

## Article

# Road Surface Defect Detection Algorithm Based on YOLOv8

Zhen Sun <sup>1</sup> , Lingxi Zhu <sup>1</sup>, Su Qin <sup>1</sup>, Yongbo Yu <sup>2</sup>, Ruiwen Ju <sup>1</sup> and Qingdang Li <sup>3,\*</sup>

<sup>1</sup> School of Data Science, Qingdao University of Science and Technology, Qingdao 266061, China; sunzhen@qust.edu.cn (Z.S.); 4022110102@mails.qust.edu.cn (L.Z.); 4022110072@mails.qust.edu.cn (S.Q.); f022110003@mails.qust.edu.cn (R.J.)

<sup>2</sup> College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China; 4022110093@mails.qust.edu.cn

<sup>3</sup> College of Sino-German Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China

\* Correspondence: cdiwiing@qust.edu.cn

**Abstract:** In maintaining roads and ensuring safety, promptly detecting and repairing pavement defects is crucial. However, conventional detection methods demand substantial manpower, incur high costs, and suffer from low efficiency. To enhance road maintenance efficiency and reduce costs, we propose an improved algorithm based on YOLOv8. Our method incorporates several key enhancements. First, we replace conventional convolutions with a module composed of spatial-to-depth layers and nonstrided convolution layers (SPD-Conv) in the network backbone, enhancing the capability of recognizing small-sized defects. Second, we replace the neck of YOLOv8 with the neck of the ASF-YOLO network to fully integrate spatial and scale features, improving multiscale feature extraction capability. Additionally, we introduce the FasterNet block from the FasterNet network into C2f to minimize redundant computations. Furthermore, we utilize Wise-IoU (WIoU) to optimize the model's loss function, which accounts for the quality factors of objects more effectively, enabling adaptive learning adjustments based on samples of varying qualities. Our model was evaluated on the RDD2022 road damage dataset, demonstrating significant improvements over the baseline model. Specifically, with a 2.8% improvement in *mAP* and a detection speed reaching 43 FPS, our method proves to be highly effective in real-time road damage detection tasks.



**Citation:** Sun, Z.; Zhu, L.; Qin, S.; Yu, Y.; Ju, R.; Li, Q. Road Surface Defect Detection Algorithm Based on YOLOv8. *Electronics* **2024**, *13*, 2413. <https://doi.org/10.3390/electronics13122413>

Academic Editor: José Carlos Castillo

Received: 16 May 2024

Revised: 17 June 2024

Accepted: 18 June 2024

Published: 20 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** road surface; defect detection; YOLOv8; nonstrided convolution; multiscale fusion

## 1. Introduction

The highway transportation system is an important factor in economic development and regional stability and an important part of people's production and life. In China, with the continuous development of basic economic construction, highway mileage is increasing year by year. According to 2022 statistics, the total length of asphalt roads nationwide reached 5.3548 million km at that time, representing an increase of 74,100 km compared to the previous year. Since most roads are asphalt pavement, which has poor temperature stability, they are affected by natural and human factors such as wind and sun, resulting in many defects such as potholes and cracks, which seriously affect people's transportation and economic development. By the end of 2022, the maintenance mileage of highways accounted for 99.9% of the total road mileage. As highway maintenance mileage increases, reasonable highway maintenance not only can ensure road safety, improve traffic efficiency, and reduce traffic accidents but also holds significant importance for the development of road infrastructure [1].

Early research on road surface defect detection primarily utilized digital image processing and traditional machine learning to assess road conditions. Digital image processing technology uses imaging equipment to collect high-resolution images and then designs characteristic conditions to detect pavement defects through characteristic patterns, such

as edges and textures. Common methods include threshold segmentation, edge detection, and texture analysis. Ayenu-Prah et al. [2] introduced bidimensional empirical mode decomposition (BEMD) for pavement crack evaluation. By combining BEMD with the Sobel edge detector, they effectively removed noise from crack images, enhancing edge detection. Their approach outperformed the traditional Canny edge detector, showcasing BEMD's potential for improved crack detection in pavement analysis. Wang S. et al. [3] aimed to improve the performance of automated pavement crack detection, leveraging multiscale and local optimum threshold methods, demonstrating their method's enhanced effectiveness and robustness compared to conventional segmentation approaches. Huang W. et al. [4] proposed a novel method for pavement crack detection utilizing image processing technology. They first performed image preprocessing and then connected the cracks through the proposed algorithm. Experiments showed that the algorithm is very efficient [5]. The recognition method of machine learning involves manually extracting features and then inputting these features into a classifier for classification. Lin J. et al. [6] detected road surface potholes using a nonlinear support vector machine (SVM), and experiments proved that this algorithm has a high recognition rate. Zou Q. et al. [7] developed CrackTree, a method that utilizes geodesic shadow removal and tensor voting to detect cracks in pavement images, providing a new technical approach for pavement defect detection. Li H. et al. [8] proposed a crack detection algorithm based on unsupervised methods, which effectively addresses the challenges in pavement crack detection, as demonstrated through extensive experimentation on public datasets, showing superior performance compared to existing methods. However, the above two detection methods require manual feature extraction, which increases both manpower and time costs, resulting in low detection efficiency.

As a result of the widespread adoption of deep learning [9], the accuracy and efficiency of road surface defect detection have been greatly improved. Deep learning models are capable of automatically learning and extracting features without human intervention, greatly simplifying task workflows. Compared to traditional methods, they can identify various types of road surface defects more accurately and have stronger generalization ability. Based on the steps and complexity involved in the detection process, deep learning detection methods are divided into one-stage and two-stage approaches. In the one-stage approach, effective features are automatically learned from raw data and directly utilized during the prediction process. Typical representatives are the YOLO series [10–12] and SSD [13] object detection algorithm. The two-stage model is divided into two steps: first generating candidate areas and then classifying. The typical representative is the R-CNN series [14,15] algorithm. The first-stage model is simple, efficient, and fast but has poor detection accuracy for small objects. The second-stage model has higher accuracy but has a high computational cost and slow speed. Opara J.N. et al. [16] used YOLOv3 to detect damaged asphalt pavement. Through analysis, they found that this method has high detection accuracy and provides a method for formulating road inspection procedures. Dong J. et al. [17] introduced an enhanced two-stage algorithm for segmenting cracks in asphalt pavement; the *mAP* of the improved model reaches 95.2%, significantly improving the detection efficiency of road cracks. Du F.J. et al. [18] proposed a BV-YOLOv5S algorithm, utilizing a bidirectional feature pyramid network (BiFPN) for multiscale feature fusion and varifocal loss to enhance road defect detection accuracy, which demonstrates superior performance compared to existing models and is suitable for high real-time and flexibility requirements in road safety detection projects. Xu Y. et al. [19] proposed YOLOv5-PD for asphalt pavement defect detection, integrating big kernel convolution and a channel attention mechanism for improved performance. It achieved 73.3% *mAP* and 41FPS inference speed, surpassing existing models. Huang P. et al. [20] proposed a lightweight pavement defect detection model based on an improved YOLOv7 architecture, achieving 91% average accuracy with significant reductions in parameters and computations, making it suitable for edge terminal devices. Liu Y. et al. [21] proposed an optimized road defect detection model based on YOLOv5s, enhancing speed and precision in detecting on the GRDDC

dataset while reducing the model size. Cano-Ortiz S. et al. [22] addressed the issue of scarce road defect data by proposing a model that synthesized rare defects and trained it on YOLOv5, achieving efficient road defect detection. Ranyal E. et al. [23] combined crack detection with crack width estimation and devised an efficient method based on vehicle crack detection to facilitate infrastructure maintenance, offering a comprehensive set of solutions. Zhao M. et al. [24] proposed MED-YOLOv8s, an efficient road damage detection model based on YOLOv8s with a MobileNetv3 backbone and efficient channel attention (ECA) [25], achieving a 95.2% *mAP@0.5* with 46.2% reduction in model complexity. Wang H. et al. [26] proposed the YOLOv8-D-CBAM algorithm for pavement crack detection, integrating depthwise separable convolution and a convolutional block attention module (CBAM) [27] to enhance recognition rates, achieving high precision and recall rates above 98%, with over 90% recognition for each crack category.

