RESEARCH ARTICLE

LBSD-YOLO: A Lightweight YOLOv10-Based Network with Multi-Attention Enhancement for Bridge Surface Defect Detection

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

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

Abstract

Bridge surface defect detection plays a critical role in ensuring traffic safety and facilitating infrastructure maintenance. A lightweight object detection network based on YOLOv10, termed LBSD-YOLO, is developed to achieve high detection accuracy while maintaining high efficiency for deployment on resource-constrained devices. The proposed framework consists of three main components: a feature extraction backbone, a feature fusion neck, and a detection head. In the backbone, the C2f_FEMA (C2f with Feature Enhancement and Multi-branch Attention) module and the LAEDS (Lightweight Adaptive Encoder–Decoder for Sampling) spatial attention module are incorporated to enhance multi-scale feature representation.The neck incorporates multi-scale feature fusion with an Efficient Multi-scale Attention (EMA) mechanism. In the detection head, a lightweight DP-Head structure is developed, variant integrated with the DAMF_CA coordinate attention for improved channel and spatial focus. Experiments are conducted on the self-built BDD-1234 dataset, which contains 6,617 high-resolution images covering six common bridge defect categories (cracks, spalling, exposed reinforcement, rust stains, efflorescence, and delamination). Compared to the baseline YOLOv10s, LBSD-YOLO reduces model size from 16.6 MB to 9.6 MB (42.2% reduction), computational complexity from 21.4 GFLOPs to 17.3 GFLOPs (19.2% reduction), and parameters from 7.2 M to 4.6 M (36.1% reduction), while achieving comparable detection performance (mAP@50 of 64.1% vs. 65.5%). The results demonstrate that LBSD-YOLO offers an efficient and accurate solution for real-time bridge defect detection on portable devices.

Keywords: bridge surface defect detection, lightweight object detection, LBSD-YOLO.

1 Introduction

Bridge defect detection plays a crucial role in traffic safety monitoring and maintenance management. In recent years, deep learning technology has

Submitted: 16 November 2025
Accepted: 22 January 2026
Published: 10 March 2026

Vol. 3, **No.** 1, 2026.
10.62762/TSCC.2025.718989

***Corresponding author:**
✉ Liyun Zhuang
zly428@126.com

Citation
Ji, R., Xu, Y., Wang, X., Zhuang, L., Zhang, X., Tang, X., & Shi, J. (2026). LBSD-YOLO: A Lightweight YOLOv10-Based Network with Multi-Attention Enhancement for Bridge Surface Defect Detection. *ICCK Transactions on Sensing, Communication, and Control*, 3(1), 39–53.

made significant progress in the field of target detection, providing a new technological approach for automated bridge defect detection. The existing bridge defect detection methods can be divided into traditional image processing methods and methods based on deep learning. Although traditional methods such as edge detection [1] and texture analysis have low computational cost, their accuracy is low under complex backgrounds and varying illumination, failing to meet practical engineering requirements. With the development of deep learning, CNN-based object detection framework is widely used in bridge defect detection. Among them, two-stage detectors such as FasterR-CNN [3] have high detection accuracy, but the computational complexity is high, and it is difficult to run in real time on resource-constrained devices ; single-stage detectors such as SSD [2] and YOLO series [4, 5] improve the detection speed, but the detection ability of small target defects and dense defects needs to be improved.

Current bridge defect detection methods suffer from several limitations: Firstly, most deep learning models have a large number of parameters and high computational complexity, and it is difficult to deploy on portable devices in real time such as mobile terminals [6, 7]; secondly, although the existing lightweight methods such as MobileNet and ShuffleNe reduce the complexity of the model [8], when they are directly applied to bridge defect detection, they often lead to a decrease in the detection accuracy of special defects such as fine cracks [9]; third, there are various types of bridge defects (such as exposed reinforcement, rust stain, crack, spalling, efflorescence, delamination), and it is difficult for the existing models to take into account the detection performance of different types of defects at the same time. Lastly, the field detection environment is complex and changeable, and the existing models are not robust enough under different light and weather conditions [11].

Within the BDD-1234 dataset, small-target defects such as cracks and spalling constitute a notably high proportion. Specifically, cracks with an area of less than 10 cm^2 account for 60% of all crack instances, leading to particularly poor performance of existing detection models on small targets. Furthermore, long-tail categories (e.g., exposed reinforcement and spalling), although comprising only about 15% of the total defect instances, are critically important in practical applications. The detection performance of existing models on these categories is significantly lower than that on common defects such as cracks. In experiments, the model achieves an mAP of 75% under sunny conditions, which drops to 65% in overcast environments. These issues highlight the practical challenges in bridge defect detection, especially the requirement for robustness under varying environmental conditions. Therefore, existing detection methods necessitate more robust and efficient models to address these challenges.

This manuscript is structured as follows. Section 2 systematically reviews the evolution of real-time object detection algorithms, with a particular focus on the YOLO series. It also examines related research and identifies existing challenges in the field of bridge and concrete defect detection. Section 3 elaborates on the proposed LBSD-YOLO network, detailing its overall architecture, the C2f_FEMA and LAEDS modules integrated into the backbone, the EMA_attention and DAMF_CA variants within the neck, and the lightweight v10Detect_DP-Head. Section 4 describes the experimental setup, presents ablation studies on the BDD-1234 dataset, and provides a comparative analysis and discussion of the results. Finally, Section 5 concludes the paper and outlines directions for future research.

2 Related Work

2.1 Evolution of YOLO Series for Real-Time Object Detection

Many classical object detection algorithms, such as Fast R-CNN [12] and R-CNN [13], employ a two-stage strategy. This workflow first extracts features from the input image and generates several candidate regions that may contain the target object. Subsequently, a classification network evaluates each candidate region to determine its authenticity and assign a corresponding category confidence score. The YOLO (You Only Look Once) series has become the mainstream method for object detection; however, each version still has limitations. For example, YOLOv4 exhibits limited capability in detecting small targets. YOLOv5 struggles to balance detection accuracy against model complexity/speed, while YOLOv8 may struggle to balance detection accuracy with computational resource consumption.

YOLOv10 is a relatively new version of the YOLO object detection algorithm family. On the basis of inheriting the efficient detection ability of the previous generation model, it further optimizes the network structure and computational efficiency [14]. Compared with the previous generation model,

YOLOv10 mainly has the following characteristics: Firstly, YOLO adopts a more efficient feature extraction backbone network that significantly reduces computational overhead by optimizing convolution operations [15, 16]; secondly, YOLOv10 introduces an improved detection head design, which improves the detection accuracy by separating positioning and classification tasks [17]; Finally, YOLOv10 is designed with hardware-aware optimization, enabling efficient deployment across diverse computing platforms [18, 19].

