Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

MADF-YOLOv8: A Lightweight Model for Road Distress Detection Based on Adaptive Multiscale Feature Fusion

Tao OuYang, Haohui Yu, Guanlin Pan, Yan Cui*, Qingling Chang, Xiulong Fu

Wuyi University, Jiangmen 529020, Guangdong, China

**Author to whom correspondence should be addressed.*

Abstract: Efficient road distress detection is crucial for transportation safety. To address the challenge of balancing detection accuracy, efficiency, and multi-scale feature fusion in existing methods, this paper proposes a lightweight model named MADF-YOLOv8. The model enhances multi-scale feature extraction capability by introducing the Multi-Scale Ghost Residual Convolution (MSGRConv) and the Multiscale Adaptive Feature Processing Module (MAFP). Furthermore, it constructs a Multi-scale Dynamic sampling Bidirectional Feature Pyramid Network (MD-BiFPN) and incorporates the C2f-Faster module to optimize feature fusion efficiency. Experiments on the RDD2022 dataset demonstrate that the proposed model achieves a mean Average Precision at 0.5 Intersection over Union (mAP@0.5) of 88.6% with only 2.312 million parameters. Its overall performance surpasses various mainstream detectors, achieving an exceptional balance between accuracy and efficiency.

Keywords: Road distress detection; Multi-scale feature fusion; YOLOv8

Online publication: December 16, 2025

1. Introduction

As critical infrastructure, highways inevitably develop cracks that progressively deteriorate, significantly increasing maintenance costs and safety risks. Current road distress detection methods face two persistent challenges: the inherent trade-off between detection accuracy and processing speed, and ineffective multi-scale feature integration leading to high false-negative rates. To address these limitations, this study proposes MADF-YOLOv8, an enhanced detection model incorporating a novel Multiscale Adaptive Feature Processing Module (MAFP) and a Multi-scale Dynamic Sampling Bidirectional Feature Pyramid Network (MD-BiFPN). The architecture integrates several key improvements: the C2f-Faster module replaces standard components to reduce computational redundancy, while MSGRConv enhances multi-scale feature extraction in the backbone network. These innovations collectively improve feature representation and fusion efficiency, particularly for fine crack details. Experimental results on the RDD2022 dataset demonstrate that our method achieves a state-of-the-art mAP@0.5 of 88.6% with only 2.312M parameters. This demonstrates its potential for enabling efficient and intelligent road inspection systems.

2. Methodology

2.1. Overview of methodology

This paper proposes MADF-YOLOv8, an improved YOLOv8-based model, with the network architecture depicted in **Figure 1**. MSGRConv replaces standard convolutions in the backbone, enhancing multi-scale feature extraction while maintaining lightweight design. Following that, the proposed MAFP module supplants the SPPF layer to refine directional and fine-grained feature extraction. A novel C2f_FasterBlock then restructures the bottleneck using FasterNet blocks, reducing computational redundancy. Finally, the proposed MD-BiFPN replaces the standard neck to better integrate multi-scale information through dynamic sampling and grouped residual convolution, significantly improving cross-scale feature fusion.

