Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 1, 2024

Research on Pavement Defect Detection Algorithm Based on SEM-YOLOv8n

Jingwei Deng[1], Li Yang[1, 2]

[1]School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China
[2]Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjin 300222, China

Abstract: Automated detection and identification of pavement distresses is essential for timely pavement repair. Subtle pavement defects and multiple defects detection is a challenging task under complex background. With the deepening of the deep learning network, some subtle features tend to disappear and are more difficult to detect under the influence of the complex background. To solve the above problems, this paper proposes the SEM-YOLOv8n pavement defect detection algorithm. Firstly, SPD-Conv is used to replace the traditional convolution, which is conducive to retaining more defect detail information in the image and improving the detection ability of subtle defects; then an efficient multi-scale attention mechanism is added to the fusion network, so that the network suppresses the background information and focuses more on the defect information. Finally, MPDIoU is introduced as a loss function, which optimizes the minimum perpendicular distance between the predicted bounding box and the real bounding box and improves the localization ability, thus improving the accuracy of the network. Finally, the effectiveness of the proposed network is verified on the IRRDD dataset, and the results show that the method achieves 91.9% (Precision), 91.3% (Recall), and 71.3% (mAP) for the classification and detection of road multi-scale minor defects, which meets the demand of real-time road defect detection.

Keywords: Road defect; deep learning; YOLOv8.

1. Introduction

Defects in pavement directly affect the quality of the pavement. Cracks and potholes are the most common causes of damage to pavements, which are usually caused by improper operation or inferior materials during construction, excessive pressure on the pavement during long periods of heavy traffic, or the influence of specific climatic and environmental factors in certain areas. Specifically, poor pavement condition, abnormal pavement or severe damage may hinder traffic, guide incorrect driving behavior, or even cause traffic accidents and casualties. Therefore, regular comprehensive pavement inspections to detect pavement anomalies and damage in a timely manner are essential to ensure the convenience, correctness and safety of the associated traffic or driving behavior. In the early days, pavement crack inspection was mainly carried out by trained workers through on-site field investigations. However, this solution was inefficient, labor-intensive, and even hindered traffic, resulting in missed inspections. Later, with the development of science and technology, such as the use of ground-penetrating radar for pavement defect detection [1], this method, although the accuracy has been improved to some extent, is costly and slow, and cannot meet the huge number of pavement inspections in today's society. Several image processing techniques, such as edge detection [2], threshold segmentation [3], and mathematical morphology [4], have been used to detect pavement defects in the past decades. Due to the complex background interference such as multi-texture, multi-targets, and variable background illumination, which make pavement images the most difficult targets to recognize, traditional detection methods can no longer satisfy the need for fast and accurate detection of pavement defects.

With the rapid development of deep learning, computer vision-based defect detection methods have attracted great interest from academia and industry due to their advantages of safety, cost, efficiency and objectivity. Deep learning techniques have been successfully applied to target detection [5] and image classification tasks with good experimental results. Deng et al [6] applied a faster region-based convolutional neural network to real-world images taken from concrete bridges with complex backgrounds, and experimentally proved that the network meets the detection requirements. Wang et al [7] proposed and validated an effective crack length measurement method. The method consists of a detection module based on the target detection algorithm and a length calculation module, and the experiment proves the effectiveness of the methods. Zheng et al[8] proposed a new automatic road crack detection algorithm for the problems of low efficiency of the current real-time road crack detection research results, and low storage and computation capacity of the edge devices, and the experiment proves that the method effectively solves these problems. Luo et al[9] proposed a road crack automatic detection architecture STrans-YOLOX, which solves the problems that convolutional neural network cannot adequately simulate the long-term dependency between pixels in complex scenes, and is prone to lose the edge detail information. Although all the above methods target specific datasets and improve the accuracy of pavement defect detection, the computational volume is still large and the accuracy of detecting subtle pavement defects is low.

