

Research

Road damage detection and classification using deep neural networks

Yiwen Jiang¹

Received: 22 April 2024 / Accepted: 26 July 2024

Published online: 01 August 2024

© The Author(s) 2024 **OPEN**

Abstract

In addressing the challenges of enhancing road damage detection efficiency and accuracy, this paper introduces an optimized YOLOv8 model suitable for embedded systems. The model significantly enhances precision, recall, and mean Average Precision (mAP), achieving 65.7% mAP on the RDD2022 dataset, thereby surpassing models such as Faster R-CNN and SSD. This advancement is attributed to the integration of a Deformable Attention Transformer, a GSConv-powered slim-neck module, and the MPDIoU loss function. These innovations not only contribute to the model's high performance but also set a new benchmark in road damage detection technology, thereby paving the way for future enhancements in the field.

1 Introduction

Road networks are vital to economic stability, supporting daily travel and commercial and industrial development. Road surface cracks significantly affect road and traffic safety. High-quality road maintenance is essential for ensuring safety, making prompt detection of road damage crucial to preventing further deterioration and ensuring traffic safety.

Current methods of road damage detection are marred into manual inspection, automated inspection, and image processing techniques [1]. Manual inspection, prevalent in developing countries, is marred by issues such as poor safety, low efficiency, high costs, and inconsistent judgment due to reliance on inspectors' experience. Technological advancements have led to increased use of automated road inspection, employing vehicles equipped with infrared or sensor equipment. However, these methods face challenges regarding recognition accuracy, speed, and high hardware costs in complex road environments [2].

Image processing technology, known for its high efficiency and low cost, has seen increasing accuracy with technological advancements, leading many researchers to employ these techniques for pavement damage detection. However, traditional image processing methods face challenges in handling complex road environments and improving model generalization [3, 4].

Deep learning-based image processing techniques have introduced new possibilities for road damage detection [5, 6]. Convolutional Neural Networks (CNNs) excel in extracting and learning features from images, effectively identifying various road damages such as cracks and potholes [7]. Deep learning models vary in architecture, from simpler networks like VGG [8] to more complex and efficient structures like YOLO [9] and SSD [10], each balancing accuracy and computational efficiency differently.

In this paper, we present a refined road damage detection algorithm that capitalizes on the enhanced capabilities of the YOLOv8 architecture [11]. This algorithm is specifically designed for embedded systems, offering a substantial

✉ Yiwen Jiang, jiangyiwen@ccit.edu.cn | ¹School of Intelligent Equipment, Changzhou College of Information Technology, Changzhou 213164, China.



improvement in the detection and analysis of road surface imperfections. The salient innovations of our approach are outlined as follows:

1. The model incorporates a Deformable Attention Transformer (DAT) [12] within the YOLOv8's C2f module, establishing a novel attention mechanism that amplifies the model's focus on pertinent image regions. This strategic integration is designed to streamline computational expenditure, thereby enhancing the model's interpretative and representational prowess with respect to image data.
2. Addressing the imperatives of embedded deployment, our approach introduces the GSConv, a lightweight convolutional technique, into the model's Neck architecture via a slim-neck module [13]. This strategic redesign significantly reduces the model's architectural complexity without compromising its detection capabilities, thus delivering improved computational efficiency.
3. We propose the adoption of the MPDIoU loss function [14] in place of the conventional VFL loss function [15]. By defining the loss with a four-point coordinate system, our method employs a more sophisticated metric for bounding box evaluation. This innovation not only accelerates the model's convergence but also improves its regression precision.

The structure of this paper is as follows: Sect. 2 examines existing research on road defect detection. Section 3 shifts focus to the YOLOv8 model, detailing the specifics of the enhanced framework proposed in this study. Section 4 discusses the experiments conducted and interprets the results obtained. Finally, Sect. 5 summarizes the conclusions drawn from this research.

2 Related work

The evolution of road damage detection methodologies has transitioned from basic manual inspections to advanced automated and image processing techniques [16]. Initially, manual inspections constituted the primary approach, relying on the subjective assessment of trained personnel. While straightforward, this method was fraught with issues including high variability, safety risks, and inefficiency, particularly over extensive road networks [17]. The advent of automated inspection methods marked a significant leap forward, utilizing sensor-equipped vehicles to systematically scan road surfaces. These methods enhanced consistency and safety but were impeded by high operational costs and the complexities of sensor data interpretation in diverse conditions.

The integration of image processing techniques represented a pivotal shift towards more scalable and economically viable solutions. Early methods utilized basic digital image analysis, including thresholding and edge detection, to identify road damages. However, these techniques faced challenges with accuracy in complex environmental conditions [18, 19]. Subsequent advancements introduced more refined image segmentation and feature extraction methods, improving the detection of various damage types. Despite their improvements, these techniques were still limited by their reliance on manually designed features and elevated false-positive rates in complex road scenarios.

Deep learning has emerged as a transformative force, mitigating many limitations of traditional image processing methods. Models such as VGG, known for their deep architectures, have significantly advanced the field with their robust feature extraction capabilities [20]. However, the computational intensity of such models restricts their applicability in real-time scenarios. The YOLO framework introduced a paradigm shift by executing detection tasks at remarkable speeds, making it suitable for real-time applications, although sometimes at the expense of precision in detecting small or subtle damages [21]. YOLO-LWNet provided a lightweight solution for road damage detection on mobile terminal devices, focusing on a streamlined architecture and efficient performance [22]. SSD adapted deep learning to the constraints of mobile and embedded devices, offering an efficient yet simpler solution [23]. Despite these advancements, the field continues to face challenges in achieving a balance between accuracy and computational efficiency, especially in diverse and variable environmental conditions. Models experience difficulty in generalizing across different road types and conditions, often requiring extensive fine-tuning and calibration [24]. The integration of these models into practical, embedded systems also necessitates consideration of hardware limitations and the imperative for optimized performance without sacrificing detection quality. In addition to the conventional approaches, emerging techniques under constrained conditions such as small samples and weak supervision are gaining traction. Generative adversarial networks combined with convolutional neural networks have shown promise in small-sample crack detection [25]. Additionally, image-to-image translation has been effective for night-time detection [26]. Techniques using class activation maps have also advanced

detection with minimal supervision, improving accuracy in complex scenarios [27, 28]. These innovations enhance the adaptability and effectiveness of road damage detection systems.

