

ORIGINAL ARTICLE

Open Access



Efficient and accurate road crack detection technology based on YOLOv8-ES

Kaili Zeng¹, Rui Fan² and Xiaoyu Tang^{1*}

Abstract

Road damage detection is an important aspect of road maintenance. Traditional manual inspections are laborious and imprecise. With the rise of deep learning technology, pavement detection methods employing deep neural networks give an efficient and accurate solution. However, due to background diversity, limited resolution, and fracture similarity, it is tough to detect road cracks with high accuracy. In this study, we offer a unique, efficient and accurate road crack damage detection, namely YOLOv8-ES. We present a novel dynamic convolutional layer (EDCM) that successfully increases the feature extraction capabilities for small fractures. At the same time, we also present a new attention mechanism (SGAM). It can effectively retain crucial information and increase the network feature extraction capacity. The Wise-IoU technique contains a dynamic, non-monotonic focusing mechanism designed to return to the goal-bounding box more precisely, especially for low-quality samples. We validate our method on both RDD2022 and VOC2007 datasets. The experimental results suggest that YOLOv8-ES performs well. This unique approach provides great support for the development of intelligent road maintenance systems and is projected to achieve further advances in future applications.

Keywords: Road crack detection, Object detection, Attention mechanism, Dynamic convolutional layer

1 Introduction

Autonomous driving science and uncrewed aerial vehicle (UAV) technology are experiencing significant growth [1–3]. While the primarily former depends on terrestrial transportation, uncrewed vehicles must regularly navigate diverse and challenging terrains, including rocky and arduous routes. Anticipating probable ground faults will significantly enhance the vehicle's ability to make precise driving judgments [4], making detecting such defects extremely vital. Moreover, unmanned vehicles function as transport platforms for road monitoring [5], and their operation effectively gathers data on road conditions. Simultaneously, UAVs, with their distinctive airborne vantage and adaptability, can offer significant data assistance for terrestrial inspections. In summation, the signif-

icance of ground defect inspection is apparent, serving as the foundation for the safe and efficient operation of uncrewed technology and a crucial element in advancing intelligent transportation systems to a higher echelon.

The primary considerations in road damage identification are efficiency and accuracy, whereas conventional approaches depend on labor-intensive and time-consuming human inspections [6]. The manual inspection involves capturing photographs of identified faults stored on a hard disk. An operator subsequently tracks, marks, and analyzes the cracks manually. This method will negatively impact the pavement damage assessment procedure due to inspector inexperience and visual inaccuracies [7]. To mitigate these restrictions, researchers have commenced trials involving road inspection vehicles [4, 5] and cell phones [8]. Nonetheless, they are costly and challenging to market. Currently, UAVs, characterized by their compact size, affordability, versatility, high mobility, and capability for multi-channel parallel inspections, are garnering

*Correspondence: tangxy@scnu.edu.cn

¹School of Data Science and Engineering, Xingzhi College, South China Normal University, Guangzhou, 516600, China

Full list of author information is available at the end of the article

heightened interest in the structural health assessment of civil engineering infrastructures [9, 10].

While these conventional approaches can be efficient, they necessitate intricate procedures and are vulnerable to variations in light and shadow. In recent years, deep learning-based crack detection has garnered heightened interest in the identification of road damage. Deep learning networks offer enhanced speed and precision in object detection tasks, exhibiting significant robustness and generalization abilities. By circumventing manual feature extraction processes, deep learning can reduce the likelihood of misclassification or the omission of essential target characteristics during feature pre-sampling. This research presents a novel target detection network. It is constructed and verified using the RDD2022 and VOC2007 datasets to enhance the efficiency and effectiveness of road defect detection. The contributions to this work are enumerated below:

1. In this research, we propose a Selective Global Attention Mechanism (SGAM). It can efficiently store information and magnify the interaction features of the global dimension, hence boosting the accuracy of the network model.
2. A new Enhanced Dynamic Convolution Module (EDCM) is proposed to enhance the performance and usefulness of the road damage detection system.
3. The Wise-IoU technique introduces a dynamic non-monotonic focusing mechanism that provides more accurate regression to the object bounding box, especially for low-quality samples. This change in the loss function considerably enhances the efficiency of the object detection algorithm, making it more dependable in real-world road traffic scenarios.

2 Related work

Road damage has been thoroughly examined for segmentation [11–15], detection [16–23]; this article focuses on object detection. Object identification methodologies are primarily classified into single-stage and two-stage algorithms. Representative instances of two-stage algorithms encompass R-CNN [23] and Faster R-CNN [24]. Lin et al. [16] proposed importance weighting based on Faster R-CNN to assess intermediate image-level alignment across sample domains and introduced aggregated ROI-wise features with multiscale contextual information to restore crack details for progressive domain alignment at the instance level. Nonetheless, both techniques were constrained by the intricacy of the technological model and the substantial resource consumption prevalent at that time. To tackle this issue, researchers examined single-stage object detection algorithms, which exhibit superior detection speed relative to two-stage algorithms, wherein the initial detector employs a unified neural network (e.g., SSD [25], EfficientDet [26], and the YOLO family [17, 19–21]) to correlate image pixels with predictions

directly. Naddaf-Sh et al. [27] used different scales of EfficientDet and many data augmentation policies for pavement crack detection. The YOLO series has achieved numerous successful applications in traffic-related domains due to its limited training memory and optimized accuracy and speed. Li et al. [19] introduced an enhanced YOLOX-RDD model derived from YOLOX. It adaptively modifies the receptive field based on object size. Additionally, by incorporating the Feature Enhancement Attention (FEA) module, fusing dark2 with the three output features of the neck map, and implementing two-level adaptive spatial feature fusion (ASFF), the model significantly enhances the detection capabilities for multiscale targets. However, there remains potential for further improvement in detection accuracy. Wang et al. [18] integrated SE and CA modules into YOLOv5 for comprehensive learning. Xiong et al. [17] integrated GAM and Wise-IoU on top of YOLOv8 to significantly increase the detection accuracy of the produced models. Ye et al. [28] incorporated the self-attention mechanism module from the Swin Transformer, designated as a self-study module (Myswin), and additionally integrates a self-study module (FEEM) built upon the foundation of YOLOv7.

