ISSN 1392-3730/eISSN 1822-3605

JOURNAL of CIVIL ENGINEERING and MANAGEMENT

2026
Volume 32
Issue 1
Pages 119–132
https://doi.org/10.3846/jcem.2026.26166

YOLOv11n-CDL: ACCURATE AND LIGHTWEIGHT PAVEMENT DEFECT DETECTION VIA ENHANCED MULTI-SCALE ATTENTION AND FEATURE FUSION

Jun DAI[1]✉, Yanyang GAO[2]

[1]Enterprise Technology Center, Zhongbei Transportation Construction Group Co., Lt, 710075 Xi'an, China
[2]School of Highway, Chang'an University, 710064 Xi'an, China

Article History:
- received 5 August 2025
- accepted 10 December 2025

Abstract. Pavement defect detection requires both high accuracy and real-time performance in complex road environments, yet existing lightweight models often struggle with blurred textures, background interference, and small cracks. To address these limitations, this study proposes YOLOv11n-CDL, an enhanced lightweight detector integrating three targeted improvements. First, the ConvSmart module expands the receptive field and strengthens multi-scale feature extraction, improving the representation of defects of varying sizes. Second, a Double-Stage Attention (DSA) mechanism, embedded at the deepest backbone stage, iteratively highlights discriminative crack patterns while suppressing shadows, markings, and texture noise. Third, a P2-level small-object detection path provides high-resolution features that significantly improve sensitivity to fine cracks and micro-potholes. Experiments on IRRDD show that YOLOv11n-CDL achieves 75.3% mAP@0.5 and 44.6% mAP@0.5:0.95, outperforming the baseline by 3.0 and 1.1 percentage points, and exceeding YOLOv8n and YOLOv7-tiny in both precision and recall. Additional results on RDD2022 and low-power devices confirm strong generalization and real-time deployability. These improvements demonstrate that YOLOv11n-CDL offers an effective balance between accuracy, robustness, and efficiency for practical pavement inspection applications.

Keywords: pavement maintenance, repair, YOLOv11n, attention mechanism, pavement defect detection, multi-scale perception, lightweight convolution module, small object detection.

✉Corresponding author. E-mail: *daijun@ccnc.cc*

Abbreviations

DSA – Double Stage Attention Model;
FPS – Frames Per Second;
mAP – Mean Average Precision;
IoU – Intersection over Union.

1. Introduction

Traditional pavement defect detection methods, such as model-based and system-based inspection techniques (Dong et al., 2025b), largely rely on manual inspection. However, these approaches suffer from inherent limitations, including low efficiency, high labor costs, and insufficient accuracy. As traffic volumes continue to grow and road networks expand, manual inspection methods can no longer meet the demands of modern infrastructure management.

In recent years, artificial intelligence technologies have achieved remarkable advancements. From the introduction of Fast R-CNN (Girshick, 2015) to the evolution of Faster R-CNN (Ren et al., 2017), deep learning algorithms have rapidly progressed in the fields of image recognition and object detection, delivering significant breakthroughs. For instance, Chen et al. (2020) proposed the GF-CNN convolutional network, leveraging the sensitivity of Gabor filters to texture features in order to enhance the capability of CNNs in extracting pavement crack features. However, this approach was limited to defect classification and localization tasks. Lang et al. (2021) developed a CNN-based model named PCCNet, which comprises seven convolutional layers with activation functions and six pooling layers, demonstrating stable performance in detecting fine cracks. Furthermore, Ibragimov et al. (2022) improved the Faster R-CNN framework to achieve higher precision in detecting larger-scale pavement cracks. Wang et al. (2022) enhanced crack disease detection performance by training feature fusion layers based on the U-Net++ architecture. Similarly, Zhang et al. (2024b) refined the U-Net mod-

Copyright © 2026 The Author(s). Published by Vilnius Gediminas Technical University

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (*http://creativecommons.org/licenses/by/4.0/*), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

el by employing a ResNet50 backbone and embedding CNSN to improve crack feature enhancement. However, the increased parameter size limited real-time seg-mentation capabilities. Duan et al. (2024) introduced an effective method called EMFCHBNet, which integrates multi-scale feature fusion and cascaded optimization to reduce irrelevant noise interference and capture complex and diverse pavement crack patterns.

The YOLO (You Only Look Once) algorithm (Redmon et al., 2016) has emerged as a highly efficient object detection approach, characterized by its fast detection speed and high accuracy (Dong et al., 2025a). Compared to conventional two-stage object detection methods, YOLO can recognize transportation-related objects in images more rapidly and accurately, making it particularly suitable for real-time road defect detection scenarios. Du et al. (2021) constructed a large-scale dataset and conducted comparative experiments, concluding that the YOLO model demonstrates a detection efficiency nine times faster than Faster R-CNN and operates at only 70% of the computational cost of SSD. Since the introduction of the YOLO algorithm, its applications in pavement defect detection have garnered widespread research interest, with many scholars exploring different versions of YOLO for this task. For instance, Chen et al. (2019) employed infrared image preprocessing techniques and trained their infrared dataset using the Fast YOLOv1 model, demonstrating superior performance in both detection speed and accuracy compared to traditional digital image-based methods. Sun et al. (2020) compared a modified Faster R-CNN model enhanced with the VGG16 backbone to YOLOv2, aiming to evaluate the improved model's performance in detecting sealed pavement cracks. Miao et al. (2023) proposed a YOLOv3-based method incorporating the CBAM attention mechanism within a leaf-sweeping robotic system, exploring new directions for object detection under complex backgrounds. Qiu and Lau (2023) introduced a YOLOv4-tiny-based method integrated into unmanned aerial vehicles (UAVs), achieving remote, reliable, and rapid crack detection. Bai et al. (2025) further optimized the YOLO model to suit UAV-captured imagery, thereby validating the feasibility of UAV-based defect detection solutions. Shen et al. (2023) enhanced YOLOv5l by integrating the CA-plus attention mechanism and the ESPP feature fusion module, thereby improving crack detection accuracy; however, its model compactness remains limited. Zuo and Niu (2023) introduced the CBAM attention module into YOLOv5s to capture richer target features, though the model suffered from limited generalization capability. Xing et al. (2024) proposed replacing the conventional Intersection over Union (IoU) loss with the more efficient MPDIoU loss function, and substituted traditional convolution layers with the GCC3 module to improve detection performance on edge computing devices. He et al. (2024) similarly based their work on YOLOv5s, designing an SPPF-CSPC module to acquire receptive fields at various scales and enhance feature fusion, leading to improved performance in conventional crack inspection. Yao et al. (2024) focused on optimizing the neck network of YOLOv8n by integrating the efficient SCConv module into the C2f block and introducing the SimAM attention mechanism. Their approach significantly enhanced detection accuracy and speed in industrial applications; however, further validation is required to assess its scalability to complex transportation scenarios.

Zhou et al. (2024) redesigned the neck module of YOLOv8n by introducing the DysnakeConv convolutional operation to enhance the C2f-Dysnake module and iteratively optimize the neck structure using the RDFPN module. Additionally, the integration of the MPCA coordinate attention mechanism significantly improved the model's performance in high-precision pavement defect detection. Ding et al. (2024) decomposed large convolutional kernels into smaller ones and employed shift operations to strengthen feature extraction capabilities in the backbone network, thereby enhancing YOLOv8n's ability to detect both longitudinal and transverse cracks. Qin et al. (2025) introduced an enhanced feature pyramid network (EFFPN) into the YOLOv8n framework and optimized the detection heads, achieving improved accuracy while maintaining high computational efficiency. Wang et al. (2024d) addressed the relatively high parameter count of YOLOv8s by incorporating a deep feature pyramid network (DFPN) and designing lightweight detection heads, which improved detection precision at the expense of slightly increased inference time. Hu et al. (2024) based their approach on the YOLOv9 architecture, utilizing the Mamba aggregation feature extraction layer as the core component. They further refined Mamba into a lightweight algorithm, enhancing the model's capacity to extract global defect information across various pavement scales. Although this improved detection accuracy, the model's robustness in complex environments still requires further validation. Wang et al. (2024a) proposed a novel end-to-end architecture in YOLOv10, achieving a more favorable balance between detection accuracy and computational efficiency.