For subtle defects with small size and low resolution, the role of Space-to-depth (SPD) [10] in convolutional neural networks is to split the incoming feature maps into multiple bands, each band is convolved using a different convolution kernel, and the results are weighted and summed up according to the band they belong to, which can better retain the detailed information in complex background images and small objects,

to Improve the detection accuracy. At the same time, due to the complex background of the pavement, the attention mechanism can be considered to properly suppress the interference of the background, so that the network is more focused on the detection of minor defects on the pavement, in view of this, this paper uses the efficient multi-scale attention (EMA) [11] as the attention mechanism in this paper. The main contributions of this paper are as follows:

(1) In the backbone network of YOLOv8n, the original conventional convolutional layer with a step size of 2 is replaced with an SPD layer, followed by a stepless convolutional layer, and finally an SPD-Conv is formed, which can better retain the detail information in the complex background image, and thus efficiently extract the feature information of the fine defects at multiple scales.

(2) The EMA attention module was introduced to enhance the model's feature extraction capability for pavement defects in complex environments, allowing the network to focus on the pavement defects, thus reducing the influence of the background, such as the shadows of the trees resembling the shape of the cracks.

(3) MPDIoU is introduced as the loss function of the network to improve the localization ability of the model by considering the minimum vertical distance between the predicted bounding box and the real bounding box, thus improving the convergence speed and accuracy of the model.

2. Method

In this section, the overall network structure of SEM-YOLOv8n is first introduced, followed by the SPD-Conv module, the EMA attention module, and the MPDIoU loss function, respectively.

2.1. Structure of network

The network structure consists of four basic components: the input, the backbone network, the neck network, and the detection head, as shown in Figure 1. The inputs are 640×640×3 sized images which are preprocessed and fed into the backbone network. The backbone network is similar to YOLOv8n and utilizes SPD-Conv instead of traditional convolution, which allows for the extraction of more detailed pavement defect feature information while reducing the loss of subtle pavement defect features. The fusion network retains the original structure of YOLOv8n, fuses the shallow, medium and deep pavement defect feature information extracted from the backbone network, and adds the EMA attention mechanism in the small target branch to make the network more focused on subtle defects. Finally, the three feature layers output from the fusion network are further trained in the detection head network, and these outputs are integrated to achieve multi-scale detection of targets, utilizing the corresponding detection heads according to different defect sizes.