Notwithstanding the considerable advancements in target recognition methodologies for road damage identification, persistent issues remain: detecting small and obscured ambiguous targets continues to pose substantial challenges under intricate and dynamic road conditions. The efficacy of current models in detecting these scenarios must be enhanced. Numerous studies have concentrated on enhancing the efficacy of attention mechanisms in picture classification tasks. Squeeze-and-Excitation Networks (SENet) [29] pioneered the application of channel attention and channel feature fusion to mitigate the influence of insignificant channels. Nevertheless, it was less effective at suppressing insignificant pixels. Subsequent attention mechanisms take into account both spatial and channel dimensions. The Convolutional Block Attention Module (CBAM) [30] implements channel and spatial attention operations sequentially, whereas the Bottleneck Attention Module (BAM) [31] executes them in parallel. Nevertheless, both overlook channel-space interactions, resulting in the loss of cross-dimensional information. Recognizing the significance of cross-dimensional interactions, researchers have conducted studies to tackle this issue, and Zhang et al. [32] introduced a novel Efficient Pyramid Squeezing Attention (EPSA) block. It adeptly extracts multiscale spatial information at a more granular level and enhances remote channel dependencies. Nonetheless, attentional activities are executed simultaneously on two dimensions rather than all three (channel, spatial width, and spatial height). To enhance cross-dimensional interactions, Liu et al. [33] suggested an attentional technique that identifies significant aspects across all three dimensions. Li

et al. [34] introduced a multidimensional attention mechanism and a parallel approach to acquire complementary attention of the convolutional kernel across all four dimensions of the kernel space in any convolutional layer. Likewise, Qi et al. [35] introduced a multi-perspective feature fusion technique that enhances attention to significant features from many viewpoints. However, existing convolutional layers still have limitations in high-dimensional data processing.

The loss function of Boundary Box Regression (BBR) is essential for target detection. A precise definition will yield substantial performance enhancement for the model. YOLOv1 [36] formulates a loss function that incorporates the BBR loss, classification loss, and objectivity loss. Nevertheless, this type of loss function fails to mitigate the impact of bounding box size and hence offers minimal localization performance for the model. To resolve the concerns above, the researchers conducted the following study: Intersection over Union (IoU) [37]. It is employed to quantify the extent of overlap between the anchor box and the target box in a target detection job. It effectively conceals the impact of bounding box size proportionally, enabling the model to adeptly balance the learning of large and small items when L_{IoU} is employed as a BBR loss. Nonetheless, L_{IoU} possesses a critical deficiency; in the absence of overlap between bounding boxes, the gradient of L_{IoU} during backpropagation diminishes to zero. Consequently, the width of the overlapping region remains unaltered during the training period. Current research examines several geometric parameters associated with bounding boxes. Ma et al. [38] encompassed all pertinent elements addressed in the current loss functions, including overlapping or non-overlapping regions, centroid distances, and variations in width and height. Zhang et al. [39] originally introduced the joint intersection (EIoU) loss, which directly quantifies the discrepancies in three geometric parameters in BBR: overlapping regions, centroids, and edge lengths. Subsequently, Focal-EIoU v1 utilizing non-monotonic FM was introduced. However, Focal-EIoU v1 fails to acknowledge that the quality evaluation of the anchor frames is manifested in the intercomparison. It fails to exploit the capabilities of non-monotonic FM fully. Moreover, most previous research presumes that the training data comprises high-quality examples, neglecting the impediment that subpar cases present to the learning efficacy of the target detection model, hence yielding constrained performance improvements. Tong et al. [40] introduced a dynamic non-monotonic focus frame-based loss function. It allocates a reduced gradient increment to low-quality anchor boxes, thereby successfully mitigating the detrimental impact of low-quality instances on the BBR.

This work addresses the issues above by implementing EDCM and SGAM and enhancements to the loss function,

thereby considerably augmenting the model's detection accuracy for road damage.

3 Method

3.1 Overview of our network

This study introduces a unique damage detection system, termed YOLOv8-ES, to tackle road damage issues in adverse weather conditions. This approach integrates three essential elements into the YOLOv8 framework, and the comprehensive network architecture presented in this research is illustrated in Fig. 1.

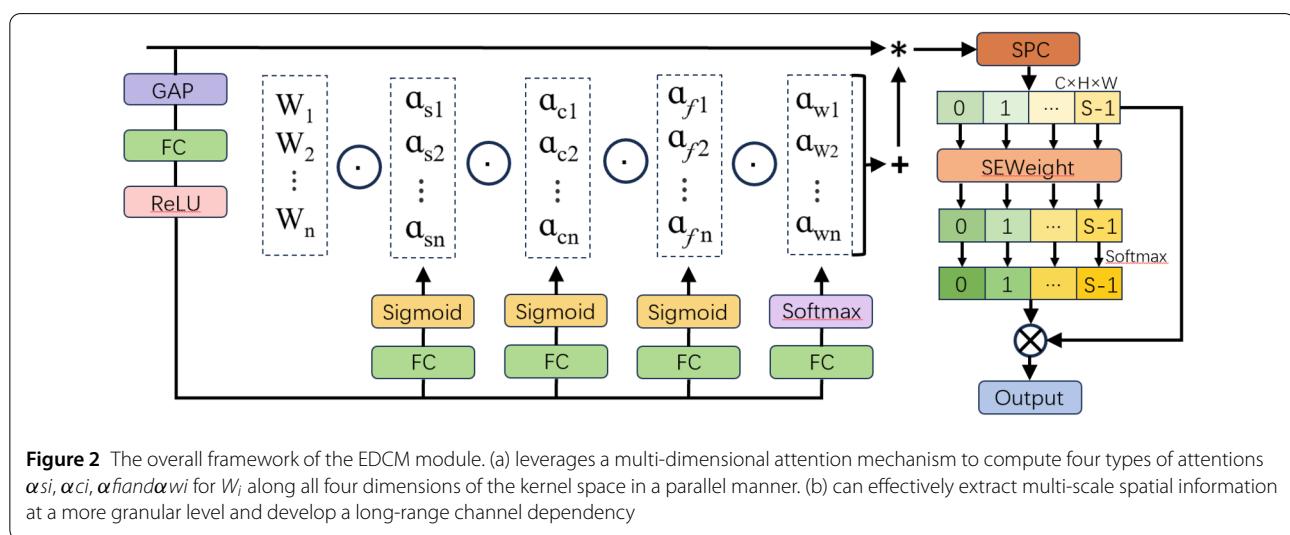
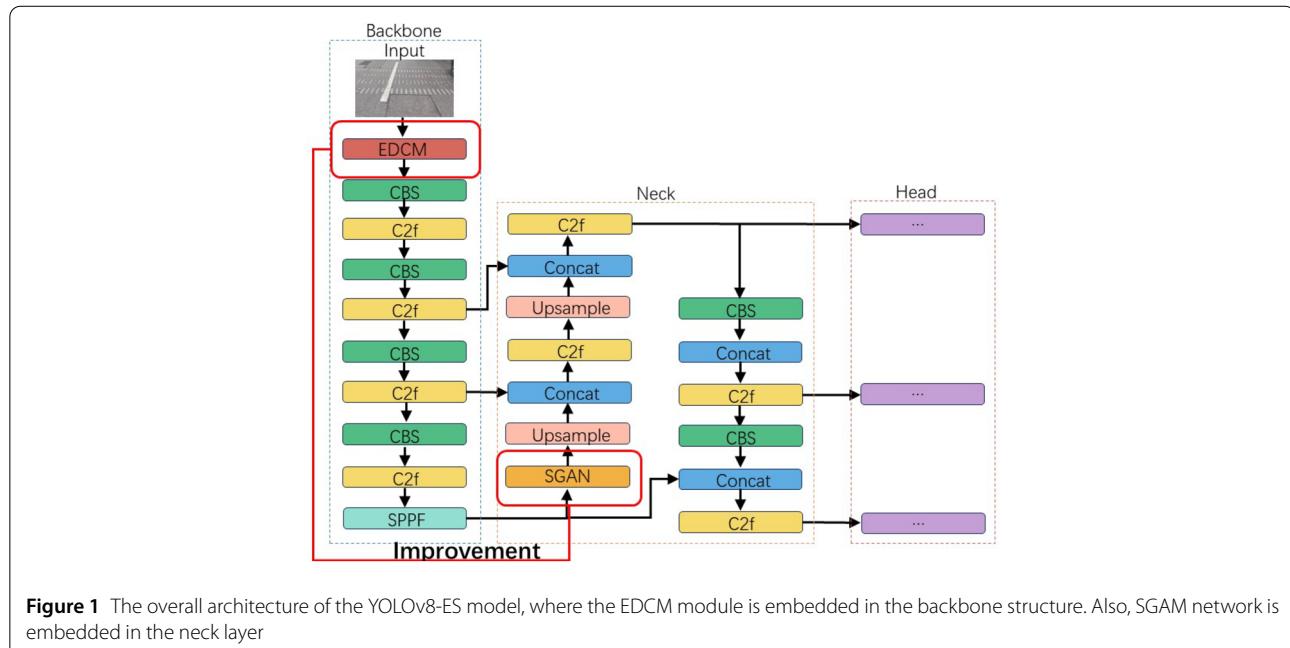
3.2 EDCM