In recent years, lightweight detectors tailored for edge devices have continuously emerged, providing critical insights for efficient bridge defect detection. For instance, PP-PicoDet [20] achieves an excellent balance between accuracy and speed with minimal parameters, serving as a strong baseline for mobile deployment. The YOLO series itself has seen rapid architectural evolution: YOLOv7 [22] introduced a "bag-of-freebies" strategy that significantly enhanced performance without increasing inference cost, while the latest YOLO-World [21] extends the paradigm to open-vocabulary detection, demonstrating remarkable generalization. Beyond the YOLO family, transformer-based detectors like RT-DETR [23] have achieved state-of-the-art real-time performance, offering a novel and competitive architectural perspective. These advances collectively enrich the design space for developing specialized, lightweight models for bridge inspection.

2.2 Bridge and Concrete Defect Detection: Current Progress and Limitations

The design of efficient attention mechanisms remains a highly active research area, crucial for balancing model performance and computational overhead. Recent innovations include the Efficient Multi-scale Attention (EMA) module [15], which improves cross-spatial feature integration, and FcaNet [24], which incorporates frequency domain analysis to enhance channel attention. For scenarios demanding extreme parameter efficiency, the SimAM module [25] offers a parameter-free, yet effective, attention mechanism derived from neuroscience principles. These works build upon the foundational concept of combining spatial and channel attention, as exemplified by the widely adopted Convolutional Block Attention Module (CBAM) [26]. The ongoing refinement of these mechanisms provides a robust toolkit for enhancing feature representation in resource-constrained vision tasks like bridge defect detection.

As an important means to improve the performance of bridge defect detection, attention mechanism is widely used in the field of object detection. Squeeze-and-Excitation Networks (SENet) proposes a Squeeze-and-Excitation structure, which obtains channel attention weights through global average pooling, which helps to enhance the expression of linear features such as cracks; Convolutional Block Attention Module (CBAM) organically combines channel attention and spatial attention mechanisms, and is embedded in CNN network as a plug-and-play module, which effectively improves the detection ability of area defects such as spalling; Coordinate Attention (CA) effectively captures position information through the coordinate attention mechanism, avoiding information loss caused by feature compression, and is especially suitable for dealing with complex textures of bridge surfaces .

DAMF_CA and EMA_attention modules are designed, where DAMF_CA enhances the capture of defect details, such as bridge cracks and spalling, by combining discriminative and average features, while EMA_attention focuses on maintaining the original morphological feature information of defects. For the detection head design, we adopt the v10Detect_DP-Head structure to achieve a balanced trade-off between detection accuracy and computational cost is achieved by adjusting the number of channels. At the same time, the CIB module is introduced to collaboratively optimize deep and shallow features, which significantly improves the detection robustness of the model in complex lighting, different weather and other environments.

Motivated by these observations, a YOLOv10s-based lightweight detection algorithm, LBSD-YOLO, is developed with a focus on architectural efficiency and optimized computational resource utilization. Compared with the original model, the size of the model is reduced by 42.2% and the amount of calculation is reduced by 19.2% while maintaining the detection accuracy, which is especially suitable for real-time deployment on portable detection equipment, and provides a new practical solution for the field of bridge defect detection.

To address the poor real-time performance of existing bridge defect detection algorithms, a lightweight approach is designed to significantly reduce computational cost and parameter count while maintaining detection accuracy. Based on the

YOLOv10 benchmark framework, A lightweight and efficient bridge defect detection algorithm LBSD-YOLO is proposed, The algorithm is designed with two key objectives: lightweight network design and model compression, which makes it more suitable for real-time processing of portable detection equipment. The main innovative work of this algorithm is as follows:

A feature collaborative awareness multi-attention mechanism is proposed, which enhances the feature extraction capability for different types of bridge defects. This is achieved through the synergy of DAMF_CA and EMA_attention modules, and improves the accuracy and robustness of defect detection in complex environments.

A lightweight and efficient LBSD-YOLO framework is designed. Through the innovative design of modules such as C2f_FEMA, LAEDS and v10Detect_DP-Head, the parameter quantity and computational complexity of the model are significantly reduced while maintaining detection accuracy. It is more suitable for deployment on devices with limited computing power.

Rather than introducing a new basic module, this work focuses on system-level redesign and optimization of the YOLOv10s architecture to balance lightweight deployment and detection accuracy in bridge defect detection. This work builds upon existing efficient modules (such as coordinate attention and multi-scale attention), and creatively combines and adapts them into the detection framework to form a dedicated LBSD-YOLO model.

3 Proposed Method: LBSD-YOLO

3.1 Overall Architecture

With its excellent hardware adaptability and lightweight characteristics, the YOLOv10 network has lower parameters and calculation complexity than YOLOv5 and YOLOv8, making it especially suitable for bridge defect detection tasks. The overall LBSD-YOLO framework is illustrated in Figure 1, which is mainly composed of three parts: feature extraction Backbone network (Backbone), feature fusion Neck network (Neck) and Detection Head.

The backbone network is enhanced with two key components: an improved C2f_FEMA module and a novel LAEDS spatial attention module. Specifically, the C2f_FEMA module (C2f with Feature Enhancement and Multi-branch Attention) boosts feature extraction by employing feature shunting and adaptive fusion mechanisms. Concurrently, the LAEDS module augments feature representation across spatial dimensions.The neck network adopts a multi-level fusion architecture, enhances feature expression through multi-scale feature pyramid structure and EMA_attention mechanism, and innovatively introduces Spatial Channel Down-sampling (SC Down) to achieve feature dimensionality reduction, cooperating with the residual learning structure of C2fCIB module to ensure the effective transfer of features. The detection head employs a lightweight v10Detect_DP-Head structure with 256 channels. It integrates the DAMF_CA channel attention mechanism and a sensitivity-based adaptive channel pruning strategy. Multi-scale prediction is performed on three feature layers (P3, P4, and P5), balancing computational efficiency and detection accuracy.

Through multi-level feature extraction and fusion mechanism, the network structure achieves 64.1% mAP50 detection accuracy on the BDD-1234 dataset, and shows stable detection performance for different types and scales of bridge defects (such as six kinds of bridge defects such as cracks, spalling and rust stain). High detection speed can be achieved on portable detection devices, requiring only 9.6 MB of storage space and 17.3 GFLOPs of calculation, significantly lowering the deployment threshold. The optimized network structure is especially suitable for real-time detection in complex field environment, and provides an efficient and reliable technical solution for daily maintenance, regular detection and safety monitoring of bridges, and has important engineering application value. Figure 1 shows the overall network structure diagram.

3.2 Backbone Enhancements

3.2.1 C2f_FEMA Module

To address inadequate feature extraction and information loss in bridge defect detection, this study designs C2f_FEMA module based on YOLO network. As shown in Figure 2, this module adopts an innovative tandem structure, including key components such as ConvBNSiLU, split operation.