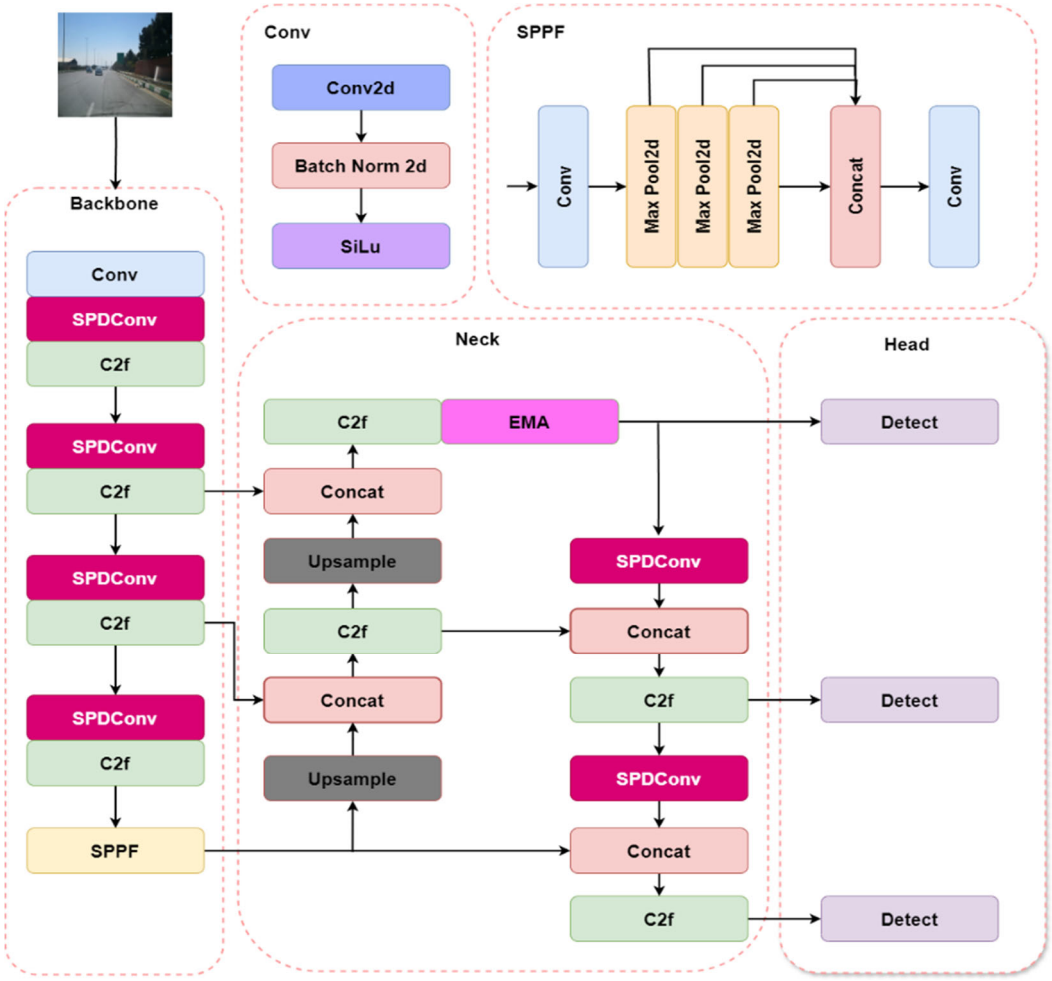


Figure 1. Structure of SEM-YOLOv8n

2.2. Structure of SPD-Conv

In this paper, the SPD-Conv structure is used to replace all the 3×3 convolutional layers in the YOLOv8n model, which prevents the loss of fine-grained information of the

small targets in the image processing process, especially in the downsampling process. The SPD-Conv consists of two main convolutional operations: the space-to-depth (SPD) and the non-spanned rows of convolutional layers as shown in Fig.2, and the feature mapping is processed by the SPD-Conv module. Firstly, the input feature maps are preprocessed from space to depth, and the input feature maps are divided into four classes in the spatial dimension, and the four vectors are spliced in the spatial dimension. Then the preprocessed feature maps are subjected to standard convolution to generate four feature maps of size $\frac{S}{2}\times\frac{S}{2}\times C_1$, which are spliced along the C_1 dimension to obtain a feature map of size $\frac{S}{2}\times\frac{S}{2}\times 4C_1$, and finally go through a non-spanning convolutional layer to obtain $\frac{S}{2}\times\frac{S}{2}\times C_2$.

SPD-Conv is utilized instead of traditional step convolution to mitigate the loss of detailed information that occurs when only a small fraction of pixels is occupied during small target detection. SPD-Conv enhances the model's ability to process spatial information, facilitating the differentiation of individual spermatozoa in dense clusters while it deepens the feature mapping so that the model is better able to account for complex backgrounds and dense objects. Thus utilizing SPD-Conv can significantly improve the accuracy of small target detection while retaining more detailed information.