EDCM enhances detection accuracy by leveraging the benefits of PSA (Pyramid Squeeze Attention) and ODConv (OMNI-DIMENSIONAL DYNAMIC CONVOLUTION), hence capturing target features with more precision.

In road damage identification, the road damage closely resembles the background, rendering typical convolutional networks incapable of reliably localizing the damage location. Consequently, we propose a novel dynamic convolutional EDCM and include it in the backbone network of YOLOv8n to enhance model accuracy without incurring significant processing overhead. The comprehensive procedure is illustrated in Fig. 2. Traditional dynamic convolution attributes the dynamic properties of convolution kernels solely from one dimension of the kernel space (the number of convolution kernels), neglecting the other three dimensions (the spatial size of each convolution kernel, the number of input channels, and the number of output channels). Conventional convolution is limited to adequately capturing local information and needs to establish long-range channel dependencies. Conversely, EDCM employs multidimensional attention to acquire complementary focus across the four dimensions of the kernel space via a parallel approach, refining the local fine-grained features of the image while adeptly extracting multiscale spatial information at a more granular level and fostering long-range channel dependence to efficiently capture the salient features of the image, thereby enhancing overall model performance.

3.3 SGAM

SGAM through the incorporation of Squeeze-and-Excitation (SE), Global Attention Mechanism (GAM), and Coordinate Attention (CA), the network acquires the ability to leverage global information to selectively accentuate salient characteristics while diminishing the significance of less pertinent ones. Information is preserved to improve cross-dimensional interactions and strengthen the representation of global interaction. Additionally, by incorporating positional information into the channel attention, the network can focus on a broader area, thereby enhancing the model's feature extraction capability without significant computational costs. We integrate SGAM into the



neck of YOLOv8n. The entire procedure is illustrated in Fig. 3.

3.4 Wise-IoU loss

Wise-IoU is universally applicable as it incorporates a dynamic non-monotonic focusing mechanism to allocate gradient gain more judiciously. This effectively addresses low-quality training data, thereby enhancing the accuracy and robustness of the target detection model, which has extensive applications [41–43].

In road damage detection, obstacles arise from incomplete target boundaries and subpar quality samples. The conventional geometric loss function CIoU frequently im-

poses excessive penalties when addressing ambiguous target borders, thereby diminishing the model's generalization capability. Consequently, to mitigate the influence of low-quality samples on the boundary loss function in target identification and enhance the accuracy of the network model, we propose WIoU. WIoU is a dynamic non-monotonic frequency modulation that enhances the emphasis on high-quality anchor boxes, mitigates the influence of low-quality samples, and offers a more thorough training framework for the model. This work utilizes WIoUv3, one of three versions of WIoU. Wise-IoUv3 enhances Wise-IoUv1 by incorporating a non-monotonic focusing coefficient 'r' derived from the anomalous parame-

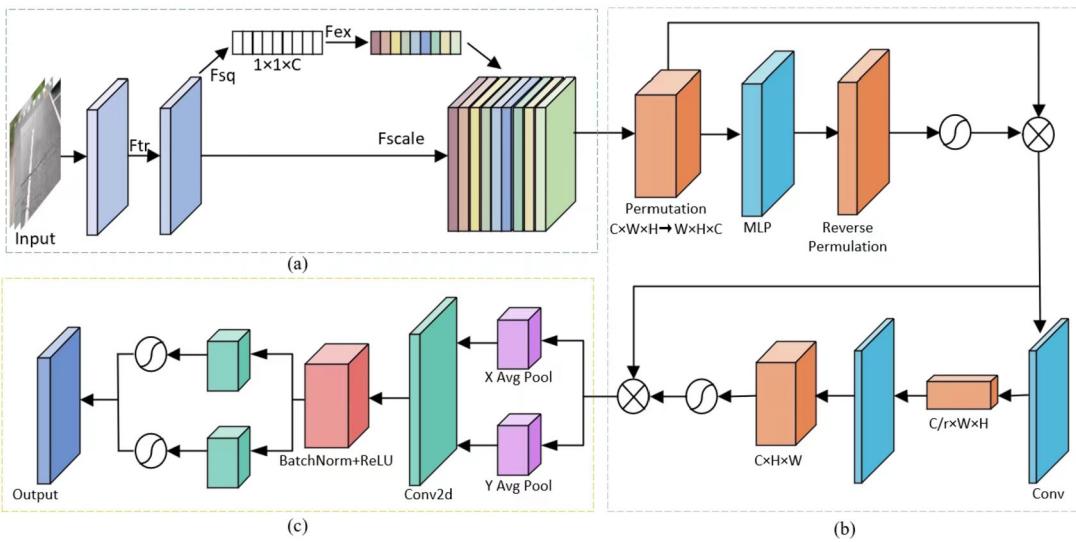


Figure 3 The overall framework of the SGAM module. (a) adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels. (b) reserves information to magnify the “global” cross-dimension interactions. (c) captures cross-channel, direction-aware and position-sensitive information

ter β . The formula is presented below [40]:

$$L_{WIoUv3} = rL_{WIoUv1}, \quad (1)$$

$$L_{WIoUv1} = R_{WIoU} L_{IoU}, \quad (2)$$

$$R_{WIoU} = \exp \left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right), \quad (3)$$

$$\beta = \frac{L_{IoU}^*}{\overline{L_{IoU}}}, \quad (4)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}}. \quad (5)$$

