Article

Not peer-reviewed version

YOLOv8n-DSRS:Road Disease Detection Based on Multiscale Dual-Path Feature Reorganization with Selective Kernel Networks

Yizong Wang , Zhengrong Xiao , Xuegang An , Fei Li , Jiya Tian *

Posted Date: 30 June 2025

doi: 10.20944/preprints202506.2429.v1

Keywords: Road disease detection; Dynamic dual path; Multi-scale pooling; Multi-path Feature extraction; YOLOv8n-DSRS

 doi:10.20944/preprints202506.2429.v1

Article

YOLOv8n-DSRS:Road Disease Detection Based on Multiscale Dual-Path Feature Reorganization with Selective Kernel Networks

Yizong Wang †,‡, Zhengrong Xiao ‡, Xuegang An, Fei Li and Jiya Tian *

School of Information Engineering,Xinjiang Institute of Technology
* Correspondence: jytian2025@163.com(J.T.)
† Current address: School of Information Engineering,Xinjiang Institute of Technology.
‡ These authors contributed equally to this work.

Abstract

Aiming at the problems of variable target scale, complex background and low model calculation efficiency in road diseases, a road disease detection method YOLOv8n-DSRS based on multi-scale dual-path feature recombination and selective kernel network is proposed. Firstly, a dynamic dual-path down-sampling DS_ module is designed. Through the parallel fusion of spatial rearrangement strategy and standard convolution path, the spatial resolution is reduced while retaining detailed features, and the detection accuracy of small-scale diseases is improved. Then, a multi-scale pooling SPPF_WD module is proposed, which combines serial maximum pooling and lightweight convolution LightConv to enhance multi-scale feature perception. Finally, the RCRep2_FRFN module is introduced. The redundant feature map is generated by GhostConv and the feature refinement FRFN strategy is combined to strengthen local and global information fusion and reduce the missed detection rate.Experimental results show that the improved model YOLOv8n-DSRS achieves precision, recall, mAP50%, and mAP50-95% of 95.6%, 92.8%, 96.3%, and 77.3%, respectively, representing improvements of 7.4%, 12.2%, 7.8%, and 15.2% compared to the baseline model YOLOv8n. The number of parameters and GFLOPS are slightly lower than those of the original network model, but the overall performance is superior to other YOLOv8 series and mainstream algorithms, featuring lightweight, high accuracy, and low computational requirements. This comprehensively validates the effectiveness and superiority of this method.This provides strong support for applications in intelligent transportation and unmanned inspection.

Keywords: road disease detection; dynamic dual path; multi-scale pooling; multi-path feature extraction; YOLOv8n-DSRS

1. Introduction

Roads, as the primary venues for pedestrian and vehicular traffic, inevitably develop various defects due to prolonged use and the combined effects of natural environmental factors.These defects not only significantly reduce the service life of roads but also diminish driving comfort and safety, and may even lead to accidents, posing a threat to public safety.Therefore, timely, efficient, and accurate detection of road defects is of critical importance. Traditional manual road surface inspection methods suffer from low efficiency, high costs, strong subjectivity, and inadequate safety, making them unsuitable for current road maintenance requirements. In recent years, with the rapid development of computer vision and deep learning technologies, automated defect detection techniques based on image recognition have increasingly become the mainstream. These include target detection algorithms such as Convolutional Neural Networks (CNN)[1], Deep Residual Networks (ResNet)[2], and YOLO[3], which have significantly improved recognition efficiency and accuracy.

 doi:10.20944/preprints202506.2429.v1

With the widespread application of deep neural networks in the field of computer vision, various strategies for road surface defect detection have emerged.Lv B *et al.*[4] proposed a structure-aware and feature attention network (SFFAN) that effectively improves detection accuracy and computational efficiency without significantly increasing complexity; Hou Y *et al.*[5] proposed a test time adaptive framework (TTA) to address feature distribution biases across different scenarios, while achieving similar performance gains and reducing reliance on high-quality labeled data; Li J *et al.*[6] utilized a semi-supervised instance segmentation technique based on deep transfer learning, employing SAM's interactive segmentation method to achieve efficient crack detection across different scenarios with limited labeled data; addressing shortcomings in key stages such as road defect database construction, data collection, and preprocessing, Liu Z *et al.*[7] utilized a specialized road condition inspection vehicle along China's S315 highway to collect and construct a large-scale, high-resolution road condition defect detection dataset, Pave Distress, providing a robust data foundation for research on automated road condition monitoring and management systems; Compared to traditional convolutional neural networks [1] and Transformers [8], Huang Q W *et al.*[9] proposed the LTPLN detection method, which improves the Transformer model and employs an iterative training strategy with maximized label distillation, achieving automated pavement defect detection;Mahdy *et al.*[10] demonstrated the importance of balancing cost and efficiency through an instance segmentation method based on low-cost PCI sensors combined with Deep Neural Networks (DNNs); Haohui Y and Junfei Z [11] utilized drones to establish the UAV-PDD2023 dataset, using its diversity and large-scale characteristics as a benchmark to evaluate the performance of different algorithms in tasks such as object detection and image classification; Cancan Y *et al.*[12] developed an efficient PDD method based on an improved YOLOv7 model for end-to-end object detection suitable for high-speed detection tasks; Wu, Lingxiao *et al.*[13] integrated the improved lightweight attention module SCBAM into the backbone network and replaced the spatial pyramid with SPPF for feature fusion, achieving a 3.8% improvement in mAP@0.5 compared to the original model; Chu, Yinze *et al.*[14] simultaneously introduced the Ghost and C3Ghost modules into YOLOv5S and enhanced feature expression performance through mosaic data augmentation, The proposed method improved accuracy by 4.01% compared to the original model; Yang Zhen *et al.*[15] proposed a lightweight algorithm PDNet based on an improved YOLOv5 to achieve model compression, accelerated detection, and precise recognition under multi-scale conditions. The results showed that compared to the baseline model, F1 and speed were improved by 10% and 77.5%, respectively; to address the challenge of damage detection in road maintenance, Hou Yun *et al.*[16] adopted a new method, FS-Net, which integrates the FReLU structure and strip pooling method, to rapidly identify and detect road conditions, providing extensive data support for maintenance decision-making by road maintenance departments; Du Yuchuan *et al.*[17] constructed a large-scale dataset PD and applied fine-grained annotation techniques using the YOLO network for road damage detection. The results showed that this method is 9 times faster than Faster R-CNN and only 70% of SSD.

To further integrate multi-sensor data and cross-modal detection techniques, Sun Pengyuan *et al.*[18] developed a new framework called DSWMamba by combining deep fusion and selective scanning block DFSS with the VSD module. This framework demonstrated significantly higher detection accuracy than CNN and Transformer models on three different datasets, enabling high-precision detection of asphalt pavement defects; Abdelkader *et al.*[19]introduced the EGYPDD dataset to address the limited availability of real-world datasets for training deep learning models, enabling precise pavement defect detection and classification, and providing new samples for international pavement defect detection and classification research; He J *et al.*[20] utilized high-altitude unmanned aerial vehicles (UAVs) to construct a large-scale pavement damage database, HighRPD, addressing the scarcity of publicly available drone-based road pavement damage datasets with limited data volume, and providing rich data support for real-time detection on drone platforms; Zhao Yiyue *et al.*[21] achieved efficient pavement defect detection through a combined drone and ground vehicle detection scheme, not only reducing the cost of fixed equipment investment but also achieving the detection