EMA_attention mechanism, multiple Faster Blocks, and feature fusion. This module not only inherits the advantages of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures, But also enhance that detection ability of six categories of defects including crack, spalling, delamination . As shown in Figure 2, the input features are first

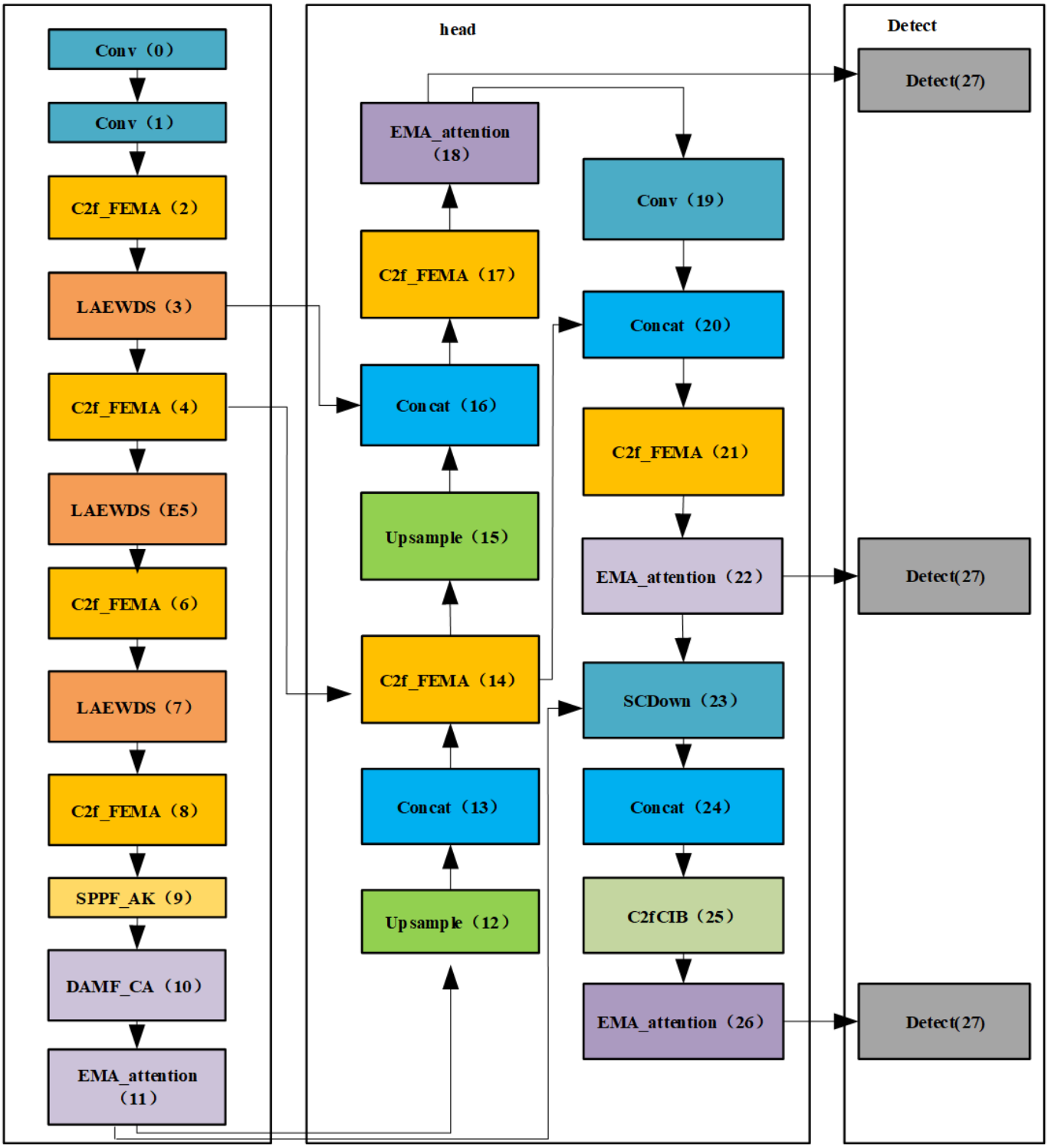

Figure 1. Overall network structure diagram.

processed by a ConvBNSiLU layer. Subsequently, a split operation divides the features into two streams. One stream is fed into the EMA_attention module, while the other is processed by multiple parallel FasterBlocks. This shunt design can maintain the original feature information and enhance the feature expression ability at the same time.

Multi-branch parallel processing: Figure 2 shows three parallel Faster Blocks, which act together on the characteristics after split; Each FasterBlock can independently extract feature information of different scales and levels; The parallel structure significantly improves the feature extraction ability and processing efficiency of the module.

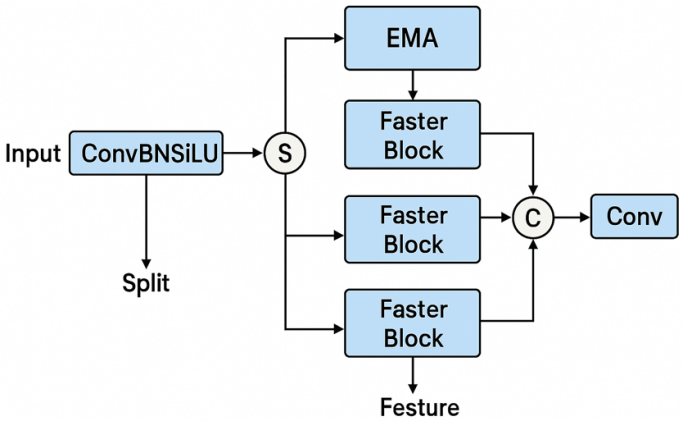

Figure 2. C2f_FEMA Module.

EMA_attention and feature fusion: As shown in Figure 2, the EMA module receives the feature input of split and enhances feature expression through the attention mechanism; All branches (including Faster Block output and EMA output) finally undergo feature fusion at the Concat node ("C" in Figure 2); The fused features are integrated again through Convulsion to form the final output.

3.2.2 LAEDS Spatial Attention Module

After the above module processing, the feature map has fully considered the feature representation of bridge defects. In the baseline YOLOv10s network structure, The direct use of step convolution or maximum pooling for downsampling limits the ability to capture defect features, These include surface spalling and efflorescence at different scales.

The LAEDS module in this study, as a key component in the network backbone, is deployed at three feature levels: P3 (step size is 8), P4 (step size is 16) and P5 (step size is 32) respectively. Through the unique dual-branch structure design, it not only realizes efficient downsampling function, but more importantly, it can adaptively process feature maps of different scales, which helps to enhance the bridge defect feature representation at each scale.

The LAEDS module is one of the key components of the LBSD-YOLO algorithm. The module is mainly composed of two parallel branches, namely defect attention branch and defect feature transformation branch. In the defect attention branch, by learning the attention weights on the feature maps of different scales, The salient features related to defects can be highlighted, Irrelevant background information can be effectively suppressed. This attention mechanism can accurately capture and enhance defect features at various scales, providing key clues for subsequent detection.