In the above equations, the superscript “*” denotes the segregation of W_g and H_g from the computational map, hence enhancing convergence efficiency. β is defined as the outlier degree, with a lower outlier degree indicating a superior-quality anchor frame. A minimal gradient gain is allocated to concentrate the bounding box regression on standard quality anchor frames. Anchor boxes with more outlier degrees are assigned a diminished gradient gain to effectively mitigate the impact of low-quality data on creating substantial adverse gradients. Given that the quality requirements for both L_{IoU} and anchor boxes are dynamic, WIoUv3 can consistently allocate the gradient gain approach that is most appropriate for the prevailing circumstances. L_{IoU} functions as a moving average with momentum m , and it is continuously adjusted to sustain a consistently elevated β level. This effectively addresses the issue of sluggish convergence during the latter phase of training.

4 Experiment

In the experimental part, we initially delineate the attributes of the RDD2022 dataset. We will outline the configuration of the experimental setting and the specific procedures, encompassing data preprocessing, model selection, model training, and evaluation techniques. Ultimately, we examine the model’s performance outcomes and assess the algorithm’s efficacy via ablation studies.

4.1 Material

4.1.1 Dataset

The Road Damage Dataset RDD2022 comprises 47,420 road photographs from six countries: Japan, India, the Czech Republic, Norway, the USA, and China. These pictures have been annotated with over 55,000 incidents of roadway damage. The dataset encompasses four categories of road damage: longitudinal cracks, transverse cracks, alligator cracks, and potholes. The annotated dataset is intended to develop deep learning systems to identify and categorize road defects autonomously. The RDD2022 dataset offers extensive training data for deep learning models, aiding researchers and developers in enhancing and optimizing target detection methods for intricate road damage identification.

4.1.2 Experimental environment

This paper’s software component utilizes the Pytorch framework version 2.2.1, Cuda version 11.8, and Python version 3.8. The gear comprises an RTX 4060 graphics card, an R7-7435H CPU, and 16 GB of video memory.

4.1.3 Details of the experiment

Step 1 (data preprocessing): To thoroughly assess the efficacy of the target recognition algorithm for road damage, we pick the RDD2022 dataset. We picked a total of 1977 photos from the China MotorBike part of RDD2022, allocating 1187 images for training, 395 images for testing, and 395 images for validation. Confronted with photos that are absent labels or erroneous annotations, we manually rectify and tag them utilizing methods like LabelImg to guarantee the dataset's accuracy and completeness.

Step 2 (Model Selection): Numerous advanced algorithms have been created in the domain of target recognition and progressively implemented in traffic scene detection tasks. Each algorithm possesses distinct designs and characteristics, each with certain advantages and downsides. When choosing an algorithm for our research, it is crucial to evaluate both its intrinsic performance and its adaptability to a specific dataset. This study examines eight cutting-edge target identification algorithms: YOLOv5, SSD [25], YOLOv7 [20], EfficientDet [26], Faster R-CNN [24], YOLOv8, Li [19] and Wang [18]. This work also presents and integrates a novel target identification approach, YOLOv8-ES. While these algorithms have exhibited outstanding performance in their specific areas and practical applications, our primary objective is to differentiate their performance in traffic scenarios. To attain this objective, we evaluate them against a harmonized dataset to determine the most appropriate model for object detection in traffic scenarios.

Step 3 (Model Training): In the context of our research, the YOLOv8-ES model is subjected to a stringent training protocol utilizing a transfer learning methodology. We initialize the model with commonly utilized pre-training weights from the COCO dataset. Subsequently, the model is refined using our harmonized dataset. Table 1 presents a comprehensive enumeration of configuration parameters related to all the models.

Our model has exceptional performance across various measures, as illustrated in Fig. 4. In particular, our model achieves high-precision object detection in complex road scenarios and maintains stable performance when dealing with blurred backgrounds and objects of different proportions.

Step 4 (Model assessment): To thoroughly and systematically assess the model's performance in target recognition inside traffic scenarios, we utilize a range of defined assessment metrics, including precision, recall, and F1. These

measures offer a quantitative assessment of the model's overall performance in intricate traffic scenarios. Should the model exhibit subpar performance in specific areas, we may contemplate adjusting the hyperparameters or augmenting the training dataset to improve its efficacy. Maintaining a balance between overfitting and underfitting is essential for the model's generalizability. In the following section, we will examine the particular evaluation metrics employed in this study.

This research study adopts the evaluation metrics of Mean Average Precision(mAP), recall and F1 to scrutinize the capabilities of the developed YOLOv8-ES model (see Eqs. (6)-(8)) [44].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (7)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (8)$$

where TP (true positive) is the number of true positive detections, FP (false positive) is the number of bounding boxes that the model incorrectly predicts, and FN (false negative) is the number of existing objects that are not detected by the model. Precision indicates whether detection results are correct, while recall indicates whether targets are detected. Due to the conflict between them, F1 is usually used to be a balanced measure. AP is calculated based on the precision-recall curve, which is commonly used in object detection tasks. mAP denotes the average value of AP for each category [45].

$$\text{AP} = \int_0^1 P(R)dr, \quad (9)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i. \quad (10)$$

4.2 Result

To assess the improved efficacy of the proposed model, eight noTabletarget detection networks (YOLOv5, SSD [25], YOLOv7 [20], EfficientDet [26], Faster R-CNN [24], YOLOv8, Li [19] and Wang [18]) were comparatively analyzed within the same configuration framework using the RDD2022 dataset. Faster R-CNN has a two-stage detection methodology, whereas SSD, Li [19], Wang [18], YOLOv7, EfficientDet, and YOLOv5 are all single-stage detection algorithms. The outcomes of this comparison are presented in Table 2. The results indicate that the model sizes of the algorithms created in this study are considerably less than those of YOLOv7, SSD, and Faster R-CNN while substantially surpassing these models in

Table 1 Parameter configuration of all the models

Training parameters	Details
Epochs	100
batch-size	16
image-size (pixels)	512 × 512
Initial learning rate	0.01