 doi:10.20944/preprints202506.2429.v1

objectives of wide coverage and fast feedback, addressing the challenges posed by complex road structures that ground vehicles struggle to handle; Li Yuxuan *et al.*[22] designed a target detection network named Crack Convolution (Crack YOLO) to address the issue of road damage automatic detection being influenced by shadows caused by inter-row trees, weeds, soil, and differences in the scale of damaged objects, for extracting road cracks on rural roads. The results showed that it outperformed the current road crack detection models YOLO-LWNet, Faster R-CNN, and YOLOv7 by 9.99%, 12.79%, and 4.61%, respectively; Hu Xiaowei *et al.*[23] addressed the issue of insufficient accuracy in existing road surface defect detection methods by proposing a lightweight road surface defect detection method that integrates a state space model, known as YOLOM. The results demonstrated that this method has significant advantages in terms of small model size, low complexity, and high detection accuracy, with strong learning capabilities and generalization performance, providing support for intelligent road detection; Han Zhibin *et al.*[24] addressed the issues of high cost, long processing time, and low accuracy in traditional methods by proposing an enhanced road defect recognition algorithm, MS-YOLOv8, to improve detection accuracy and adaptability to different road conditions, further expanding the application of drones and deep learning in road defect recognition; addressing practical issues such as uneven data distribution and insufficient sample size, Wang Sicheng *et al.*[25] designed an enhanced YOLOv8 network utilizing synthetic data.This network, combined with a synthetic data detection method based on texture background modeling, significantly alleviates the impact of data scarcity and reduces reliance on the collection of a large number of defect samples. Liu Wenchao *et al.*[26] combined GPR with deep learning to the proposed bidirectional object detection model effectively improves the detection accuracy of small-sized hazardous objects. Compared with the baseline model, the average accuracy for small objects increased by 17.9%. The results demonstrate that this algorithm can effectively improve the efficiency of defect detection in GPR images.

However, there are still some challenges in the current road defect detection. Existing methods are prone to losing texture information of small defects due to excessive downsampling or insufficient receptive fields during the feature extraction stage, leading to positioning errors; traditional multi-scale strategies rely on stacked convolutional layers, resulting in a sharp increase in the number of parameters and limited receptive field expansion, making it difficult to balance efficiency and accuracy; existing models are easily affected by background noise in complex road conditions, and the integration of local and global information is insufficient.

To address these issues, this paper proposes the following improvements: 1) Utilizing a publicly available road defect dataset captured by drones to overcome the challenges posed by complex road structures that ground-based detection vehicles struggle to address; 2) Introducing the Dynamic Dual-Path Feature Reorganization Mechanism (DS_Module) to to address feature loss in small-scale defect detection; 3) Designing a lightweight multi-scale pooling module (SPPF_WD) to reduce computational complexity while enhancing multi-scale feature fusion capabilities; 4) Constructing a multi-path feature extraction enhanced architecture (RCRep2A_FRFN) to improve model detection stability in complex backgrounds.

2. Pavement Disease Data Set and Experimental Environment Configuration

2.1. Pavement Disease Dataset

This study uses the public road damage dataset https://drive.google.com/file/d/16rAC8E52oUAWZ4YXb3O3_25qMUTRhdfV/view?usp=sharing, which contains a total of 2401 road disease images with a resolution of 512×512, covering six typical diseases:Transverse crack, Longitudinal crack, Alligator crack, Oblique crack, Repair, Pothole. Before introducing the data augmentation strategy, the data set is divided into training set, validation set and test set according to the ratio of 8:1:1. The experimental results are shown in Fig.1. The experimental results show that the baseline model without data augmentation improves mAP@0.5 slowly in the early stage of training, and the performance is limited after convergence, and the final accuracy is maintained at 61.2%.In order to optimize the adaptability of the model to complex scenes, image enhancement methods such as data

 doi:10.20944/preprints202506.2429.v1

translation, rotation, occlusion and noise are performed on the original data. After data enhancement, the scale of the data set is expanded to 12005, and the same proportion is re-divided.

After analyzing Figure 1, after the data is amplified, the mAP@0.5 of the model reaches 88.5% after 200 training cycles, which is significantly improved by 27.3% compared with the unenhanced model, which verifies that rotation, scaling and noise disturbance have an optimization effect on the accuracy of target detection in complex scenes. Taking mAP@0.5 as 41% as the baseline, the amplification model mAP@0.5 needs 22 training cycles to achieve this accuracy, which is 28 cycles less than the unamplified model, and the convergence speed is increased by about 2.28 times, indicating that data diversity significantly accelerates feature learning efficiency. In the middle and late stages of training (100-200epoch), the fluctuation range of the accuracy curve of the amplified model is gentle, and there is no oscillation compared with the unamplified model. It shows that image amplification methods such as geometric transformation translation enhance the robustness of the model to occlusion and low contrast diseases, and further improve the model's resistance to complex background interference.

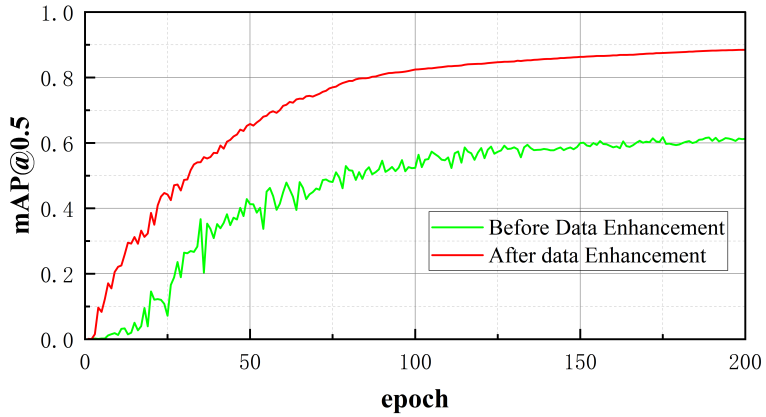


Figure 1. Comparison of mAP50 before and after data enhancement

2.2. Experimental Environment Configuration

The experiments in this experiment are completed on the local high-performance computing platform based on the PyTorch deep learning framework. The experimental parameters epochs are 200, batch is 72, images size is 640, workers is set to 12, and cache is False. The specific hardware and software environment configuration is shown in Table 1.

Table 1. TExperimental environment configuration table.

Experimental environment	Item	Specific information
Hardware environment	GPU	NVIDIA GeForce RTX 3090
	CPU	Intel(R) Core(TM) i9-13900K
	video memory	24GB
	memory	16GB
Software environment	Python	3.8.10
	PyTorch	1.11.0 + cu113
	CUDA	11.3

3. Model Introduction and Improvement

3.1. YOLOv8 Network Model

As an updated version of YOLOv5, YOLOv8 network has been optimized and improved on the basis of inheriting th2e advantages of the previous version.The overall architecture includes three parts : Backbone, Neck and Head, and its network structure is shown in Figure 2[27]. In order to improve the feature extraction ability and reduce the size of the network, the Backbone part uses the C2f module as the basic component unit. Compared with the original model C3 module, the C2f module improves the

 doi:10.20944/preprints202506.2429.v1

computational efficiency while reducing the redundant parameters through efficient structural design, which further enriches the feature expression ability. The Neck part enhances the model's ability to detect targets of different sizes by fusing feature maps from different levels, and optimizes the anchor frame allocation and feature map expression capabilities to achieve efficient fusion of multi-scale features.