Our research contributes to addressing these challenges by introducing an enhanced YOLOv8 model tailored for embedded systems. We incorporate a Deformable Attention Transformer to improve detection accuracy while maintaining computational efficiency. The introduction of a GSConv-powered slim-neck module diminishes architectural complexity, facilitating deployment in real-time scenarios. Additionally, our novel MPDIoU loss function enhances bounding box precision, crucial for accurate damage localization. These innovations collectively advance the state-of-the-art in road damage detection, offering a more precise, efficient, and robust solution for maintaining road safety.

3 Proposed method

3.1 YOLOv8n network framework enhancement

YOLOv8 emerges as a state-of-the-art, single-stage object detection architecture, carefully crafted to enhance input image processing through a series of strategically defined segments. Initially, the input segment employs mosaic data augmentation, adaptive anchor optimization, and grayscale padding to prepare the image for subsequent feature extraction. At the core of YOLOv8, the backbone and neck segments utilize an array of Conv and C2f modules, extracting feature maps across multiple scales. The C2f module, a refined iteration of its C3 predecessor, employs the ELAN structure from YOLOv7 [29] to advance the Bottleneck module, thereby enriching gradient flow while maintaining model efficiency. Following the backbone, the SPPF module [30] synergizes feature maps through a variety of pooling kernels, setting the stage for the neck's feature fusion process. An integrated FPN and PAN framework within the neck segment orchestrates a seamless blend of high-level semantic and low-level localization features via up and down sampling techniques. This integration is crucial for the model's proficiency in detecting objects of disparate scales. In the detection head, YOLOv8 distinguishes between classification and bounding box regression, conducting loss computations and refining detection box accuracy. The Task Aligned Assigner is pivotal in this process, facilitating precise sample classification and enabling decoupled heads to concurrently predict classification scores and regression coordinates. These predictions are quantified within a two-dimensional matrix indicating object presence at the pixel level, and a four-dimensional matrix detailing the object's center deviations. In the culmination of its processing, YOLOv8 employs the task-aligned assigner to formulate a metric that synergizes classification accuracy with IoU values, thereby optimizing both classification and localization. This metric plays a critical role in eliminating inferior prediction boxes, thereby solidifying YOLOv8 as a highly accurate and expedient object detection system.

In this work, we enhance the YOLOv8 framework for road damage detection, integrating the Deformable Attention Transformer (DAT) to improve image focus and enhance computational efficiency. The architecture is optimized with GSConv and a slim-neck module for embedded deployment, decreasing complexity and increasing cost-effectiveness. We introduce the MPDIoU loss function, employing four-point coordinates for precise bounding box predictions, which expedites convergence and enhances regression accuracy. These advancements, illustrated in Fig. 1, present a scalable solution for efficient and accurate road damage detection.

3.2 Deformable attention mechanism

Deformable Attention, emerging within the Deformable DETR framework, revolutionizes the conventional Transformer attention by addressing its inefficiencies in uniformly attending to all spatial locations. Central to this approach is the focused attention on key sampling points in the input features. The Deformable Attention module operates by initially processing a query feature through a linear transformation to calculate sampling offsets. These offsets are then utilized to direct the attention of multiple heads, each creating a set of attention weights via a softmax layer. These weights are applied to specific values sampled from the input feature map, aggregating them to form a refined representation of the input. The aggregated information undergoes another linear transformation to produce the final output. This methodology enables dynamic and flexible attention allocation, focusing on the most informative parts of the input for enhanced feature processing. This mechanism is mathematically represented by:

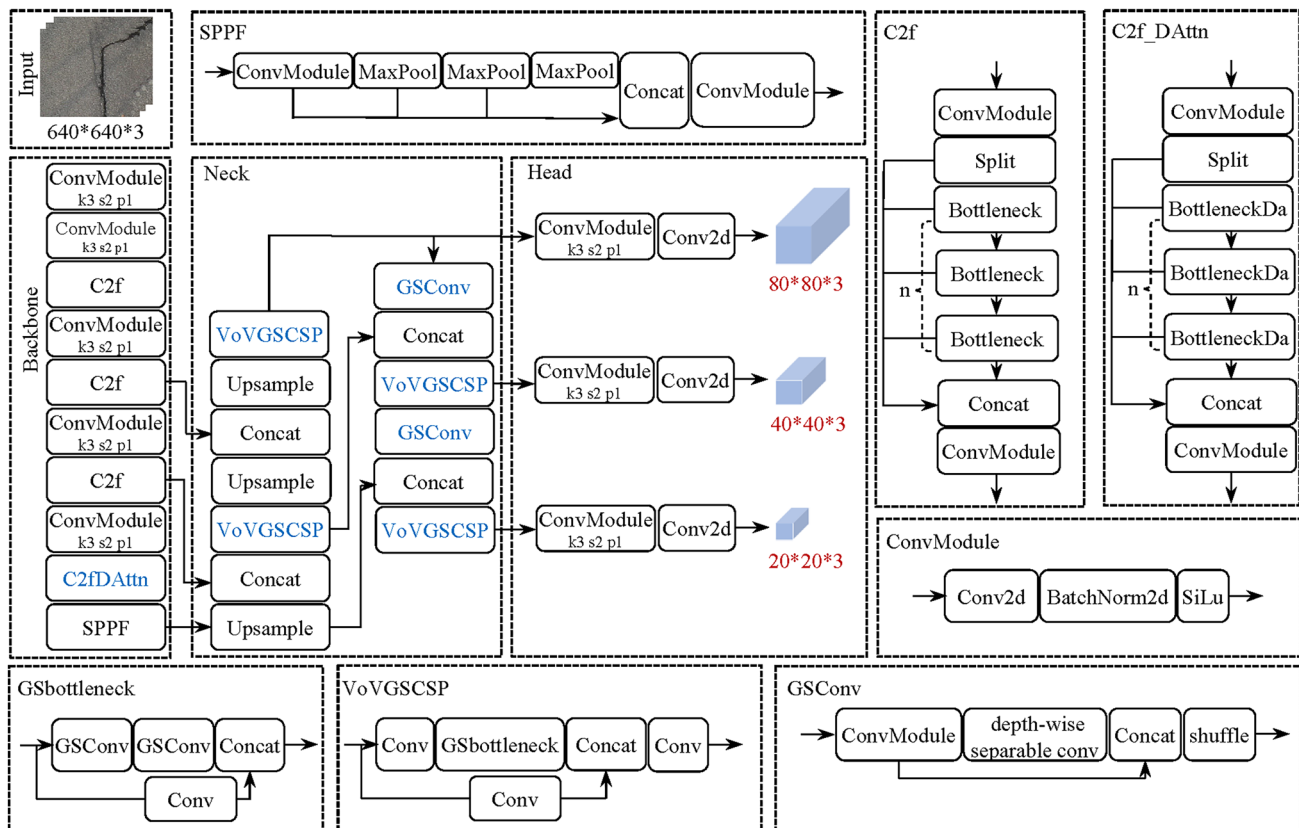


Fig. 1 Structure of improved YOLOv8n

$$A(x) = \sum_{i=1}^N W_i \cdot f(x + \Delta p_i) \quad (1)$$

where $A(x)$ denotes the output of the attention mechanism, W_i represents the learned weights, f is the feature map, and Δp_i are the learnable offsets, crucial for enabling the attention to adapt dynamically based on the input context. This selective and adaptable attention mechanism significantly enhances the flexibility and efficiency of the model, especially in processing large or complex image datasets. The choice of the Deformable Attention Transformer (DAT) was driven by its superior ability to handle irregular geometrical transformations within images—a common challenge in road damage detection. Unlike standard attention mechanisms that assume fixed geometric structures, DAT adapts more flexibly to the varied and unpredictable patterns of road damage.

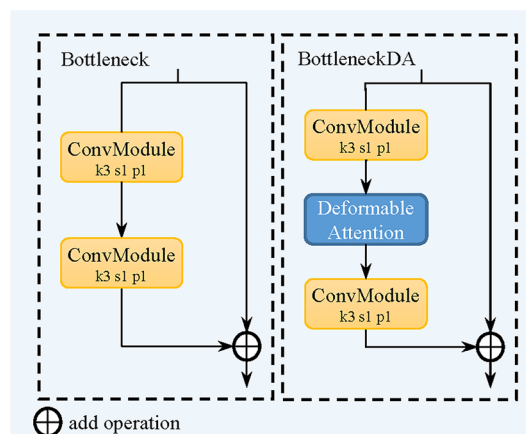
Incorporating this attention model into convolutional frameworks, such as the C2f module, results in the innovative C2f_DAttention. This module combines Deformable Attention with traditional convolutional layers, as illustrated in Fig. 2. C2f_DAttention enhances feature processing by integrating a DAttention layer, designed with specific attributes such as channel count, query size, and the number of attention heads. The Deformable Attention mechanism within C2f_DAttention adapts to the most informative segments of the input, offering a contrast to the uniform focus in classic bottleneck models.

The integration of Deformable Attention into C2f results in a more versatile and effective network. This adaptability is particularly beneficial in tasks that require acute spatial sensitivity, including object detection and image segmentation. C2f_DAttention illustrates the symbiosis of deformable attention and convolutional operations, resulting in a network skilled at navigating complex visual environments.

3.3 The role of slim-neck

The Neck component in the YOLOv8 architecture plays a critical role in bridging the backbone and the detection head. It is carefully designed to facilitate the fusion and refinement of multi-scale features, which are essential for enhancing

Fig. 2 Bottleneck module with attention mechanism



the model's detection capabilities. This part of the model integrates low-level detail-oriented features with high-level abstract features, ensuring a comprehensive understanding of the input data. Its sophisticated design enables efficient and effective feature propagation, which is crucial for maintaining accuracy while detecting diverse and complex objects in varied environments.

Slim-neck, an innovative architectural component, is specifically engineered to enhance the efficiency of deep learning models, particularly in the realm of object detection as exemplified by YOLOv8 [31]. This advanced component, serving as an augmentation to the standard neck structure, aims to minimize both the size and computational demands of the model. Notably, it achieves this without compromising its capability to extract and amalgamate features effectively. The design of Slim-neck is informed by methodologies that bolster the learning capacity of CNNs. As a result, Slim-neck significantly reduces both the computational complexity and inference time while maintaining accuracy. The integration of GSConv, or Grouped Separable Convolutions, within the Slim-neck framework further reinforces this balance. GSConv, by grouping and separating convolutional operations, substantially cuts down on computational overhead yet preserves crucial feature extraction abilities. The combination of GSConv with Slim-neck in models like Scaled-YOLOv4 [32] and YOLOv5 [33] demonstrates a marked reduction in computational complexity. Moreover, the incorporation of VoVGSCSP [34], an advanced form of the Cross Stage Partial Network (CSPNet), into the Slim-neck structure, exemplifies the blend of innovation and efficiency. This amalgamation significantly enhances the model's feature representation capabilities, optimizing both computational efficiency and accuracy. The Slim-neck architecture, thus, integrates GSConv and VoVGSCSP, creating a more efficient and effective system for handling complex feature extractions, which is crucial in applications like road damage detection.

Our proposed improvement involves integrating GSConv in place of the standard convolutions in the Neck of YOLOv8, and replacing the ConvModule with VoVGSCSP. This alteration aims to harness the efficiency of GSConv and the advanced feature fusion capabilities of VoVGSCSP, thereby enhancing the overall performance of the model. The expected outcome of this integration is twofold: firstly, a significant reduction in the computational load, making the model more agile and efficient for real-time applications. Secondly, an improvement in the accuracy and reliability of road damage detection, even under challenging conditions. This innovative approach is poised to set a new benchmark in the field of object detection, particularly in the context of identifying and assessing road damage using advanced deep learning techniques.

3.4 Integration of MPDIoU loss function

In the domain of road damage detection using YOLOv8, the precision of bounding box predictions is critical for accuracy. While YOLOv8 originally employs VFL for evaluating the accuracy of predictions, this study introduces a significant shift towards the Minimum Points Distance Intersection over Union (MPDIoU) loss function. This transition represents a significant enhancement, addressing some of the inherent limitations of the VFL loss function.