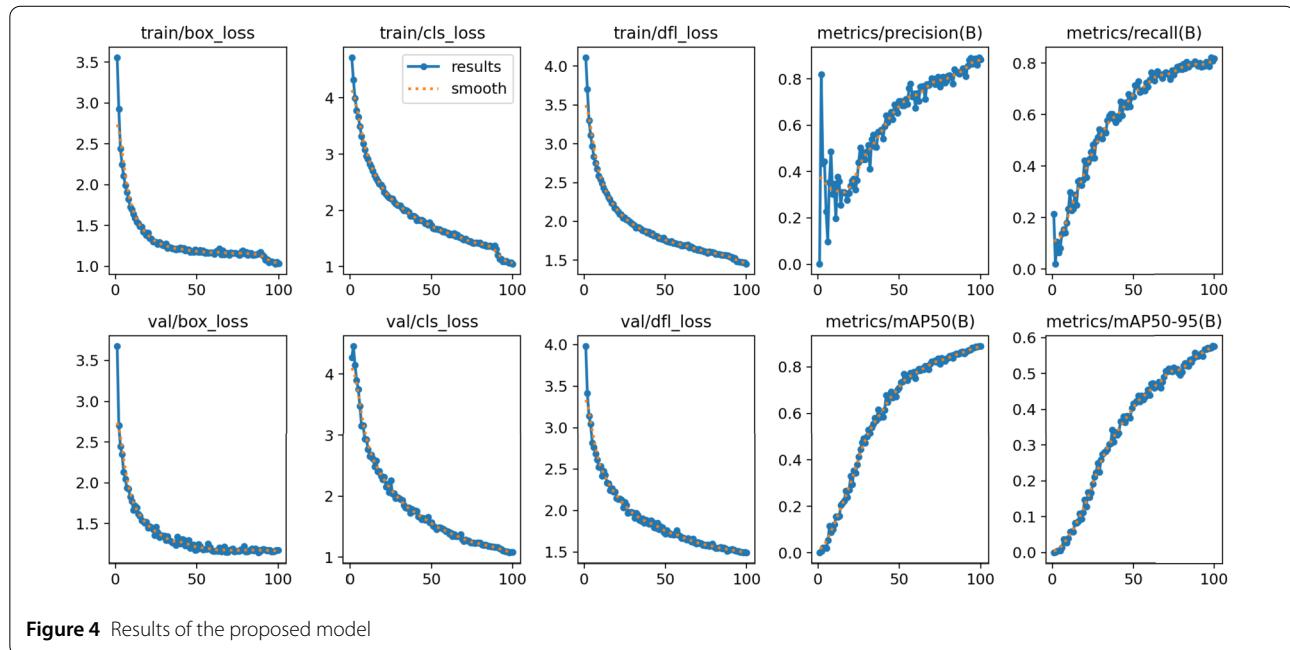


Table 2 Comparison of experimental results of YOLOv8-ES with other object detection algorithms on RDD2022

Model	Venue	F1	Recall	mAP50	FPS	GFLOPs
SSD	ECCV'16	0.548	0.467	0.664	55	62.7
Faster R-CNN	ICCV'15	0.389	0.310	0.523	20	93.4
EfficientDet	CVPR'20	0.407	0.315	0.575	36	7.5
YOLOv5	/	0.739	0.652	0.731	66	4.3
YOLOv7	CVPR'23	0.779	0.753	0.809	46	13.3
YOLOv8	/	0.781	0.762	0.805	71	8.2
Li [19]	T-ITS'24	0.773	0.749	0.779	27	8.4
Wang [18]	T-ITS'24	0.785	0.890	0.891	77	7.5
YOLOv8-ES(ours)	-	0.850	0.809	0.888	79	9.8

mAP50 performance. Moreover, our proposed model exhibits superior GFLOPs performance and significantly outperforms Li [19], YOLOv5, EfficientDet, and YOLOv8 in mAP50, which is essential for effectively addressing the pavement crack detection task and meeting contemporary accuracy requirements in complex environments. Our suggested approach is marginally less practical than Wang [18] in mAP50, whereas it has superior detection speed. The YOLOv8-ES model demonstrates superior and consistent detection results compared to one-stage and two-stage deep learning models. This comparison clearly emphasizes the superior performance of YOLOv8-ES in damage detection. In conclusion, YOLOv8-ES exhibits superior performance and provides a distinct advantage in detecting roadway damage in traffic situations. It substantially enhances the achievement of safer and more efficient urban transportation possibilities.

To assess the generalization and robustness of the proposed algorithm, eight noTablet target detection networks

(YOLOv5, SSD [25], YOLOv7 [20], EfficientDet [26], Faster R-CNN [24], YOLOv8, Li [19] and Wang [18]) were evaluated within a consistent configuration framework on the VOC2007 dataset. The outcomes of this comparison are presented in Table 3. The proposed model surpassed most others in mAP50 (58.3%), recall (53.4%), and F1 (0.580), indicating the significant potential of YOLOv8-ES to manage diverse data types effectively. In contrast to Faster R-CNN and SSD, the proposed model exhibits a diminished mAP50 value; however, it surpasses these models in recall, FPS, F1, and GFLOPs, underscoring its superiority in scenarios where computational resources are limited and real-time detection is essential. While YOLOv7 surpasses the proposed model in mAP50, recall, and F1, it falls short in FPS and GFLOPs. This suggests that our proposed model can achieve high accuracy even in scenarios with limited computational resources and the need for real-time detection.

Table 3 Comparison of experimental results of YOLOv8-ES and other object detection algorithms on VOC2007

Model	Venue	F1	Recall	mAP50	FPS	GFLOPs
SSD	ECCV'16	0.556	0.481	0.680	21	62.7
Faster R-CNN	ICCV'15	0.508	0.487	0.699	13	93.4
EfficientDet	CVPR'20	0.387	0.283	0.537	19	7.5
YOLOv5	/	0.555	0.469	0.511	22	4.3
YOLOv7	CVPR'23	0.646	0.599	0.621	24	13.3
YOLOv8	/	0.566	0.519	0.562	78	8.2
Li [19]	T-ITS'24	0.519	0.483	0.525	13	8.4
Wang [18]	T-ITS'24	0.599	0.548	0.572	57	7.5
YOLOv8-ES(ours)	-	0.580	0.534	0.583	60	9.8

Table 4 Experimental results for the components

Model	EDCM	SGAM	WIoU	Precision	Recall	mAP50	mAP50-95
YOLOv8				0.802	0.762	0.805	0.511
	✓			0.823	0.751	0.820	0.511
		✓		0.855	0.757	0.826	0.528
			✓	0.867	0.817	0.862	0.554
		✓	✓	0.859	0.796	0.847	0.544
	✓	✓		0.852	0.767	0.850	0.555
	✓		✓	0.825	0.762	0.825	0.514
	✓	✓	✓	0.895	0.809	0.888	0.576