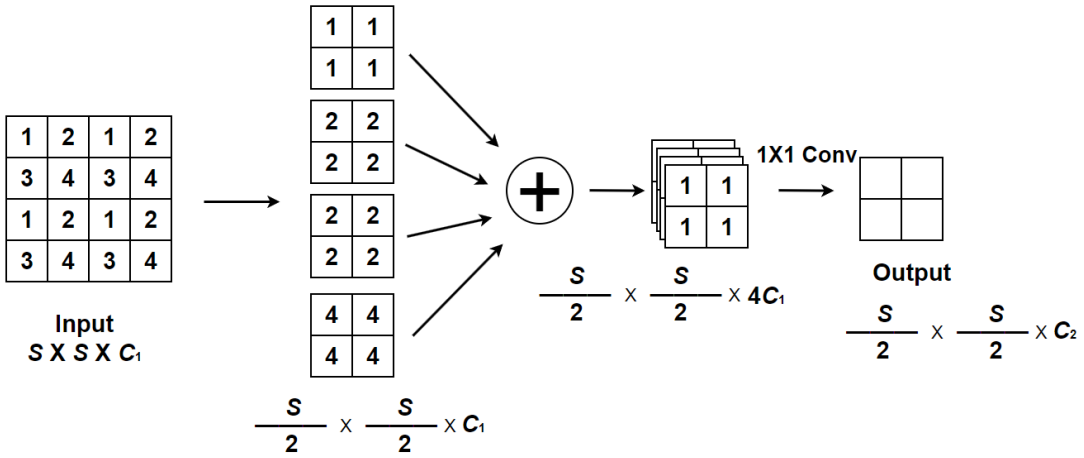


Figure 2. Structure of SPD-Conv

2.3. Structure of attention

Given the differences in defect scales and shapes, it is particularly important to further improve the multi-scale feature extraction capability of the network. Meanwhile, the complex background of the pavement is prone to have an impact on the defect detection, so this paper introduces the EMA attention mechanism to make the network more focused on the fine defects of the pavement. The EMA attention mechanism is an efficient multiscale attention mechanism based on cross-space learning proposed by Ouyang et al [11] in 2023, which prevents the loss of channel feature information and reduces computational overheads by reshaping part of the channels into bulk dimensions and grouping the channel dimensions without the need of a dimensionality reduction operation, which can prevent the loss of channel feature information and reduce the computational overhead with high accuracy and small number of parameters [12].The structure of EMA attention mechanism is shown in Fig. 3. The workflow is as follows: first, for any input $X\in R^{C\times H\times W}$, EMA slices it into G sub-features, such as $X=[X_0,X_i,...,X_{G-1}],X\in R^{C//G\times H\times W}$, in the channel dimension to obtain different semantics. Next, EMA uses 3 routes to extract the attention weight descriptors of the grouped feature graphs respectively [13].