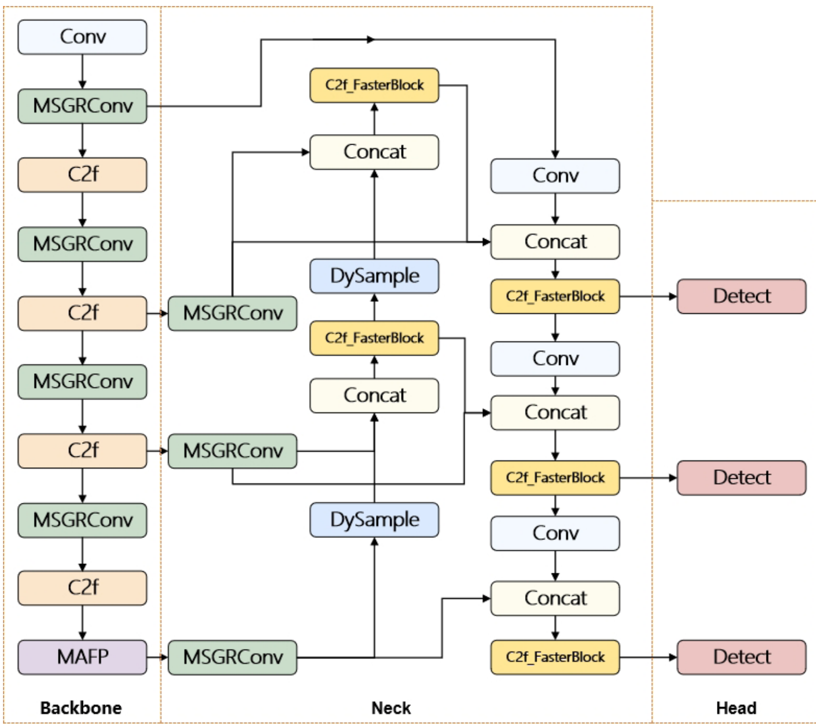

Figure 1. Network architecture of MADF-YOLOv8.

2.2. Multi-scale ghost residual convolution

Traditional convolutional neural networks face inherent limitations in multi-scale feature extraction due to their fixed-size kernels. Expanding the receptive field typically requires increasing network depth or kernel size, which substantially raises computational costs and parameters. While layer stacking can achieve multi-scale extraction, it often induces gradient vanishing, restricts model depth, and diminishes adaptability to scale variations.

In road distress detection, standard convolutions' single-scale nature and high computational load hinder the detection of small targets like fine cracks and complex morphological features. To address this, we propose the Multi-Scale Ghost Residual Convolution (MSGRConv) module, which integrates multi-scale features while reducing computational redundancy, significantly improving detection performance. As shown in **Figure 2**, MSGRConv processes input features through a 1×1 convolution, then splits and routes them to 3×3 and 5×5 Ghost modules before merging outputs. A residual block with DSConv prevents gradient issues. Ghost modules reduce feature redundancy through group convolutions, generating lightweight features that emulate original distributions while substantially decreasing parameters and computations.

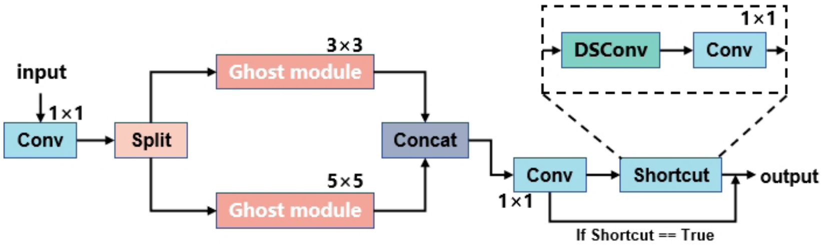

Figure 2. Architecture of the multi-scale ghost residual convolution module.

2.3. Multiscale adaptive feature processing module

The SPPF layer in YOLOv8 aggregates multi-scale contextual information through successive max-pooling operations. However, this structure presents notable limitations for road-distress detection. The repeated pooling steps substantially reduce spatial resolution, leading to the loss of fine-grained crack details that are crucial for reliable identification. In addition, the use of isotropic square pooling windows makes it difficult to represent orientation-dependent patterns, such as linear, oblique, or branched cracks. Moreover, the static nature of the pooling operation restricts its ability to perform adaptive feature selection, which in turn weakens its capacity to distinguish subtle crack features from complex background textures and noise. To overcome these issues, we propose the MAFP module, which employs grouped asymmetric convolutions and dual-path adaptive spatial attention to enhance global context modeling and crack feature representation. The structure is shown in **Figure 3**.