Although the abovementioned methods have made some improvements in enhancing the capability of road surface defect detection, it is difficult to completely cover all situations due to the complexity and diversity of pavement defects. At the same time, existing methods may face some challenges, such as poor detection of small-sized defects and insufficient robustness in complex backgrounds. In response to these issues, this paper aims to improve the accuracy, speed, and versatility of road surface defect detection by enhancing the YOLOv8 model. Here are the four main focuses of the work discussed in this article:

1. Using SPD-Conv to replace conventional convolutions in the backbone enables better capturing of the details and features of the object, thereby enhancing the performance of small object detection.
2. Utilizing the neck of the ASF-YOLO network to enhance the feature fusion method thereby improves the feature fusion capability and enhances the detection effectiveness of road surface defects.
3. Introducing the FasterNet module to transform C2f reduces unnecessary calculations and improves detection speed.
4. Employing WIoU to enhance the loss function, adjusting the importance of different samples, improves the performance and robustness of the model in the bounding box regression task.

## 2. Materials and Methods

### 2.1. YOLOv8 Algorithm

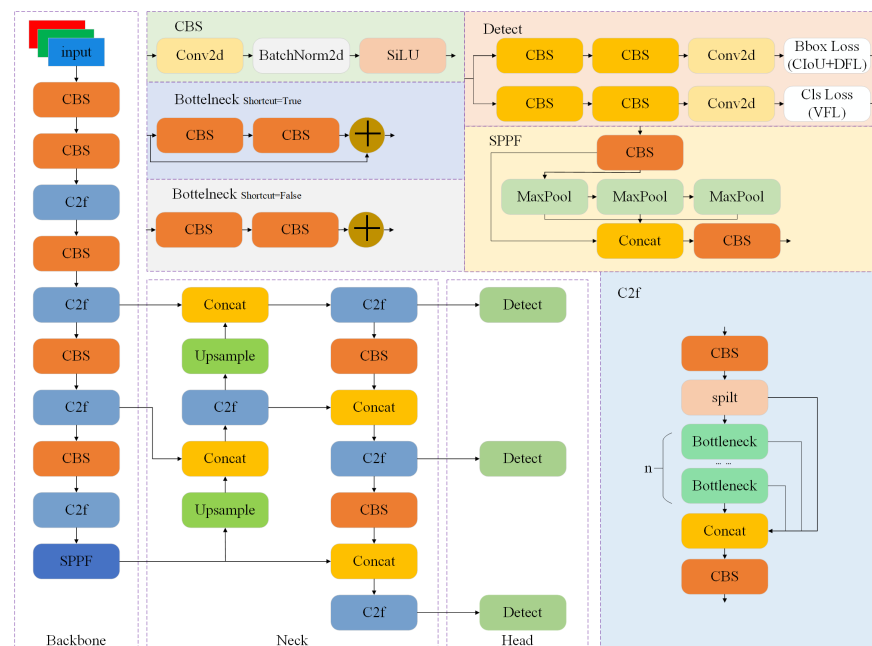
YOLOv8 is an open source model developed by Ultralytics in 2023, and its accuracy has been significantly improved compared to the previous-generation YOLOv5. It is a unified model that combines tasks such as detection, segmentation, and classification, consisting of five different depths and widths of networks, denoted as n, s, and m. Since this research is aimed at detecting defects on road surfaces and the hardware equipment is limited, YOLOv8s, which is smaller in size and higher in accuracy, is used as the basic model for improvement. Figure 1 presents the model structure of YOLOv8.

YOLOv8 comprises three primary components: the backbone, neck, and head.

The backbone consists predominantly of CBS, C2f, and spatial pyramid pooling with features (SPPF), tasked with extracting features from input images. CBS comprises two-dimensional convolution, a batch normalization layer, and a SiLU activation function and mainly performs convolution operations on feature maps. C2f is a new module proposed by YOLOv8. It draws on the design ideas of C3 and the efficient layer aggregation network (ELAN), adds multiple jump links and segmentation operations, and delivers rich gradient flow information. SPPF consists of CBS and MaxPool. Each pooling layer participates in the final splicing and extracts features of different receptive fields.

The neck is mainly composed of CBS and C2f, and it is responsible for the feature fusion of three effective feature maps passed by the backbone. The neck adopts the feature fusion idea of a feature pyramid network (FPN) [28]–path aggregation network (PAN) [29], introducing lateral connections and cascaded operations to effectively merge

features from different levels, constructing a feature pyramid that encompasses multiscale information effectively.



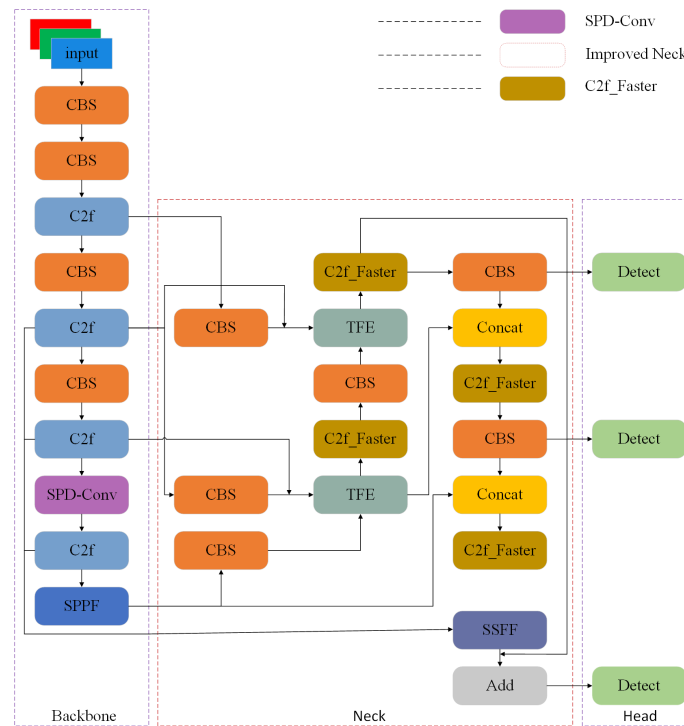
**Figure 1.** Network structure diagram of YOLOv8 algorithm.

The head is mainly composed of convolutional layers and fully connected layers, which are responsible for generating object detection results. The head adopts the design approach of a decoupled head, separating the detection task into distinct processes for handling the object class probabilities and the bounding box position information. The localization method adopts an anchor-free approach, eliminating predefined anchors and enabling direct detection on feature maps, thus reducing the complexity and cost of parameter tuning.