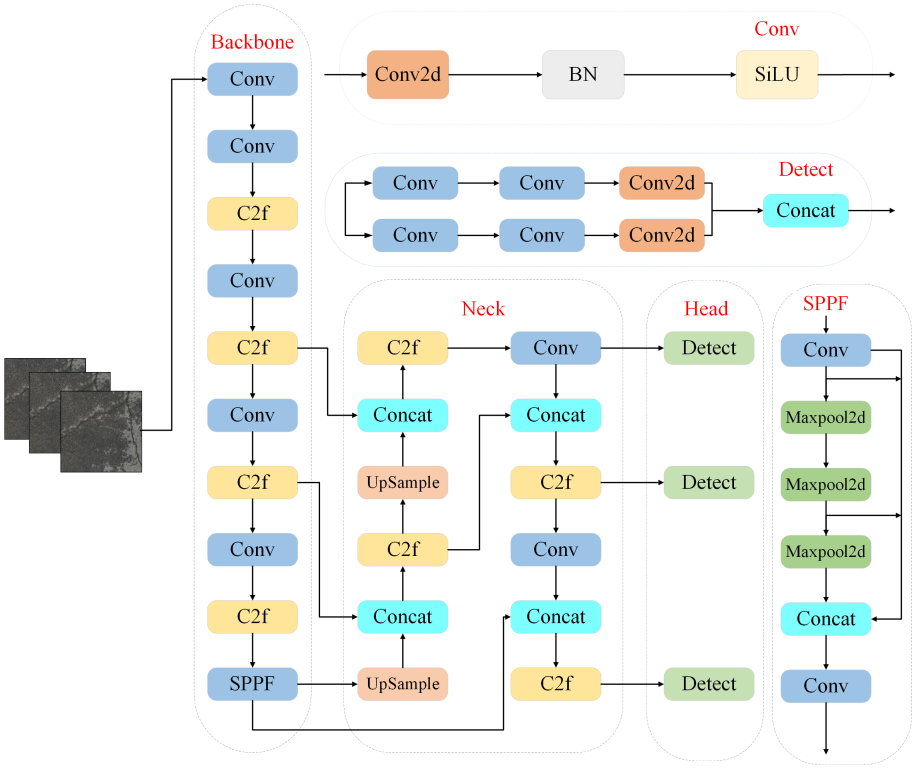


Figure 2. YOLOv8n network structure

3.2. Dynamic Dual-Path Down-Sampling DS_ Module

Since cracks, potholes and other diseases usually show the characteristics of small scale, slender shape and blurred edges, the traditional convolution downsampling structure is easy to lose key details in the feature extraction process, resulting in a decrease in the detection accuracy of small target diseases. In order to solve the above problems, this paper designs a lightweight dynamic dual-path downsampling DS_ module, which combines spatial rearrangement strategy and standard convolution path. Through the dual-path structure, the spatial resolution is compressed while the structural information and detail features of the input image are retained to the greatest extent, so as to improve the network 's perception of small disease targets. The module performs two operations on the input feature map, one is the spatial rearrangement path operation, and the other is the downsampling using the convolution module, as shown in Figure 3.

In order to reduce the loss of spatial information, the DS_ module adopts the spatial rearrangement strategy. The module divides the input feature map into four sub-regions (X1,X2,X3,X4) through spatial rearrangement. The information is extracted from different spatial locations and spliced on the channel dimension to achieve the reduction of spatial resolution and the expansion of the number of channels, which enhances the expression ability of local detail features. Then, the 1×1 convolution is used to fuse the rearranged features, and the fine-grained context information is further extracted, and the output is Y1.

 doi:10.20944/preprints202506.2429.v1

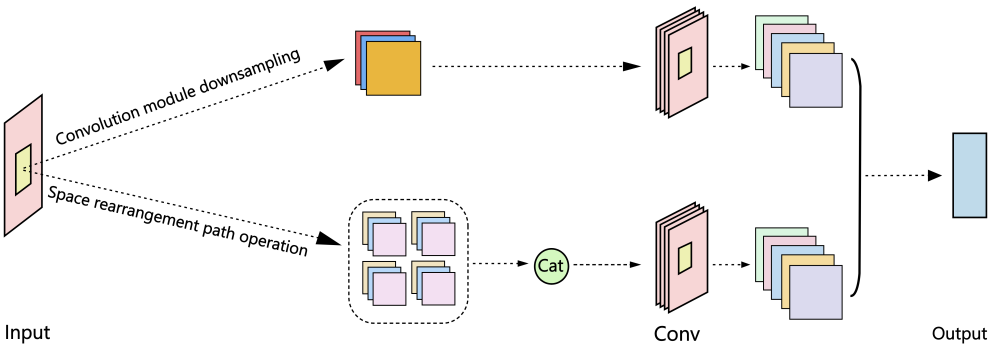


Figure 3. Dynamic dual-path down-sampling DS_ module diagram

DS_module introduces a standard convolution path in parallel, directly down-sampling the input feature map, extracting global features with higher semantic level, and the output is Y2. Finally, by splicing the output Y1 and Y2 of the two paths, a comprehensive feature map integrating local and global information is formed. The output features after splicing integrate high-precision detail information and semantic-level down-sampling features, and realize the dual perception of local texture and global structure.

3.3. Multi-Scale Pooling SPPF_WD Module

To address issues such as inconsistent target scales and varying background complexities in road defect detection, which lead to low accuracy, this paper proposes a lightweight module based on Spatial Pyramid Pooling (SPPF) called SPPF_WD. This module enhances the model's ability to perceive features of different sizes through multi-scale pooling and feature fusion techniques, while significantly reducing computational complexity through lightweight convolutions, thereby improving the detection accuracy of defects at different scales. The SPPF_WD module consists of lightweight convolutions (Lightconv) and multi-scale pooling (SPPF). Compared to the traditional SPPF structure, it features lightweight, serialized, and multi-scale feature enhancement structures, as shown in Figure 4.

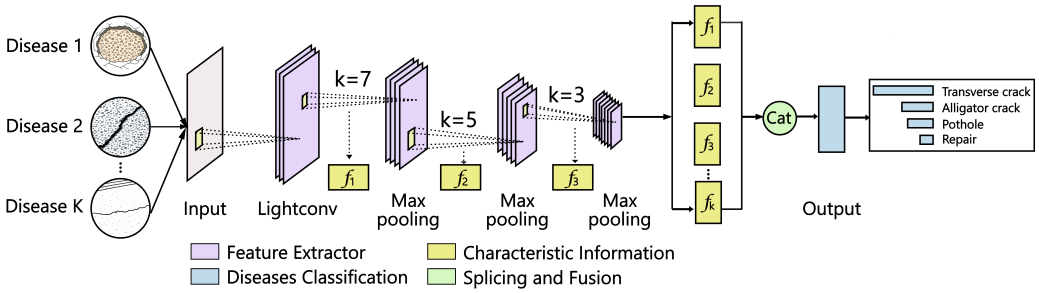


Figure 4. SPPF_WD Module Diagram

The Lightconv module [28] is a deep convolutional module designed for sequential modeling. It consists of an initial linear transformation (Linear), gate-controlled linear units (GLU), lightweight separable convolutions (LConv), and a final linear fusion layer. By utilizing local convolutional operations and weight normalization techniques, it significantly reduces the number of parameters and computational complexity while preserving the model's expressive power, as shown in Figure 5.

 doi:10.20944/preprints202506.2429.v1

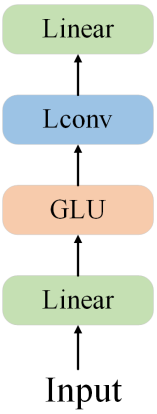


Figure 5. lightconv module diagram

First, the Linear module projects the input features through a fully connected layer to construct a more abstract feature space for GLU activation and convolution operations. Second, the GLU mechanism uses dynamic gating to effectively highlight key regions such as cracks and fissures while suppressing interfering information like shadows and lighting changes.Its mathematical form is expressed in Equation (1).Then, the LConv module replaces the traditional self-attention mechanism with separable convolutions. significantly reducing computational complexity while maintaining local context modeling capabilities. This structure enables fast inference while preserving high-resolution spatial details. Finally, the Linear module performs feature fusion again.

$$\mathrm{GLU}(A, B) = A \otimes \sigma(B) \tag{1}$$

where A and B are the two branches of the linear layer output, $\otimes$ denotes element-wise multiplication, and σ is the sigmoid function. The SPPF module employs serial stacked max pooling operations (K=7, K=5, and K=3) to fuse feature information at different scales, enhancing the model's ability to recognize targets at various scales while maintaining low computational costs.This enables simultaneous detection of both small cracks and large-scale defects, as shown in Figure 6.