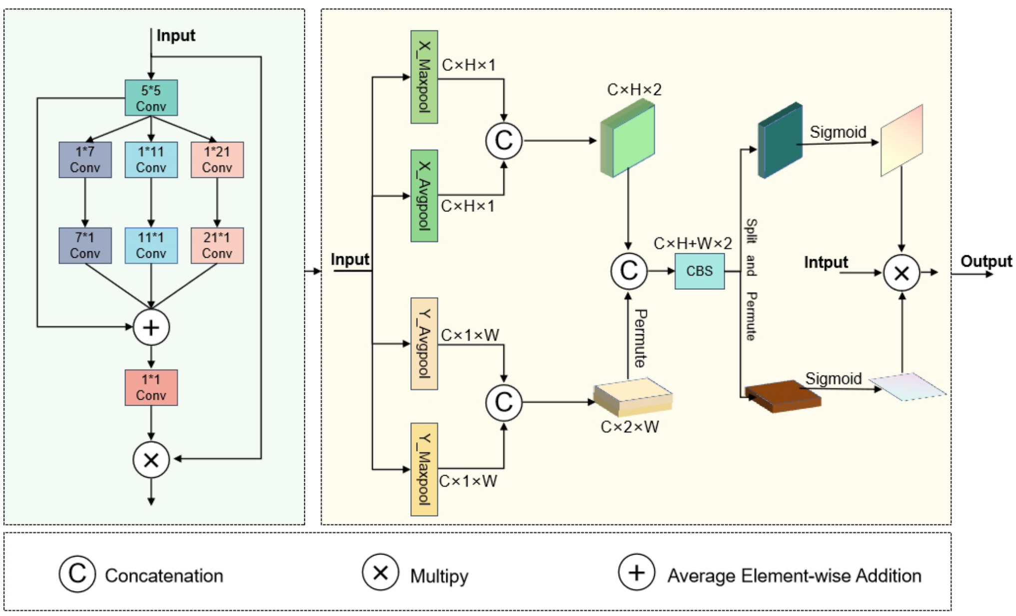

Figure 3. Multiscale adaptive feature processing module architecture.

The MAFP module processes the backbone feature map by first applying a 5×5 grouped convolution for initial feature extraction. Three parallel asymmetric convolutional branches then process the features: a short-range branch (sequential 1×7 and 7×1 convolutions), a medium-range branch (1×11 and 11×1), and a long-range branch (1×21 and 21×1). This design preserves spatial details while enhancing directional feature responses. All branch outputs are fused through element-wise summation to form multi-scale direction-sensitive features, followed by channel integration via a 1×1 convolution to produce the enhanced features, expressed as:

$$F_{\mathrm{branch}_k} = W_{k,1} * (W_{k,2} * (W_0 * X)) \quad (1)$$

$$F_{\mathrm{sum}} = \sum_{k=0}^{2} F_{\mathrm{branchk}} + W_0 * X \quad (2)$$

$$F_1 = (W_3 * F_{\mathrm{sum}}) \odot X \quad (3)$$

Here, W_0 and W_3 denote the weights of the 5×5 and 1×1 convolutions respectively, while $W_{k,1} \in R^{1\times S}$ and $W_{k,2} \in R^{1\times S}$ ($s \in \{7,11,21\}$) represent the convolutional kernels of each asymmetric branch. denotes the initially enhanced features. A dual-path spatial attention mechanism then suppresses background noise and amplifies crack responses. Specifically, adaptive average and max pooling along the height and width dimensions of generate directional attention weights. These weights are normalized via Sigmoid to [0,1], dynamically enhancing crack pixel activations while suppressing noise. Spatial calibration is achieved through element-wise multiplication with the original features, while a residual connection with preserves spatial details and mitigates background interference.

2.4. Multi-scale dynamic sampling bidirectional feature pyramid network

Road distress detection faces significant challenges due to the scale, morphological, and textural diversity of damage types, compounded by complex backgrounds and noise interference. While YOLOv8's PANet facilitates cross-level information flow, its feature fusion overly depends on preceding outputs and contains redundant nodes, limiting original feature utilization and increasing computational costs. To address these limitations, we propose the MD-BiFPN integrating Dynamic Sampling and MSGRConv.

As shown in **Figure 4**, MD-BiFPN introduces several key enhancements: a streamlined topology with skip connections between same-scale nodes enhances feature fusion while maintaining computational efficiency; adaptive weight learning optimizes multi-scale feature propagation; MSGRConv modules employ multi-scale grouped residual convolutions to capture diverse distress patterns while reducing parameters; and Dynamic Sampling enables adaptive feature alignment to resolve spatial mismatches. These innovations collectively improve detection accuracy for slender cracks and small defects [1].