In the defect feature transformation branch, An adaptive feature transformation mechanism is adopted, It intelligently adjusts the resolution of feature maps to better match defects of different scales. This dynamic adjustment enables the LAEDS module to more fully extract the discriminant features of multi-scale defects, significantly enhancing the model's perception of various size defects.

$$F_{\text{defect\_pool}} = \text{AvgPool2d} \tag{1}$$

$$F_{\text{defect\_channel}} = \text{Conv3x3} \tag{2}$$

where $\mathbb{R}^{C\times H\times W}$ is the input bridge defect feature map, and the attention weight is formed after Rearrange and Softmax operations.

The outputs of the two branches are fed to the detection head after feature fusion. This multi-scale feature enhancement and fusion mechanism ensures comprehensive and accurate capture of various defect features. This applies to the LBSD-YOLO algorithm, laying a solid foundation for improving detection accuracy. The detailed structure of the LAEDS module is illustrated in Figure 3.

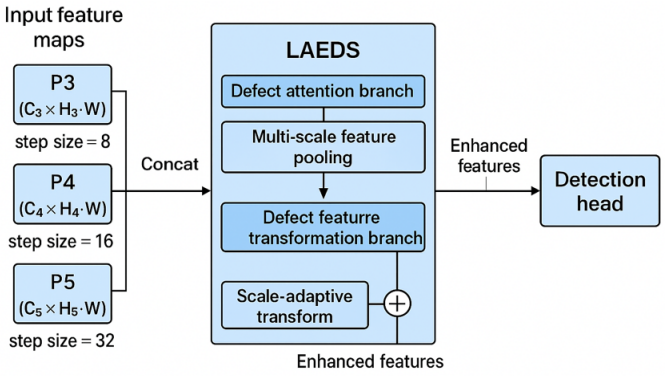

Figure 3. LAEDS module.

From the theoretical basis, this LAEDS module design optimized for bridge defect detection is based on the attention mechanism theory in deep learning. Through feature pooling and adaptive weight learning, it effectively improves the model's ability to perceive defect areas.

3.3 Neck Enhancements

3.3.1 EMA_attention Mechanism

In recent years, efficient channel attention mechanisms have gained significant attention in lightweight object detection. For instance, ECA-Net [10] employs adaptive 1D convolution to achieve extremely

low-parameter cross-channel interaction while avoiding the dimensionality reduction loss typical of traditional SE modules, thereby substantially improving computational efficiency on edge devices. The EMA-attention module focuses on maintaining the original characteristic information of bridge defects, expanding the receptive field range. As shown in Figure 4, the module receives input features with dimensions $C \times H \times W$, divides the features into C/G Groups through Groups operation, and then processes them through three parallel branches, each of which is designed for feature extraction of different scales. The specific structure includes the following key parts:

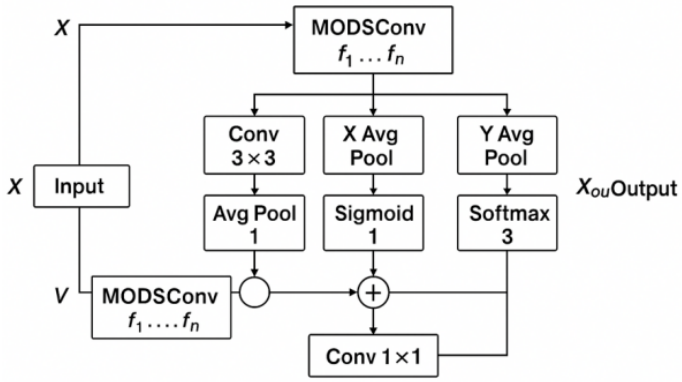

Figure 4. EMA_attention module.

Feature grouping and parallel processing:

- **Left branch:** Feature extraction by Conv(3×3) and Avg Pool 1, and finally normalization by Softmax 1.
- **Middle branch:** using X Avg Pool followed by Concat + Conv (1×1), through Sigmoid 1 and Avg Pool 2, and finally through Softmax 2 treatment.
- **Right branch:** After using Y Avg Pool, it goes through a similar processing flow, passing through Sigmoid 2 and Avg Pool 3, and finally going through Softmax 3 processing.

The module continuously extracts the input features through multi-layer MODSConv $(f_1, ..., f_n)$, which keeps the original dimensional information of the defect features. Especially when dealing with slender defects such as surface spalling, the information loss caused by the reduction of feature dimensions is avoided. The original surface texture information is maintained by residual connection $(X_1 + X)$, and finally the feature integration is carried out by 1×1 convolution operation f to generate the final enhanced feature X_{out}. This design enables the module to better maintain the complete morphological characteristics of bridge defects, It also improves the detection accuracy.

By integrating the DAMF_CA and EMA_attention modules, the proposed neck network achieves a synergistic enhancement of feature representation. Specifically, DAMF_CA captures salient defect features and their contextual information, while EMA_attention, through its multi-branch parallel structure (as shown in Figure 4), enriches feature expression across different scales and expands the receptive field. This collaborative mechanism significantly improves the model's adaptability to complex bridge surface conditions, leading to more robust and accurate defect detection in practical applications, particularly for deployment on mobile and edge devices.

3.3.2 DAMF_CA Coordinate Attention Variant

In bridge defect detection, the six categories of defects (e.g., cracks, spalling, exposed reinforcement) vary greatly in shape and scale, and are susceptible to interference from surface textures and lighting. This necessitates a detection model with exceptionally robust feature extraction capabilities. To solve these challenges, an adaptive attention module (DAMF_CA) is incorporated, which modifies coordinated attention with double pooling to replace the original PSA module for more focused feature extraction. As can be seen in Figure 5, the DAMF_CA module, as a key component of the network, realizes multi-dimensional enhancement of input features by simultaneously processing features in the height and width directions during the feature extraction process.

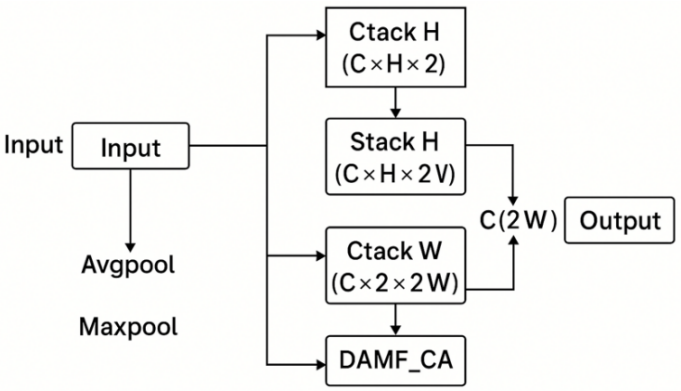

Figure 5. DAMF_CA module.