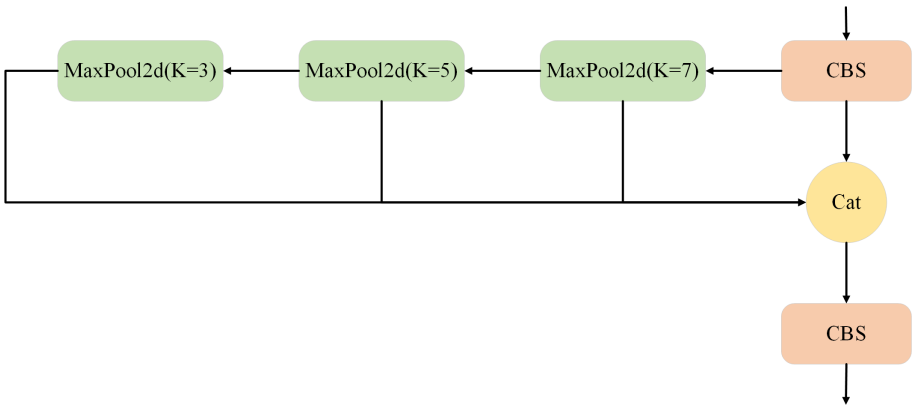


Figure 6. SPPF Module Diagram

The CBS module compresses channel dimensions, standardizes, and enhances nonlinear expressive capabilities. then uses maximum pooling (MaxPool) at multiple scales to pool the input feature maps into multiple regions of different sizes. Compared to traditional parallel pooling in SPP, serial pooling gradually expands the receptive field, extracting both global and local information, which is beneficial for handling low-contrast defects and improving robustness under complex road conditions such as strong light, shadows, and rain interference.Finally, the Cat module concatenates the original input features with the three-layer pooled features to output a fused feature map containing multi-scale

 doi:10.20944/preprints202506.2429.v1

contextual information, enhancing detection capabilities.The CBS module then fuses and concatenates the features again to output a fixed-dimensional feature map.

3.4. Multi-Path Feature Extraction and Enhancement RCRep2A_FRFN Module

To address the complexity of road defects in terms of morphology and distribution, as well as issues such as high or limited computational costs and insufficient feature representation capabilities, this paper introduces an efficient structural module—the RCRep2A_FRFN module—that integrates multi-path feature extraction with deep feature enhancement. This module aims to enhance the model's ability to perceive and represent defect regions. It primarily integrates the Ghostconv module and the feature refinement FRFN module, offering efficient feature extraction capabilities, multi-path feature fusion mechanism, and fine-grained feature enhancement advantages, as shown in Figure 7.

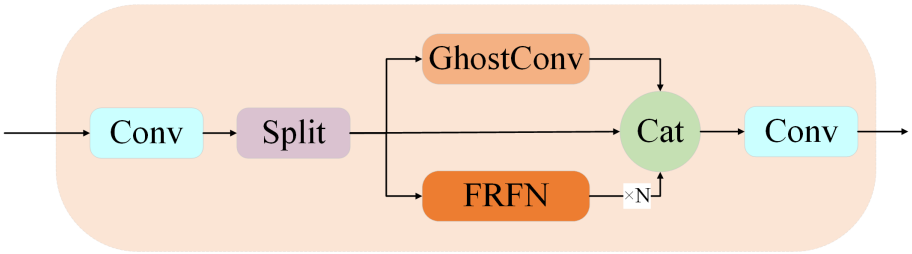


Figure 7. RCRep2A_FRFN image

The GhostConv module was proposed by Han et al.[29] in 2020. Its core idea is to generate more "ghost" feature maps from existing feature maps through low-cost linear transformations, thereby improving the computational efficiency of networks. As shown in Figure 8, the module primarily consists of three steps: regular convolution, ghost generation, and feature map concatenation.

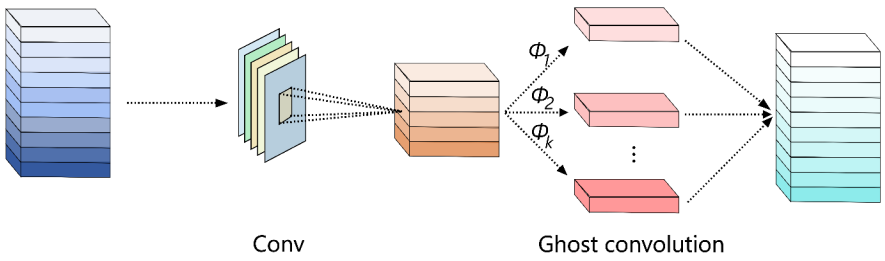


Figure 8. Ghostconv module diagram

The GhostConv module first performs a regular convolution to generate an initial feature map.The module then generates additional feature maps through inexpensive operations, denoted as Φ_1, Φ_2, ..., Φ_k. Finally, in the ghost module, multiple Φ operations generate supplementary feature maps to enhance the network's representation capability of images while performing low-cost operations.The additional generated feature maps are concatenated with the original feature map to form the final output. This module generates more feature maps from existing ones through linear transformations, significantly reducing the required computational load compared to additional convolutional layers. The calculation process is shown in formulas (2) and (3).

 doi:10.20944/preprints202506.2429.v1

$$t_s = \frac{h \times w \times c \times H \times W \times n}{\frac{n}{s} \times H \times W \times k \times k \times c + (s-1) \times n/s \times H \times W \times d \times d} = \frac{c \times k \times k}{\frac{1}{s} \times c \times k \times k + \frac{(s-1)}{s} \times d \times d} \approx s \tag{2}$$

$$t_s = \frac{n \times c \times k \times k}{\frac{n}{s} \times c \times k \times k + (s-1) \times n/s \times d \times d} = \frac{s \times c}{s + c - 1} \approx s \tag{3}$$

where:

- h, w, c: height, width, and channels of the input feature map
- H, W: height and width of the output feature map
- n: number of convolutional kernels
- k: size of convolutional kernel
- d: kernel size in linear transformation
- s: number of transformations
- t_s, t_c: computational and parameter ratios between regular convolution and GhostConv

The Feature Refinement and Fusion Network (FRFN) module [30] enhances the network's ability to capture both local and global information by progressively refining features layer by layer.This module effectively processes detailed information in complex backgrounds, particularly improving the recognition of subtle defects such as cracks and potholes in road surface images.The architecture of the FRFN module is illustrated in Figure 9.