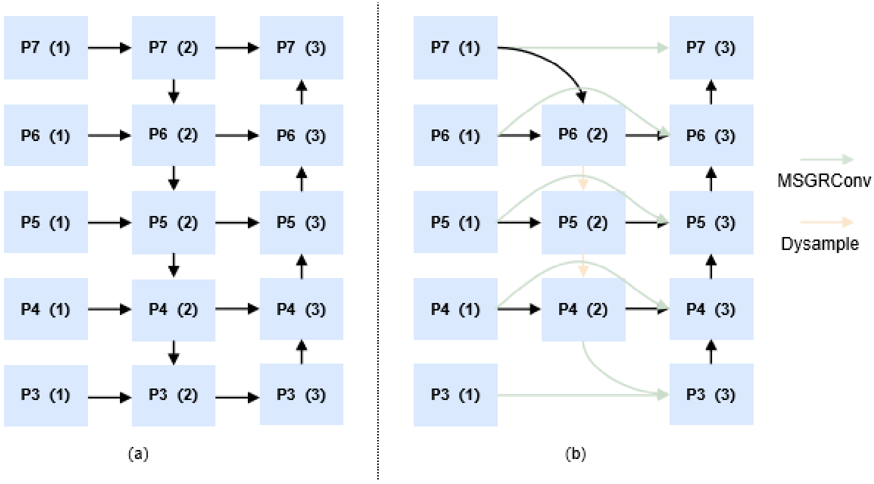

Figure 4. Comparative diagram of feature fusion network architecture (a) PANet; (b) MD-BiFPN.

2.5. C2f-FasterBlock

To enhance feature extraction efficiency and reduce network complexity, this paper improves the C2f module in YOLOv8. The original C2f module strengthens feature representation through multi-layer convolution and bottleneck stacking but suffers from high computational cost and parameter redundancy. We address this by integrating the FasterNet block into the C2f architecture, forming the novel C2f_Faster module [2]. This design preserves representational capacity while significantly lowering computational complexity and memory usage, enabling more efficient feature extraction. The structure is shown in **Figure 5**.

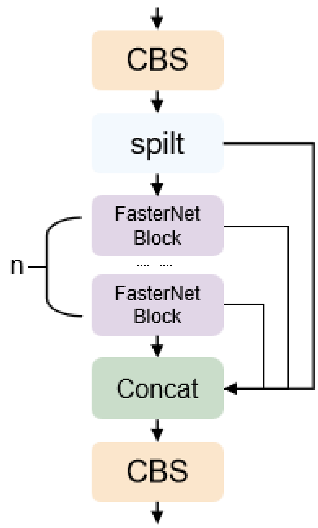

Figure 5. C2f_FasterBlock module architecture.

In deep neural networks, feature extraction channels often exhibit semantic or structural similarities, creating computational redundancy. To address this, Chen *et al.* proposed Partial Convolution (PConv), which applies standard convolution to only a subset of input channels while preserving the remainder. As shown in **Figure 6** (left), this approach maintains spatial feature extraction while substantially reducing computation and memory access.

The FasterNet block, illustrated in **Figure 6** (right), comprises PConv and Pointwise Convolution (PWConv). It first employs PConv for efficient local spatial feature extraction, followed by PWConv for cross-channel feature integration. Batch normalization and ReLU activation enhance nonlinear representation, with subsequent PWConv refining features. A residual connection maintains information flow and gradient stability, achieving computational efficiency while preserving representation capacity.

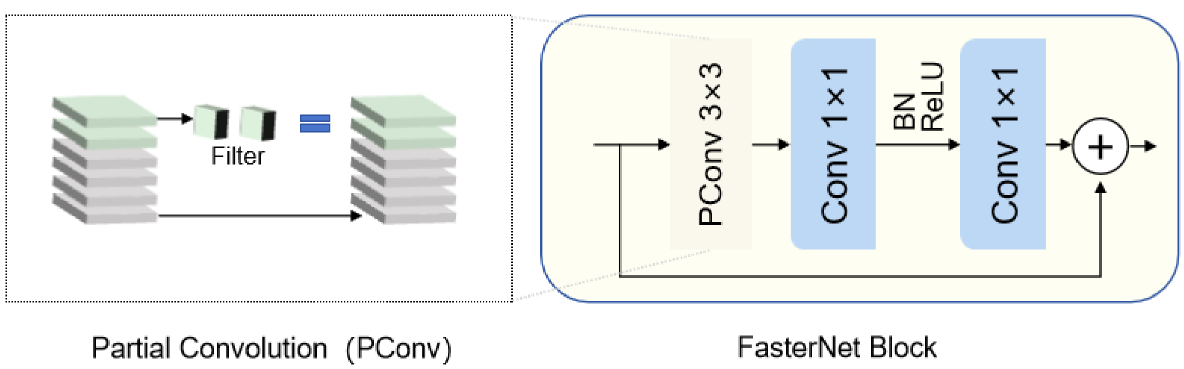

Figure 6. FasterNet block architecture.