In summary, YOLO-based algorithms have demonstrated significant advantages in pavement defect detection by reducing computational complexity and storage requirements while maintaining high detection accuracy. Their application enables fast and precise identification of pavement defects, thereby improving both detection efficiency and reliability. Nevertheless, practical implementation reveals that factors such as viewpoint variations, complex backgrounds, and image noise considerably affect model training. Furthermore, the intrinsic variability of pavement defects introduces substantial challenges in feature extraction, demanding more robust and adaptive detection frameworks. To address the aforementioned challenges, this study introduces a series of optimizations to the YOLOv11 framework, aiming to improve its robustness and accuracy in pavement defect detection under complex conditions. The main contributions are summarized as follows:

(1) The convolutional modules in both the backbone and head of YOLOv11n are redesigned. Specifically, parallel 3×3 and 5×5 convolutional kernels are employed to capture features from different receptive fields. The outputs are concatenated and subsequently fused via residual connections, enhancing the network's capability to represent multi-scale features and improving its adaptability to varying defect patterns.

(2) A double-stage attention mechanism is proposed, comprising both channel attention (SE Attention) and spatial attention (Spatial Attention) in each stage. This architecture is designed to concurrently mine "important channel information" and "critical spatial location information". By guiding feature refinement across both dimensions and progressively integrating the attention outputs, the model effectively enhances its ability to distinguish complex defect features from noisy backgrounds, thereby improving recognition accuracy in real-world scenarios.

(3) A small object detection branch is introduced to better handle diminutive targets. Given the larger downsampling stride in YOLOv11n, deeper feature maps often lose critical information regarding small-sized defects. To compensate, an additional detection head is constructed using shallower feature maps, which are rich in fine-grained spatial details. These are then fused with deeper features to enhance detection precision and recall for small targets, reducing false negatives and improving performance under occlusions and scale variation.

2. Materials and methods

2.1. Baseline model

YOLOv11 (Ultralytics, 2025) is a next-generation object detection model introduced by the Ultralytics team on September 27, 2024, during the YOLO Vision 2024 event. The architecture of YOLOv11 is composed of four key components: the Input module, Backbone, Neck, and Head. In the Input stage, YOLOv11 leverages Mosaic data augmentation, which combines multiple images into a single composite image to enhance data diversity and improve model generalization. The Backbone consists of a sequence of Conv, C3k2, and SPPF modules, followed by a C2PSA block. Notably, the C3k2 module represents a core architectural innovation of YOLOv11. It inherits design ideas from the C2f module in YOLOv8 and the ELAN structure in YOLOv7, while incorporating residual structures from C3 modules. This integration enables the network to maintain lightweight properties while enhancing gradient flow and feature richness.

The SPPF (Spatial Pyramid Pooling – Fast) module effectively fuses multi-scale features, thereby improving the model's detection accuracy across varied object sizes. The C2PSA module, introduced in the Neck, integrates the C2f block with a PSA (Pointwise Spatial Attention) mechanism to strengthen the model's ability to focus on significant spatial information. The Neck adopts a combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN), which enhances semantic representation and localization capability across multiple feature scales. This structural design allows the model to perform robustly under complex environmental conditions. For the Head, YOLOv11 employs a decoupled detection head, separating classification and regression tasks to improve learning specialization. Moreover, an anchor-free detection paradigm is adopted, eliminating reliance on predefined anchor boxes and consequently accelerating inference. Compared with YOLOv8, YOLOv11 further integrates depthwise separable convolutions in the classification branch of the head to reduce both parameter count and computational overhead. The overall architecture of YOLOv11 is illustrated in Figure 1.

2.2. YOLOv11n-CDL model

While aiming to improve the accuracy of pavement defect detection and reduce the false detection rate, this study also focuses on enhancing the detection performance for small-scale targets and optimizing computational efficiency. In this study, a pavement defect detection model, termed YOLOv11n-CDL, is developed on the basis of the YOLOv11n architecture. The overall network structure is illustrated in Figure 2. To enhance multi-scale feature extraction, the conventional standard convolutional blocks are redesigned into the ConvSmart module, which strengthens the model's capability of capturing defect patterns of varying sizes and thereby improves the detection performance across diverse target scales. Furthermore, a dual-stage attention mechanism (DSA) is integrated into the backbone network to enhance the model's focus on edge and crack regions while preserving essential feature representations along the information pathways. In addition, the original detection head of YOLOv11n is redesigned to incorporate an additional small-object detection layer, which helps reduce missed detections and occlusion-related errors. This structural enhancement significantly improves the model's ability to accurately detect small-scale defects in complex road environments.

2.3. ConvSmart model

As a lightweight version in the YOLOv11 series, the YOLOv11n model still exhibits certain limitations in its feature extraction capability, particularly in detecting low-contrast targets such as small-sized and blurry pavement cracks or potholes with indistinct edges. To further reduce the number of parameters while preserving detection accuracy, and to enhance the model's feature fusion capability, this study draws inspiration from the design principles of multi-branch fusion structures (Wang et al., 2024b). Accordingly, a ConvSmart convolutional module is proposed to replace the standard convolution blocks, enabling improved receptive-field structuring and more effective spatial feature integration within the network.

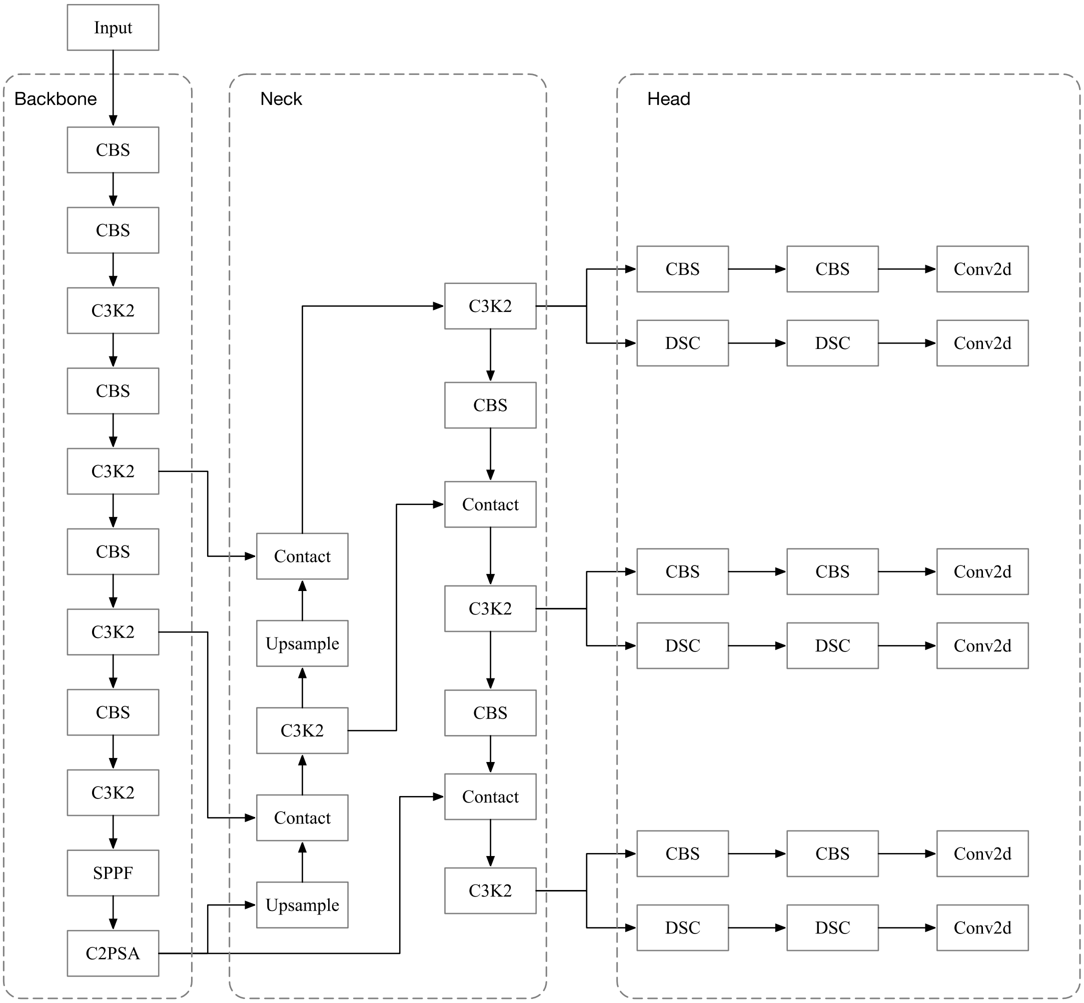


Figure 1. Model structure of YOLOv11

ConvSmart employs a combination of multi-scale convolution fusion, residual connections, and nonlinear activation functions to improve the model's ability to recognize objects of various sizes, while maintaining a controlled parameter overhead. This design is particularly effective in adapting to fine-grained pavement textures and edge features of road defects.