The MPDIoU loss function innovatively combines a minimum distance metric with the standard Intersection over Union calculation, refining the bounding box regression process. This function is particularly adept at incorporating overlapping and non-overlapping regions, centroid distances, and width-height deviations, thereby streamlining the computational process. The mathematical expression for MPDIoU is as follows:

$$MPDIoU = IoU - \frac{d_{12}}{2h^2 + w^2} - \frac{d_{22}}{2h^2 + w^2} \quad (2)$$

$$L_{MPDIoU} = 1 - MPDIoU \quad (3)$$

where $d_{12} = (x_1 - x_{10})^2 + (y_1 - y_{10})^2$ and $d_{22} = (x_2 - x_{20})^2 + (y_2 - y_{20})^2$ denote the squared distances between the corresponding corners of the predicted and actual bounding boxes, while w and h represent the width and length of the image, respectively.

Integrating the MPDIoU loss function into YOLOv8 confers multiple advantages compared to the VFL LOSS. Primarily, it enhances the precision of bounding box alignment through the direct minimization of corner distances between predicted and actual bounding boxes. This approach results in a more accurate and precise measurement, critical in complex detection scenarios. Additionally, the MPDIoU loss function accelerates the model's training process by providing a more discriminating gradient signal. This leads to faster convergence, enabling the model to rapidly adapt and improve its detection capabilities. Furthermore, the MPDIoU formula ensures a stable and bounded loss function, offering consistency in the loss calculation and preventing the issues of scale sensitivity and instability that can arise with VFL LOSS.

4 Experiments and analysis

4.1 Dataset and environment

In our study, we utilized the RDD2022 dataset [35], a comprehensive compilation of road damage images from a global perspective, encompassing six countries: Japan, India, the Czech Republic, Norway, the United States, and China. The RDD2022 dataset, forming an integral part of the Crowd sensing-based Road Damage Detection Challenge (CRDDC2022), includes 47,420 images, annotated with over 55,000 instances of road damage. A unique aspect of the RDD2022 dataset is the variation in damage categories across different countries. In our training and testing processes, we focused on the four types of road damage consistently present across all countries in the dataset. These types are designated as D00 (longitudinal cracks), D10 (transverse cracks), D20 (alligator cracks), and D40 (potholes). Each of these categories is represented with a substantial number of samples, offering a nuanced view of the diverse road damage scenarios. The RDD2022 dataset's diversity presents distinct challenges and opportunities for our improved YOLOv8 model. The variation within each damage category necessitates a model capable of distinguishing subtle differences in damage patterns, a challenge compounded by the similarities between different types of damages and the varying lighting and road surface conditions.

To demonstrate the breadth and complexity of road damages captured in the RDD2022 dataset, select samples of each damage type are displayed in Fig. 3, highlighting the dataset's utility in training models for diverse and complex road damage detection tasks.

Given YOLOv8's fixed input size, we then augmented the data using image rotation, translation, brightness adjustment, and scaling. The images are divided into training, validation, and test sets in an 7:1:2 ratio. The initial learning rate is 0.01, and we use the cosine annealing strategy to reduce the learning rate. The batch size is set to 32, and a total of 300 iterations are trained. The experiment training software environment is Linux4.15.0-142-generic Ubuntu 18.04, the YOLOv8.0.202 version. The hardware includes an Intel quad-core i7-6850 K, 3.60 GHz CPU, 32 GB memory and NVIDIA GeForce GTX1080Ti (11 GB).

4.2 Qualitative results analysis

In this study, we selected three prominent object detection methods, namely Faster R-CNN [36], SSD [37], and YOLOv5 [33], for a comparative analysis with the approach introduced in this paper. Faster R-CNN employs a sophisticated two-stage network, integrating a region proposal network for advanced object detection accuracy. SSD achieves efficient detection by predicting multiple bounding boxes and class scores in a single pass, while YOLOv5 is tailored for various optimizations including data processing, network training, and loss functions. The comparative analysis of recognition effects between our method and these established techniques is visually represented in Fig. 4.

As depicted in Fig. 4, the method presented in this paper excels in the precise detection of defects across a range of types and sizes, demonstrating superior overall performance. While Faster R-CNN performs reliably across diverse

Fig. 3 Examples of different types of defects



detection scenarios, it is hindered by slower execution speeds. In contrast, SSD and YOLOv5, although faster, show limitations, particularly in the accurate detection of smaller-sized defects, with instances of both false positives and missed detections. This contrast underscores the robustness of our method, especially in challenging detection conditions where SSD and YOLOv5 tend to falter, while also acknowledging the trade-off between accuracy and speed in Faster R-CNN.

4.3 Quantitative result analysis

The evaluations were carried out to verify the effectiveness of the proposed model. The main indicators are selected: Precision (P), recall (R), average precision (AP), and mean average precision (mAP). The corresponding equations are as follows:

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (4)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (5)$$