The DAMF_CA module is an improvement of coordinate attention, designed to consider both the salient features of bridge defects and the environmental information of the surrounding

structures. As shown in Figure 5, the input feature dimension processed by the module maintains the same output dimension after processing. The whole processing process is divided into two key parts:

- **Two-way feature pooling:** The left side of Figure 5 shows that the input features undergo average pooling (Avgpool) and max pooling (Maxpool) in the height and width dimensions, respectively, to generate feature layers with spatial dimensions of $C \times H \times 1$ and $C \times 1 \times W$. This dual-way pooling design helps to simultaneously capture the local details of linear defects such as cracks, as well as the overall texture information of the concrete surface.
- **Dimension conversion and stacking:** As shown in Figure 5, Through Stack H and Stack W operations, features of the same dimension are stacked and grouped. This generates feature layers of $C \times H \times 2$ and $C \times 2 \times W$, thus realizing the effective integration of height and width direction features.

3.4 Lightweight Detection Head: v10Detect_DP-Head

The core design of convolutional neural networks lies in their unique local perception mechanism and parameter reuse strategy. This design not only significantly reduces the parameters of the model, but also enables the network to effectively learn hierarchical feature representations. Aiming at the specialized field of bridge defect detection, the v10Detect_DP-Head detection head not only inherits this basic concept, It also innovatively introduces the fusion convolution structure and dual-channel concurrent architecture. This achieves diversified detection of exposed reinforcement and rust stain, etc. Accurate identification of bridge damage. As shown in Figure 6, the detection head follows a complete pipeline architecture. It first processes input features with an initial 1×1 GN module, then channel alignment through the channel alignment module, and finally feature enhancement using shared convolutional feature enhancement. An innovative dual-channel concurrency structure is adopted in the core part of the network, including a Multi-target Branch that handles multi-target detection tasks and a Single-target Branch that focuses on single target detection. Both branches contain Position Prediction and Class Prediction modules. Feature extraction is carried out through BoxHead and GeoHead, and the Feature Fusion module is used to achieve effective fusion of multi-layer features, and the Distribution Expression module further enhances feature expression. Finally, the detection head decodes features through GFL Decoder, and outputs Output Features as the detection result. This innovative dual-channel architecture design can not only handle different types of detection tasks at the same time, but also improve the efficiency of the model through feature reuse and parallel computing. The feature enhancement module and multi-level fusion mechanism further improve the model's detection accuracy of various bridge defects. Through experimental verification, the architecture shows superior performance in various bridge defect detection tasks.

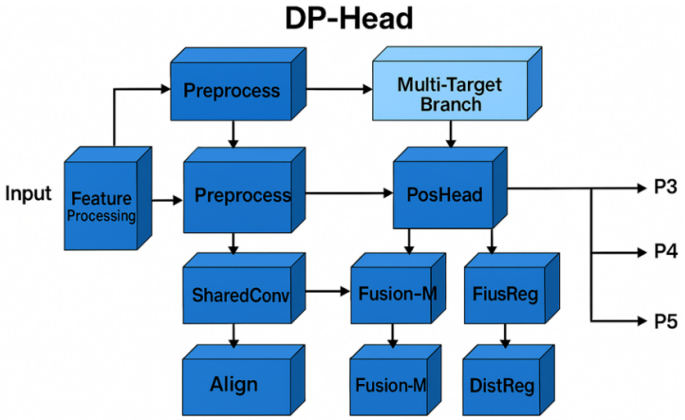

Figure 6. DP-Head detection head.

When the input features of P3/P4/P5 scale enter the network, they are first preprocessed by Features Preprocessing, and then channel adjustment is performed by Initial 1×1 convolution and GN module. The input feature dimensions are expressed as:

$$F_i \in \mathbb{R}^{H \times W \times C} \tag{3}$$

where F_i is the input feature map, H and W represent the height and width of the feature map respectively, and C is the number of channels.

After Feature Enhancement, the network shown in Figure 6 is divided into two parallel processing branches: Multi-target Branch and Single-target Branch. The mapping process of multi-objective branches is expressed as:

$$M = \sigma_m(F_s) \tag{4}$$

$$B_m = \Psi_b(M) \tag{5}$$

$$C_m = \Psi_c(M) \tag{6}$$

where σ_m is a multi-objective feature mapping function, and Ψ_b and Ψ_c are position location and

category recognition functions respectively, which is especially suitable for dealing with dense crack detection on bridge surface. Equation (4) represents multi-object feature mapping, focusing on dense object detection; Equation (5) represents bounding box positioning using bounding box prediction function; Equation (6) represents category identification using a category prediction function.

As shown in Figure 6, the features of the two branches are fused in the Feature Fusion module (Fusion M and Fusion O), then Distributed Regression is performed, and finally Output Features P3/P4/P5 are generated through the DFL Decoder. The DFL decoding process is expressed as:

$$D(x) = \sum (s_i \otimes p_i) \tag{7}$$

where it represents position coding and prediction distribution. This decoding method improves the positioning accuracy of bridge defects.

It can be clearly seen from Figure 6 that DP-Head's feature processing strategy constructs a complete feature extraction and prediction pipeline. Feature enhancement is achieved by Channel Adjustment and Shared Conv Block, and the dual-branch structure handles multi-object and single-object detection tasks respectively. Finally, accurate defect detection is achieved by feature fusion and distributed regression. This multi-level feature processing and dual-branch parallel architecture design not only optimize the computational efficiency, but also significantly enhance the detection head's ability to recognize multi-scale bridge defects, laying a foundation for realizing high-precision defect detection.

4 Experiments

In this study, only six defect categories defined in Section 4.1 are used for evaluation. Benchmark evaluations are conducted on the BDD-1234 bridge defect detection dataset. This data set contains a large number of defect images collected in real bridge environment, covering many typical defect types, such as surface crack, crack, concrete carbonation, component structure corrosion, concrete fracture damage, mixed metal component corrosion, concrete unqualified quality, concrete erosion, surface spalling and salinization. The data comes from bridges in different seasons and environmental conditions, which ensures the diversity and authenticity of defect images. YOLOv10, which exhibits good hardware compatibility, is selected as the benchmark framework and YOLOv10 provides three models with different scales according to the width and depth of the network: YOLOv10n, YOLOv10s, YOLOv10l. These models have different balances between detection accuracy, inference speed, and computational resource consumption. Through preliminary experiments, Experimental results indicate that YOLOv10s achieves a good balance among deployment speed, defect detection accuracy and model size of portable detection devices, and is especially suitable for real-time detection on mobile terminals and embedded devices. Therefore, this study takes YOLOv10s as the basic model, improves and optimizes it for the special needs of bridge defect detection, and combines more accurate feature extraction and efficient reasoning strategies to improve the performance and deployment efficiency of the model in practical applications.