Specifically, the ConvSmart module incorporates two standard convolution kernels of different sizes – 3×3 and 5×5 – to simultaneously extract local details and broader con-textual features. The number of output channels from both convolutions is summed to match the target output dimension, followed by channel concatenation to effectively integrate both local and global information. To maintain channel consistency between the input and output, a 1×1 convolution-based identity mapping path is added to form a lightweight residual connection. This design enhances the network's nonlinear representation capacity and improves gradient propagation efficiency. At the terminal stage of the structure, an optional MaxPooling operation is introduced to perform spatial downsampling and feature compression. After merging the main and residual branches, batch normalization (BatchNorm) and a SiLU activation function are applied to further stabilize feature representation and boost nonlinearity. The structural layout of Con-vSmart is illustrated in Figure 3.

The ConvSmart module comprises four components: multi-scale convolutional fusion, residual pathway mapping, residual fusion and downsampling, as well as normalization and activation.

1. Multi-Scale Convolutional Fusion

To capture features at different receptive fields, two standard convolutions with kernel sizes 3×3 and 5×5 are applied in parallel to the input feature map *X*. The mathematical formulation is as follows:

$$F_1 = Conv_{3\times3}X,\ F_2 = Conv_{5\times5}X, \tag{1}$$

where *X* denotes the input image, $X \in R^{\mathrm{B}\times C_{in}\times\mathrm{H}\times\mathrm{W}}$ represents a four-dimensional real-valued tensor. Here, B is the batch size indicating the number of input images; C_{in} denotes the number of input channels; H is the height of the feature map and W is its width. Furthermore, C_{out} denotes the number of output channels produced by the convolutional operation. $F_1,\ F_2 \in R^{\mathrm{B}\times\frac{C_{out}}{2}\times\mathrm{H}\times\mathrm{W}}$.

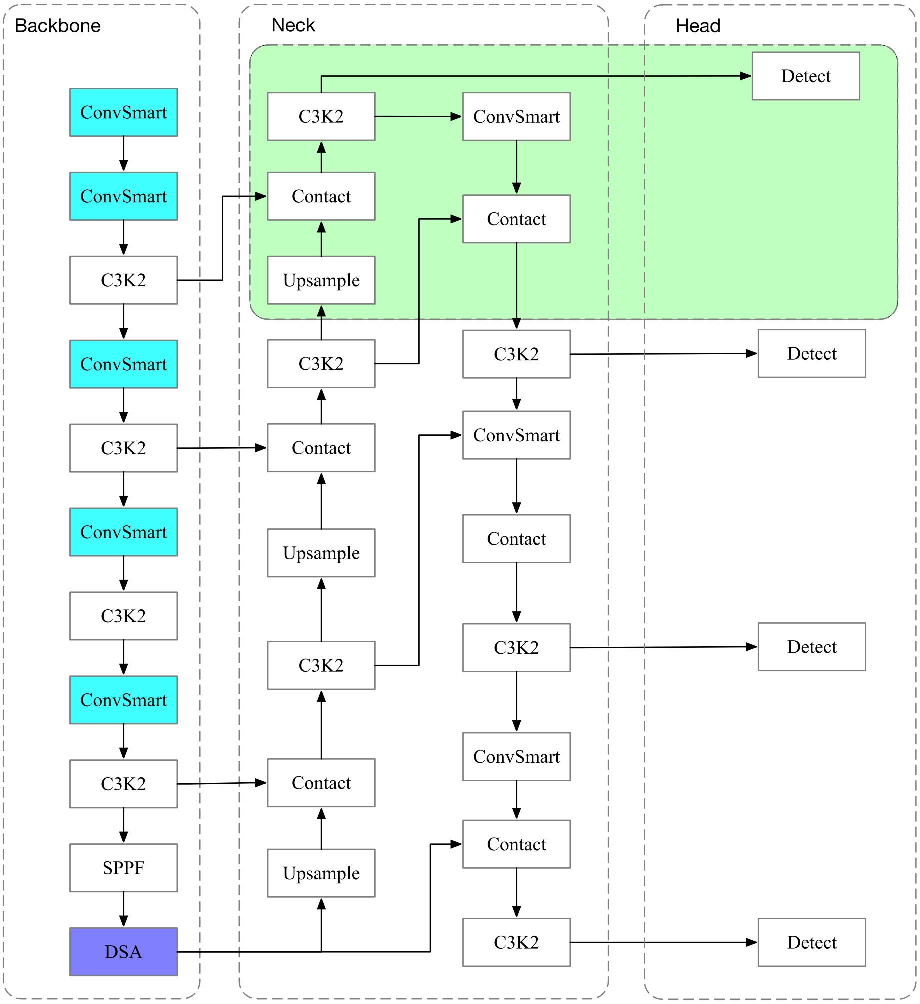


Figure 2. Model structure of YOLOv11-CDL

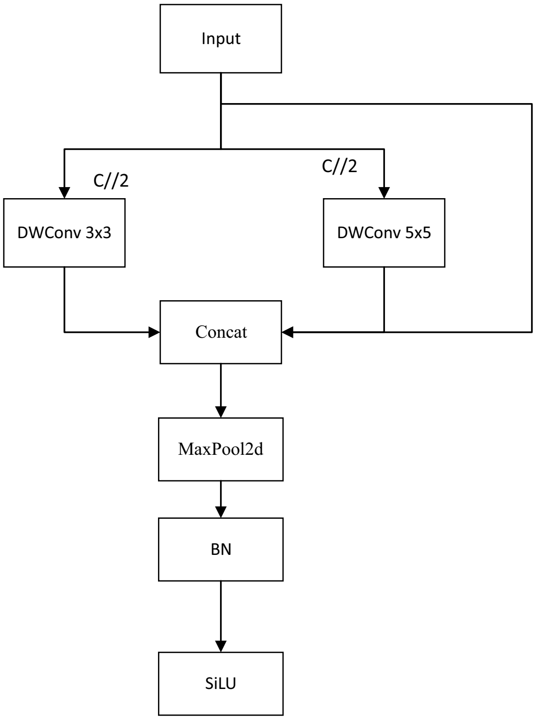


Figure 3. Structure of ConvSmart

These two are concatenated along the channel dimension:

$$F_{merge} = Concat\left(F_1, F_2\right) \in R^{B \times C_{out} \times H \times W}. \qquad (2)$$

2. Residual Pathway Mapping

To preserve identity features and ensure dimension alignment, a 1×1 convolution is applied directly to the input:

$$F_{id} = Conv_{1\times1}\left(X\right) \in R^{B \times C_{out} \times H \times W}. \qquad (3)$$

3. Residual Fusion and Downsampling

The main path and residual path are element-wise added:

$$F_{sum} = F_{merge} + F_{id}. \qquad (4)$$

If downsampling is required, a max-pooling operation with stride > 1 is applied:

$$F_{pool} = \begin{cases} MaxPool\left(F_{sum}\right), \ stride > 1 \\ F_{sum}, \ otherwise \end{cases}. \qquad (5)$$

4. Normalization and Activation

Finally, the output feature is normalized and activated via Batch Normalization and the SiLU function:

$F_{out} \in R^{B\times C_{out}\times H'\times W'}$.

Here H', W'denote the height and width of the output feature map, which may be reduced if pooling is applied.

$$F_{out} = SiLU\left(BN\left(F_{pool}\right)\right). \tag{6}$$

By replacing the original convolutional (Conv) layers in the backbone network with the proposed ConvSmart modules, the model significantly enhances its capacity to perceive defects of varying sizes. As a composite of multi-scale convolutional operations, ConvSmart is particularly effective in handling complex pavement scenarios where both fine cracks and large potholes coexist. The integration of residual connections and nonlinear activation functions alleviates the gradient vanishing problem commonly encountered in deep networks. This architectural enhancement not only improves the network's nonlinear representation capacity but also contributes to better training stability and feature expressiveness.

2.4. Double stage attention model (DSA)

In recent years, feature-guided attention mechanisms have been widely adopted in lightweight defect detection tasks to enhance target saliency and suppress background interference (Zhu et al., 2025). Inspired by these advances, this study designs and integrates a Double-Stage Attention (DSA) module into the YOLOv11n architecture, aiming to further strengthen the model's capability in representing multi-scale and texture-varying defect features. The DSA module guides feature selection from both the channel and spatial dimensions, progressively fusing information to refine feature representations and effectively improve the recognition of complex pavement defects.

The proposed DSA module is embedded into the backbone of the network and consists of two sequential stages. Each stage includes a channel attention branch implemented via SE Attention (Hu et al., 2020) and a spatial attention branch based on Spatial Attention mechanisms (Mnih et al., 2014; Xue et al., 2021). This dual-branch design simultaneously captures "critical channel information" and "salient spatial locations", contributing to a more comprehensive and focused feature representation. The channel attention branch preserves important semantic features along the transmission path, while the spatial attention branch highlights structurally relevant areas. Together, these complementary attention mechanisms reinforce the model's ability to detect pavement defects across various scales and textures. The detailed architecture is illustrated in Figure. The channel attention branch preserves important semantic features along the transmission path, while the spatial attention branch highlights structurally relevant areas. Together, these complementary attention mechanisms reinforce the model's ability to detect pavement defects across various scales and textures. The detailed architecture is illustrated in Figure 4.