3. Experiment

3.1. Details of experiment

This study employs the Global Road Damage Detection Challenge 2022 (GRDDC2022) dataset, containing road images from India, Japan, the United States, China, and the Czech Republic with high diversity and representativeness [3]. We utilized 4,398 Chinese road images, including 2,401 UAV-captured and 1,977 vehicle-mounted images, focusing on five distress types: longitudinal cracks (D00), transverse cracks (D10), alligator cracks (D20), potholes (D40), and repair patches. The dataset is split into training, validation, and test sets in an 8:1:1 ratio. Experimental configurations are detailed in **Table 1**.

Table 1. Experimental detail

Configuration	Version
Operating system	Windows 11
CPU	12th Intel(R) Core ™i7-12700
GPU	Nvidia RTX3060
Language	Python 3.8
Framework	Pytorch1.12.1+CUDA11.3.1
Optimizer	SGD
Epochs	200
Learning rate	0.01
Image size	640×640
Batch size	16

3.2. Evaluation of experiment

This study employs standard object detection metrics: Precision, Recall, and mean Average Precision (mAP). As the principal comprehensive metric, mAP calculates the mean Average Precision (AP) across all categories, where AP measures precision-recall balance. We primarily use mAP@0.5 (IoU threshold=0.5) to evaluate detection accuracy under different matching criteria, with calculation formulas as follows:

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$mAP = \frac{1}{m}\sum_{i=1}^{m} AP_i \tag{6}$$

$$IoU = \frac{Area(I)}{Area(U)} \tag{7}$$

In the evaluation metrics: TP (True Positives) indicates correctly identified positive samples; TN (True Negatives) represents correctly identified negative samples; FP (False Positives) denotes negative samples misclassified as positive; FN (False Negatives) refers to positive samples misclassified as negative. Here, m represents the total category count, where the Average Precision (AP) for the i-th category contributes to the overall mAP. The IoU threshold is set to 0.5, considering detections with IoU > 0.5 as correct. mAP@0.5 serves as the primary metric, while parameters, GFLOPs, and FPS act as auxiliary indicators, collectively assessing accuracy, complexity, and efficiency.

3.3. Ablation experiment

Ablation studies on YOLOv8n validate our improvements for road distress detection [4]. As **Table 2** shows, integrating MD-BiFPN improves mAP50 by 1.9% while reducing parameters and computation by 23% and 22% respectively, attributed to enhanced fusion of semantic and detail features. Adding MSGRConv further increases mAP50 by 1.0% while maintaining efficiency. Replacing SPPF with MAFP raises mAP50 to 88.2% (3.3% overall improvement) through better multi-scale perception. Finally, implementing C2f-FasterBlock achieves optimal performance (88.6% mAP50, 2.312M parameters) by optimizing feature extraction while preserving inference speed.

The ablation study confirms that MD-BiFPN, MAFP, MSGRConv, and C2f-FasterBlock collectively enhance road distress detection performance. Their incremental integration synergistically optimizes detection accuracy, parameter efficiency, and computational cost. Each module demonstrates complementary benefits in feature fusion, semantic perception, and lightweight design, providing a viable approach for efficient road distress detection models.

Table 2. The comparison of quantification for each model

MD-BiFPN	MAFP	MSGRConv	C2f-FasterBlock	mAP50(%)	MD-BiFPN	MAFP	MSGRConv
				84.9	79.7	3.006	8.1
√				86.8	81.3	2.509	6.3
	√			87.1	81.8	2.932	7.2
		√		87.4	82.1	2.643	6.8
			√	85.7	80.4	2.487	7.4
√		√		87.8	82.9	2.512	6.1
√	√	√		88.2	83.3	2.477	6.3
√	√	√	√	88.6	83.5	2.312	6.3

3.4. Quantitative evaluation

MADF-YOLOv8 is comprehensively evaluated against representative detectors on road distress detection, including single-stage models (YOLOv7n, YOLOv7-tiny, YOLOv9-t, YOLOv10n, YOLOv11n, Road-EfficientDet, BL-YOLOv8n) and the two-stage Faster-RCNN [5–11]. All experiments use identical settings to ensure comparability. As detailed in Table 3, our model achieves state-of-the-art accuracy with 88.6% mAP@50 and 83.5% mAP@50:95, improving the YOLOv8n baseline by 3.7 and 1.3 percentage points respectively. With only 2.312M parameters and 6.3 GFLOPs, it maintains lightweight characteristics comparable to YOLOv11n but significantly more efficient than YOLOv9-t (60.8M/266.1GFLOPs). MADF-YOLOv8 outperforms BL-YOLOv8n with merely 0.07M additional parameters, exceeds YOLOv7-tiny in accuracy with 52.3% lower computation, and surpasses YOLOv10n by 4.2% mAP@50 at similar complexity.