4.1 BDD-1234 Bridge Defect Dataset

The BDD-1234 dataset is a self-collected dataset for bridge defect detection, containing 6,617 high-resolution images. Each image is annotated with bounding boxes for six common types of bridge surface defects: cracks, spalling, exposed reinforcement, rust stains, delamination, and efflorescence. The images were captured under varied lighting and weather conditions from multiple angles to simulate real inspection scenarios. Annotations follow the standard YOLO format. To enhance the model's robustness in complex environments, the dataset also includes a small number of images containing non-bridge structures as background. The dataset is split into training and validation sets in an 8:1 ratio. All images are resized to a uniform resolution of 640 × 640 pixels, and bounding box coordinates are normalized. Although the experiments primarily focus on the BDD-1234 dataset to validate the proposed method, evaluating the generalization ability of the model on publicly available bridge defect datasets will provide additional insights. This direction represents an important avenue for future research.

4.2 Experiment

4.2.1 Implementation Details

To verify the effectiveness of the proposed algorithm, a corresponding experimental platform is established. The operating system of the experimental platform is Linux, the CPU is Intel (R) Xeon (R) CPU E5-2699 v4 @ 3.78 GHz (88 cores), the memory is 378GB, and the GPU is NVIDIA GeForce RTX 4090 (24GB video memory), the deep learning framework is PyTorch (CUDA 12.2). The specific experimental parameter

configuration is shown in Table 1.

Table 1. Experimental parameter.

Enter image size	640×640
Momentum	0.9
Batch Size	4
Training rounds	250
Optimizer	SGD
Learning rate	0.01

4.2.2 *Evaluation Metrics*

To comprehensively evaluate the performance of the proposed method, several standardized evaluation indicators are used. Three detection accuracy indexes are preferred: Precision (P), Recall (R) and Weighted Average Precision (wAP).

$$P = \frac{TP}{TP + \alpha \times FP} \tag{8}$$

$$\int_0^1 w(R) \times p(R)\, dR \tag{9}$$

In the above formula, TP (True Positive) represents the number of targets correctly identified by the model, FP (False Positive) represents the number of false detections, and FN (False Negative) represents the number of missed detections. α and β are balancing factors (default value is 1), which are used to adjust the weights of precision and recall in different application scenarios. $w(R)$ is the weight function of the recall rate and w_i is the weight coefficient of the i-th class.

4.3 Ablation Studies

Sixteen sets of ablation experiments were conducted to evaluate the contribution of each module, which are evaluated from two aspects: accuracy improvement and model lightweight. Based on baseline, C2f_FEMA, LAEDS, DAM + FMA and DP-HEAD modules were added or removed in different combinations, and the independent and synergistic effects of each module were analyzed. The evaluation indicators include mAP50%, mAP50:95%, parameter quantity, calculation quantity and model size (MB), which provide guidance for framework optimization. The experiment is carried out under the same hardware environment and parameter configuration, and the control variable method is used to ensure the objectivity of evaluation, which comprehensively verifies the advantages of LBSD-YOLO in improving detection accuracy and optimizing computing resources.

Through the systematic analysis of experiments A to Q, the effectiveness of each module and their combined optimization effects are systematically evaluated. Experiment A shows that after introducing the C2f_FEMA module, the model parameters decrease from 7.2×10^6 to 6.2×10^6, the computation decreases from 21.4 GFLOPs to 16.4 GFLOPs, and the detection accuracy (mAP50%) slightly improves from 65.5% to 66.1%, indicating that the module effectively reduces computational cost while maintaining detection performance. Experiment B adjusts the neck structure and introduces the LAEDS module, where mAP50% becomes 65.9%. Although the parameters (7.3×10^6) and computation (21.8 GFLOPs) increase compared with Experiment A, the detection accuracy remains stable, showing the module's contribution to feature extraction.

A–Q represent the ablation experimental configuration with different module combinations, and ✓ indicates that this module is used.

After introducing the C2f_FEMA module in Experiment A, the amount of model parameters decreased from 7.2×10^6 of the baseline to 6.2×10^6, the amount of calculation decreased from 21.4 GFLOPs to 16.4 GFLOPs, and the detection accuracy (mAP50%) increased slightly from 65.5% to 66.1%, indicating that the module effectively reduces the computational cost while maintaining the detection performance. Experiment B demonstrated the enhancement effect of this module on the feature extraction ability. It combines the LAEDS module to make the mAP50% reach 65.9%, the parameter amount is 7.3×10^6, and the calculation amount is 21.8 GFLOPs. The results of experiments C to N show that different combinations of each module have a significant effect on the model performance. Among them, experiment K combined C2f_FEMA and LAEDS, and DAM + EMA module, obtained 61.4% mAP50% and 40.9% mAP50-95%, with a parameter amount of 7.7×10^6, a computational amount of 28.1 GFLOPs, and a model size of 23.0 MB. Experiment N combines the C2f_FEMA module and the DAM + EMA module, as well as the DP-Head detection head, and obtains an mAP50% of 61.5%, the amount of parameters is reduced to 5.2×10^6, the amount of computation is 17.3 GFLOPs, and the model size is 10.7 MB.

The final selected Q scheme (LBSD-YOLO) integrates all four modules: C2f_FEMA, LAEDS, DAM + EMA, and DP-Head, with mAP50% reaching 64.1%, mAP50-95% reaching 44.1%, parameter volume $4.6 \times$

Table 2. Ablation experiment.

EXP	C2f_FEMA	LAEDS	DAM+EMA	DP-Head	mAP50%	mAP50:95%	Params/10^6	GFLOPs	Size/MB
Baseline	×	×	×	×	65.5	45.6	7.2	21.4	16.6
A	✓	×	×	×	66.1	45.7	6.2	16.4	14.5
B	×	✓	×	×	65.9	45.6	7.3	21.8	16.6
C	×	×	✓	×	66.9	45.2	7.3	30.9	18.6
D	×	×	×	✓	65.9	45.6	6.8	24.0	14.0
E	✓	✓	×	×	63.6	43.1	6.2	16.8	14.6
F	✓	×	✓	×	61.2	40.3	6.3	25.8	16.5
G	✓	×	×	✓	64.7	44.5	5.8	18.9	12.0
H	×	✓	✓	×	65.6	44.5	7.4	31.3	18.7
I	×	✓	×	✓	66.1	44.9	7.4	31.3	18.7
J	×	×	✓	✓	63.3	41.6	6.1	25.3	12.5
K	✓	✓	✓	×	61.4	40.9	7.7	28.1	23.0
L	✓	✓	×	✓	61.1	40.6	5.8	19.3	12.0
M	×	✓	✓	✓	61.4	39.9	6.5	24.5	13.4
N	✓	×	✓	✓	61.5	40.9	5.2	19.4	10.7
Q	✓	✓	✓	✓	64.1	44.1	4.6	17.3	9.6

Figure 7. Plot of prediction results on BDD-1234.