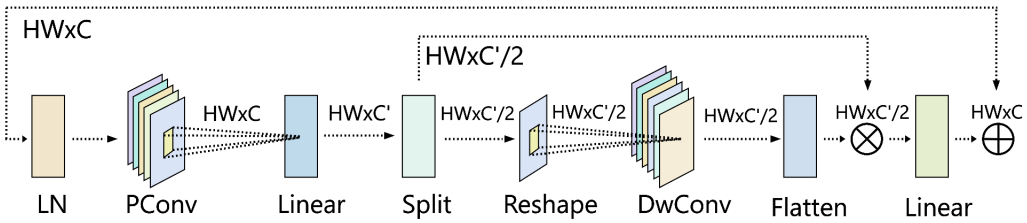


Figure 9. FRFN Module Diagram

The FRFN module employs a series of carefully designed processing steps:

1. **Layer Normalization (LN)**: Stabilizes the training process and accelerates convergence.
2. **Pointwise Convolution (PConv)**: Performs efficient inter-channel information fusion and spatial feature extraction, enhancing feature representation capability.
3. **Linear Layer**: Maps input features to meet the requirements of subsequent network layers.
4. **Split and Reshape Operations**: Enhance feature fusion capabilities across different levels and channels, enabling extraction of subtle features from multiple dimensions.
5. **Depthwise Convolution (DWConv)**: Performs independent convolution operations on each input channel, effectively reducing computational load while preserving spatial information of feature maps.
6. **Flatten Operation**: Transforms processed feature maps into a flattened format suitable for the linear layer.
7. **Final Linear Layer**: Further maps features to improve detection accuracy for complex defects.

Through this sequence of refined feature processing steps, the FRFN module significantly enhances the model's ability to identify subtle defects in road surfaces. The module's key advantages include:

 doi:10.20944/preprints202506.2429.v1

- Progressive refinement of features layer by layer
- Effective handling of complex background details
- Enhanced fusion of features across different levels and channels
- Significant improvement in recognizing minute defects (cracks, potholes)
- Balanced computational efficiency and feature representation

As shown in Figure 7, the specific operation of the RCRep2A_FRFN module involves using conventional convolution (conv) to perform convolution processing on the input feature map to extract information. Next, the Split operation divides the input feature map into different parts to achieve multi-path feature processing, thereby enhancing the model's flexibility and expressive diversity. Then, the core Ghostconv module performs inexpensive operations on redundant feature maps, significantly reducing computational complexity while maintaining high expressive capability. effectively improving model efficiency. The FRFN module then performs multi-level stacking and refinement processing on the generated feature maps to enhance the network's discriminative ability, further improving the model's discriminative capability in complex tasks. Finally, the Cat module concatenates and fuses the multiple feature groups processed by the GhostConv and FRFN modules, further achieving deep feature fusion and multi-scale modeling, thereby comprehensively expressing multi-scale information.The RCRep2A_FRFN module achieves optimized model efficiency while maintaining performance through feature partitioning, redundant feature generation, refinement enhancement, and multi-channel fusion, thereby improving the model's overall accuracy.

3.5. YOLOv8n-DSRS Network Model

To achieve more accurate and efficient detection of road defects, the DS_ module was designed in the backbone network (Backbone) section of YOLOv8n, the SPPF_WD module was adopted, and the RCRep2A_FRFN module was introduced. The improved network model was named YOLOv8n-DSRS. Compared to YOLOv8, this network model achieves more efficient detection performance while reducing computational overhead through lightweight operations and refined modules. it achieves more efficient detection performance while reducing computational overhead. Its network structure is shown in Figure 10.

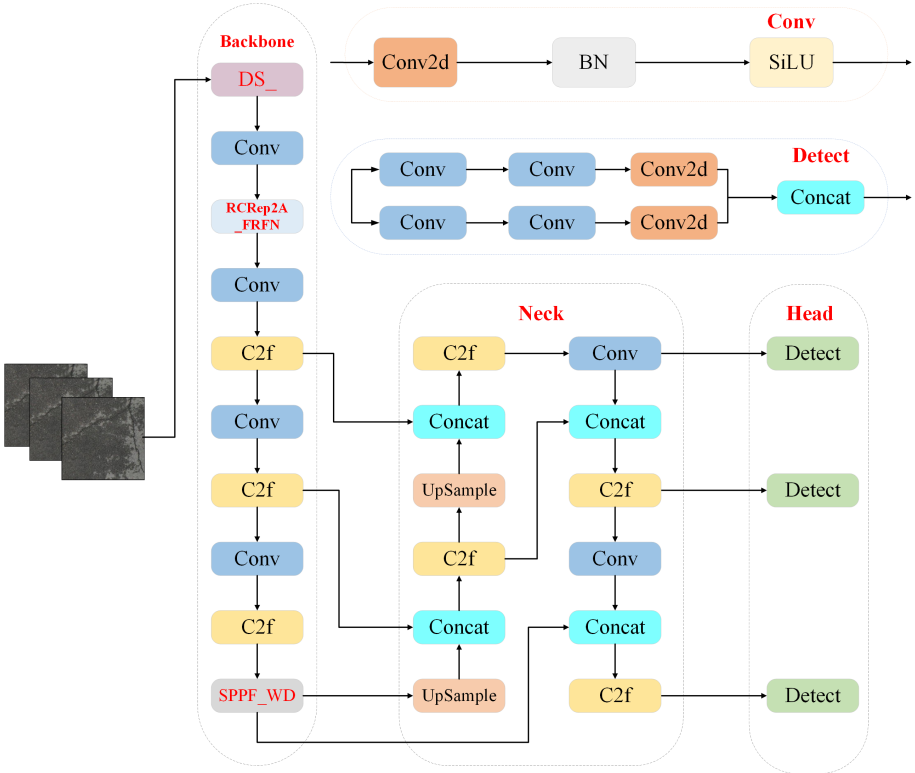


Figure 10. YOLOv8n-DSRS Network Model

 doi:10.20944/preprints202506.2429.v1

4. Experimental Results and Analysis

4.1. Evaluation Criteria

When evaluating model performance, multiple metrics are often used to comprehensively assess its effectiveness. This paper employs the following evaluation metrics: mean average precision (mAP), F1-score, recall rate (Recall), number of parameters (Parameters), number of network layers (Layers), and the number of billion floating-point operations (GFLOPs). Among these, mAP is used to assess the model's precision in information retrieval or ranking tasks, reflecting its average performance across multiple queries or classifications.The F1-score is the harmonic mean of precision and recall, providing a balanced metric for the model's performance in predicting positive and negative classes. Recall evaluates the model's ability to identify true positive samples. Parameters indicate the number of model parameters, reflecting its complexity and computational requirements. GFLOPs measure the computational complexity of the model, specifically the number of floating-point operations required per second. The larger the values of Params and GFLOPs, the higher the hardware requirements and performance needed. Layers indicate the depth of the model; the larger the number of layers, the more features the model captures.

4.2. Melting Experiments and Analysis

To validate the effectiveness of the three proposed modules on the YOLOv8n base network, six experiments were conducted, with the specific results shown in Table 2. As shown in Table 2, introducing the lightweight dynamic dual-path downsampling module DS_ into the original YOLOv8n network model improved the model's mAP50 from 88.5% to 92.4% while keeping the number of parameters and computational complexity nearly unchanged. an increase of 4.4% compared to the original model, validating that the DS_ module can effectively compress spatial resolution while retaining the detail information of the input image, thereby enhancing the model's ability to detect small-scale target anomalies. Further introduction of the FRFN module resulted in the YOLOv8n+FRFN model's mAP50 improving to 94.1%, and mAP50-95 increasing from 62.1% to 70.6%, representing increases of 6.3% and 13.7% respectively compared to the original model. The FRFN module enhances feature refinement layer by layer, strengthening the capture of local and global information, and improving detail extraction and information fusion in complex backgrounds. Although the number of network layers increased by 8.3% compared to the original model, the number of parameters and GFLOPs remained nearly stable, with a significant improvement in accuracy, validating the model's effectiveness in complex scenarios. When both the DS_ and FRFN modules are introduced, mAP50 further improves to 95.7%, and mAP50-95 reaches 76.1%. Experiments show that the combined use of the two modules has a good synergistic effect, enhancing the ability to recognize small objects while maintaining low computational overhead and a small model size. Finally, by introducing SPPF_WD, the mAP50 of the YOLOv8nDSRS network model reached 96.3%, and mAP50-95 improved to 77.3%. This demonstrates that the SPPF_WD module enhances the model's perception of diseases at different scales through multi-scale pooling and feature fusion techniques, while effectively reducing computational complexity. The introduction of this module further optimizes the model's robustness when handling multi-scale targets. In summary, the ablation experiments validate the effectiveness of each module in the original network.