$$AP = \int_0^1 P(r) dr. \quad (6)$$

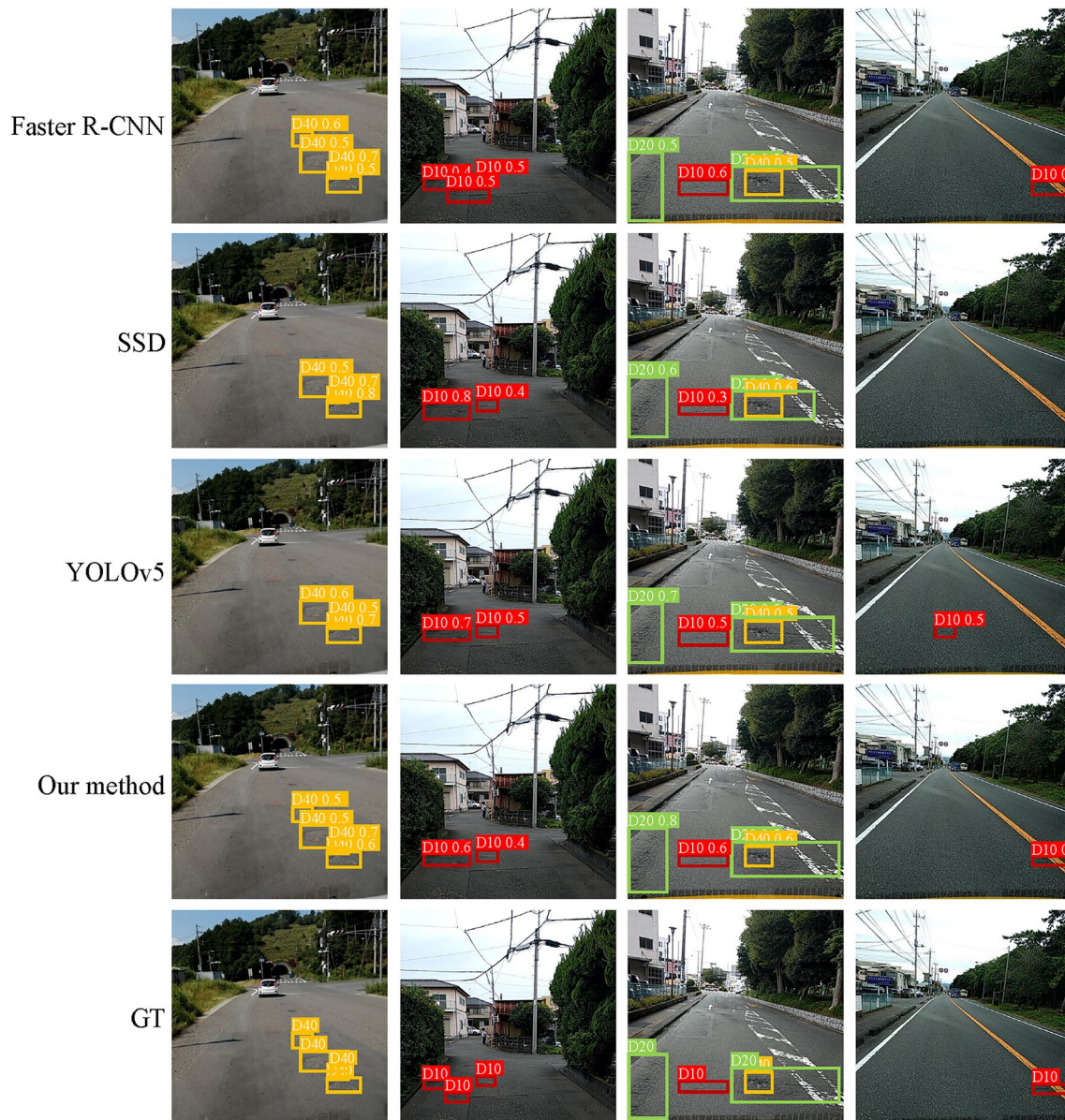


Fig. 4 Sample defect images

where T_p denotes the quantity of defects accurately identified by the detection model, F_p signifies the count of incorrect or unrecognized defects and F_N designates the number of falsely detected targets. To validate the effectiveness of the suggested approach, ablation studies were carried out, with the objective assessments and their associated outcomes presented in Table 1.

Table 1 Ablation experiment

Metric	Test set	mAP@.5 (%)	mAP@.5:.95 (%)	Parm (MB)
YOLOv8n	4171	61.8	32.3	12.9
+ C2fDAtn	4171	64.4	34.4	14.1
+ Slim Neck	4171	62.3	32.4	9.7
+ WPDiou	4171	62.7	33.9	13
Ours	4171	65.7	34.8	11.8

In our study, we assessed the performance of the YOLOv8n model on a test set comprising 4171 images, observing marked improvements through various enhancements. The original YOLOv8n model achieved a mean Average Precision (mAP) at 0.5 of 61.8% and mAP at 0.5:0.95 of 32.3%, with a model size of 12.9 MB. The introduction of the C2fDAttn attention mechanism module not only resulted in a noticeable increase in performance, with the mAP at 0.5 rising to 64.4% and mAP at 0.5:0.95 to 34.4%, but also maintained a competitive inference speed. Despite the slightly increased model size of 14.1 MB, the enhanced YOLOv8 model achieved an inference speed of 395 FPS, demonstrating that the addition of attention layers can significantly boost accuracy without critically compromising the model's real-time processing capabilities. Comparative tests with other attention mechanisms like CBAM [38] and Squeeze-and-Excitation (SE) [39] underscored the unique advantages of C2fDAttn. While CBAM and SE improved the base YOLOv8 model's performance, achieving mAP scores slightly lower than C2fDAttn, they did not match its adaptability in processing diverse and irregular road damage patterns. CBAM reached a mAP at 0.5 of 63.2% and SE achieved 62.7%, both underperforming in irregular damage scenarios compared to C2fDAttn. This adaptability of C2fDAttn to dynamically focus on complex damage patterns, crucial for accurate road damage detection, highlights why it is more suited for our model designed for varying real-world conditions. It provides a more robust performance, making it the optimal choice for enhancing detection accuracy in embedded systems focused on road safety.

Furthermore, the implementation of the slim neck architecture modification led to a moderate increase in the mAP at 0.5:0.95 to 32.4%, while reducing the model size to 9.7 MB. This indicates that the Slim Neck design contributes to model efficiency without significantly compromising the detection accuracy. Replacing the loss function with WPDiou also showed improvement, with the mAP at 0.5:0.95 increasing to 33.9% and the model size being 13 MB. This adjustment underscores the effectiveness in enhancing the model's predictive accuracy through a more sophisticated metric for bounding box evaluation, which is crucial for precise localization. Our model achieved an mAP at 0.5 of 65.7% and mAP at 0.5:0.95 of 34.8%, with a model size of 11.8 MB. This indicates that the combined modifications not only improve the model's accuracy significantly but also maintain a balance in model size, making it an efficient and effective solution for detecting and classifying images in the dataset.