The Dual-Stage Attention (DSA) module is composed of three key phases: an initial stage that extracts attention features in parallel from the original input, a second stage that refines attention features based on intermediate outputs, and a final fusion stage that generates enhanced feature maps for downstream detection tasks. The computational structure is formally defined as follows:

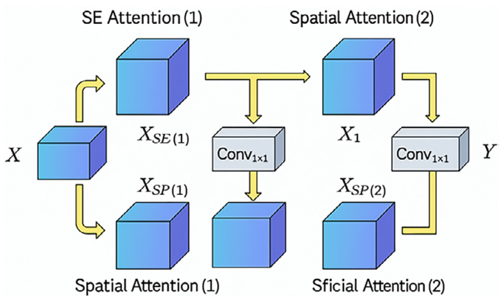


Figure 4. Double stage attention module

Stage 1. Initial Attention Extraction:

$$X_{SE}^{(1)} = SE\ Attention\left(X\right),\ X_{SP}^{(1)} = Spatial\ Attention\left(X\right). \tag{7}$$

Stage 2. Fusion Output:

$$X_1 = Conv_{1\times1}\left(\left[X_{SE}^{(1)}, X_{SP}^{(1)}\right]\right). \tag{8}$$

Stage 3. Attention-Guided Refinement:

$$X_{SE}^{(2)} = SE\ Attention\left(X_1 + X_{SP}^{(1)}\right), X_{SP}^{(2)} = Spatial\ Attention\left(X_1 + X_{SE}^{(1)}\right). \tag{9}$$

Final Attention Fusion:

$$X' = Conv_{1\times1}\left(\left[X_{SE}^{(2)}, X_{SP}^{(2)}\right]\right), \tag{10}$$

where X denotes the input image, $X \in R^{B\times C_{in}\times H\times W}$ represents a four-dimensional tensor in the real-valued space. Here, B is the batch size corresponding to the number of input images, C_{in} denotes the number of input channels, H indicates the height of the feature map, and W represents the width of the feature map.

The integration of the DSA module into the YOLOv11n architecture demonstrates superior performance in detecting road surface defects under complex background conditions. Pavement cracks and small potholes are often more prominent when considered in the context of global structural information. However, the presence of variable illumination and background interference in road defect images poses significant challenges for conventional detectors. With the assistance of the DSA mechanism, the network can automatically focus on critical discriminative cues. Moreover, the discrepancy-aware representation enhances its ability to suppress background interference and highlight abnormal regions, consistent with the findings reported in Cai et al. (2023). This capability significantly improves detection robustness and reduces the rates of both false positives and missed detections. In addition, the DSA module introduces approximately 0.08 million extra parameters, accounting for only about 2% of the total model size. Considering the corresponding improvement in mAP@0.5, this parameter overhead is relatively minor and remains acceptable for real-time pavement inspection applications. As such, it enhances the model's feature selection capacity without compromising efficiency, making it particularly well-suited for real-world pavement inspection tasks in visually cluttered environments.

2.5. Improvement design of the small object detection layer

In practical scenarios such as pavement defect detection, small targets, such as fine cracks, shallow potholes, and minor spalling, are characterized by their small size, blurry contours, and complex edge structures. However, due to the relatively large downsampling factor used in the YOLOv11n architecture, deeper feature maps often struggle to capture detailed representations of such small objects. To enhance the model's perception of small-scale targets, this study designs a dedicated small-object detection pathway and optimizes the multi-scale feature fusion process. Similarly, Wang et al. (2023) proposed the YOLO-MSAPF model, which incorporates multiscale alignment fusion (MSAF) and a parallel feature filtering (PFF) module, effectively improving the detection performance of complex defects in industrial inspection scenarios.

Specifically, the second layer of the backbone network (P2) is selected to extract shallow features, which are then concatenated with the upsampled features from the P3 layer. By combining semantic context from deeper layers and texture details from earlier layers, the merged features are further fused using a lightweight C3k2 module to generate a high-resolution feature map (160×160), denoted as P2/4. This feature map effectively addresses the absence of detection heads in traditional YOLO models at lower levels and significantly improves detection performance for targets smaller than 16×16 pixels.

To ensure spatial consistency across all detection branches, this study further constructs a dedicated downsampling pathway. By leveraging the improved convolutional structure of the ConvSmart module, the network gains multi-scale receptive fields and residual fusion capabilities, effectively preserving shallow texture information while enhancing semantic representation. This design reduces model complexity and strengthens structural semantic modeling across feature levels. Previous studies have similarly introduced graph convolution networks (GCNs) to improve local perceptual modeling, achieving a favorable balance between accuracy and efficiency (Wang et al., 2024c). Through this pathway, a four-scale detection head configuration is established, consisting of P2 (160×160), P3 (80×80), P4 (40×40), and P5 (20×20), as illustrated in Figure 5. This architecture significantly improves the model's sensitivity to small-scale targets. In particular, it enhances the recall rate for fine-grained objects, such as hairline cracks, short linear surface damage, and low-contrast abrasions. Moreover, the integration of P2 features, often overlooked in traditional detection heads, enables more accurate localization of small-scale defects. The proposed multi-scale fusion structure demonstrates superior detection robustness and stability, particularly under challenging imaging conditions, such as nighttime, low-light environments, or rainy weather, where image quality is typically degraded.

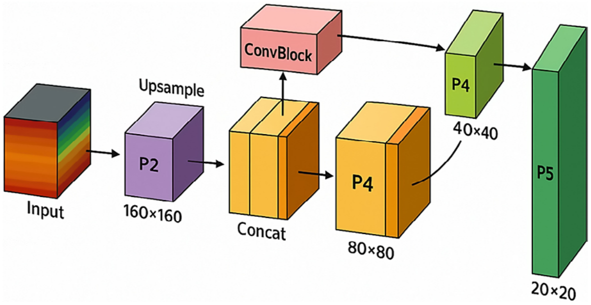


Figure 5. Structure of small object detection layer

3. Experiments

3.1. Experimental environment and dataset

The experiments in this study were conducted on an Ubuntu 22.04 operating system. The hardware configuration included an Intel(R) Xeon(R) Platinum 8457C CPU, 150 GB of RAM, and an H20-NVLink GPU with 96 GB of memory. The programming language used was Python 3.10.13, and the deep learning framework was PyTorch 2.4.0, with GPU acceleration enabled by Cuda 12.4.1.

For the training process, the following hyperparameters were configured: Mosaic data augmentation was employed for image preprocessing, and the AdamW optimizer was used for parameter updates. The learning rate was adjusted using a cosine annealing schedule. The total number of training epochs was set to 100, with an input image size of 640×640 pixels and a batch size of 16.

The dataset utilized in this study is the publicly available IRRDD dataset, which is an Iranian Road Roughness and Damage dataset comprising 25,000 images and four types of road surface damage. All annotations are provided in the YOLO format. The dataset was split into training, testing, and validation sets at a ratio of 8:1:1 to ensure effective training support and reliable performance evaluation.

The types of road surface damage in the dataset include:

1. D00 (Longitudinal crack);
2. D10 (Lateral crack);
3. D20 (Alligator crack); and
4. D40 (Pothole).

These four defect categories represent the main forms of pavement damage. Notably, the distribution of these defect types within the dataset is non-uniform and irregular, which helps to maintain the objectivity and generalization capability of the training results.

3.2. Evaluation metrics

In this study, several evaluation metrics were adopted to comprehensively assess the detection performance of the proposed model. The primary detection performance indicators include Mean Average Precision (mAP), Precision (Pr), Recall (Re) and Frames Per Second (FPS). In addition, model efficiency is evaluated using the number of parameters (Params).