Table 2. Ablation Study

Algorithm	P%	R%	mAP50%	mAP50-95%	Layers	Parameters	GFLOPs
YOLOv8n	88.2	80.6	88.5	62.1	72	3,006,818	8.1
YOLOv8n+DS_	90.3	85.4	92.4	68.7	73	3,006,578	8.0
YOLOv8n+FRFN	92.2	86.7	94.1	70.6	78	3,006,138	8.1
YOLOv8n+DS_+FRFN	95.4	89.7	95.7	76.1	79	3,005,898	8.0
YOLOv8n+SPPF_WD	91.6	87.8	93.2	69.0	77	3,007,074	8.1
YOLOv8n-DSRS	95.6	92.8	96.3	77.3	84	3,006,154	8.0

 doi:10.20944/preprints202506.2429.v1

4.3. Comparison and Analysis of the Original YOLOv8 Model and the Improved Model

Figure 11 shows the comparison curves of various evaluation metrics between the original network model and the improved algorithm.

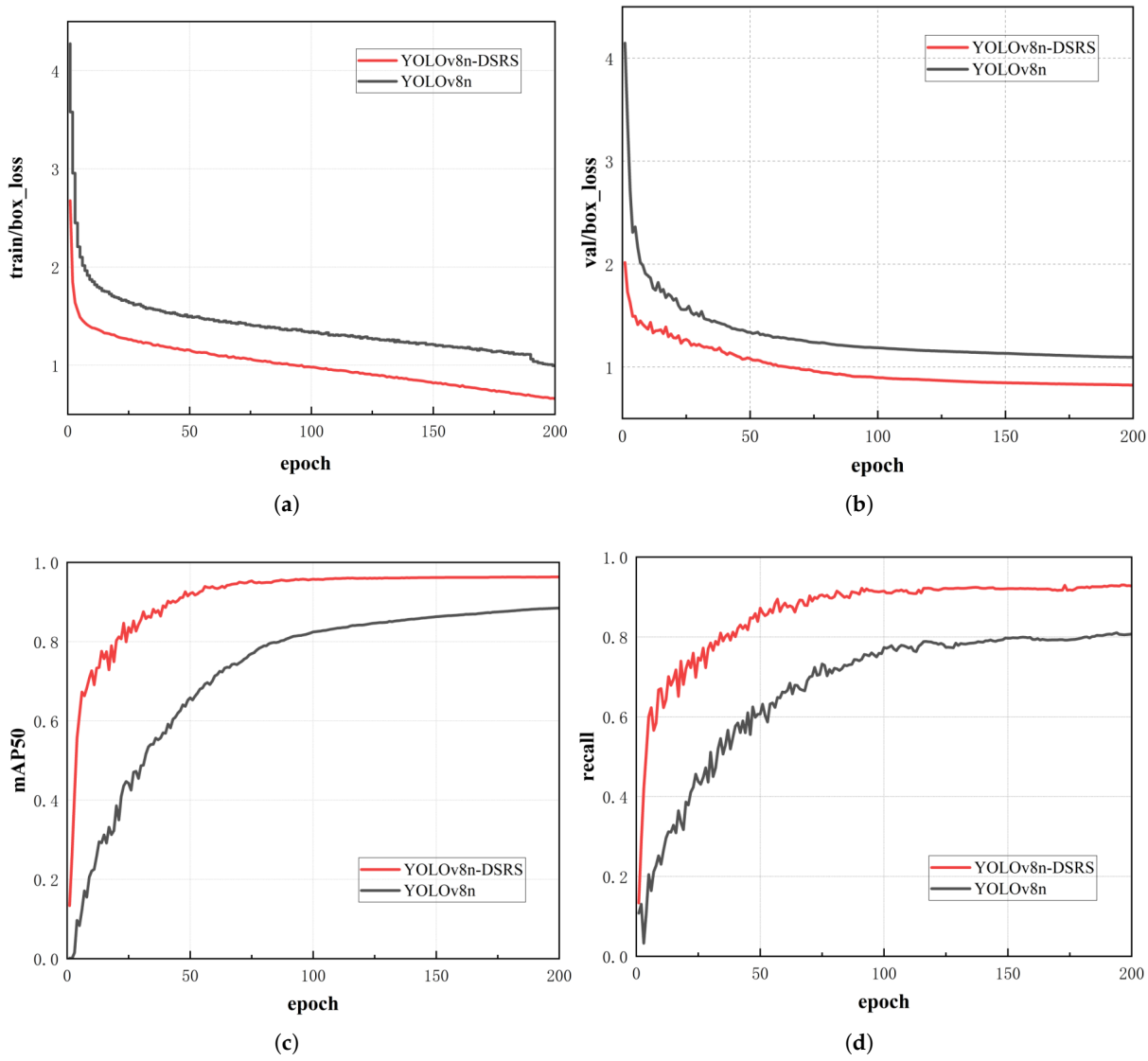


Figure 11. Comparison of evaluation metrics between YOLOv8n and YOLOv8n-DSRS: (**a**) Training set loss function comparison chart. (**b**) Comparison chart of validation set loss functions. (**c**) mAP50 comparison chart. (**d**)recall comparison chart

As shown in Figure (a), YOLOv8n-DSRS demonstrates significantly better box loss performance than YOLOv8n during training. In the early stages of training (0–50 epochs), the box loss decreases at a faster rate, improving by approximately 22.6% compared to the baseline, and continues to decrease steadily in subsequent training, ultimately reaching a low loss level. while the original network model exhibits a slower decrease in bounding box loss and slight fluctuations in the later stages. The improved algorithm maintains a smooth decreasing trend, demonstrating its excellent convergence and stability. It reduces unnecessary computations and localization interference while also achieving lightweight optimization by reducing the number of parameters and computational complexity.

Figure (b) shows the trend of bounding box loss on the validation set for both models, systematically verifying the improved algorithm's positive effect on target localization performance. YOLOv8n-DSRS achieved a 19.1% improvement in loss reduction rate compared to the original model during the early stages (0-50 epochs), and after training to 200 epochs, the improved model's final

 doi:10.20944/preprints202506.2429.v1

loss value was 24.7% lower than the original model's. This verifies its enhanced role in improving the localization accuracy of multi-scale targets, especially those in complex backgrounds. Compared to the earlier fluctuations, the improved algorithm did not show any fluctuations in the middle and late stages (100-200 epochs).This may be because the RCRep2A_FRFN module in the improved model filters key feature layers and enhances the expression of target edges and textures, significantly reducing the interference of redundant calculations on the localization task, effectively balancing the model capacity and generalization ability, and suppressing the risk of overfitting.

As shown in Figure (c) for the mAP50 of the two network models, the improved algorithm achieves a detection accuracy close to 1.0 after 200 training cycles, an 8.8% improvement over the baseline model. This indicates that the improved model optimizes multi-scale feature fusion capabilities through the SPPF_WD module, significantly enhancing detection robustness in complex scenes. In the early training phase (0–50 epochs), the improved model's accuracy increased from 13% to 92%, while the original model's accuracy increased from 0% to 66%, verifying that the DS_ module accelerates the learning efficiency of key features while reducing gradient redundancy. In the mid-to-late training phase (100–200 epochs), the improved model maintained a steady increase with a fluctuation amplitude of 0.59%, a 6.7% reduction compared to the original algorithm. This indicates that the improved algorithm effectively balances model capacity and generalization ability by suppressing redundant feature interference through the RCRep2A_FRFN module. Therefore, the YOLOv8n-DSRS algorithm demonstrates superior performance in detection tasks compared to the original algorithm.