## 2.2. Improved YOLOv8 Algorithm

Faced with challenges such as small objects, dense distribution, and complex backgrounds in the pavement defect dataset, the original YOLOv8s has problems with false positives, missed negatives, low detection accuracy, and slow detection speed. Therefore, the following improvements were made: Firstly, by replacing conventional convolutions with SPD-Conv in the backbone network, it enhances the learning of subtle details and local features in the data, thereby enabling more accurate differentiation of different categories or scenarios. Secondly, utilizing the neck of the ASF-YOLO network as the neck of the improved network combines effective information from different scales, enabling a more comprehensive understanding of image content and more accurate localization and recognition of objects across different scales. Then, a FasterNet block is used to transform C2f, and the improved C2f\_Faster replaces C2f in the neck network, reducing redundant computation and storage access and extracting defect feature information more effectively. Finally, employing WIoU to guide parameter adjustments during the training process effectively optimizes model training and enhances its ability to handle samples of different qualities. Figure 2 presents the model structure of the improved YOLOv8.

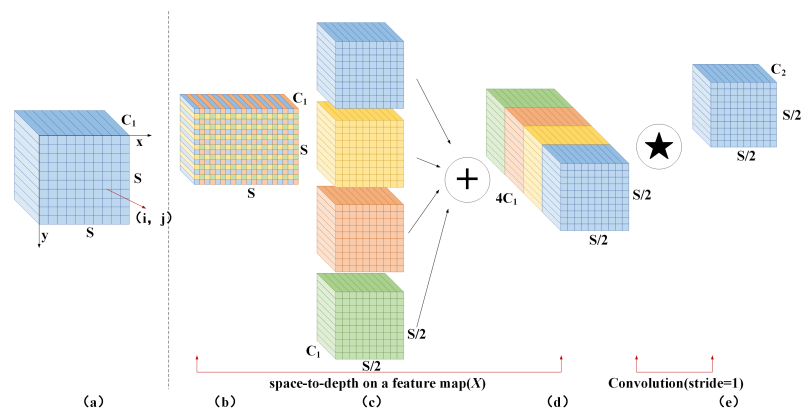




**Figure 2.** Network structure diagram of improved Yolov8 algorithm.

### 2.2.1. SPD-Conv

In convolutional neural networks, cross-row convolution or pooling operations are commonly used to reduce the size of feature maps, decrease computational complexity, and extract higher-level abstract features. However, this may lead to information loss, which can result in the model having an incomplete understanding of the input data, thus impacting the accuracy of detection tasks. To address this, Sunkara R. et al. [30] proposed SPD-Conv. Figure 3 presents the structure of the SPD-Conv module.



**Figure 3.** SPD-Conv module structure. (a) the original feature map. (b) the feature map segmented according to a certain ratio. (c) the four sub-feature maps after segmentation. (d) the feature map concatenated in the channel dimension. (e) the feature map obtained by  $1 \times 1$  convolution.

SPD-Conv comprises two components: the space-to-depth (SPD) layer and the non-strided convolutional layer. The SPD layer divides the feature graph  $X$  into a series of subfeature graphs according to a certain proportion and then splices these subfeature graphs along the channel dimension to realize the downsampling on the spatial dimension of the feature graph. The non-cross-row convolution layer carries out a step-1 convolution operation on the feature graph obtained by the SPD layer so that more information in

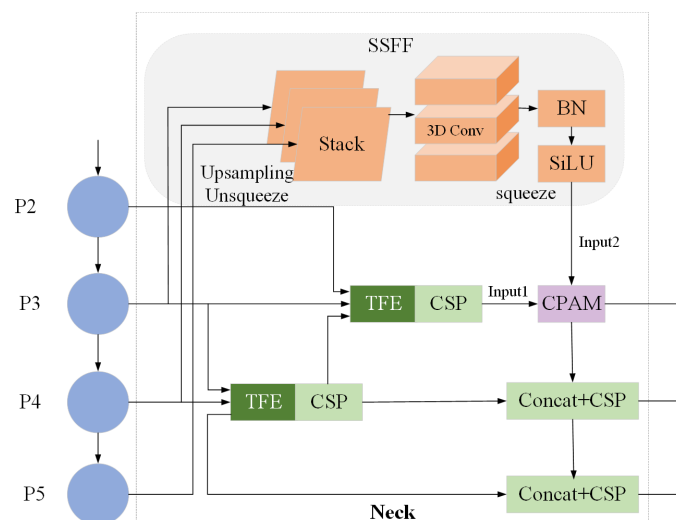
the feature graph is retained, which helps the network to learn more rich and detailed feature representation.

### 2.2.2. Improved Neck

ASF-YOLO is a YOLO framework based on attention scale sequence fusion proposed by Kang M. et al. [31]. This framework fuses scale and spatial features, enabling swift and precise segmentation tasks. ASF-YOLO proposed a new network design on the neck, as shown in Figure 4.

The ASF-YOLO model's neck comprises three components: the scale sequence feature fusion (SSFF) module, the triple feature encoding (TFE) module, and the channel and position attention mechanism (CPAM) module.

Traditional feature fusion typically involves methods like multiscale convolution to fuse features of different scales, but this often leads to redundant computation or loss of information. For this purpose, the SSFF module is proposed. SSFF enables more effective fusion of features from different scales, avoiding unnecessary redundant computations, allowing the model to comprehensively capture characteristics and contextual information of road defects. The module design is shown at the top of Figure 4. Since P3-level feature maps can provide more comprehensive and accurate information during detection, SSFF is designed based on P3. First, adjust the channel number of the two highest-level feature maps to 256, and resize their spatial dimensions to match those of the P3 level. Then, use the unsqueeze method to expand the tensor's shape and concatenate it along the depth dimension. Finally, utilize three-dimensional convolution, batch normalization, and SiLU to accomplish the feature extraction of the SSFF module. The spliced feature maps have the same resolution and different scales, so they constitute a scale sequence.

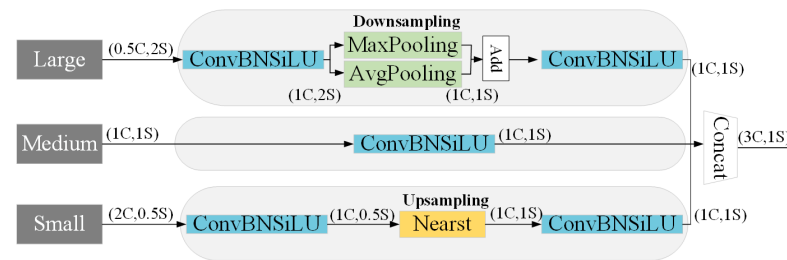


**Figure 4.** Neck network design for ASF-YOLO.

FPN typically constructs a multiscale pyramid layer by layer from top to bottom through upsampling and cascading operations during feature propagation. However, this approach may lead to the loss or blurring of fine details in the small-scale feature maps. To address this concern, the TFE module is proposed to achieve more comprehensive feature extraction. Figure 5 presents the detailed construction of TFE. TFE separates the three different-sized feature maps for individual operations. Firstly, convolutional operations are applied to the large-scale and small-scale feature maps. Subsequently, downsampling or upsampling operations are conducted. Finally, all-sized feature maps undergo another convolutional operation and are concatenated along the channel dimension.