The formula for calculating the Average Precision (*AP*) is given by:

$$AP = \int_0^1 p(r)dr,\ mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i\,, \tag{11}$$

where *p*(*r*) is the interpolated function of the precision-recall (PR) curve, and *N* represents the total number of defect categories in this study.

mAP@0.5 refers to the mean average precision when the Intersection over Union (IoU) threshold is set at 0.5.

mAP@0.5:0.95 denotes the mean average precision averaged over multiple IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05.

$$\mathrm{IoU} = \frac{\text{Area of Overlap between the predicted box and the ground truth box}}{\text{Area of Union between the predicted box and the ground truth box}}. \tag{12}$$

3.3. Ablation experiment

To verify the effectiveness of the three improvement strategies proposed in this paper, namely the ConvSmart module, the Double Stage Attention (DSA) mechanism, and the enhanced small object detection layer structure, four groups of ablation experiments were designed. Specifically, the model YOLOv11n-C denotes the baseline YOLOv11n where the standard convolution modules in the backbone are replaced with the ConvSmart modules. The model YOLOv11n-CD indicates the integration of both the ConvSmart modules and the DSA attention mechanism. The model YOLOv11n-CDL represents the full version of the proposed approach, which combines the ConvSmart modules, the DSA mechanism, and the improved small object detection layer. The experimental results for each configuration are summarized in Table 1.

The results of the ablation experiments clearly demonstrate that applying different improvement strategies to the YOLOv11n model leads to a notable increase in detection accuracy, with each scheme achieving varying degrees of performance enhancement compared to the baseline YOLOv11n model. Specifically, replacing the Conv modules in the backbone with the proposed ConvSmart modules improves the detection accuracy by 1.1%. Combining both the ConvSmart modules and the DSA attention mechanism results in a 2.6% increase in detection accuracy. Finally, integrating all proposed improvements into the YOLOv11n architecture yields the best performance, achieving an increase of 3.0% in mAP@0.5%. To verify the generality and transferability of the proposed CDL module, it was further integrated into two widely adopted lightweight detection architectures, namely YOLOv5s and YOLOv8n. Transfer experiments were conducted under identical datasets and training configurations, and the results are summarized in Table 1. Across both architectures, the inclusion of the CDL module consistently improves Precision, Recall, mAP@0.5, and mAP@0.5:0.95. These results demonstrate that CDL functions as a plug-and-play enhancement module**,** whose effectiveness does not rely on any specific YOLO backbone design.

The contribution of the small-object detection layer is not clearly reflected in the overall results presented in Table 1. Therefore, a dedicated experiment was conducted to individually evaluate its effectiveness. A total of 500 images containing fine cracks or small potholes were extracted from the original dataset to form a small-defect-focused test subset. This experiment simulates the model's generalization ability in recognizing micro-scale defects under limited sample conditions. Similarly, Zhang et al. (2024a) demonstrated that sparse feature representations can substantially enhance the identification of micro-structural anomalies in low-shot anomaly detection tasks. The corresponding experimental results are summarized in Table 2.

On the dataset dominated by fine cracks and small potholes, the contribution of the small-object detection layer becomes considerably more evident. As shown in Table 2, the Recall increases by 3.6%, while mAP@0.5 improves by 2.9%, indicating a notably enhanced capability in identifying miniature defects. The introduction of the P2 small-object detection branch increases the model size from 3.3M to 3.4M parameters (+3.0%); however, this additional branch provides measurable performance gains on the IRRDD dataset, including a 0.4% improvement in mAP@0.5 and a 0.3% increase in mAP@0.5:0.95, and more importantly, substantially boosts the recall of small-scale defects such as hairline cracks and micro-pits.

Table 1. Ablation experiment

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FPS	Params (10^6)
YOLOv11n	71.3	67.0	72.3	43.5	159.6	2.6
YOLOv11n-C	71.6	67.2	73.4	44.3	134.0	3.3
YOLOv11n-CD	72.3	68.7	74.9	44.3	131.5	3.3
YOLOv11n-CDL	72.7	69.6	75.3	44.6	227.3	3.4
YOLOv5s	70.0	64.2	71.2	42.2	112.6	7.8
YOLOv5s-CDL	71.1	69.1	74.1	43.6	204.1	8.5
YOLOv8n	68.3	62.3	68.7	40.9	128.5	2.7
YOLOv8n-CDL	73.2	70.0	75.6	45.0	208.3	3.6

Table 2. Impact of the small-object branch on different object scales

Model	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLOv11n-CD	66.3	72.7	42.4
YOLOv11n-CDL	69.9	75.6	44.7

In practical pavement maintenance scenarios, missing such small yet safety-critical defects often result in more severe consequences than a slight increase in model size. Therefore, the additional 0.1M parameters introduced by the P2 branch are considered a worthwhile trade-off, as they significantly enhance the model's sensitivity to small targets and contribute to improved overall robustness.

3.4. Comparative experiments

To comprehensively validate the effectiveness of the proposed YOLOv11n-CDL algorithm for pavement defect detection, a series of comparative experiments were conducted against several mainstream object detection algorithms, including Faster R-CNN, Cascade R-CNN, NanoDet, PP-YOLOE and representative versions from the YOLO series such as YOLOv5s, YOLOv7-tiny, and YOLOv8n.

As shown in Table 3, YOLOv11n-CDL achieves a more optimal balance between detection accuracy and model lightweight design. Compared to classical object detection algorithms like Faster R-CNN and Cascade R-CNN, the proposed YOLOv11n-CDL significantly reduces parameter count and computational complexity while delivering better performance in terms of precision, recall, mean average precision (mAP), parameter size, and inference speed. Specifically, compared with YOLOv5s, YOLOv11n-CDL shows an improvement of 4.1% in mAP@0.5 and 2.4% in mAP@0.5:0.95, while reducing the parameter count by approximately 57% and maintaining superior detection accuracy. For YOLOv7-tiny, YOLOv11n-CDL achieves an 11.5% increase in recall, an 11.3% increase in mAP@0.5, and a 9.6% improvement in mAP@0.5:0.95, demonstrating stable model performance with a compact architecture. In comparison with the representative lightweight detectors NanoDet and PP-YOLOE, YOLOv11n-CDL achieves approximately 4% higher Recall and mAP@0.5, demonstrating its competitive advantage in detecting pavement defects. When further compared with the more advanced lightweight model YOLOv8n, YOLOv11n-CDL still exhibits consistently superior performance across all evaluation metrics. Moreover, the FPS comparison indicates that YOLOv11n-CDL not only delivers higher detection accuracy but also achieves a substantial improvement in inference speed. These results show that the proposed architectural enhancements effectively strengthen the model's ability to identify defects under multi-scale variations and complex background conditions, thus providing superior real-time processing capability and deployment feasibility.

3.5. Detection results visualization

To clearly compare the performance improvements of the proposed model, representative images from the dataset were selected for visual verification. Table 4 illustrates the comparison of detection results between the standard YOLOv11n algorithm and the optimized YOLOv11n-CDL model.

As shown in the Table 4, when the crack color is similar to the background or lane markings interfere near the cracks, the standard YOLOv11n model exhibits missed detections, while the proposed YOLOv11n-CDL model effectively identifies the road crack information. In cases where the cracks are extremely fine, YOLOv11n-CDL demonstrates a clear advantage in small-object detection by successfully recognizing tiny cracks. For scenarios involving multiple dense defect types, YOLOv11n-CDL achieves better detection performance than YOLOv11n, showcasing its improved capability in identifying complex pavement defects.

As illustrated in the visualization results presented in Table 5, the proposed YOLOv11n-CDL achieves noticeably better detection performance compared with the baseline YOLOv11n model. YOLOv11n-CDL is able to more effectively overcome object occlusion, capture subtle and fine-scale crack patterns, and suppress background noise. The highlighted regions produced by YOLOv11n-CDL exhibit a higher degree of overlap with the actual defect areas in the original images, indicating a stronger capability for focusing on true defect regions. Consequently, the improved YOLOv11n-CDL demonstrates a clearly enhanced ability to extract pavement defects, particularly under challenging and noise-prone conditions.