4.2.1 Ablation experiment

This section shows a series of ablation experiments to validate the effectiveness of the algorithm improvement, with comparison results displayed in Table 4. Each phase of the enhanced algorithm demonstrates substantial performance improvement in complex road damage detection. The updated YOLOv8n model with EDCM exhibited a 1.5% enhancement, achieving 82% in mAP50 compared to YOLOv8n. EDCM mitigates the constraints of conventional convolutional layers in high-dimensional data processing by implementing a multidimensional attention mechanism and a parallelization strategy, thereby enhancing the efficiency and efficacy of convolutional neural networks in capturing multidimensional features and intricate dependencies. The mAP of the YOLOv8n model, following the incorporation of SGAM, demonstrates a relative enhancement of 2.1%, reaching 82.6% compared to the original YOLOv8n. The module improves the model's localization and recognition precision by recalibrating channel feature responses, preserving essential information, and enhancing global cross-dimensional interactions. It also efficiently captures cross-channel, orientation-aware, and position-sensitive data. The incorporation of SGAM successfully mitigates the issues related to feature extraction concerning the morphology and characteristics of intricate objects. The YOLOv8 model, with modifications to the loss function, enhances the mAP of the original YOLOv8n by approximately 5.7%, reaching 86.2%. The balanced gradient assignment helps the model focus on average-quality anchor boxes. This prevents overfitting to high-quality samples and excessive penalization

Table 5 Performance comparison of convolution layer

Method	F1	Recall	mAP50	mAP50-95
YOLOv8	0.781	0.762	0.805	0.511
+ODConv	0.778	0.729	0.815	0.513
+EPSA	0.782	0.747	0.818	0.512
+DSCNet	0.785	0.752	0.814	0.517
+EDCM	0.786	0.751	0.820	0.511

of low-quality ones. It improves the model's adaptability to real-world situations and boosts overall detection performance. Ultimately, our model surpasses the original YOLOv8n by 10.3%, achieving a performance of 88.8%, illustrating the efficacy of the numerous changes implemented. These findings underscore algorithmic improvement's significance and practical utility in intricate road damage identification.

Table 5 presents the efficacy of EDCM in delineating intricate characteristics in crack images within the framework of crack detection tasks. The experimental findings on the RDD2022 dataset utilize YOLOv8n as the baseline. The results are concisely summarized below. The F1, Recall, mAP50, and mAP50-95 of EDCM are 0.786, 75.1%, 82%, and 51.1%, respectively. The performance surpasses the other three noTableconvolutional layers, substantiating the efficacy of EDCM in road damage identification.

Table 6 presents the experimental findings on the RDD2022 dataset utilizing YOLOv8 as the baseline, contrasting the suggested SGAM with the salient attention mechanism. The F1, Recall, mAP50, and mAP50-95 of SGAM are 0.803, 75.7%, 82.6%, and 52.8%, respectively.

Table 6 Performance comparison of attention mechanisms

Method	F1	Recall	mAP50	mAP50-95
YOLOv8	0.781	0.762	0.805	0.511
+SE	0.795	0.743	0.814	0.516
+CBAM	0.782	0.747	0.818	0.512
+CA	0.787	0.751	0.815	0.527
+GAM	0.803	0.753	0.825	0.515
+SGAM	0.803	0.757	0.826	0.528

Table 7 Performance comparison of the loss function

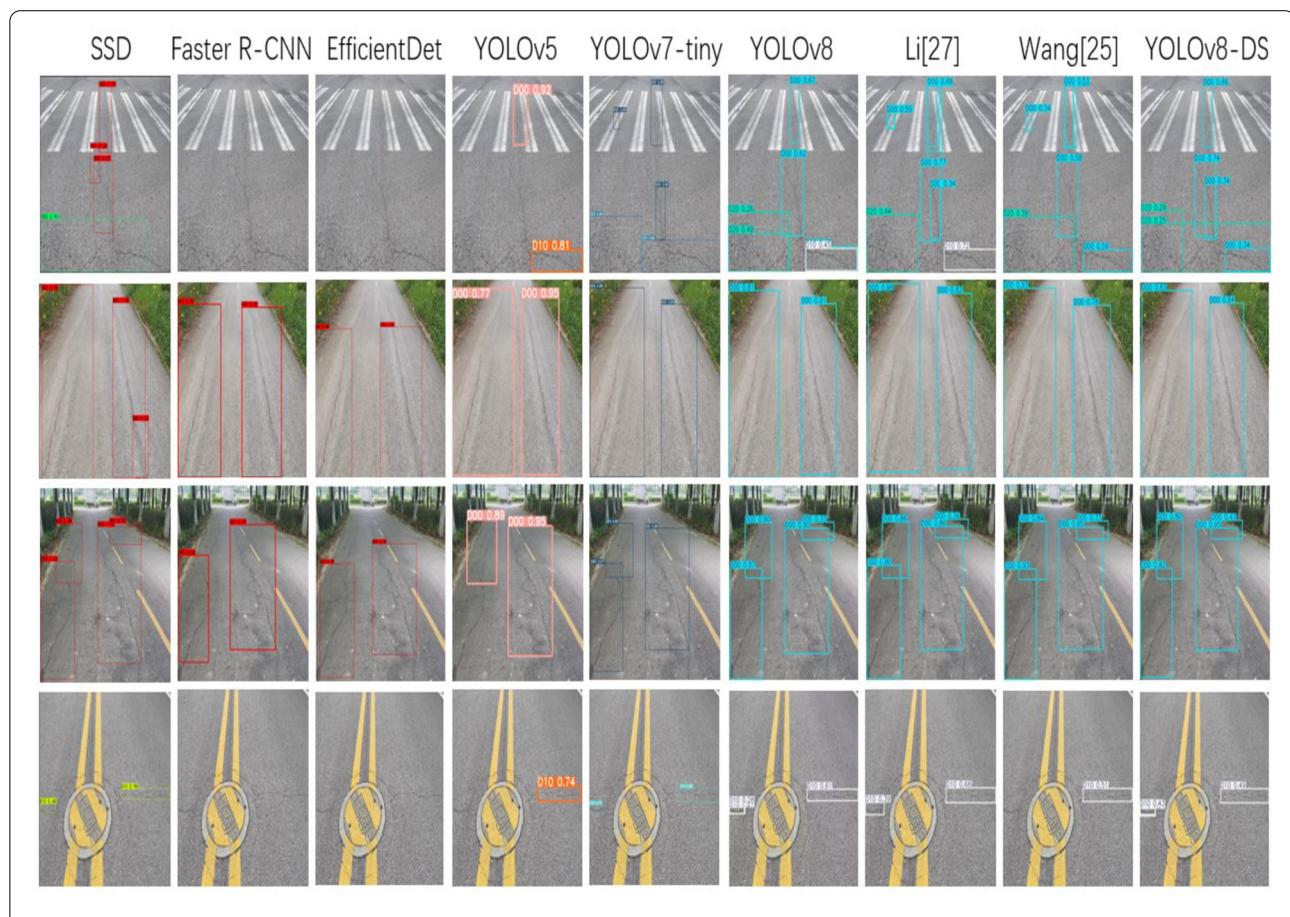
Method	F1	Recall	mAP50	mAP50-95
YOLOv8	0.781	0.762	0.805	0.511
+Focal-EIoU	0.822	0.781	0.843	0.544
+MPDIoU	0.803	0.796	0.842	0.549
+EIoU	0.813	0.768	0.836	0.537
+WIoU	0.841	0.817	0.862	0.554