This article focuses on detecting defects on road surfaces without the need for detailed pixel-level labels or extracting key location information from each pixel. Therefore,

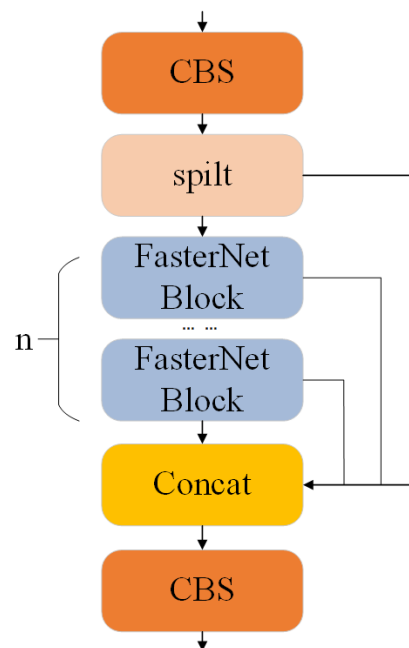
the CPAM attention mechanism is replaced with the Add operation to reduce unnecessary computational operations.



**Figure 5.** TFE module structure.

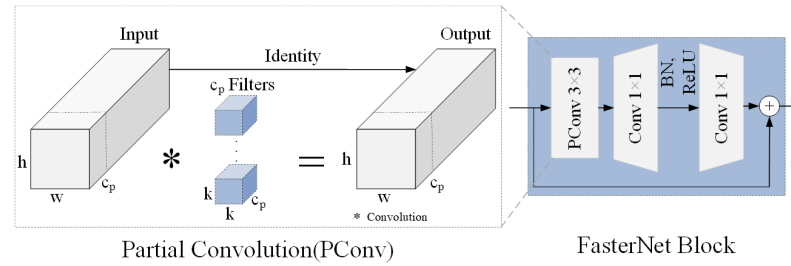
### 2.2.3. C2f\_Faster

C2f is a newly proposed module of YOLOv8, primarily composed of multiple convolution and bottleneck modules. However, multiple bottleneck stacking results in a large amount of calculation, increasing the model's complexity. To solve this problem, the C2f module incorporates the FasterNet block from the FasterNet network, forming the new C2f\_Faster module. Figure 6 presents the detailed construction of C2f\_Faster.



**Figure 6.** C2f\_Faster module structure.

In deep neural networks, each channel performs different types of feature extraction on the input data. However, in some cases, the features extracted by these different channels may be similar either in semantics or in form. To reduce computational complexity and memory access, Chen et al. [32] proposed the partial convolution (PConv) method. The left side of Figure 7 presents the PConv module. PConv selectively applies conventional convolution to some channels while keeping the others unchanged. The outputs are then concatenated, enabling spatial feature extraction while minimizing redundant computations and memory accesses. The FasterNet block primarily consists of PConv and pointwise convolution (PWConv). The right side of Figure 7 presents the FasterNet block module. Firstly, feature extraction is conducted using PConv, followed by information propagation across all channels using PWConv. Subsequently, batch normalization, ReLU activation function, and PWConv are employed for further feature extraction. Finally, the processed results are merged with the input of this module to complete the feature extraction of the FasterNet block.



**Figure 7.** PConv and FasterNet block module structure.

#### 2.2.4. Wise-IoU (WIoU)

The loss function serves as a crucial metric for model learning and optimization, and its precise definition can significantly enhance detection performance. The YOLOv8 model employs CIoU [33] as its bounding box loss function, which not only focuses on the overlap between bounding boxes but also considers their spatial positions and size relationships, thereby providing a rational approach for measuring the similarity between two bounding boxes. However, it fails to consider the potential impact of samples with varying qualities on model training. In road surface defect data, there exists a certain proportion of low-quality samples. If only high-quality samples are reinforced while neglecting low-quality ones, it will lead to the model lacking robustness when facing unseen data. Therefore, the loss function WIoU [34] based on the dynamic nonmonotonic focusing mechanism is employed as the bounding box regression loss function for our model in this paper. WIoU assesses anchor box quality based on the outlier degree and allocates more appropriate gradient weights to different-quality anchor boxes, thereby achieving more precise road surface defect detection. WIoU mainly contains three parts: the nonmonotone focusing coefficient ( $r$ ), distance loss ( $R_{\text{WIoU}}$ ), and IoU loss ( $L_{\text{IoU}}$ ). The calculation of WIoU is shown in Equation (1), and the calculation of  $r$ ,  $R_{\text{WIoU}}$ , and  $L_{\text{IoU}}$  is shown in Equations (2), (3), and (4), respectively:

$$L_{\text{WIoU}} = r R_{\text{WIoU}} L_{\text{IoU}}, \quad (1)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}}, \quad \beta = \frac{L_{\text{IoU}}^*}{L_{\text{IoU}}} \in [0, +\infty), \quad (2)$$

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

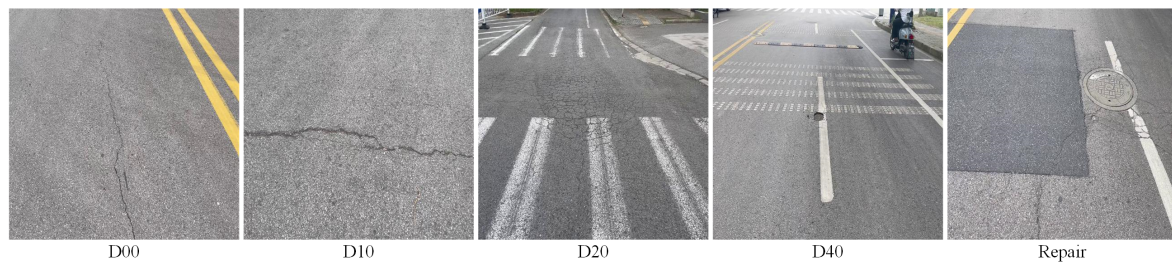
$$L_{\text{IoU}} = 1 - \text{IoU}, \quad (4)$$

In Formula (2),  $\alpha$  and  $\delta$  represent hyperparameters, and  $\beta$  represents the outlier degree of the anchor box. Outlier degree is used to measure the quality of anchor boxes, with smaller scores indicating higher quality and larger scores indicating lower quality. Assigning smaller gradient weights to high-quality anchor boxes helps ensure more reliable bounding box regression. Assigning lower gradient weights to poorer anchor boxes similarly helps alleviate the risk of model overfitting on low-quality samples.  $\overline{L_{\text{IoU}}}$  represents the exponential running average with momentum  $m$ , and  $*$  represents separation from the calculation graph. In Formula (3), the center of the anchor box is denoted as  $(x, y)$ , and the center of the object box is denoted as  $(x_{\text{gt}}, y_{\text{gt}})$ .  $w_g$  and  $H_g$  represent the width and height, respectively, of the minimum rectangle formed by the anchor box and the object box. When  $R_{\text{WIoU}} \in [1, e)$ , one can amplify the  $L_{\text{IoU}}$  of regular-quality anchor boxes. In Formula (4), IoU represents intersection over union, which is the ratio of the intersection area of the ground-truth bounding box and the predicted bounding box to their union area, indicating their overlap. When  $L_{\text{IoU}} \in [0, 1]$ , the  $R_{\text{WIoU}}$  of high-quality anchor boxes can be reduced.