Figure (d) verifies the improved model's optimization effect on false negatives. Experimental results show that after 200 epochs of training, the recall rate of the improved algorithm reached 92.7%, while the original model achieved 80.7%, representing a 12% improvement and a significant reduction in false negatives. The fluctuation range of the improved model during the mid-term training phase was reduced by 18.45% compared to the baseline model's 9.08%, significantly lowering performance fluctuations during training and demonstrating superior stability and convergence.

4.4. Comparison and Analysis of Different Model Experiments

To further evaluate the superiority of the improved algorithm YOLOv8n-DSRS proposed in this paper, comparative experiments were conducted with the current mainstream algorithms YOLOv5, YOLOv6, the YOLOv8 series, and YOLOv9 under the same environment and dataset. The experimental results are shown in Table 3.

Table 3. Comparison of Mainstream Algorithms

Algorithm	P%	R%	mAP50%	mAP50-95%	Layers	Params	GFLOPs
YOLOv5	83.7	81.8	87.0	57.7	157	1.8M	4.2
YOLOv6	69.7	60.1	64.0	40.6	142	4.2M	11.8
YOLOv8-powerneck	83.6	76.5	84.6	58.9	209	3.4M	9.3
YOLOv8-C2f_Atten_iRMB	93.2	88.2	94.6	72.3	187	3.5M	8.4
YOLOv8-P6	91.0	83.8	92.3	66.4	220	4.8M	8.1
YOLOv8-ghost-P6	87.8	82.2	90.1	64.1	409	2.7M	5.0
YOLOv8-word	91.1	86.9	92.3	68.8	195	3.5M	10.0
YOLOv9	91.3	87.1	92.4	69.8	604	50.7M	236.7
YOLOv8n-DSRS	95.6	92.8	96.3	77.3	84	3.0M	8.0

[1] All experiments conducted on the same dataset and hardware configuration. YOLOv8n-DSRS achieves state-of-the-art performance with minimal computational overhead.

In terms of detection accuracy, YOLOv8n-DSRS significantly outperforms all comparison models with a precision rate (P) of 95.6% and a recall rate (R) of 92.8%. A visual comparison of the algorithms is shown in Figure 12. Compared to the YOLOv5 algorithm, it achieves improvements of 11.9% and 11.0% in precision and recall, respectively. Additionally, it outperforms YOLOv9 by 3.9% and 7.5% in mAP50% and mAP50-95%, respectively, demonstrating the algorithm's superior robustness and higher accuracy in high-difficulty detection scenarios.

 doi:10.20944/preprints202506.2429.v1

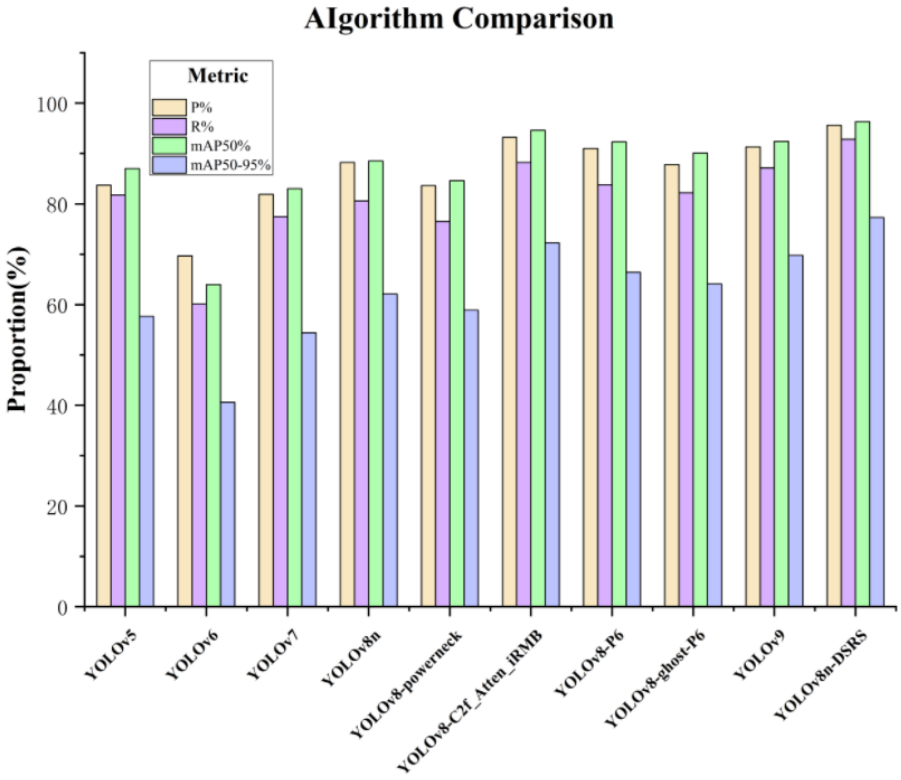


Figure 12. Visual analysis chart comparing different algorithms

Compared to the YOLOv9 model, this algorithm reduces the number of parameters and the number of network layers by 98% and 86.1%, respectively, significantly lowering memory and computational resource consumption. Compared to similar algorithms, this algorithm has the fewest network layers, with a ratio of 20.5% compared to YOLOv8-ghost-P6, and a parameter count ratio of 62.9% compared to YOLOv8-P6. This demonstrates the algorithm's lightweight characteristics, as visualized in Figure 13. In summary, the YOLOv8n-DSRS algorithm proposed in this paper possesses high accuracy, lightweight design, and low computational complexity.

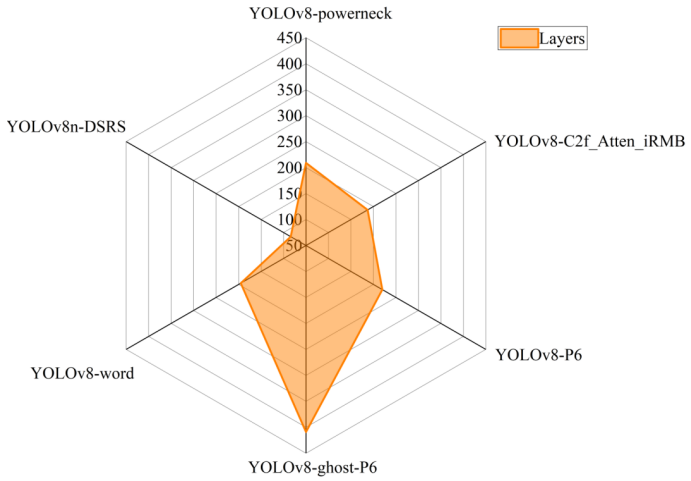
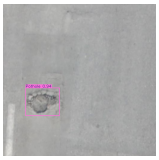
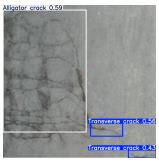
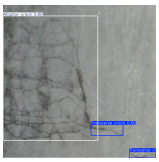
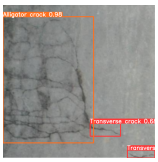
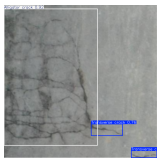
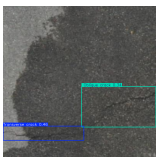
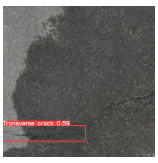
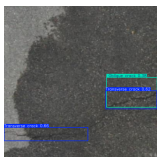
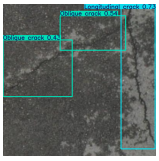
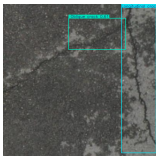
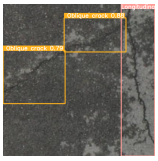
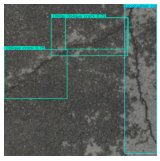
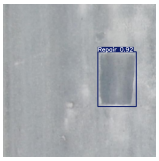
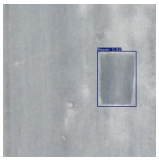
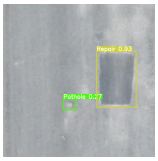
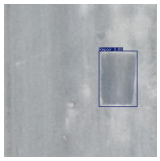
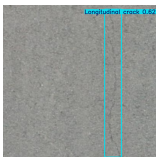
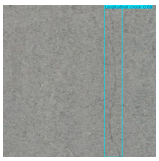
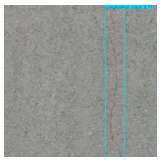