Table 3. Experimental comparison of mainstream algorithms

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FPS	Params (10^6)
Faster-RCNN	25.1	41.1	50.4	26.5	87.5	165.0
Cascade-RCNN	31.7	40.6	66.4	37.0	56.1	69.2
NanoDet	72.1	65.8	71.2	41.7	68.7	68.7
PP-YOLOE	69.7	65.9	71.1	42.5	91.6	91.6
YOLOv5s	70.0	64.2	71.2	42.2	112.6	7.8
YOLOv7tiny	69.2	58.4	64.0	35.0	119.3	6.0
YOLOv8n	68.3	62.3	68.7	40.9	128.5	3.2
YOLOv11n-CDL	72.4	69.9	75.3	44.6	227.3	3.3

Table 4. Comparison of detection results

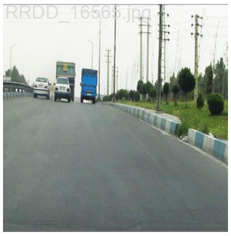
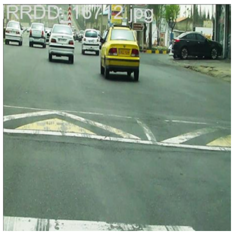
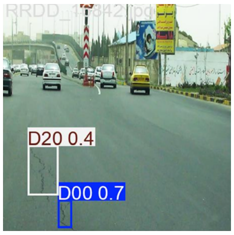
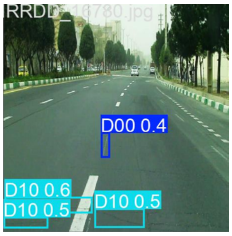
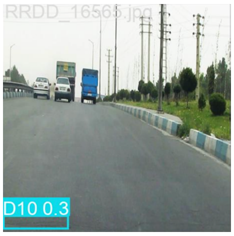
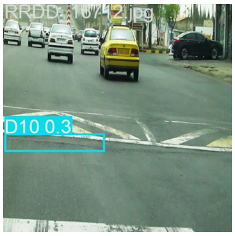
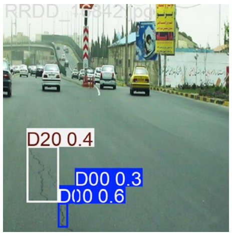
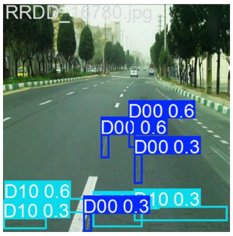
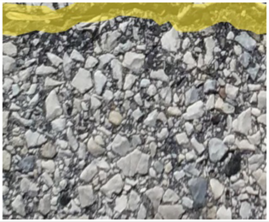
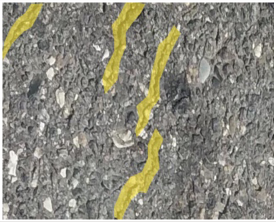
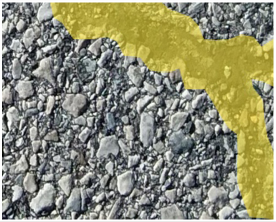
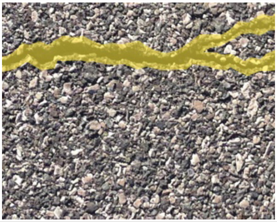
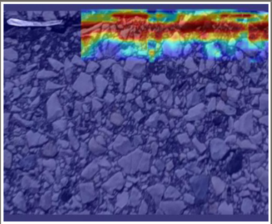
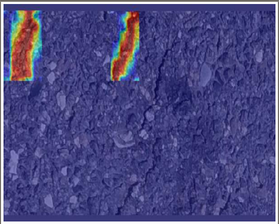
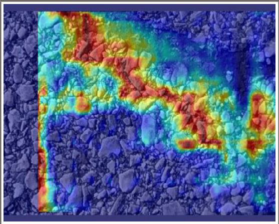
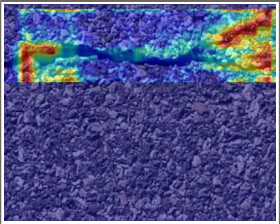
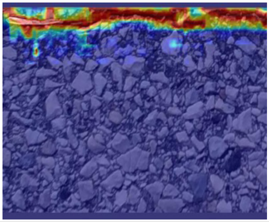
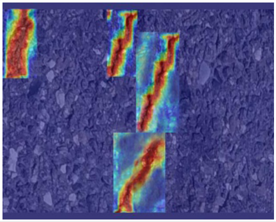
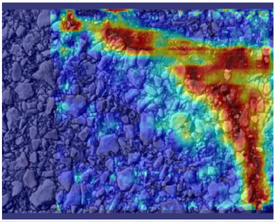
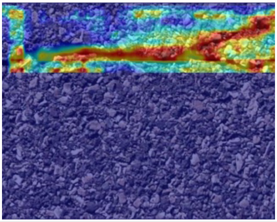
YOLOv11n				
YOLOv11n-CDL				 D00 0.6 D00 0.6 D00 0.3 D10 0.6 D10 0.3 D00 0.3 D10 0.3
	Background Color Interference	Lane Marking Interference	Tiny Cracks	Multiple Dense Defects

Table 5. Visualization of localization results

Test Image				
GT				
YOLOv11n				
YOLOv11n-CDL				
	Object Occlusion	Tiny Cracks	Background Color Interference	Typical Pavement Cracks

In addition, a comparative analysis of the key training metrics between YOLOv11n and the proposed YOLOv11n-CDL was conducted.

Figure 6 illustrates the training curves for Precision, Recall, mAP@0.5, and mAP@0.5:0.95, with the blue curves representing the baseline YOLOv11n model and the red curves corresponding to the improved YOLOv11n-CDL model.

As shown in the figure, under the same number of training epochs, the red curves consistently remain above

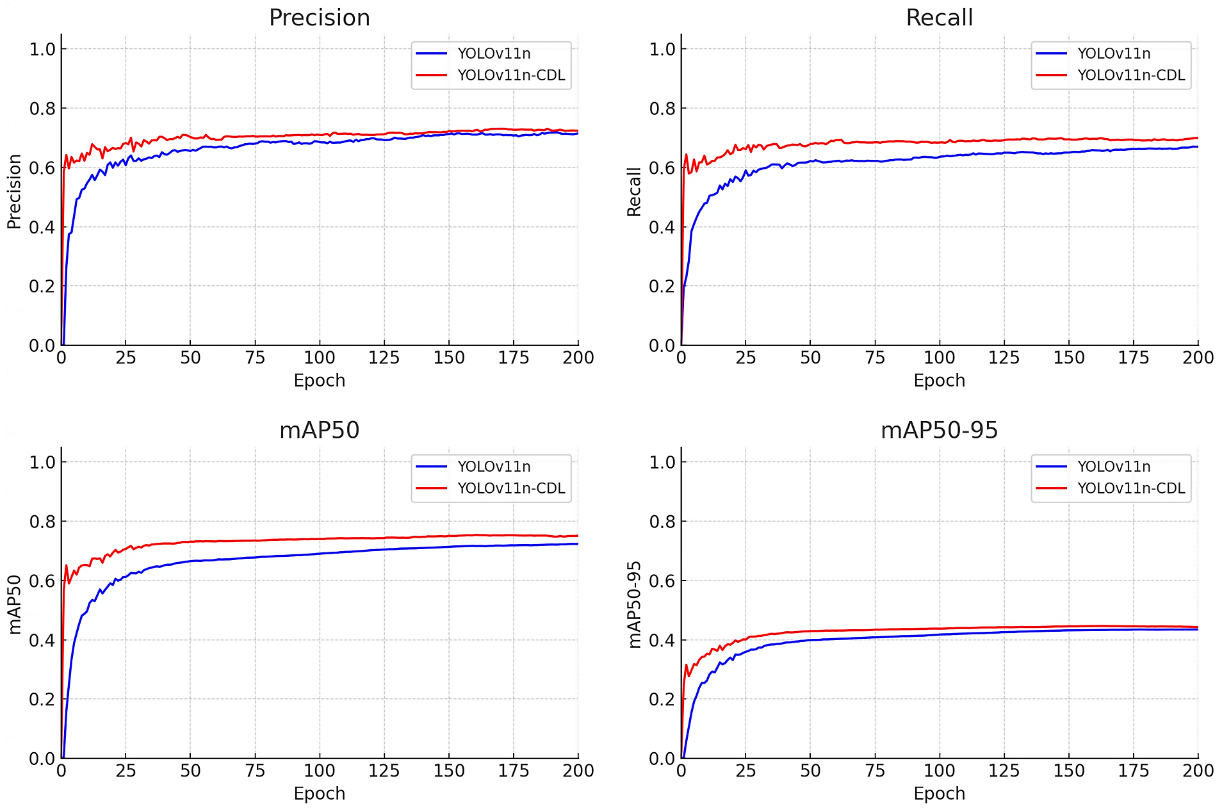


Figure 6. Comparison of model improvement effects

the blue curves across all metrics, indicating superior performance. Moreover, the YOLOv11n-CDL model exhibits a faster convergence rate, demonstrating enhanced training efficiency.

These results further validate the effectiveness of the proposed architectural improvements in boosting the overall detection performance and learning stability of the model.

3.6. Comparison on a different dataset

To further evaluate the generalization capability of the proposed YOLOv11n-CDL model, additional experiments were conducted on the RDD2022 dataset. After preprocessing and consolidation, the dataset contains 23,767 images. Following the same defect taxonomy as in the previous experiments, road surface defects were categorized into four major classes: D00 (longitudinal cracks), D10 (lateral cracks), D20 (alligator cracks), and D40 (potholes). The dataset was divided into training, validation, and test sets in an 8:1:1 ratio.