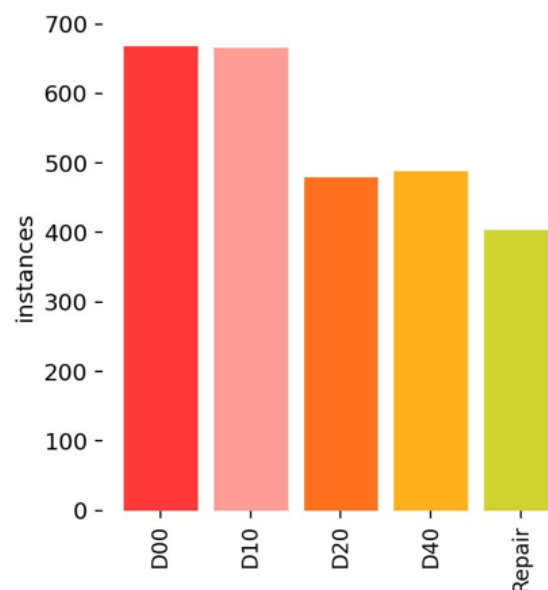
### 3. Results and Discussion

#### 3.1. Dataset

This experiment uses the public road damage detection dataset RDD2022 [35] as the experimental dataset, which collects road damage images from six countries. Based on the structure and characteristics of Chinese road surfaces, we select 2000 Chinese road damage images from the RDD2022 dataset for our study. The dataset contains five types of detection objects, including vertical cracks (D00), horizontal cracks (D10), alligator cracks (D20), potholes (D40), and repair. Figure 8 presents five types of data. The distribution of the five types of data is presented in Figure 9. The dataset was obtained in two ways, namely, by a camera mounted on a motorcycle and by a drone. The resolution of the images is  $512 \times 512$ .



**Figure 8.** Five types of detection object.



**Figure 9.** Distribution of the five types of data.

#### 3.2. Experimental Environment and Parameter Settings

The computer used in this experiment is equipped with an NVIDIA GeForce RTX 3090 graphics card, an Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz processor, 64 GB of memory, PyTorch 1.8.1 as the deep learning framework, CUDA version 11.1, and Python version 3.8.0. Table 1 presents the detailed experimental parameter settings.

**Table 1.** Training parameter settings.

Parameter	Configuration
epochs	150
batchsize	4
workers	4



Table 1. Cont.

Parameter	Configuration
imgsz	640
optimizer	SGD
lr0	0.01

### 3.3. Evaluation Metrics

Our experiments employ the following evaluation metrics to comprehensively assess the algorithm's performance across different dimensions.

Precision ( $P$ ) reflects the accuracy of the model by measuring the proportion of samples predicted as positive that truly belong to the positive class. The calculation formula is

$$P = \frac{TP}{TP + FP}, \quad (5)$$

Recall ( $R$ ) measures how many of the positive-class samples are correctly predicted as positive by the model, reflecting the model's ability to capture all positive instances. The calculation formula is

$$R = \frac{TP}{TP + FN}, \quad (6)$$

The Mean Average Precision ( $mAP$ ) quantifies the average accuracy of predictions across all classes in object detection tasks, providing a more comprehensive reflection of the model's performance across different classes. The calculation formula is:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i, \quad (7)$$

Parameters quantify the learnable elements within a model, reflecting the complexity and capacity of the model.

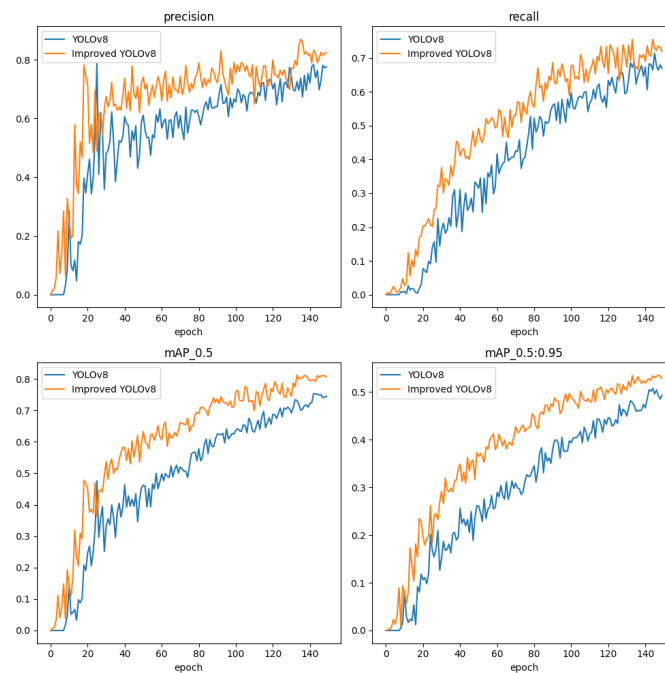
Frames per second (FPS) gauges the model's video processing efficiency, reflecting its processing speed and real-time performance.

In the above equation,  $TP$  stands for the number of correctly detected objects,  $FP$  represents the number of non-objects erroneously labeled as objects, and  $FN$  indicates the number of objects missed in detection.  $AP$  denotes the average precision at various thresholds, while  $N$  represents the total number of categories.

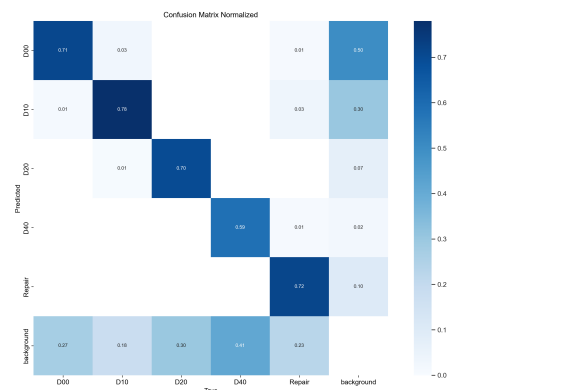
### 3.4. Experimental Process and Results

First, we convert the selected road defect dataset into the YOLO format required for training and split the dataset into a training set, validation set, and test set in a 7:2:1 ratio. Next, we use YOLOv8 as the baseline model and make improvements to YOLOv8 based on the modules introduced in Section 2.2. Finally, we train the models in the experimental environment described in Section 3.2. The training processes for both the original YOLOv8 model and the improved YOLOv8 model are shown in Figure 10. Figures 11 and 12 show the confusion matrices obtained from the detection results of the YOLOv8 model and the improved YOLOv8 model, respectively.

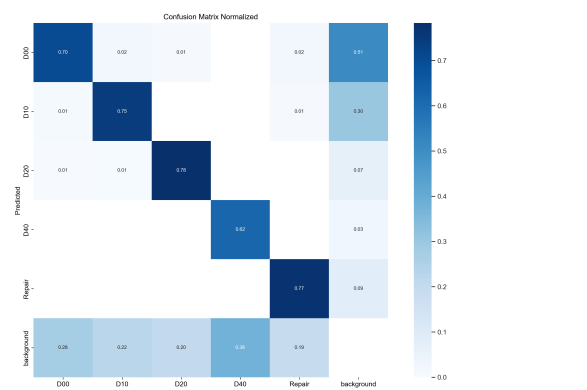
In Figure 10, the precision and recall curves of the improved YOLOv8 are higher than those of the original YOLOv8, demonstrating that the improved YOLOv8 has a significant advantage in road defect detection. In Figures 11 and 12, the values on the main diagonal of Figure 12 are generally higher than those in Figure 11, while the values in the lower-left and upper-right corners are generally lower than those in Figure 11. This indicates that the improved YOLOv8 performs better across multiple road defects.