In model efficiency, MADF-YOLOv8 maintains excellent lightweight characteristics with 2.312M parameters and 6.3 GFLOPs, comparable to YOLOv11n yet significantly lower than complex models like YOLOv9-t (60.8M/266.1GFLOPs), demonstrating superior parameter utilization efficiency.

Regarding accuracy-efficiency trade-off, MADF-YOLOv8 outperforms BL-YOLOv8n with only a 0.07M parameter increase. Compared to YOLOv7-tiny, it achieves higher accuracy with 52.3% lower computational load; versus YOLOv10n, it improves mAP@50 by 4.2 percentage points under similar complexity, demonstrating superior overall performance.

Table 3. Performance comparison of different detection algorithms on road distress datasets

Methods	Parameters/M	FLOPS (G)	mAP50/%	mAP50-90 (%)
Faster-RCNN	60.1	108.6	77.3	73.4
YOLOv8n	3.0	8.1	84.9	82.2
YOLOv7-tiny	6.2	13.2	86.2	82.1
YOLOv9-t	60.8	266.1	85.9	81.3
YOLOv10n	2.7	8.2	84.4	82.5
YOLOv11n	2.58	6.3	84.2	81.2
Road-EfficientDet	6.58	6	84.7	80.6
BL-YOLOv8n	2.241	7.3	87.2	82.9
MADF-yolov8	2.312	6.3	88.6	83.5

4. Conclusion

To overcome road distress detection challenges, including the accuracy-efficiency trade-off and high small-target miss rates, we propose MADF-YOLOv8, an enhanced lightweight YOLOv8n-based model. Key innovations include: MSGRConv in the backbone for efficient multi-scale feature extraction; MAFP with grouped asymmetric convolutions and dual-path attention for detailed feature perception; MD-BiFPN incorporating dynamic sampling for optimized cross-level fusion; and C2f-Faster for neck lightweighting. On RDD2022, our model achieves 88.6% mAP@50 and 83.5% mAP@50:95 with only 2.312M parameters and 6.3 GFLOPs, outperforming mainstream detectors while maintaining superior accuracy-efficiency balance. While MADF-YOLOv8 performs well on standard datasets, its accuracy in challenging real-world scenarios (e.g., extreme weather, low-light conditions) needs improvement. Additionally, though currently focusing on distress localization and classification, it lacks geometric attribute quantification (e.g., crack width, pothole area), limiting precise maintenance applications. Future work will focus on developing generalized features with adversarial training and domain adaptation for enhanced robustness; incorporating instance segmentation or pixel-level analysis for quantitative assessment of distress morphology and severity.

Funding

This research was supported by the National Key Research and Development Program of China under Grant No. 2022YFA1602003, entitled "Intelligent Monitoring of Taishan Neutrino Detector".

Disclosure statement

The authors declare no conflict of interest.

References

[1] Liu W, Lu H, Fu H, et al., 2023, Learning to Upsample by Learning to Sample. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.

[2] Chen J, Kao S, He H, et al., 2023, Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

[3] Arya D, Maeda H, Ghosh S, et al., 2024, RDD2022: A Multi-National Image Dataset for Automatic Road Damage Detection. ArXiv. https://doi.org/10.48550/arXiv.2209.08538

[4] Glenn J, Alex S, Jirka B, 2023, Ultralytics Yolov8, 8.

[5] Wang C, Bochkovskiy A, Liao H, YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.

[6] Wang C, Yeh I, Mark L, 2024, Yolov9: Learning What You Want to Learn Using Programmable Gradient Information. Proceedings of the European Conference on Computer Vision, 2024.

[7] Wang A, Chen H, Liu L, et al., 2024, Yolov10: Real-Time End-to-End Object Detection, 37(10): 7984–8011.

[8] Khanam R, Hussain M, 2024, Yolov11: An Overview of the Key Architectural Enhancements.

[9] Naddaf S, Naddaf S, Kashai A, et al., 2020, An Efficient and Scalable Deep Learning Approach for Road Damage Detection. Proceedings of the 2020 IEEE International Conference on Big Data

[10] Wang X, Gao H, Jia Z, et al., 2023, BL-YOLOv8: An Improved Road Defect Detection Model based on YOLOv8, 23(20): 8361.

[11] Ren S, He K, Girshick R, et al., 2016, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 39(6): 37–49.