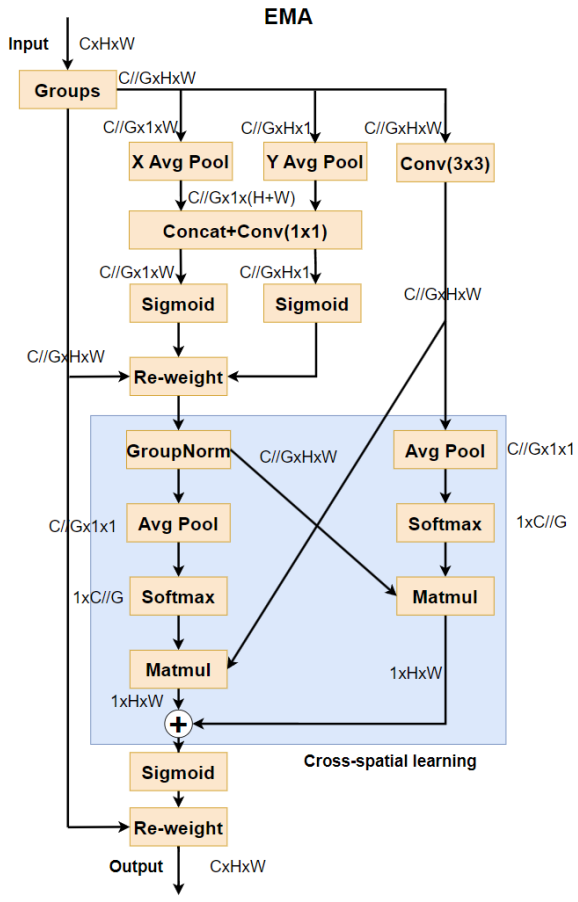


Figure 3. Structure of EMA

2.4. Loss function

While the complete intersection over union (CIoU) loss function used by YOLOv8n, SEM-YOLOv8n uses the minimum point distance intersection over union (MPDIoU) loss function. It improves the localization ability of the model by taking into account the minimum vertical distance between the predicted bounding box and the real bounding box, especially when the bounding boxes are highly overlapped or partially overlapped. MPDIoU [14] compensates for this by introducing an additional distance metric, i.e., the minimum vertical distance of the vertices between the predicted box and the real box, which allows the loss function to be more concerned with the exact alignment of the bounding boxes during the optimization. alignment, which can be mathematically expressed as.

$$MPDIoU = IoU - \frac{d_1^2 + d_2^2}{h^2 + w^2} \tag{1}$$

where d_1 and d_2 represent the Euclidean distances between the diagonals of the predicted bounding box and the real bounding box respectively. h and w are the height and width of the bounding box, respectively, and IoU denotes the intersection and concurrency ratio between the predicted bounding box and the real bounding box.

3. Experiment

In order to verify the performance of SEM-YOLOv8n proposed in this paper for road defect detection, the algorithm was trained and tested using Iranian road disease dataset and SEM-YOLOv8n was compared with other mainstream detection algorithms and its actual detection was visualized, which verified that the algorithm proposed in this paper shows good detection performance for subtle road defects with complex background. Finally, ablation experiments are conducted to verify the effectiveness of each improved module.

3.1. Dataset

The experimental results of the proposed SEM-YOLOv8n were evaluated on the Iranian Road Disease Dataset (IRRDD). This dataset collects local Iranian road damage dataset including different environmental conditions such as different shadows, lighting levels and daylight hours. In this dataset, there are 25,000 images of urban roads with a resolution of 640×640, and it is divided into training, testing and validation sets with 17,500, 5,000 and 2,500 images in a 7:2:1 ratio, respectively. The defect categories are categorized into four, namely: transverse cracks, longitudinal cracks, mesh cracks and potholes.

3.2. Evaluation metrics

The classification of the model is evaluated by precision (P) and recall (R). In addition, mean Average Precision (mAP) is used to evaluate the defect detection results. The number of parameters is the metric used to evaluate the complexity of the model and is defined as follows:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P(x)dx \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

where P is the average precision; TP is the count of "cracked" examples predicted by the model to be "cracked"; FP is the count of "background" examples predicted by the model to be "cracked"; FN represents the number of "cracked" examples that are predicted by the model to be "background"; and mAP represents the average value of the mean precision.

3.3. Implementation details

The experiments are performed on a workstation with a CPU model Intel Core i9-12900k@3.40GHz, GPU model NVDIA GeForce RTX 3090, 24G video memory, and 128G RAM. The experiment does not use a pre-trained model so that the model architecture and parameters can be better customized for this experiment; this customization improves the performance of the model and makes the experiment more objective. The learning rate is set to 0.01 and the number of categories is set to 4. The batch size is set to 32 due to GPU memory constraints. the maximum iteration is fixed to 100, which ensures a full loop through the training data for the road images. The detailed information and hyperparameters of the experiment are shown in Table 1.

Table 1. Implementation parameters

	batch	size	epoch
Train	64	640	100
	momentum	learning rate	decay
	0.937	0.01	0.0005
Test	iou	NMS	
	0.3	0.5	