**Figure 10.** The training processes of the YOLOv8 and improved YOLOv8 models.



**Figure 11.** Confusion matrix of the YOLOv8 model.



**Figure 12.** Confusion matrix of the improved YOLOv8 model.

### 3.5. Ablation Experiments

We conducted ablation experiments by individually integrating each improved component into the original model to fully understand their respective contributions. Table 2

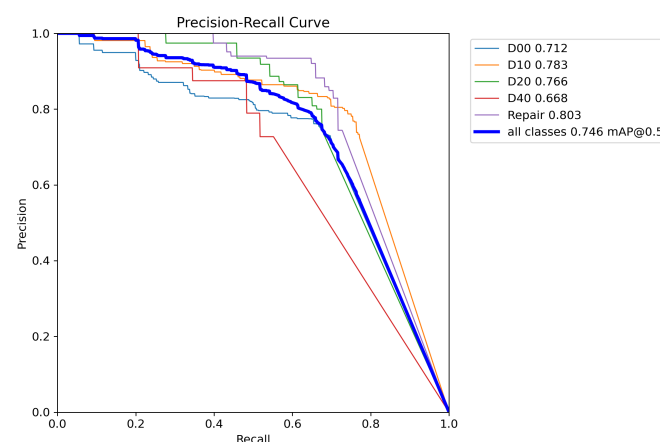
showcases the findings from the ablation experiments, in which A, B, C, and D, respectively, represent the improved YOLOv8 models with the addition of the SPD-Conv module, the improved neck, C2f\_Faster, and the WIoU loss function, while E represents the improved YOLOv8 model with all modules added, which is the model proposed in this paper.

**Table 2.** Ablation experiments.

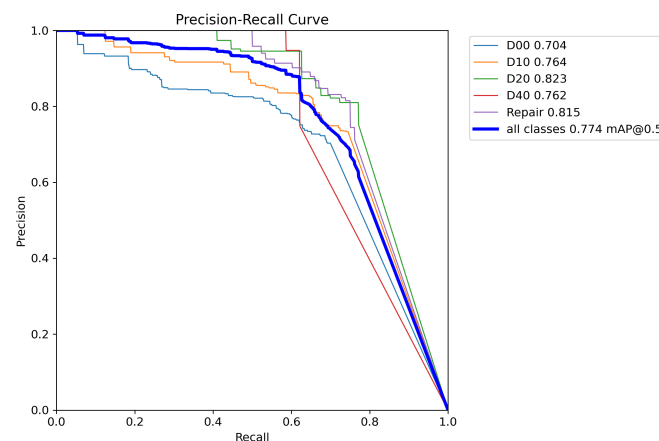
Method	SPD-Conv	Improved Neck	C2f_Faster	WIoU	P	R	mAP@0.5	Parameters/M
YOLOv8					0.754	0.667	0.746	11.13
A	✓				0.759	0.711	0.766	10.21
B		✓			0.758	0.704	0.764	10.92
C			✓		0.756	0.664	0.744	9.75
D				✓	0.733	0.701	0.758	11.13
E	✓	✓	✓	✓	0.803	0.675	0.774	8.62

As indicated by the experimental results, the A model exhibits a 2.0% improvement in  $mAP@0.5$  over the original YOLOv8 model while also boasting an 8.2% reduction in parameters, thus validating the effectiveness of the SPD-Conv module in extracting fine-grained features. For model B, in comparison to the original model, model B sees a 1.8% increase in  $mAP@0.5$ , validating that the enhanced neck effectively integrates multiscale feature mappings, providing richer and more representative feature representations for road surface defect detection. For model C, the accuracy remains nearly consistent with the original model, but the parameters have decreased by 12.4%, validating that the C2f\_Faster module efficiently extracts crucial features and optimizes the model structure. For model D, the  $mAP@0.5$  has increased by 1.2% relative to the baseline model, validating the effectiveness of the WIoU loss function in handling samples of different qualities, making the model's road surface defect detection more stable in practical scenarios. E, the method proposed by us, shows improvements over the original model with increases of 4.9% in P, 0.8% in R, and 2.8% in  $mAP@0.5$ . Additionally, parameters have been reduced by 22.6% relative to the baseline model. This validates the detection superiority of our proposed method, enabling the model to reliably identify road surface issues under various conditions, thereby contributing to enhancing road safety and maintenance efficiency.

This paper visualizes the P-R curve of both the YOLOv8 algorithm and the proposed algorithm to demonstrate the improvement of the proposed approach, as depicted in Figures 13 and 14. The horizontal axis denotes recall, while the vertical axis represents precision. The top-right corner of the graph displays the AP values for the five defect categories as well as the  $mAP$  value for all classes.



**Figure 13.** YOLOv8 P-R curve.



**Figure 14.** Improved YOLOv8 P-R curve.

From the above two figures, it can be observed that in Figure 14, the curves for the five categories mostly surpass those in Figure 13. Moreover, the area under the curve representing all categories in Figure 14 is also larger than that in Figure 13. This indicates that our method has achieved higher precision and recall rates for each category as well as overall, demonstrating superior performance in road surface defect detection tasks.

### 3.6. Comparison Experiments of Mainstream Algorithms

We conduct comparative experiments between the algorithm proposed in this paper and mainstream road surface defect detection algorithms. By comparing their performances on the same dataset, we aim to validate whether the new algorithm can achieve improvements in efficiency, accuracy, and other metrics. The experiment selected Faster-RCNN [36], SSD, YOLOv5s [37], YOLOv7-tiny, YOLOv8s [38], YOLOv9-C [39], and the improved YOLOv8 for comparison. Table 3 showcases the comparative outcomes.

**Table 3.** Comparison with mainstream algorithms.

Method	P	R	mAP@0.5	Parameters/M	FPS
Faster-RCNN	0.465	0.780	0.691	83.0	9
SSD	0.910	0.464	0.680	52.6	35
YOLOv5s	0.706	0.556	0.643	7.02	50
YOLOv7-tiny	0.606	0.338	0.271	6.01	166
YOLOv8s	0.754	0.667	0.746	11.13	38
YOLOv9-C	0.801	0.651	0.755	25.2	37
Improved YOLOv8	0.803	0.675	0.774	8.62	43

According to the table provided, it is evident that the two-stage detection model Faster-RCNN performs averagely in road defect detection, with a large number of parameters and slow processing speed, and is thus unable to promptly and effectively detect and identify road defects. In single-stage detection models, YOLOv7-tiny boasts the fastest detection speed and the least number of parameters. However, its detection accuracy is comparatively low, potentially resulting in higher rates of missed detections and false positives in road defect detection, failing to meet the requirements for precise detection. The latest detection model, YOLOv9-C, although highly accurate, has a large number of parameters, making it less suitable for deployment on mobile devices. Although the improvement algorithm in this paper may not achieve the best precision and recall rates, its overall performance is excellent, with an  $mAP@0.5$  of 0.774 and parameters totaling 8.62 M and achieving a frame rate of 43 FPS. This enables the model to effectively identify and locate road defects in practical applications, enhancing the reliability and usability of the detection system and providing robust support for road maintenance and safety management.

### 3.7. Visualization

To visually illustrate how our algorithm compares with Faster-RCNN, SSD, YOLOv5s, YOLOv7-tiny, YOLOv8s, and YOLOv9-C in localizing and identifying road defects in images, we chose three images with varied backgrounds for comparison. Figure 15 presents the detection results.



**Figure 15.** Detection result comparison. (a) Long-distance defective road surface. (b) Short-distance defective road surface. (c) Densely distributed defective road surface.