Figure 5 presents the YOLOv8 model's performance across the training and validation datasets, with initial epochs showing a sharp decrease in loss and corresponding increases in precision, recall, and mAP values. This indicates effective learning without overfitting, as evidenced by consistent metrics up to the 150th batch. Notably, after the 150th epoch, an uptick in the validation box loss suggests the onset of overfitting, where the model starts to learn noise and specificities of the training data. To counteract this, we implemented an early stopping mechanism, ceasing training after 50 epochs without observable improvement. This measure helps to preserve the model's generalizability and prevents the degradation of validation performance. The implemented early stopping criterion ensures that the model maintains a

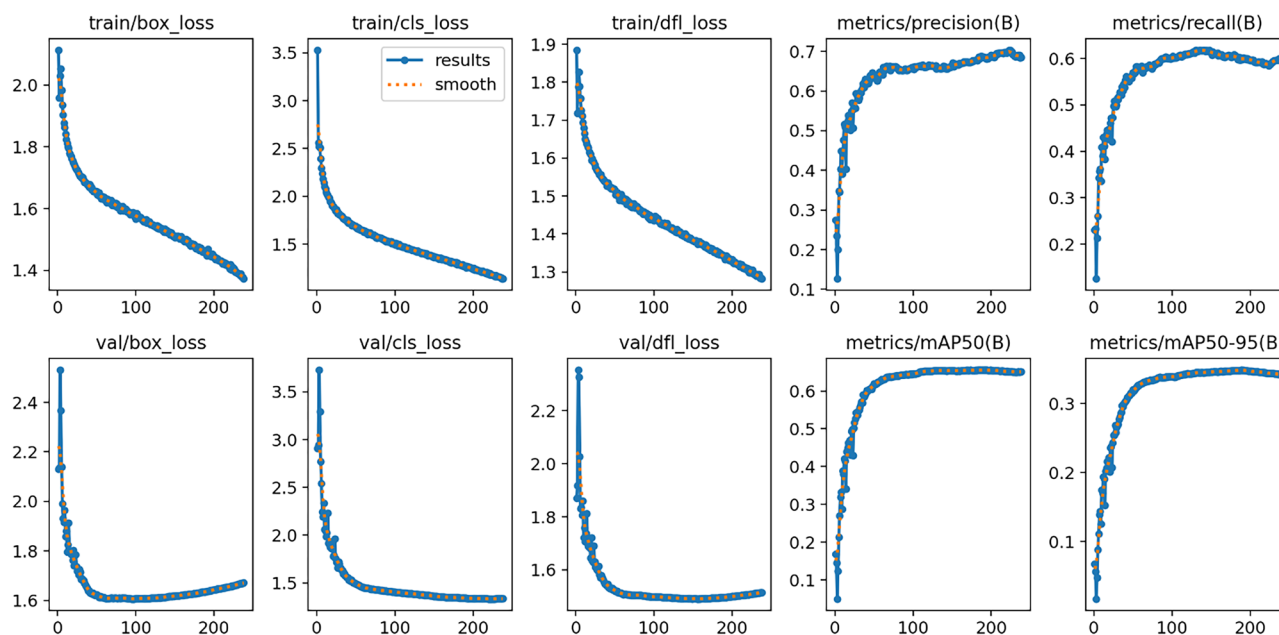


Fig. 5 Training and validation curves of the proposed method

balance between learning from the training data and retaining predictive accuracy for unseen data, thus enhancing the practical utility of the model in real-world applications.

Figure 6 shows the Precision, Recall, F1, and P-R curves of the proposed model. It can be seen that the improved model achieved good recognition results for all types of defects. The P-R curves of the corresponding categories for the four defect types in the dataset are depicted. It can be seen from the figure that the average AP value of the method reaches 65.7%, which shows excellent detection performance.

As demonstrated in Table 2, our comparison across a suite of detection algorithms on a standardized dataset highlights the superior performance of our approach. Achieving a precision of 68.8%, recall of 59.9%, and an mAP of 65.7%, our method