The performance surpasses that of the other four prominent attention mechanisms, effectively demonstrating the superiority of SGAM in road damage identification.

The part of the loss function is detailed in Table 7. The results indicate that the WIoU employed in this work achieves an mAP50 of 86.2%, surpassing EIoU (83.6%), Focal-EIoU (84.3%), and MPDIoU (84.2%). Moreover, WIoU surpasses alternative loss functions in F1, Recall, and mAP50-95. This further substantiates the appropriateness of the WIoU loss function for crack detection activities.

4.2.2 Results shown

Figure 5 illustrates the detection outcomes of the YOLOv8-ES model with eight noTabletarget detection networks (YOLOv5, SSD [25], YOLOv7 [20], EfficientDet [26], Faster R-CNN [24], YOLOv8, Li [19] and Wang [18]) detecting road damage cracks. The YOLOv8-ES model effectively identifies cracks. EfficientDet and Faster R-CNN failed to identify true-positive cracks in images characterized by poor illumination, shadows, low contrast, and irregular shapes of varying sizes. In contrast, Wang [18], SSD, YOLOv5, YOLOv8, and YOLOv7 experienced partial omissions in detecting fine cracks. Li [19] and Wang [18] observed that the model produced erroneous bounding

**Figure 5** Comparison of different models with ours

boxes for fine cracks and consistently predicted the same crack. In YOLOv5, the produced bounding boxes are often too massive or tiny, hindering precise target recognition. However, our proposed model can accurately distinguish the occluded cracks and maintains stable performance when dealing with blurred backgrounds and objects with different scales. This is a crucial performance attribute for damage identification in intricate and realistic environments. It contributes to establishing a more dependable basis for scientific inquiry. It offers enhanced and precise target detection solutions for road damage assessment and possesses significant promise for future scientific inquiry and applications.

4.2.3 Conclusion

The experimental findings unequivocally validate the efficacy of YOLOv8-ES and the model transportation system in advanced intelligent object detection. They were incorporated into the YOLOv8 framework by integrating EDCM, SGAM, and Wise-IoU methods. We have achieved substantial advancements in precisely recording intricate object shapes and characteristics, even in scenarios with occlusion and blurring. The model's versatility in detecting diverse road damage conditions significantly enhances intelligent road maintenance systems.

This study holds substantial practical importance for enhancing the safety and efficiency of urban transportation, mitigating traffic accidents, and improving the urban landscape. Through continuous efforts and research, we improve intelligent road damage identification. This will increase convenience and safety in urban transportation in the future. While our strategy yields exceptional results, it is essential to acknowledge that it incurs a more significant computational expense. Future studies will explore lightweight network designs to address this issue. Furthermore, future studies should incorporate considerations for implementing networks on hardware devices.

Acknowledgements

I would like to express my sincere gratitude to the senior brothers in the team for their correction of this work. Additionally, the authors are grateful to the reviewers for their insightful suggestions to strengthen the quality of the manuscript.

Author contributions

Zeng Kaili completed the literature research, data processing, experimental operation, and article writing. Fan Rui revised the article. The corresponding author directed the experimental ideas and was responsible for the entire project. All authors read and approved the final manuscript.

Funding

None.

Data availability

The RDD2022 data can be accessed at the GitHub repository: <https://github.com/sekilab/RoadDamageDetector>. The VOC2007 can be accessed at the following link: <http://host.robots.ox.ac.uk/pascal/VOC/>.

Declarations

Competing interests

The authors confirm that they have no competing interests with any third-party organizations related to this work.

Author details

¹School of Data Science and Engineering, Xingzhi College, South China Normal University, Guangzhou, 516600, China. ²College of Electronics and Information Engineering, Tongji University, Shanghai, 201804, China.