From the above figure, it can be observed that in images with distant defects, the proposed algorithm demonstrates a high recognition accuracy, validating its outstanding detection performance. In short-distance defect images, both the YOLOv5s and YOLOv8s algorithms mistakenly identify certain parts of vehicles as road defects, while other algorithms correctly identify the object category but with low detection accuracy. In contrast, the algorithm in this paper accurately identifies the defect category and achieves high detection accuracy, effectively reducing the false-detection rate. In densely distributed defect images, the SSD algorithm misidentifies normal pavement as repaired pavement, and none of these algorithms identify vertical cracks. Meanwhile, the proposed algorithm not only accurately identifies vertical cracks but also demonstrates high detection precision, significantly reducing the false-negative rate. In conclusion, our algorithm demonstrates outstanding detection capabilities across various scenarios, meeting the requirements of road defect detection tasks.

To demonstrate the effectiveness of the improved YOLOv8 in detecting various road surface defects, we compared it with the aforementioned algorithms in detecting specific types of road surface defects. The comparison of detection results from multiple algorithms is shown in Figure 16.

In Figure 16, the Faster-RCNN algorithm demonstrates average performance in detecting various types of pavement defects. It particularly struggles with D10 type defects, where it even misses detections. In comparison, the SSD algorithm improves overall detection accuracy for various pavement defects, although the improvement is not significant. The YOLOv5s algorithm performs relatively well in detecting multiple types of pavement defects, but its accuracy is lower compared to the YOLOv8s and YOLOv9-C algorithms. The YOLOv7-tiny algorithm has fewer parameters than the other algorithms, but its detection accuracy for the five types of pavement defects is relatively low, especially for D40 type defects, where it fails to detect the target defect. Both YOLOv8s and YOLOv9-C perform well in detecting the five types of pavement defects, but YOLOv9-C has an excessive



number of parameters. The algorithm proposed in this paper is based on improvements to YOLOv8s, achieving the highest overall accuracy and best detection performance for specific types of pavement defects.

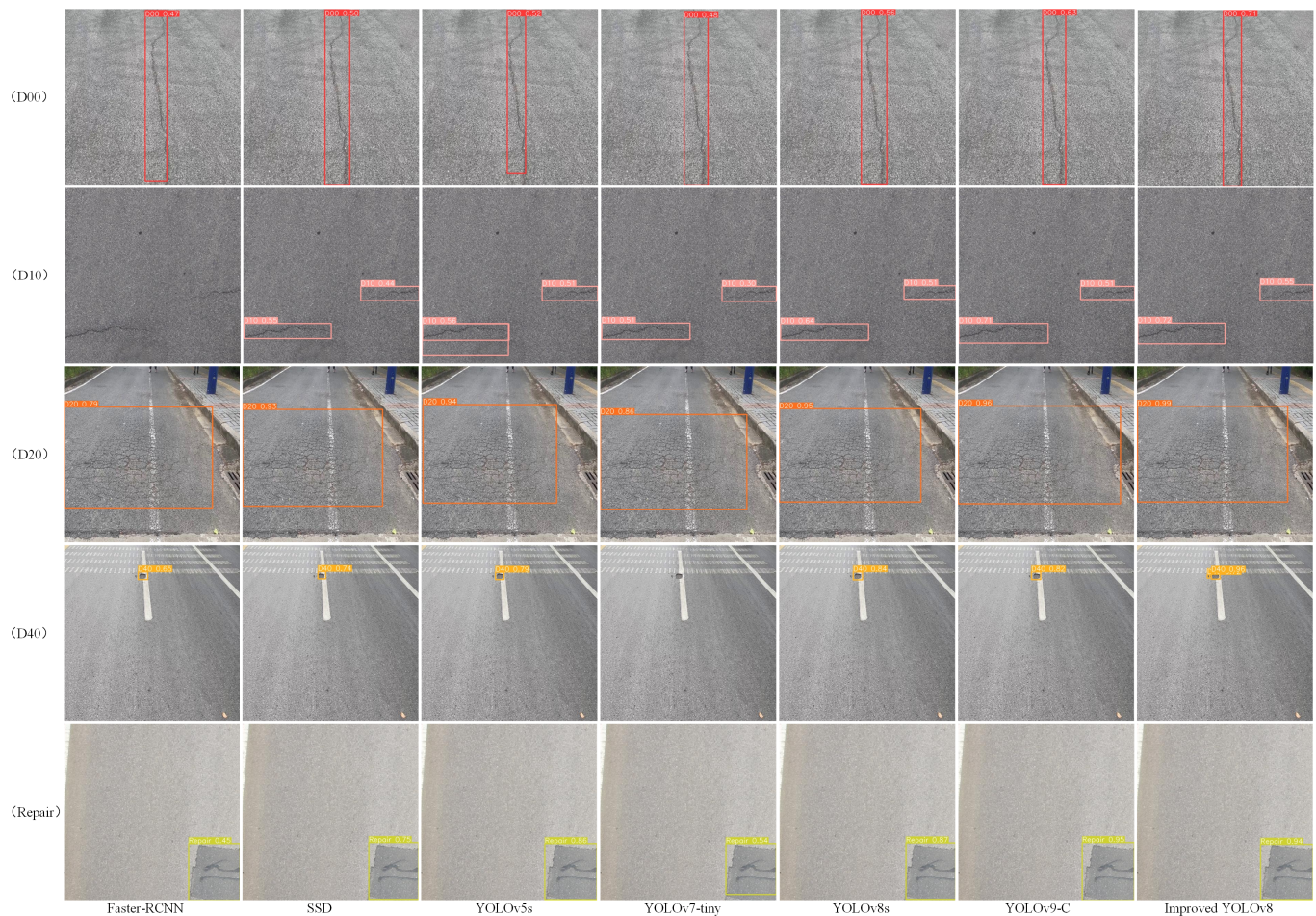


Figure 16. Detection results for different defect types.

#### 4. Conclusions

We propose an efficient algorithm for pavement defect detection, improving upon YOLOv8. Using SPD-Conv in the backbone network enhances the perception of image details, making the model more flexible and accurate in handling road surface defects in various scenarios. The ASF-YOLO network's neck is utilized to integrate features across different levels, thus better understanding and distinguishing between objects. Introducing the improved C2f\_Faster module in the neck area to adjust the network architecture reduces parameters, making the model easier to deploy and use in resource-constrained environments. The WIoU loss function is used to allocate attention to samples of varying quality, preventing overfitting, thereby making the model more robust when facing samples of different qualities and quantities. The above improvements enable the improved YOLOv8 model to more accurately capture and analyze minor issues on the road surface, enhance its ability to detect defects of various sizes and shapes, and better handle various complex road surface defect scenarios, thereby improving the efficiency and quality of road maintenance. Verified with the RDD2022 dataset, our method achieves an  $mAP@0.5$  of 77.4%, surpassing the YOLOv8 model by 2.8%, demonstrating the superiority of the improved algorithm in accuracy. Additionally, with only 8.62 M parameters, the model maintains efficiency in terms of complexity. Moreover, a detection speed of 43 FPS indicates

the model's good processing speed in real-time scenarios, which enhances its practicality in practical applications.