3.4. Results

3.4.1. Evaluation

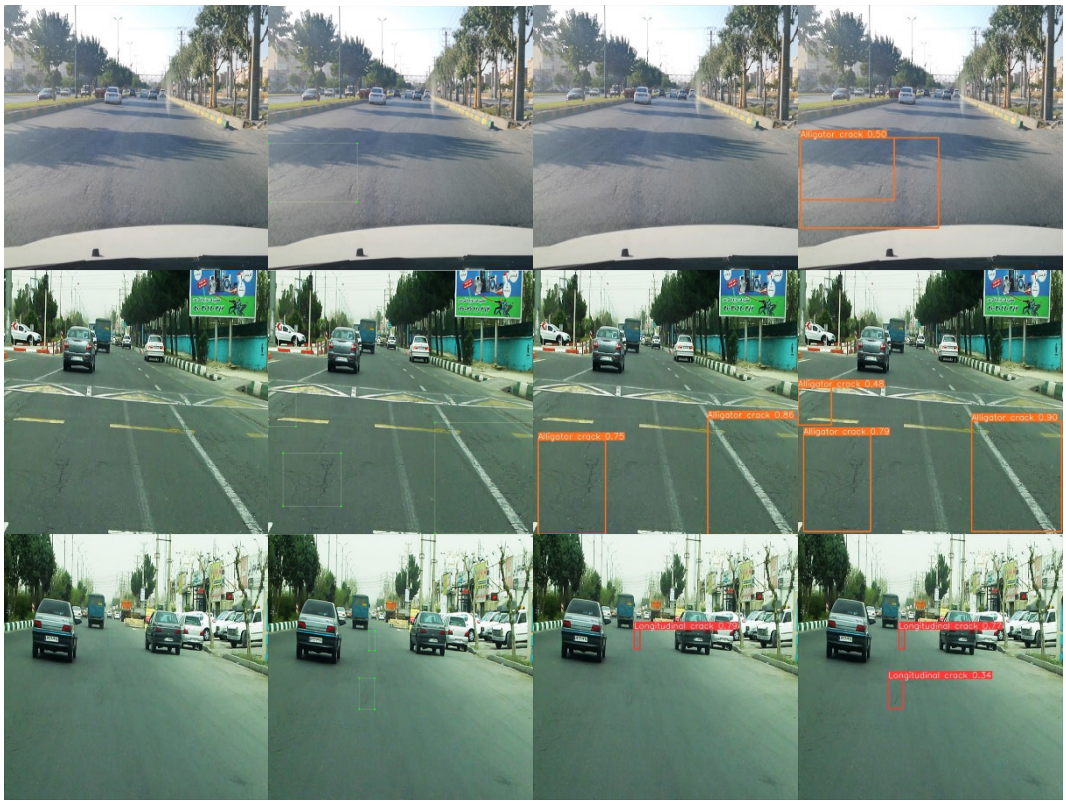
In order to validate the detection performance of the proposed model, three sets of comparison experiments were conducted on the IRRDD dataset to compare the model with the YOLOv8n, YOLOv9n, and YOLOv10n models. Table 2 shows the P, R, mAP, and parametric counts of SEM-YOLOv8n and YOLOv8n, YOLOv9n, and YOLOv10n models.

Table 2. Comparative results of the detection ability of different models

Methods	P (%)	R (%)	mAP (%)	Parameter (M)
YOLOv8n	89.1	91.0	67.4	3.01
YOLOv9n	90.5	93.0	69.4	8.18
YOLOv10n	66.1	57.6	64.7	2.70
SEM-YOLOv8n	91.9	91.3	71.3	4.76

3.4.2. Visualization

In order to visualize the detection effect of the improved modeling algorithm, an actual detection comparison experiment using YOLOv8n and SEM-YOLOv8n proposed in this paper is carried out as shown in Fig. As can be seen from the first row, YOLOv8n has missed detection due to shadow masking, while SEM-YOLOv8n can detect the defects; as can be seen from the second row, SEM-YOLOv8n has better boundary detection ability; as can be seen from the third row, SEM-YOLOv8n has better detection ability of subtle defects than YOLOv8n. Therefore, the SEM-YOLOv8n algorithm proposed in this paper can over detect pavement defects more accurately.



raw label YOLOv8n SEM-YOLOv8n

Figure 4. Comparison of YOLOv8n and SEM-YOLOv8n

3.4.3. Ablation

Ablation were conducted on the test set to verify the effectiveness of adding each module, as shown in Table 3. After replacing the ordinary convolution in YOLOv8n with SPD-Conv, the mAP is improved by 2.0%; then the EMA attention mechanism is added to the fusion network, and the mAP is improved by 2.8%; finally, applying the MPDIoU as the loss function, the line is formed into the SEM-YOLOv8n, and the mAP is improved by 3.9%, and the experiments proved the validity of the individual modules.

