Alaggio, R. Intensity and Location of Corrosion on the Reliability of a Steel Bridge. *J. Constr. Steel Res.* **2023**, *206*, 107937. [[CrossRef](#)]
4. Chen, W.; Hou, J.; Hao, S.; Zhu, Y.; Yu, M. Mapping Knowledge Structure and Themes Trends of Corrosion Detection of Steel: A Bibliometric Analysis. *Comput. Mater. Sci.* **2025**, *253*, 113889. [[CrossRef](#)]

5. Hou, J.; Chen, W.; Duan, Z.; Li, H.; Yu, M. Research on Lightweight Network for Real-Time Detection of Steel Structure Corrosion Based on Knowledge Distillation. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112356. [\[CrossRef\]](#)
6. Alshareef, I.Y.; Rahman, A.A.-H.A.; Shaameri, A.Z.; Khan, N.; Rusli, M.S.; Hassan, M.K.; Manzak, A.; Marsono, M.N. ReLU-DCT and LASP: Real-Valued Frequency-Domain Activations and Adaptive Pooling for DCT-Based CNNs. *Knowl.-Based Syst.* **2026**, *343*, 116018. [\[CrossRef\]](#)
7. Alkhamash, E.H. A Comparative Analysis of YOLOv9, YOLOv10, YOLOv11 for Smoke and Fire Detection. *Fire* **2025**, *8*, 26. [\[CrossRef\]](#)
8. Wang, D.; Tan, J.; Wang, H.; Kong, L.; Zhang, C.; Pan, D.; Li, T.; Liu, J. SDS-YOLO: An Improved Vibratory Position Detection Algorithm Based on YOLOv11. *Measurement* **2025**, *244*, 116518. [\[CrossRef\]](#)
9. Zhang, C.; Chen, X.; Liu, P.; He, B.; Li, W.; Song, T. Automated Detection and Segmentation of Tunnel Defects and Objects Using YOLOv8-CM. *Tunn. Undergr. Space Technol.* **2024**, *150*, 105857. [\[CrossRef\]](#)
10. Zhang, S.; Wang, K.; Zhang, H.; Wang, T.; Gao, X.; Song, Y.; Wang, F. An Improved YOLOv8 for Fiber Bundle Segmentation in X-Ray Computed Tomography Images of 2.5D Composites to Build the Finite Element Model. *Compos. Part A* **2024**, *185*, 13. [\[CrossRef\]](#)
11. Du, Q.; Zhang, S.; Yang, S. BLP-YOLOv10: Efficient Safety Helmet Detection for Low-Light Mining. *J. Real Time Image Process.* **2025**, *22*, 10. [\[CrossRef\]](#)
12. Tu, W.; Yu, H.; Wu, Z.; Li, J.; Cui, Z.; Yang, Z.; Zhang, X.; Wang, Y. YOLOv10-UDFishNet: Detection of Diseased Takifugu Rubripes Juveniles in Turbid Underwater Environments. *Aquac. Int.* **2025**, *33*, 125. [\[CrossRef\]](#)
13. Liu, C.; Tian, L.; Wang, P.; Yu, Q.-Q.; Song, L.; Miao, J. Non-Destructive Detection and Quantification of Corrosion Damage in Coated Steel Components with Different Illumination Conditions. *Expert Syst. Appl.* **2025**, *282*, 127854. [\[CrossRef\]](#)
14. Phannakham, N.; Hashimoto, K.; Sato, Y.; Ueda, N. Application of Unsupervised AI-Assisted Acoustic Wave Sound Analysis for Non-Destructive Detection of Steel Corrosion Induced Deterioration. *Case Stud. Constr. Mater.* **2026**, *24*, e05702. [\[CrossRef\]](#)
15. Li, F.; Min, Y.; Zhang, Y.; Zuo, H.; Bai, F.; Zhang, Y. Towards General and Efficient Fault Diagnosis: A Novel Framework for Multi-Fault Cross-Domain Diagnosis of Lithium-Ion Batteries in Real-World Scenarios. *Energy* **2025**, *334*, 137825. [\[CrossRef\]](#)
16. Cai, J.; Wang, R.; Hou, Z.; Xu, X.; Pan, Y. A DiT-Based Multi-Scenario Optical Temporal Prediction Model for Pipeline Corrosion. *Process Saf. Environ. Prot.* **2026**, *210*, 108726. [\[CrossRef\]](#)
17. Safonova, A.; Campanera-Moliné, A.; Silva, J.P.; Navarro Sarró, D.F.; Cardalliaguet Guerra, M.; Serra-Sagristà, J.; Santangeli, A. Deep Learning for UAV Thermal Bird Detection: Benchmarking YOLOv8–YOLOv26. *Remote Sens. Appl.* **2026**, *42*, 101981. [\[CrossRef\]](#)
18. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *arXiv* **2020**, arXiv:1910.03151. [\[CrossRef\]](#)
19. Song, S.; Li, L.; He, S.; Xue, Y.; Pan, H.; Bao, R.; Cao, X.; Liu, G. Quantitative Characterization of Fractal Features of Coal Rock Pore-Fracture Based on ResUnet-CBAM Modeling. *Measurement* **2026**, *267*, 120609. [\[CrossRef\]](#)
20. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective Kernel Networks. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2019; pp. 510–519.
21. Wang, Y.; Jiang, Y.; Xu, H.Q.; Xiao, C.L.; Zhao, K. Detection Method of Key Ship Parts Based on YOLOv11. *Processes* **2025**, *13*, 201. [\[CrossRef\]](#)
22. Lei, M.; Li, S.; Wu, Y.; Hu, H.; Zhou, Y.; Zheng, X.; Ding, G.; Du, S.; Wu, Z.; Gao, Y. YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception. *arXiv* **2025**, arXiv:2506.17733.
23. Ma, L.; Zhao, D.; Wang, S.; Lv, Z.; Wang, M. Beyond Element-Level Understanding: Explicit Relational Understanding for GUI Agents. *Pattern Recognit.* **2026**, *176*, 113262. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.