However, our method still has many areas that need improvement. Firstly, we have not considered poor lighting conditions, and the robustness of the model under low light or extreme weather conditions still needs to be enhanced. Additionally, our method still has many false-detection issues, which remains a significant challenge. In upcoming work, our primary goal will be to improve the model's accuracy, efficiency, and applicability. We will research how to achieve real-time monitoring and continuous learning mechanisms to adapt to diverse environments and the emergence of new types of defects, further improving the practicality and reliability of the model.

**Author Contributions:** Conceptualization, Z.S. and L.Z.; methodology, Z.S.; software, Z.S.; validation, Z.S., L.Z., and S.Q.; formal analysis, Y.Y.; investigation, Z.S.; resources, R.J.; data curation, S.Q.; writing—original draft preparation, Z.S.; writing—review and editing, L.Z.; visualization, Y.Y.; supervision, R.J.; project administration, Q.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Innovation Engineering Project of Shandong Province (grant No. 2017CXGC0607).

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Du, A.; Ghavidel, A. Parameterized deep reinforcement learning-enabled maintenance decision-support and life-cycle risk assessment for highway bridge portfolios. *Struct. Saf.* **2022**, *97*, 102221. [\[CrossRef\]](#)
2. Ayenu-Prah A.; Attoh-Okine N. Evaluating pavement cracks with bidimensional empirical mode decomposition. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 861701. [\[CrossRef\]](#)
3. Wang, S.; Tang, W. Pavement crack segmentation algorithm based on local optimal threshold of cracks density distribution. In *Advanced Intelligent Computing: 7th International Conference, ICIC 2011, Zhengzhou, China, 11–14 August 2011, Revised Selected Papers*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 298–302.
4. Huang, W.; Zhang, N. A novel road crack detection and identification method using digital image processing techniques. In *Proceedings of the 2012 7th International Conference on Computing and Convergence Technology (ICCT), Seoul, Republic of Korea, 3–5 December 2012*; pp. 397–400.
5. Zhao, K.; Liu, J.; Lv, X. A Unified Approach to Solvability and Stability of Multipoint BVPs for Langevin and Sturm–Liouville Equations with CH–Fractional Derivatives and Impulses via Coincidence Theory. *Fractal Fract.* **2024**, *8*, 111. [\[CrossRef\]](#)
6. Lin, J.; Liu, Y. Potholes detection based on SVM in the pavement distress image. In *Proceedings of the 2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, Hong Kong, China, 10–12 August 2010*; pp. 544–547.
7. Zou, Q.; Cao, Y.; Li, Q.; Wang, S. CrackTree: Automatic crack detection from pavement images. *Pattern Recognit. Lett.* **2012**, *33*, 227–238. [\[CrossRef\]](#)
8. Li, H.; Song, D.; Liu, Y.; Li, B. Automatic pavement crack detection by multi-scale image fusion. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2025–2036. [\[CrossRef\]](#)
9. Zhao, K. Study on the stability and its simulation algorithm of a nonlinear impulsive ABC-fractional coupled system with a Laplacian operator via F-contractive mapping. *Adv. Contin. Discret. Model.* **2024**, *2024*, 5. [\[CrossRef\]](#)
10. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
11. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
12. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. 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; Vancouver, BC, Canada, 18–22 June 2023*; pp. 7464–7475.
13. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016, Proceedings, Part I 14*; Springer: Cham, Switzerland, 2016; pp. 21–37.
14. Girshick, R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015*; pp. 1440–1448.
15. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*; pp. 2961–2969.

16. Opara, J.N.; Thein, A.B.B.; Izumi, S.; Yasuhara, H.; Chun, P.J. Defect detection on asphalt pavement by deep learning. *Geomate J.* **2021**, *21*, 87–94.
17. Dong, J.; Liu, J.; Wang, N.; Fang, H.; Zhang, J.; Hu, H.; Ma, D. Intelligent segmentation and measurement model for asphalt road cracks based on modified mask R-CNN algorithm. *Comput. Model. Eng. Sci.* **2021**, *128*, 541–564. [[CrossRef](#)]
18. Du, F.J.; Jiao, S.J. Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection. *Sensors* **2022**, *22*, 3537. [[CrossRef](#)] [[PubMed](#)]
19. Xu, Y.; Sun, F.; Wang, L. YOLOv5-PD: A model for common asphalt pavement defects detection. *J. Sens.* **2022**, *2022*, 7530361. [[CrossRef](#)]
20. Huang, P.; Wang, S.; Chen, J.; Li, W.; Peng, X. Lightweight Model for Pavement Defect Detection Based on Improved YOLOv7. *Sensors* **2023**, *23*, 7112. [[CrossRef](#)] [[PubMed](#)]
21. Liu, Y.; Duan, M.; Ding, G.; Ding, H.; Hu, P.; Zhao, H. HE-YOLOv5s: Efficient road defect detection network. *Entropy* **2023**, *25*, 1280. [[CrossRef](#)] [[PubMed](#)]
22. Cano-Ortiz, S.; Iglesias, L.L.; del Árbol, P.M.R.; Castro-Fresno, D. Improving detection of asphalt distresses with deep learning-based diffusion model for intelligent road maintenance. *Develop. Built Environ.* **2024**, *17*, 100315. [[CrossRef](#)]
23. Ranyal, E.; Sadhu, A.; Jain, K. Enhancing pavement health assessment: An attention-based approach for accurate crack detection, measurement, and mapping. *Expert Syst. Appl.* **2024**, *247*, 123314. [[CrossRef](#)]
24. Zhao, M.; Su, Y.; Wang, J.; Liu, X.; Wang, K.; Liu, Z.; Liu, M.; Guo, Z. MED-YOLOv8s: A new real-time road crack, pothole, and patch detection model. *J. Real-Time Image Process.* **2024**, *21*, 26. [[CrossRef](#)]
25. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 11534–11542.
26. Wang, H.; Han, X.; Song, X.; Su, J.; Li, Y.; Zheng, W.; Wu, X. Research on automatic pavement crack identification Based on improved YOLOv8. *Int. J. Interact. Des. Manuf.* **2024**, *2024*, 1–11. [[CrossRef](#)]
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
28. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
29. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
30. Sunkara, R.; Luo, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 19–23 September 2022; Springer Nature: Cham, Switzerland, 2022; pp. 443–459.
31. Kang, M.; Ting, C.M.; Ting, F.F.; Phan, R.C.W. ASF-YOLO: A Novel YOLO Model with Attentional Scale Sequence Fusion for Cell Instance Segmentation. *arXiv* **2023**, arXiv:2312.06458.
32. Chen, J.; Kao, S.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, Don't walk: Chasing higher FLOPS for faster neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2023; pp. 12021–12031.
33. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
34. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051.
35. Arya, D.; Maeda, H.; Ghosh, S.K.; Toshniwal, D.; Sekimoto, Y. Rdd2022: A multi-national image dataset for automatic road damage detection. *arXiv* **2022**, arXiv:2209.08538.
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*, 1–9. [[CrossRef](#)] [[PubMed](#)]
37. Yu, B.; Li, Q.; Jiao, W.; Zhang, S.; Zhu, Y. SAB-YOLOv5: An Improved YOLOv5 Model for Permanent Magnetic Ferrite Magnet Rotor Detection. *Mathematics* **2024**, *12*, 957. [[CrossRef](#)]
38. Shao, Y.; Zhang, R.; Lv, C.; Luo, Z.; Che, M. TL-YOLO: Foreign-Object Detection on Power Transmission Line Based on Improved YOLOv8. *Electronics* **2024**, *13*, 1543. [[CrossRef](#)]
39. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.