Figure 13. Similar algorithms Layers visualization analysis diagram

For illustrative purposes, the actual detection results of the mainstream algorithms YOLOv5, YOLOv8n, YOLOv9, and the improved algorithm in this paper are shown in Table 4.

 doi:10.20944/preprints202506.2429.v1

Table 4. Comparison of Detection Performance for Different Models

Defect Type and Algorithm Model	YOLOv5	YOLOv8n	YOLOv9	YOLOv8n-DSRS
Pothole				
Alligator crack				
Transverse crack				
Oblique crack				
Repair				
Longitudinal crack				

[1] Cells contain visual comparison of detection results. YOLOv8n-DSRS demonstrates superior detection accuracy for all road defect types, especially small cracks and potholes.

5. Conclusions

The working points of this study can be summarized as follows:

1) Dynamic dual-path feature reorganization mechanism: propose dynamic dual-path down-sampling DS_module, through the parallel design of spatial rearrangement strategy and standard convolutional paths, to reduce the spatial resolution while retaining the detail information, to solve the problem of feature loss in the detection of small-scale diseases, so as to improve the mAP50-95% to 77.3%, which is 15.2% higher than the baseline model YOLOv8n.

2) Multiscale pooling and lightweight feature fusion: design SPPF_WD module, using serial multiscale pooling combined with lightweight convolution LightConv, to enhance the model's ability to perceive multiscale diseases, while reducing the computational complexity GFLOPs to 8.0, which is 96.6% less than YOLOv9.

3) Adaptive Sparse Feature Enhancement Architecture: the RCRep2A_FRFN module is introduced to generate redundant feature maps through GhostConv and combine the feature refinement FRFN strategy to strengthen the fusion of local texture and global semantic information, which improves the Recall rate Recall to 92.8% and reduces the leakage rate by 12% in complex scenes.

 doi:10.20944/preprints202506.2429.v1

References

1. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11), 2278-2324.
2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
3. 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).
4. Lv B, Zhang S, Gong H, et al. Pavement Disease Visual Detection by Structure Perception and Feature Attention Network. *Applied Sciences*, 2025, 15(2):551-551.
5. Hou Y, Li Y, Du M, et al. Bridging Data Distribution Gaps: Test-Time Adaptation for Enhancing Cross-Scenario Pavement Distress Detection. *Applied Sciences*, 2024, 14(24):11974-11974.
6. Li J, Yuan C, Wang X, et al. Semi-supervised crack detection using segment anything model and deep transfer learning. *Automation in Construction*, 2025, 170:105899-105899.
7. Liu Z, Wu W, Gu X, et al. PaveDistress: A comprehensive dataset of pavement distresses detection. *Data in Brief*, 2024, 57:111111-111111.
8. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*, 2017, 30.
9. Huang Q W, Feng L, He L Y. LTPLN: Automatic pavement distress detection. *PloS one*, 2024, 19(10):e0309172.
10. Mahdy, Kamel, et al. "Pavement distress instance segmentation using deep neural networks and low-cost sensors." *Innovative Infrastructure Solutions*, 2024, 9(1):6.
11. Haohui Y, Junfei Z. UAV-PDD2023: A benchmark dataset for pavement distress detection based on UAV images. *Data in Brief*, 2023, 51:109692-109692.
12. Cancan Y, Jun L, Tao H, et al. An efficient method of pavement distress detection based on improved YOLOv7. *Measurement Science and Technology*, 2023, 34(11).
13. Wu, Lingxiao, Zhugeng Duan, and Chenghao Liang. "Research on asphalt pavement disease detection based on improved YOLOv5s." *Journal of Sensors*, 2023, 1:2069044.
14. Chu, Yinze, et al. "Pavement disease detection through improved YOLOv5s neural network." *Computational Intelligence and Neuroscience*, 2022, 1:1969511.
15. Yang, Zhen, Lin Li, and Wenting Luo. "PDNet: Improved YOLOv5 nondeformable disease detection network for asphalt pavement." *Computational Intelligence and Neuroscience*, 2022, 1:5133543.
16. Hou, Yun, et al. "The application of a pavement distress detection method based on FS-Net." *Sustainability*, 2022, 14(5):2715.
17. Du, Yuchuan, et al. "Pavement distress detection and classification based on YOLO network." *International Journal of Pavement Engineering*, 2021, 22(13):1659-1672.
18. Sun, P., et al. "DSWMamba: A deep feature fusion mamba network for detection of asphalt pavement distress." *Construction and Building Materials* 469 (2025): 140393.
19. Abdelkader, M. F., et al. "EGY_PDD: a comprehensive multi-sensor benchmark dataset for accurate pavement distress detection and classification." *Multimedia Tools and Applications* (2025): 1-36.
20. He, J., Gong, L., Xu, C., et al. "HighRPD: A high-altitude drone dataset of road pavement distress." *Data in Brief* 59 (2025): 111377-111377.
21. Zhao, Y., et al. "An efficient pavement distress detection scheme through drone-ground vehicle coordination." *Transportation Research Part A: Policy and Practice* 180 (2024): 103949.
22. Li, Y., et al. "Crackyolo: Rural pavement distress detection model with complex scenarios." *Electronics* 13.2 (2024): 312.
23. Hu, X., Yan, Y., Wang, D., et al. *A lightweight detection method for road surface defects based on the YOLOM algorithm*. Journal of Chinese Highway Engineering, 2024, 37(12): 381-391. DOI: 10.19721/j.cnki.1001-7372.2024.12.016.
24. Han, Z., et al. "MS-YOLOv8-based object detection method for pavement diseases." *Sensors* 24.14 (2024): 4569.
25. Wang, S., et al. "Automated detection of pavement distress based on enhanced YOLOv8 and synthetic data with textured background modeling." *Transportation Geotechnics* 48 (2024): 101304.
26. Liu, W., et al. "Intelligent detection of hidden distresses in asphalt pavement based on GPR and deep learning algorithm." *Construction and Building Materials* 416 (2024): 135089.
27. Li, X. & Zhang, Y. *Improved Road Damage Detection Algorithm Based on YOLOv8n*. IAENG International Journal of Computer Science, 2024, 51(11).

 doi:10.20944/preprints202506.2429.v1

28. Wu, F., Fan, A., Baevski, A., et al. *Pay Less Attention with Lightweight and Dynamic Convolutions*. CoRR, 2019, https://arxiv.org/abs/1901.10430.
29. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). GhostNet: More Features from Cheap Operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1580-1589).
30. Zhou, S., Zhang, J., Pan, J., *et al.* "Adapt or Perish: Adaptive Sparse Transformer with Attentive Feature Refinement for Image Restoration." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.