Received: 28 September 2024 Revised: 16 December 2024

Accepted: 16 January 2025 Published online: 10 February 2025

References

1. J. Bchle, J. Hringer, N. Khler, K.K. Zer, M. Enzweiler, R. Marchthaler, Competing with autonomous model vehicles: a software stack for driving in smart city environments. *Auton. Intell. Syst.* **4**(1), 1–13 (2024)
2. Q. Zhan, Y. Zhou, J. Zhang, C. Sun, R. Shen, B. He, A novel method for measuring center-axis velocity of unmanned aerial vehicles through synthetic motion blur images. *Auton. Intell. Syst.* **4**(1), 16 (2024)
3. J. Dinneweth, A. Boubezoul, R. Mandiau, S. Espié, Multi-agent reinforcement learning for autonomous vehicles: a survey. *Auton. Intell. Syst.* **2**(1), 27 (2022)
4. X. Wang, X. Qi, P. Wang, J. Yang, Decision making framework for autonomous vehicles driving behavior in complex scenarios via hierarchical state machine. *Auton. Intell. Syst.* **1**(1), 12 (2021)
5. W. Zhou, D. Chen, J. Yan, Z. Li, H. Yin, W. Ge, Multi-agent reinforcement learning for cooperative lane changing of connected and autonomous vehicles in mixed traffic. *Auton. Intell. Syst.* **2**, 5 (2022)
6. M.T. Cao, Q.V. Tran, N.M. Nguyen, K.T. Chang, Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. *Adv. Eng. Inform.* **46**, 101182 (2020)
7. L. Fan, D. Wang, J. Wang, Y. Li, Y. Cao, Y. Liu, et al., Pavement defect detection with deep learning: a comprehensive survey. *IEEE Trans. Intell. Veh.* **9**(3), 4292–4311 (2024)
8. N. Wang, X. Zhao, P. Zhao, Y. Zhang, Z. Zou, J. Ou, Automatic damage detection of historic masonry buildings based on mobile deep learning. *Autom. Constr.* **103**, 53–66 (2019)
9. C. Liu, J. Zhao, C. Zhu, X. Xia, H. Long, MECFNet: reconstruct sharp image for UAV-based crack detection. *IEEE Trans. Intell. Transp. Syst.* **25**(10), 15016–15028 (2024)
10. Y. Pan, X. Zhang, G. Cervone, L. Yang, Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**(10), 3701–3712 (2018)
11. Z. Wang, Y. Zhang, K.M. Mosalam, Y. Gao, S. Huang, Deep semantic segmentation for visual understanding on construction sites. *Comput.-Aided Civ. Infrastruct. Eng.* **37**(2), 145–162 (2022)
12. R. Fan, L.M. Road, Damage detection based on unsupervised disparity map segmentation. *IEEE Trans. Intell. Transp. Syst.* **21**(11), 4906–4911 (2020)
13. R. Fan, U. Ozgunalp, Y. Wang, M. Liu, I. Pitas, Rethinking road surface 3-D reconstruction and pothole detection: from perspective transformation to disparity map segmentation. *IEEE Trans. Cybern.* **52**(7), 5799–5808 (2022)
14. X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, X. Yang, Automatic pixel-level crack detection and measurement using fully convolutional network. *Comput.-Aided Civ. Infrastruct. Eng.* **33**(12), 1090–1109 (2018)
15. Y. Xu, D. Li, Q. Xie, Q. Wu, J. Wang, Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. *Measurement* **178**, 109316 (2021)
16. C. Lin, D. Tian, X. Duan, J. Zhou, D. Zhao, D. Cao, DA-RDD: toward domain adaptive road damage detection across different countries. *IEEE Trans. Intell. Transp. Syst.* **24**(3), 3091–3103 (2023)
17. C. Xiong, T. Zayed, E.M. Abdelkader, A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks. *Constr. Build. Mater.* **414**, 135025 (2024)
18. S. Wang, H. Jiao, X. Su, Q. Yuan, An ensemble learning approach with attention mechanism for detecting pavement distress and disaster-induced road damage. *IEEE Trans. Intell. Transp. Syst.* **25**(10), 13667–13681 (2024)
19. J. Li, Z. Qu, S.Y. Wang, S.F. Xia, YOLOX-RDD: a method of anchor-free road damage detection for front-view images. *IEEE Trans. Intell. Transp. Syst.* **25**(10), 14725–14739 (2024)

20. C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv e-prints (2022). [arXiv:2207.02696](https://arxiv.org/abs/2207.02696)
21. M.B. Prakash, K.C. Sriharipriya, Enhanced pothole detection system using YOLOX algorithm. *Auton. Intell. Syst.* **2**(1), 1–16 (2022)
22. Y.Z. Lin, Z. Nie, H. Ma, Dynamics-based cross-domain structural damage detection through deep transfer learning. *Comput.-Aided Civ. Infrastruct. Eng.* **37**, 24–54 (2021)
23. R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(1), 142–158 (2015)
24. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017)
25. L. Wei, A. Dragomir, E. Dumitru, S. Christian, R. Scott, F. Cheng-Yang, et al., *SSD: Single Shot MultiBox Detector* (Springer, Cham, 2016)
26. M. Tan, R. Pang, Q.V. Le, EfficientDet: scalable and efficient object detection, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020)
27. S. Naddaf-Sh, M.M. Naddaf-Sh, A.R. Kashani, H. Zargarzadeh, An efficient and scalable deep learning approach for road damage detection, in *2020 IEEE International Conference on Big Data (Big Data)* (IEEE, 2020), pp. 5602–5608
28. G. Ye, J. Qu, J. Tao, W. Dai, Y. Mao, Q. Jin, Autonomous surface crack identification of concrete structures based on the YOLOv7 algorithm. *J. Build. Eng.* **73**, 106688 (2023)
29. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 7132–7141
30. S. Woo, J. Park, J.Y. Lee, K.I.S. Cbam, Convolutional block attention module, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 3–19
31. P.J. Bam, Bottleneck attention module. arXiv preprint (2018). [arXiv:1807.06514](https://arxiv.org/abs/1807.06514)
32. H. Zhang, K. Zu, J. Lu, Y. Zou, D. Meng EPSANet: an efficient pyramid squeeze attention block on convolutional neural network, in *Proceedings of the Asian Conference on Computer Vision* (2022), pp. 1161–1177
33. Y. Liu, Z. Shao, N. Hoffmann, Global attention mechanism: Retain information to enhance channel-spatial interactions. arXiv preprint (2021). [arXiv:2112.05561](https://arxiv.org/abs/2112.05561)
34. C. Li, A. Zhou, A. Yao, Omni-dimensional dynamic convolution. arXiv preprint (2022). [arXiv:2209.07947](https://arxiv.org/abs/2209.07947)
35. Y. Qi, Y. He, X. Qi, Y. Zhang, G. Yang, Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 6070–6079
36. J. Redmon, You only look once: unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) [arXiv:1506.02640](https://arxiv.org/abs/1506.02640)
37. J. Yu, Y. Jiang, Z. Wang, Z. Cao, H.T. Unitbox, An advanced object detection network, in *Proceedings of the 24th ACM International Conference on Multimedia* (2016), pp. 516–520
38. S. Ma, Y. Xu, Mpdiou: a loss for efficient and accurate bounding box regression. arXiv preprint (2023). [arXiv:2307.07662](https://arxiv.org/abs/2307.07662)
39. Y.F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, T. Tan, Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **506**, 146–157 (2022)
40. Z. Tong, Y. Chen, Z. Xu, R. Yu, Wise-IoU: bounding box regression loss with dynamic focusing mechanism. arXiv preprint (2023). [arXiv:2301.10051](https://arxiv.org/abs/2301.10051)
41. B. Ma, Z. Hua, Y. Wen, H. Deng, Y. Zhao, L. Pu, et al., Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments. *Artif. Intell. Agric.* **11**, 70–82 (2024)
42. X. Du, H. Cheng, Z. Ma, W. Lu, M. Wang, Z. Meng, et al., DSW-YOLO: a detection method for ground-planted strawberry fruits under different occlusion levels. *Comput. Electron. Agric.* **214**, 108304 (2023)
43. H. Zheng, G. Wang, D. Xiao, H. Liu, X. Hu, FTA-DETR: an efficient and precise fire detection framework based on an end-to-end architecture applicable to embedded platforms. *Expert Syst. Appl.* **248**, 123394 (2024)
44. J. Zhang, X. Yang, W. Li, S. Zhang, Y. Jia, Automatic detection of moisture damages in asphalt pavements from GPR data with deep CNN and IRS method. *Autom. Constr.* **113**, 103119 (2020)
45. F. Guo, Y. Qian, Y. Shi, Real-time railroad track components inspection based on the improved YOLOv4 framework. *Autom. Constr.* **125**, 103596 (2021)

Publisher's Note

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

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)