10^6, computational volume 17.3 GFLOPs, and a model size of only 9.6 MB. Compared with the baseline, the amount of parameters is reduced by 36.1%, the amount of computation is reduced by 19.2%, and the model size is reduced by 42.2%. At the same time, the detection accuracy decreases slightly from the baseline 65.5% / 45.6% (mAP50% / mAP50-95%) to 64.1% / 44.1%. The Q scheme was chosen as the final model mainly because of its optimal balance between resource consumption and performance. Although the mAP50% is only 1.4 percentage points lower than the baseline, its significant advantages in computation and model size make it more suitable for mobile deployment.

Figure 7 presents visual detection results of the proposed LBSD-YOLO on sample images from the BDD-1234 dataset, demonstrating its effectiveness in locating and classifying various bridge surface defects under different scenarios.

Based on the experimental results in Table 2, the proposed LBSD-YOLO algorithm demonstrates strong

performance in concrete structure defect detection and identification. In the concrete erosion scenario, the LBSD-YOLO algorithm can accurately detect and locate the damaged area on the concrete surface, and circle it with a bounding box, indicating that it has excellent target detection ability. At the same time, in the scene of analyzing concrete carbonation, the algorithm can also accurately detect and mark the carbonation area, providing support for engineers to evaluate the corrosion degree of concrete structures. When analyzing the images of concrete microscopic components, the algorithm may use these fine-grained visual information to further improve the ability of identifying and classifying concrete material features. Generally speaking, LBSD-YOLO algorithm shows excellent target detection and material identification capabilities, and can accurately locate and identify various defects and characteristics on the surface and inside of concrete structures. It will definitely provide strong support for the condition monitoring and fault diagnosis of concrete structures, greatly improve the safety and reliability of infrastructure, This provides reliable technical reference for the application of intelligent monitoring and prevention technology. The technology is based on computer vision in the engineering field.

4.4 Per-Class Performance Analysis

As shown in Table 3, although the overall mAP50% of LBSD-YOLO decreased slightly compared to YOLOv10s, it achieved significant improvements in detecting critical and challenging defect types. Specifically, the AP increased by 1.5% and 1.0% for "cracks" (typical small target defects) and "exposed reinforcement" (long tail category), respectively. This indicates that the proposed feature co-perception and multi-attention mechanisms effectively enhance the sensitivity of the model to these difficult-to-detect defects. The performance trade-off observed for other categories is a reasonable cost, which is consistent with the primary goal of model lightweighting, achieving a 42.2% reduction in model size and significant savings in computational resources.

4.5 Comparison with State-of-the-Art Methods

To verify the performance of the proposed algorithm, LBSD-YOLO is compared with the mainstream model and the lightweight detector PP-PicoDet-L. As shown in Table 4, compared to YOLOv10s, The proposed model reduces the number of parameters by 36.1%, computational cost by 19.2%, and model size by 42.2%, while achieving comparable accuracy (64.1% mAP50). LBSD-YOLO also showed a smaller size (9.6 MB vs. 13.0 MB) than YOLOv5s, YOLOv8s, and PP-PicoDet-L, demonstrating a better balance.

4.6 Qualitative Analysis and Failure Cases

To visually evaluate the effectiveness of the proposed LBSD-YOLO framework, qualitative comparisons with baseline methods are presented in Figure 8. The visualization results indicate that the proposed method produces more focused and discriminative activation regions when detecting bridge surface defects. In particular, the integration of the DAMF_CA and EMA_attention mechanisms enhances the model's ability to emphasize defect-related regions while suppressing background interference.

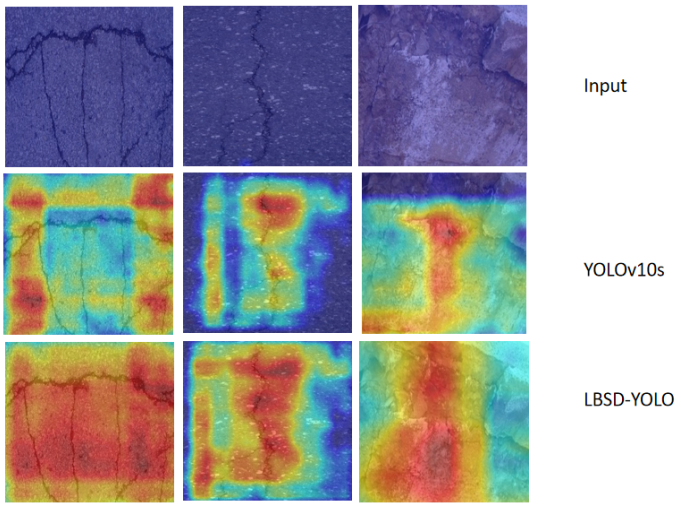

Figure 8. Comparative heat map.

Compared with baseline detectors, LBSD-YOLO demonstrates improved localization accuracy for fine-grained defects, such as cracks and spalling, especially in complex backgrounds. The attention-enhanced feature representations contribute to clearer boundary delineation and more consistent confidence responses across defect regions.

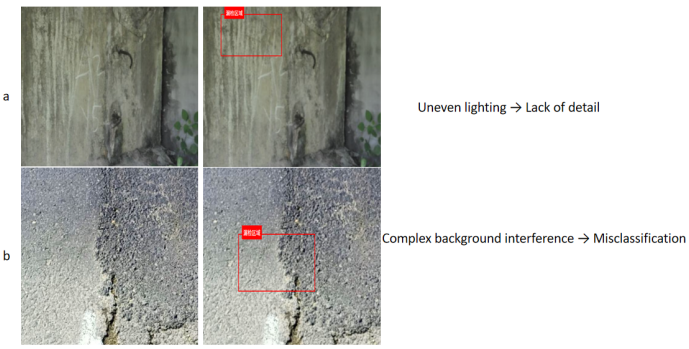

Figure 9. Comparative heat map.

Table 3. Per-class AP@0.5 comparison on BDD-1234 dataset.

Category	YOLOv10s (AP@50%)	LBSD-YOLO (AP@50%)	Improvement (ΔAP)
Crack	65.0	66.5	+1.5
Spalling	67.5	64.5	-3.0
Exposed Reinforcement	62.5	63.5	+1.0
Rust Stain	66.5	63.5	-3.0
Delamination	64.5	61.5	-3.0
Efflorescence	67.5	65.1	-2.4
Mean (mAP50)	65.5	64.1	-1.4

Table 4. Comparative experiments of different algorithms.