On the RDD2022 dataset, YOLOv11n-CDL was compared with several representative lightweight object detectors, including NanoDet, PP-YOLOE, YOLOv3-tiny, YOLOv5s, YOLOv8n, and YOLOv11n. The results, summarized in Table 6, show that YOLOv11n-CDL achieves 65.07% Precision, 54.39% Recall, 59.94% mAP@0.5, and 29.99% mAP@0.5:0.95. Despite its compact size of 3.37M parameters, YOLOv11n-CDL delivers the highest accuracy among all evaluated lightweight models.

In terms of computational efficiency, YOLOv11n-CDL also demonstrates strong real-time performance, achieving 109.9 FPS and 23.5 GFLOPs, which highlights its deployment-friendly characteristics. Compared with the baseline YOLOv11n, YOLOv11n-CDL obtains a 4.44% improvement in mAP@0.5, while introducing only 0.79M additional parameters and an incremental increase of 1.9 GFLOPs, effectively balancing accuracy and efficiency.

Table 6. Experimental comparison of mainstream algorithms for the RDD2022 Dataset

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FPS	Params (10^6)	GFLOPs
NanoDet	62.4	49.6	54.4	28.0	120.94	0.95	0.72
PP-YOLOE	59.4	49.0	53.0	27.2	100.71	7.93	17.36
YOLOv3-tiny	59.1	48.3	50.2	23.3	115.68	12.13	18.9
YOLOv5s	64.7	52.8	57.8	29.6	107.32	7.82	23.8
YOLOv8n	62.1	51.7	55.9	28.6	124.15	3.00	8.1
YOLOv11n	63.6	50.7	55.5	28.4	117.48	2.58	6.3
YOLOv11n-CDL	65.07	54.39	59.94	29.99	243.90	3.37	23.5

Table 7. Performance of YOLOv11n-CDL on Different GPUs

GPU	GPU memory usage during training (G)	training time for 200 epochs (h)	GPU memory usage during inference (G)	FPS
RTX 3090 24G	10.2	0.931	4.72	277
RTX 3060 12G	10.2	1.888	2.91	158
RTX 1060Ti 6G	6G	34.83	3.11	47

These results indicate that the proposed YOLOv11n-CDL is not only well-suited for the IRRDD dataset but also exhibits robust cross-dataset transferability and promising practical deployment potential.

3.7. Edge deployment analysis

To not only enhance detection accuracy but also maintain a lightweight architecture, the proposed model is designed with deployment on devices of limited computational capability in mind. Such adaptability is essential for promoting its practical application in real-world pavement inspection scenarios. To further assess the suitability of YOLOv11n-CDL across hardware platforms with differing computational power, three devices were selected for comparative testing: a workstation equipped with an NVIDIA GeForce RTX 3090, a PC server equipped with an RTX 3060, and a laptop integrating an RTX 1060 Ti GPU.

The Crack500 dataset, consisting of 1,796 images, was used for this evaluation, with the training, validation, and test sets divided according to an 8:1:1 ratio. The comparison considered GPU memory consumption during training, the time required for 200 training epochs, GPU memory usage during inference, and the frames per second (FPS) achieved during detection. As summarized in Table 7, YOLOv11n-CDL maintains consistently low memory consumption and stable efficiency across all three hardware platforms. Notably, the model achieves an inference speed of 47 FPS on the RTX 1060 Ti laptop, indicating that it can process a larger number of images within the same time frame while imposing lower computational demands. This makes the model particularly suitable for deployment on edge devices or in resource-constrained environments.

4. Discussion

Although the proposed YOLOv11n-CDL model achieves substantial improvements in detection accuracy, robustness, and real-time performance, several important challenges remain and provide opportunities for future research.

1. Enhancing generalization across diverse environments.

While additional experiments on RDD2022 have demonstrated that YOLOv11n-CDL possesses certain cross-dataset transferability, road conditions in real applications vary widely in terms of pavement materials, surface aging patterns, traffic loads, and regional maintenance standards. Moreover, rare or region-specific defects remain underrepresented in most public datasets. Future work should incorporate broader multi-region, multi-material datasets, particularly those collected under varying traffic and climatic conditions, to further improve the model's robustness and generalization capacity.

2. Improving robustness under extreme weather, illumination, and motion conditions.

Despite the integration of ConvSmart and the double-stage attention (DSA) module, challenging real-world scenes, such as heavy rain, nighttime glare, fog, shadows, and motion blur caused by high-speed vehicle-mounted cameras, continue to influence the clarity and consistency of road-surface textures. Future research may explore adaptive feature enhancement mechanisms, such as weather-invariant representation learning, temporal multi-frame compensation, or physically informed augmentation strategies to further reinforce performance under adverse conditions.

3. Advancing lightweight design while preserving accuracy.

Although YOLOv11n-CDL maintains a compact model size and exhibits strong deployment feasibility across devices with different computational capacities, further reduction of model parameters and GFLOPs remains desirable for embedded-edge hardware such as Jetson Nano, mobile SoCs, or real-time inspection drones. Future studies may investigate model compression, neural architecture search (NAS), mixed-precision optimization, or low-bit quantization to achieve higher efficiency without compromising detection quality.

4. Broadening defect categories and incorporating structural semantics.

Current experiments focus on four dominant surface defect types. However, many structurally relevant forms of pavement deterioration, such as rutting, joint dislocation, manhole cover anomalies, or crosswalk fading, are essential for comprehensive maintenance assessment. Future work will consider expanding the defect taxonomy and incorporating higher-level geometric or contextual cues, potentially through multi-task learning that integrates classification, segmentation, and structural reasoning.

In summary, while YOLOv11n-CDL demonstrates strong performance and practical potential, future research should focus on expanding data diversity, improving robustness under extreme conditions, further advancing lightweight architecture design, and integrating structural or multi-sensor information to develop more comprehensive, scalable, and deployment-ready pavement inspection solutions.

5. Conclusions

This study presents YOLOv11n-CDL, a lightweight and high-performance pavement defect detection framework incorporating three targeted enhancements: the ConvSmart convolution module, the Double-Stage Attention (DSA) mechanism, and a small-object detection pathway. Based on comprehensive evaluations across multiple datasets and hardware platforms, the main conclusions are as follows:

1. Enhanced multi-scale feature representation with minimal parameter overhead.

   The proposed ConvSmart module expands the effective receptive field and strengthens spatial feature fusion while maintaining lightweight complexity. This design results in improved detection of defects of varying sizes and contributes to notable gains in Precision and mAP compared to the baseline YOLOv11n.

2. Improved robustness to complex backgrounds through DSA integration.

   By embedding the DSA mechanism at the deepest stage of the backbone, the model effectively emphasizes discriminative crack patterns and suppresses background noise such as shadows, lane markings, and texture clutter. The DSA module introduces only ~2% additional parameters yet yield clear gains in both recall and overall detection accuracy.

3. Significant enhancement in small-defect detection capability.

   The introduction of the P2 small-object detection layer substantially improves the model's sensitivity to microcracks and small potholes. Experiments on a dedicated small-defect subset demonstrate increases of 3.6% in Recall and 2.9% in mAP@0.5, confirming the effectiveness of the added high-resolution detection branch.

4. YOLOv11n-CDL achieves strong generalization and high deployment feasibility.

   Across IRRDD and RDD2022 datasets, the model surpasses mainstream lightweight detectors, and its components consistently improve YOLOv5s and YOLOv8n when transferred. Additionally, tests on GPUs of varying capacities demonstrate stable memory usage and real-time inference (up to 47 FPS on low-power devices), indicating strong suitability for edge computing and practical field deployment.

Overall, YOLOv11n-CDL achieves a favorable balance between accuracy, robustness, and computational efficiency, offering a practical solution for automated pavement defect detection. Future work will focus on expanding defect categories, improving detection robustness under extreme weather or illumination conditions, and exploring adaptive feature enhancement strategies to further strengthen cross-domain generalization.

Acknowledgements

The authors sincerely appreciate the helpful and valuable suggestions from all reviewers and editors. Thanks to K. Meng and J. Q. Sun for providing the guidance for this paper.

Author contributions

Conceptualization, J. D. and Y. Y. G.; methodology, J. D. and Y. Y. G.; software, J. D.; validation, J. D.; formal analysis, J. D. and Y. Y. G.; investigation, Y. Y. G.; resources, J. D.; data curation, J. D.; writing – original draft preparation, J. D.; writing – review and editing, Y. Y. G.; visualization, J. D.; supervision, Y. Y. G.; project administration, J. D.

Disclosure statement

The authors declare no conflicts of interest.

References

Bai, F., Ma, Q. L., & Zhao, M. (2025). AC-YOLO for aerial pavement crack detection. *Computer Engineering and Applications*, *61*(1), 153–164.