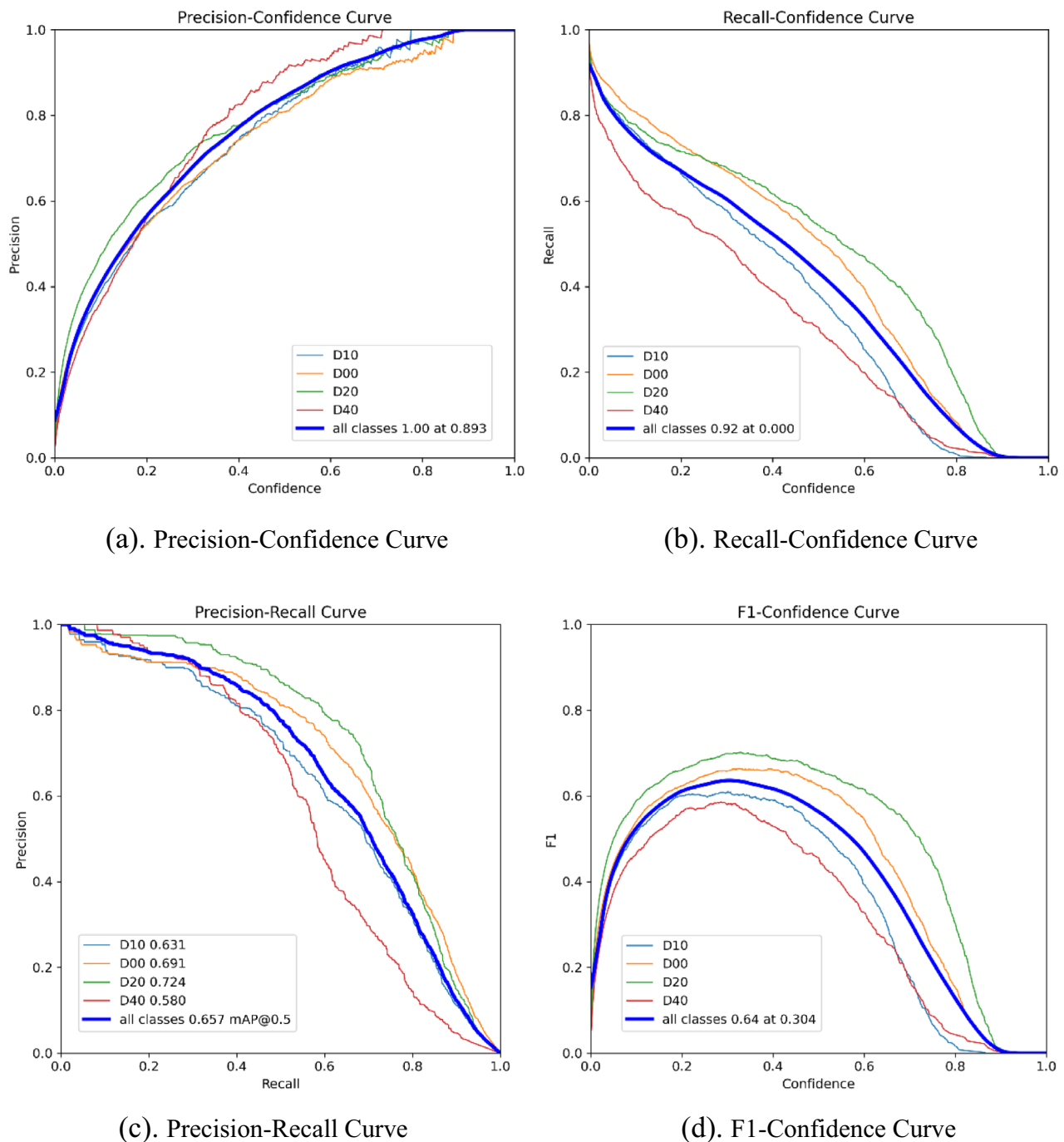


Fig. 6 Testing results of proposed method

Table 2 Comparison of efficiency and size indicators

Method	Precision (%)	Recall (%)	mAP@.5 (%)	FPS (f/s)	Param (MB)
Faster R-CNN	64.8	53.4	60.7	18	160
SSD	56.7	51.3	53.9	84	79
YOLOv5s	59.4	54.1	56.7	158	14.7
YOLOv8n	65.4	56.9	61.8	420	12.9
Our method	68.8	59.9	65.7	369	11.8

outperforms traditional models such as Faster R-CNN, SSD, and even the YOLOv8n, particularly notable for its operation at a high frame rate of 369 FPS with a compact model size of 11.8 MB. This efficiency is attributed to strategic enhancements within the YOLOv8 architecture. These strategic enhancements within the YOLOv8 architecture, particularly the integration of a Deformable Attention Transformer (DAT), GSConv in a slim-neck module, and the MPDIoU loss function, collectively refine the focus on critical image areas, streamline the architecture for embedded system deployment, and enhance bounding box precision. This results in rapid model convergence and heightened accuracy, setting our method apart as a robust solution for road damage detection, balancing performance with computational efficiency for embedded applications.

5 Conclusion and future work

This study culminates in a robust YOLOv8-based road damage detection model tailored for embedded systems, demonstrating a notable advancement in performance metrics. Our approach, distinguished by the integration of novel DAT integration, GSConv application, and MPDIoU loss function, consistently outperforms conventional models in precision, recall, and mean Average Precision (mAP), while maintaining an efficient frame rate. These contributions, particularly in enhancing targeted focus and computational efficiency, represent a significant advancement in automated road assessment technologies. Our YOLOv8 model, optimized for efficiency and accuracy, holds promise for deployment on embedded systems like NVIDIA Jetson and Raspberry Pi. These platforms, which often face resource constraints, could benefit from our model's lightweight architecture and the efficiency of the Deformable Attention Transformer. Preliminary theoretical assessments suggest that with parameter optimization, our model could deliver real-time road damage detection on these devices. Future work will focus on empirical testing to validate these capabilities and refine the model for practical use in mobile or remote sensing applications.

Author contributions Jiang yiwen independently completed all research work in this article.

Funding This work was supported by Natural Science Foundation of Changzhou College of Information Technology (CXZK202206Q) and the Key Laboratory of Industrial Internet of Things Foundation of Changzhou College of Information Technology (KYPT201803Z).

Data availability All data included in this study are available upon request by contact with the corresponding author.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bhat S, Naik S, Gaonkar M, Sawant P, Aswale S, Shetgaonkar P. A survey on road crack detection techniques. In: International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE); 2020. p. 1–6. <https://doi.org/10.1109/ic-ETITE47903.2020.67>.