Table 3. Results of ablation

SPD-Conv	EMA	MPDIoU	mAP (%)	Parameter (M)
			67.4	3.01
√			69.4	4.74
√	√		70.2	4.76
√	√	√	71.3	4.76

4. Conclusion

Aiming at the problems that the pavement defects account for few pixels in the picture, small volume, complex background and multi-target, this paper proposes an algorithm SEM-YOLOv8n.Firstly, the traditional convolution is replaced by SPD-Conv to extract the feature information of the multi-scale fine defects, and then the EMA Attention Module is added to the fusion network to improve the model's ability of feature extraction for the pavement defects in the complex environment, and finally MPDIoU is chosen as the loss function of the network to improve the

convergence speed and accuracy of the model. The experimental results show that the proposed SEM-YOLOv8n outperforms other state-of-the-art detectors in all quantitative indexes, so the deep learning model proposed in this paper has certain advantages in defect detection and provides a practical solution for the research and application of road defect detection.

References

[1] N. Ahmad, M. Wistuba and H. Lorenzl, GPR as a crack detection tool for asphalt pavements: Possibilities and limitations, 2012 14th International Conference on Ground Penetrating Radar (GPR), Shanghai, China, 2012, pp. 551-555, doi: 10.1109/ICGPR.2012.6254925.

[2] Zhao H, Qin G, Wang X. Improvement of canny algorithm based on pavement edge detection[C]//2010 3rd international congress on image and signal processing. IEEE, 2010, 2: 964-967.

[3] Peng C, Yang M, Zheng Q, et al. A triple-thresholds pavement crack detection method leveraging random structured forest[J]. Construction and Building Materials, 2020, 263: 120080.

[4] Landstrom A, Thurley M J. Morphology-based crack detection for steel slabs[J]. IEEE Journal of selected topics in signal processing, 2012, 6(7): 866-875.

[5] Yang L, Deng J, Duan H, et al. Tunnel water leakage detection method based on ECA and YOLOv5 [J]. JOURNAL OF TIANJIN UNIVERSITY OF TECHNOLOGY AND EDUCATION, 2024,34(02):19-24. DOI:10.19573/j.issn2095-0926.202402003.

[6] Deng J, Lu Y, Lee V C S. Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network[J]. Computer-Aided Civil and Infrastructure Engineering, 2020, 35(4): 373-388.

[7] Wang S, Dong Q, Chen X, et al. Measurement of Asphalt Pavement Crack Length Using YOLO V5-BiFPN[J]. Journal of Infrastructure Systems, 2024, 30(2): 04024005.

[8] Zheng X, Qian S, Wei S, et al. The combination of transformer and you only look once for automatic concrete pavement crack detection[J]. Applied Sciences, 2023, 13(16): 9211.

[9] Luo H, Li J, Cai L, et al. STrans-YOLOX: Fusing swin transformer and YOLOX for automatic pavement crack detection[J]. Applied Sciences, 2023, 13(3): 1999.

[10] SUNKARARAJA, LUOTIE. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects C. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, F-rance,2022:443-459.

[11] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[12] Chen S, Li Y, Zhang Y, et al. Soft X-ray image recognition and classification of maize seed cracks based on image enhancement and optimized YOLOv8 model[J]. Computers and Electronics in Agriculture, 2024, 216: 108475.

[13] LIU , LU A, CUI H, et al. Lightweight model for detecting lotus leaf diseases and pests using improved YOLOv8[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(19): 168-176.

[14] Ma S, Xu Y. Mpdiou: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.