Algorithm	mAP50/%	mAP50-95%	Params/10^6	GFLOPs	Size/MB
SSD	55.7	-	26.3	62.7	413.9
Faster R-CNN	53.1	-	137.1	370.2	523.0
YOLOv3-tiny	64.5	40.7	12.1	18.9	24.4
YOLOv5s	68.5	47.9	8.0	20.6	16.5
YOLOv6s	69.2	49.7	16.0	43.3	32.4
YOLOv8s	72.1	51.8	10.1	25.2	20.5
YOLOv10s	65.5	45.6	7.2	21.4	16.6
PP-PicoDet-L	63.0	42.0	5.8	16.8	13.0
LBSD-YOLO	64.1	44.1	4.6	17.3	9.6

To further understand the model's limitations, typical failure cases are analyzed. Figure 9 presents two challenging scenarios: (a) Uneven lighting → Lack of detail, where inconsistent lighting conditions cause missed detection of fine cracks; (b) Complex background interference → Misclassification, where intricate background textures lead to misclassification of cracks as background noise. These cases highlight the persistent challenges in real-world bridge inspection and point to directions for future work, such as enhancing the model's robustness to lighting variations and background noise.

Overall, the qualitative analysis confirms that the proposed attention-enhanced architecture improves feature discrimination and robustness in most scenarios, while also highlighting directions for future improvement, particularly in addressing rare and highly challenging defect cases.

5 Conclusion

This study presents LBSD-YOLO, a lightweight YOLOv10-based detection framework designed for bridge surface defect inspection under resource-constrained conditions. By incorporating multi-attention mechanisms and an optimized DP-Head structure, the proposed framework aims to achieve a balanced trade-off between detection accuracy and computational efficiency.

Experimental results on the BDD-1234 dataset demonstrate that LBSD-YOLO achieves competitive detection performance while significantly reducing model complexity and inference cost. The attention-enhanced feature extraction strategy improves the model's ability to capture fine-grained defect characteristics, contributing to more accurate localization and classification of various defect types.

Ablation studies further verify the effectiveness of each proposed component, confirming that the integration of DAMF_CA, EMA_attention, and DP-Head jointly contributes to the observed performance gains. These results indicate that the proposed design is well-suited for real-time bridge inspection applications.

Although the current experiments focus on a single dataset, future work will extend the evaluation to publicly available bridge defect benchmarks and explore further improvements in robustness under extreme environmental conditions. Overall, the proposed LBSD-YOLO framework provides a practical and efficient solution for bridge surface defect detection.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

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

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence, 39*(6), 1137-1149. [CrossRef]

[2] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, September). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Cham: Springer International Publishing. [CrossRef]

[3] Li, R., Yu, J., Li, F., Yang, R., Wang, Y., & Peng, Z. (2023). Automatic bridge crack detection using Unmanned aerial vehicle and Faster R-CNN. *Construction and Building Materials, 362*, 129659. [CrossRef]

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). [CrossRef]

[5] Meng, Q., Hu, L., Wan, D., Li, M., Wu, H., Qi, X., & Tian, Y. (2023). Image-based concrete cracks identification under complex background with lightweight convolutional neural network. *KSCE journal of civil engineering, 27*(12), 5231-5242. [CrossRef]

[6] Spencer Jr, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering, 5*(2), 199-222. [CrossRef]

[7] Guan, B., & Li, J. (2024, April). Lightweight detection network for bridge defects based on model pruning and knowledge distillation. In *Structures* (Vol. 62, p. 106276). Elsevier. [CrossRef]

[8] Luo, Y., Ling, J., Wang, J., Zhang, H., Chen, F., Xiao, X., & Lu, N. (2025). SFW-YOLO: A lightweight multi-scale dynamic attention network for weld defect detection in steel bridge inspection. *Measurement, 253*, 117608. [CrossRef]

[9] Yang, Y., Li, L., Yao, G., Du, H., Chen, Y., & Wu, L. (2024). An modified intelligent real-time crack detection method for bridge based on improved target detection algorithm and transfer learning. *Frontiers in Materials, 11*, 1351938. [CrossRef]

[10] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020, June). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11531-11539). IEEE. [CrossRef]

[11] Chen, L., Yao, H., Fu, J., & Ng, C. T. (2023). The classification and localization of crack using lightweight convolutional neural network with CBAM. *Engineering Structures, 275*, 115291. [CrossRef]

[12] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440-1448). [CrossRef]

[13] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014, June). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587). [CrossRef]

[14] Zhang, C., Peng, N., Yan, J., Wang, L., Chen, Y., Zhou, Z., & Zhu, Y. (2024). A novel YOLOv10-DECA model for real-time detection of concrete cracks. *Buildings, 14*(10), 3230. [CrossRef]

[15] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023, June). Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1-5). IEEE. [CrossRef]

[16] Xiong, C., Zayed, T., Jiang, X., Alfalah, G., & Abelkader, E. M. (2024). A novel model for instance segmentation and quantification of bridge surface cracks—The YOLOv8-AFPN-MPD-IoU. *Sensors, 24*(13), 4288. [CrossRef]

[17] Du, F. J., & Jiao, S. J. (2022). Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection. *Sensors, 22*(9), 3537. [CrossRef]

[18] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024, December). YOLOv10: real-time end-to-end object detection. In *Proceedings of the 38th International Conference on Neural Information Processing Systems* (pp. 107984-108011).

[19] Chen, G., Choi, W., Yu, X., Han, T., & Chandraker, M. (2017). Learning efficient object detection models with knowledge distillation. *Advances in neural information*

processing systems, 30.

[20] Chen, J., Wen, Y., Nanehkaran, Y. A., Zhang, D., & Zeb, A. (2023). Multiscale attention networks for pavement defect detection. *IEEE transactions on instrumentation and measurement, 72*, 1-12. [CrossRef]

[21] Xu, W., Li, X., Ji, Y., Li, S., & Cui, C. (2024). BD-YOLOv8s: enhancing bridge defect detection with multidimensional attention and precision reconstruction. *Scientific Reports, 14*(1), 18673. [CrossRef]

[22] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023, June). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7464-7475). IEEE. [CrossRef]

[23] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., ... & Chen, J. (2024, June). DETRs Beat YOLOs on Real-time Object Detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 16965-16974). IEEE. [CrossRef]

[24] Qin, Z., Zhang, P., Wu, F., & Li, X. (2021, October). FcaNet: Frequency Channel Attention Networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 763-772). IEEE. [CrossRef]

[25] Yang, L., Zhang, R. Y., Li, L., & Xie, X. (2021, July). Simam: A simple, parameter-free attention module for convolutional neural networks. In *International conference on machine learning* (pp. 11863-11874). PMLR.

[26] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3-19). [CrossRef]

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

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

Xiaoyan Wang received the Ph.D. degree from Nanjing University of Aeronautics and Astronautics in 2020. She has been engaged in research work in the field of photoelectric detection for a long time and has carried out research in spectral detection and analysis. She has published more than twenty scientific research papers in domestic and foreign journals. (Email: wxygxy@163.com)

Liyun Zhuang received the M.S. degree from the China University of Mining and Technology, Xuzhou, China, in 2009, and the Ph.D. degree in communication and information system from Shanghai University, Shanghai, China. He is currently a Lecturer with the Faculty of Electronic and Information Engineering, Huaiyin Institute of Technology. His research interests include computer vision, pattern recognition. (Email: zly418@126.com)

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

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

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