2. Oliveira H, Correia PL. Automatic road crack detection and characterization. *IEEE Trans Intell Transp Syst.* 2012;14(1):155–68. <https://doi.org/10.1109/TITS.2012.2208630>.
3. Chen C, Seo H, Jun CH, Zhao Y. Pavement crack detection and classification based on fusion feature of LBP and PCA with SVM. *Int J Pavement Eng.* 2022;23(9):3274–83. <https://doi.org/10.1080/10298436.2021.1888092>.
4. Safaei N, Smadi O, Safaei B, Masoud A. Efficient road crack detection based on an adaptive pixel-level segmentation algorithm. *Transp Res Rec.* 2021;2675(9):370–81. <https://doi.org/10.1177/03611981211002203>.
5. Hou Y, Liu S, Cao D, Peng B, Liu Z, Sun W, et al. A deep learning method for pavement crack identification based on limited field images. *IEEE Trans Intell Transp Syst.* 2022;23(11):22156–65. <https://doi.org/10.1109/TITS.2022.3160524>.
6. Chen D-R, Chiu W-M. Deep-learning-based road crack detection frameworks for dashcam-captured images under different illumination conditions. *Soft Comput.* 2023. <https://doi.org/10.1007/s00500-023-08738-0>.
7. Cao M-T, Tran Q-V, Nguyen N-M, Chang K-T. Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. *Adv Eng Inf.* 2020;46:101182. <https://doi.org/10.1016/j.aei.2020.101182>.
8. Que Y, Dai Y, Ji X, Leung AK, Chen Z, Tang Y, et al. Automatic classification of asphalt pavement cracks using a novel integrated generative adversarial networks and improved VGG model. *Eng Struct.* 2023;277:115406. <https://doi.org/10.1016/j.engstruct.2022.115406>.
9. Du Y, Pan N, Xu Z, Deng F, Shen Y, Kang H. Pavement distress detection and classification based on YOLO network. *Int J Pavement Eng.* 2020;22(13):1659–72. <https://doi.org/10.1080/10298436.1714047>.
10. Andika F, Bandung Y. Road damage classification using SSD mobilenet with image enhancement. In: *International Conference on Computer Science, Information Technology and Engineering (ICCoSITE).* 2023. p. 540–5. <https://doi.org/10.1109/ICCoSITE57641.2023.10127763>.
11. Xiao B, Nguyen M, Yan WQ. Fruit ripeness identification using YOLOv8 model. *Multimed Tools Appl.* 2024;83:28039–56. <https://doi.org/10.1007/s11042-023-16570-9>.
12. Xia Z, Pan X, Song S, Li LE, Huang G. Vision transformer with deformable attention. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition;* 2022. p. 4794–803.
13. Li H, Li J, Wei H, Liu Z, Zhan Z, Ren Q. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles. *arXiv:2206.02424*. <https://doi.org/10.48550/arXiv.2206.02424>.
14. Siliang M, Yong X. MPDIoU: a loss for efficient and accurate bounding box regression. *arXiv:2307.07662*. <https://doi.org/10.48550/arXiv.2307.07662>.
15. Li Z, Zhang Y, Wu H, Suzuki S, Namiki A, Wang WJRS. Design and application of a UAV autonomous inspection system for high-voltage power transmission lines. *Remote Sens.* 2023;15(3):865. <https://doi.org/10.3390/rs15030865>.
16. Arya D, Maeda H, Ghosh SK, Toshniwal D, Omata H, Kashiyaama T, et al. Global road damage detection: state-of-the-art solutions. In: *IEEE International Conference on Big Data (Big Data).* 2020. p. 5533–9. <https://doi.org/10.1109/BigData50022.2020.9377790>.
17. Pham V, Pham C, Dang T. Road damage detection and classification with detectron2 and faster r-cnn. In: *IEEE International Conference on Big Data (Big Data).* 2020. p. 5592–601. <https://doi.org/10.1109/BigData50022.2020.9378027>.
18. Munawar HS, Hammad AW, Haddad A, Soares CAP, Waller STJI. Image-based crack detection methods: a review. *Infrastructures.* 2021;6(8):115. <https://doi.org/10.3390/infrastructures6080115>.
19. Cao W, Liu Q, He ZJIA. Review of pavement defect detection methods. *IEEE Access.* 2020;8:14531–44. <https://doi.org/10.1109/ACCESS.2020.2966881>.
20. Czimmermann T, Ciuti G, Milazzo M, Chiurazzi M, Roccella S, Oddo CM, Dario P. Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors.* 2020;20:1459. <https://doi.org/10.3390/s20051459>.
21. Wan F, Sun C, He H, Lei G, Xu L, Xiao T. YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s. *EURASIP J Adv Signal Process.* 2022. <https://doi.org/10.1186/s13634-022-00931-x>.
22. Wu C, Ye M, Zhang J, Ma YJS. YOLO-LWNet: a lightweight road damage object detection network for mobile terminal devices. *Sensors.* 2023;23(6):3268. <https://doi.org/10.3390/s23063268>.
23. Maeda H, Sekimoto Y, Seto T, Kashiyaama T, Omata H, Engineering I. Road damage detection and classification using deep neural networks with smartphone images. *Computer Aided Civ Infrastruct Eng.* 2018;33(12):1127–41. <https://doi.org/10.1111/mice.12387>.
24. Maeda H, Kashiyaama T, Sekimoto Y, Seto T, Omata H, Engineering I. Generative adversarial network for road damage detection. *Computer Aided Civ Infrastruct Eng.* 2021;36(1):47–60. <https://doi.org/10.1111/mice.12561>.
25. Xu B, Liu C. Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples. *Measurement.* 2022;196:111219. <https://doi.org/10.1016/j.measurement.2022.111219>.
26. Liu C, Boqiang Xu. A night pavement crack detection method based on image-to-image translation. *Computer Aided Civ Infrastruct Eng.* 2022;37(13):1737–53. <https://doi.org/10.1111/mice.12849>.
27. Xu B, Liu C. Detection algorithm of structural surface cracks based on class activation map. *IABSE Congress.* 2022. p. 1216–23.
28. Liu C, Xu B. Weakly-supervised structural surface crack detection algorithm based on class activation map and superpixel segmentation. *Adv Bridge Eng.* 2023;4(1):27–24. <https://doi.org/10.1186/s43251-023-00106-0>.
29. Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2023. p. 7464–75.
30. Xue Z, Lin H, Wang FJF. A small target forest fire detection model based on YOLOv5 improvement. *Forests.* 2022;13(8):1332. <https://doi.org/10.3390/f13081332>.
31. Ye L, Chen S. GBForDet: a lightweight object detector for forklift safety driving. *IEEE Access.* 2023;11:86509–21. <https://doi.org/10.1109/ACCESS.2023.3302909>.
32. Xu H, Li H, Xu Q, Zhang Z, Wang P, Li D, et al. Automatic detection of pulmonary embolism in computed tomography pulmonary angiography using Scaled-YOLOv4. *Med Phys.* 2023. <https://doi.org/10.1002/mp.16218>.
33. Guo G, Zhang Z. Road damage detection algorithm for improved YOLOv5. *Sci Rep.* 2022. <https://doi.org/10.1038/s41598-022-19674-8>.
34. Yang S, Xing Z, Wang H, Dong X, Gao X, Liu Z, et al. Maize-YOLO: a new high-precision and real-time method for maize pest detection. *Insects.* 2023;14(3):278. <https://doi.org/10.3390/insects14030278>.

35. Arya D, Maeda H, Ghosh SK, Toshniwal D, Sekimoto Y. Rdd 2022: a multi-national image dataset for automatic road damage detection. *Geosci Data J.* 2024. <https://doi.org/10.1002/gdj3.260>.
36. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst.* 2015;28:91–9.
37. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. Ssd: single shot multibox detector. In: *Computer Vision–ECCV 14th European Conference.* 2016. https://doi.org/10.1007/978-3-319-46448-0_2.
38. Woo S, Park J, Lee J-Y, Kweon IS. Cbam: convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV).* 2018. p. 3–19.
39. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2018. p. 7132–41.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.