Cai, Y., Liang, D., Luo, D., He, X., Yang, X., & Bai, X. (2023). A discrepancy aware framework for robust anomaly detection. *IEEE Transactions on Industrial Informatics*, *20*(3), 3986–3995. https://doi.org/10.1109/TII.2023.3318302

Chen, X.-D., Jiang, N., Dong, L. tian, W., Wu, X., Zeng, P., & Yang, F. (2019). Automatic identification of asphalt pavement diseases in plateau mountainous areas based on YOLO deep learning model. *Highway Transportation Technology* (*Applied Technology Edition*), *15*(11), 75–78.

Chen, X.-D., Ai, D.-H., Zhang, J.-C., Cai, H.-Y., & Cui, K.-R. (2020). Pavement crack detection method based on Gabor filtering fused with convolutional neural networks. *Chinese Optics*, *13*(6), 1293–1301. https://doi.org/10.37188/CO.2020-0041

Ding, K., Ding, Z., Zhang, Z., Yuan, M., Ma, G., & Lv, G. (2024). SCD-YOLO: A novel object detection method for efficient road crack detection. *Multimedia Systems*, *30*(6), Article 351. https://doi.org/10.1007/s00530-024-01538-y

Dong, H. Z., Lin, S. X., & She, Y. N. (2025a). Research progress on YOLO detection technology for traffic targets. *Journal of Zhejiang University* (*Engineering Science*), *59*(2), 249–260.

Du, Y., Pan, N., Xu, Z., Zhou, Y., & Chen, Y. (2021). Pavement distress detection and classification based on YOLO network. *International Journal of Pavement Engineering*, *22*(13), 1659–1672. https://doi.org/10.1080/10298436.2020.1714047

Duan, Z. X., He, Y., Zhang, X., & Gao, J. (2024). Pavement crack detection method based on effective feature extraction and cascade optimization. *Journal of Computer-Aided Design and Graphics*, *36*(12), 2020–2028. https://doi.org/10.3724/SP.J.1089.2024.20129

Girshick, R. (2015). Fast R-CNN. In *Proceedings of 2015 IEEE International Conference on Computer Vision* (pp. 1440–1448), Santiago, Chile. IEEE. https://doi.org/10.1109/ICCV.2015.169

He, T. J., & Li, H. E. (2024). Pavement disease detection model based on improved YOLOv5. *China Civil Engineering Journal*, *57*(2), 96–106.

Hu, J., Shen, L., & Sun, G. (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

Hu, X. W., Yan, Y. X., Wang, D. W., & Zhang, Y. H. (2024). Lightweight pavement disease detection method based on YOLOM algorithm. *China Journal of Highway and Transport*, *37*(12), 381–391.

Ibragimov, E., Lee, H. J., Lee, J. J., & Lee, S. Y. (2022). Automated pavement distress detection using region-based convolutional neural networks. *International Journal of Pavement Engineering*, *23*(6), 1981–1992. https://doi.org/10.1080/10298436.2020.1833204

Lang, H., Wen, T., Lu, J., Ding, S., & Chen, S. (2021). 3D pavement crack defect detection method based on deep learning. *Journal of Southeast University* (*Natural Science Edition*), *51*(1), 53–60.

Miao, Y., Zhang, Z., Wang, H., Dai, W., Zhao, Z., Wang, X., Yang, C., & Shi, Y. (2023). Pavement leaf detection method based on AC-YOLO. *Control and Decision*, *38*(7), 1878–1886.

Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K. (2014). *Recurrent models of visual attention* (arXiv:1406.6247). arXiv. https://doi.org/10.48550/arXiv.1406.6247

Qin, L., Tan, Z. F., Lei, G. P., & Chen, Q. (2025). EMF-YOLO: Lightweight multi-scale feature extraction pavement defect detection algorithm. *Computer Engineering and Applications*. http://kns.cnki.net/kcms/detail/11.2127.tp.20250321.1618.010.html

Qiu, Q., & Lau, D. (2023). Real-time detection of cracks in tiled sidewalks using YOLO-based method applied to UAV images. *Automation in Construction*, *147*, Article 104745. https://doi.org/10.1016/j.autcon.2023.104745

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) (pp. 779–788), Las Vegas, NV, USA. IEEE. https://doi.org/10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Shen, S. Y., Hua, B., & Huang, R. W. (2023). Research on improved YOLOv5 pavement crack detection model. *Electronic Measurement Technology*, *46*(21), 132–142.

Sun, Z. Y., Pei, L. L., Li, W., et al. (2020). Pavement sealing crack detection method based on improved Faster R-CNN. *Journal of South China University of Technology* (*Natural Science Edition*), *48*(2), 84–93.

Ultralytics. (2025). *YOLOv11*: *Ultralytics YOLO models documentation*. https://docs.ultralytics.com/models/yolo11/

Wang, B. X., Bai, S. X., & Zhao, W. G. (2022). Pavement crack defect visual detection method based on feature enhancement learning. *Journal of Railway Science and Engineering*, *19*(7), 1927–1935.

Wang, G.-Q., Zhang, C.-Z., Chen, M.-S., Lin, Y.-C., Tan, X.-H., Liang, P., Kang, Y.-X., Zeng, W.-D., & Wang, Q. (2023). YOLO-MSAPF: Multiscale alignment fusion with parallel feature filtering model for high accuracy weld defect detection. *IEEE Transactions on Instrumentation and Measurement*, *72*, Article 5022914. https://doi.org/10.1109/TIM.2023.3302372

Wang, A., Chen, H., Liu, L., Li, M., & Zhao, X. (2024a). *YOLOv10*: *Real-time end-to-end object detection* (arXiv:2405.14458). arXiv. https://doi.org/10.48550/arXiv.2405.14458

Wang, G., Chen, M., Lin, Y. C., Tan, X., Zhang, C., Yao, W., Gao, B., Li, K., Li, Z., & Zeng, W. (2024b). Efficient multi-branch dynamic fusion network for super-resolution of industrial component image. *Displays*, *82*, Article 102633. https://doi.org/10.1016/j.displa.2023.102633

Wang, G.-Q., Zhang, C.-Z., Chen, M.-S., Lin, Y. C., Tan, X.-H., Kang, Y.-X., Wang, Q., Zeng, W.-D., & Zhao, W.-W. (2024c). A high-accuracy and lightweight detector based on a graph convolution network for strip surface defect detection. *Advanced Engineering Informatics*, *59*, Article 102280. https://doi.org/10.1016/j.aei.2023.102280

Wang, X. Q., Gao, H. B., & Jia, Z. M. (2024d). Pavement defect detection algorithm based on improved YOLOv8. *Computer Engineering and Applications*, *60*(17), 179–190.

Xing, Y., Han, X., Pan, X., Wang, H., & Li, C. (2024). EMG-YOLO: Road crack detection algorithm for edge computing devices. *Frontiers in Neurorobotics*, *18*, Article 1423738. https://doi.org/10.3389/fnbot.2024.1423738

Xue, M., Chen, M., Peng, D., & Chen, S. (2021). One spatio-temporal sharpening attention mechanism for light-weight YOLO models based on sharpening spatial attention. *Sensors*, *21*(23), Article 7949. https://doi.org/10.3390/s21237949

Yao, J. L., Cheng, G., & Wan, F. (2024). Lightweight bearing defect detection algorithm based on improved YOLOv8. *Computer Engineering and Applications*, *60*(21), 205–214.

Zhang, F., Zhu, H., Cen, Y., Kan, S., Zhang, L., Vadakkepat, P., & Lee, T. H. (2024a). Low-shot unsupervised visual anomaly detection via sparse feature representation. *IEEE Transactions on Neural Networks and Learning Systems*, *36*(85), 7903–7917. https://doi.org/10.1109/TNNLS.2024.3420818

Zhang, M. X., Xu, J., Liu, X. P., Zhang, Y., Zhang, C., & Ning, X. (2024b). Pavement crack detection method based on improved U-Net. *Computer Engineering and Applications*, *60*(24), 306–313.

Zhou, J. X., Zhang, Y., Jia, Z. H., & He, Y. (2024). Pavement defect detection algorithm based on improved YOLOv8. *Electronic Measurement Technology*, *47*(19), 146–154. https://doi.org/10.19651/j.cnki.emt.2416638

Zhu, Y., Lin, Y. C., Tan, X. H., & Li, S. X. (2025). A lightweight defect detection network for titanium strip via efficient convolution and feature-driven mechanism. *Engineering Applications of Artificial Intelligence*, *158*, Article 111503. https://doi.org/10.1016/j.engappai.2025.111503

Zuo, H., & Niu, X. W. (2023). Pavement defect detection algorithm based on improved YOLOv5. *Information Technology and Informatization*, (1), 50–53.