IAENG International Journal of Computer Science

Road Defect Detection Model Based on YOLOv8

Shengqiang Cong, Chunna Zhang*, Yang Yu, Xiaoping Yue,
Jinchi Zhao and Yuming Shen

***Abstract*—The detection of road defects is crucial for ensuring vehicular safety and facilitating the prompt repair of roadway imperfections. Existing YOLOv8-based models face the following issues: extraction capabilities and insufficient feature representation in complex scenarios, slow bounding box regression speed. To address these challenges, we propose the YOLOv8-DSW model. Firstly, we incorporate the Dilation-wise Residual (DWR) module into the C2f module in the neck, improving the detection performance of multi-scale feature maps. Secondly, We incorporate the SENetv2 attention mechanism into the neck of the YOLOv8 model to augment feature expression capabilities. Finally, we introduce Wise-IoU (WIoU) to enhance the speed of bounding box regression. Experimental results indicate that YOLOv8-DSW enhances the mean Average Precision at 50% (mAP50) by 2.5% in comparison to the YOLOv8n, significantly improving detection accuracy and provide valuable method for the development of road defect detection.**

***Index Terms*—YOLOv8; road defect detection; deep learning; attention mechanism**

I. Introduction

THE detection of road defects is crucial for preventing traffic accidents and ensuring public safety [1]. Conventional detection techniques, such as edge detection, color segmentation, and texture analysis, effectively identify cracks and potholes on road surfaces, thereby establishing a robust foundation for the advancement of road defect detection [2]. However, due to its low efficiency and significant errors, manual detection lacks practicality. The emergence of deep learning has made to groundbreaking advancements in object detection algorithms [3]. Alzraiee et al. [4] utilized Faster R-CNN to annotate pavement defects, enhancing identification confidence and thereby improving the method's practicality. Xu et al. [5] proposed a tunnel pavement detection method utilizing Mask R-CNN. The model's robustness and accuracy in defect detection and segmentation were validated through the incorporation of a feature-enhanced pyramid network (PAFPN). However, they are hindered by slow detection speeds and large model parameters, rendering them unsuitable for applications on lightweight mobile devices [6]. To mitigate this limitation, single-stage target detection algorithms have been introduced, including You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD). Du et al. [7] utilized bidirectional feature pyramid networks for multi-scale fusion in the YOLOv5s model, and Varifocal Loss to alleviate the problem of data imbalance. Yi et al. [8] introduced an enhanced YOLOv7 method by integrating the SimAM attention mechanism and Ghost modules, as well as substituting the original loss function with SIoU. Increase computational speed and reduce latency. Yan et al. [9] integrated deformable convolution into the backbone network of SSD, thereby enhancing accuracy. Due to YOLOv8's superior accuracy and speed compared to previous YOLO versions, we adopt YOLOv8 as the baseline model.

II. Related Principles

A. YOLOv8 model

YOLO is currently one of the most popular real-time object detection algorithms [10]. YOLOv8 inherits the efficient real-time detection capabilities of the YOLO series while introducing significant improvements in model architecture, training processes and multi-task handling capabilities [11]–[16]. YOLOv8 has five distinct models, each featuring different parameter sizes to accommodate various application requirements [17]. In this paper, we propose the YOLOv8-DSW, an improved YOLOv8n model. YOLOv8 references the C3 module of YOLOv5 and the ELAN module of YOLOv7, and proposes the C2f module, enhancing the ability of feature fusion [18]. Subsequently, the Spatial Pyramid Pooling - Fast (SPPF) module from YOLOv5 is incorporated, and the model parameters are fine-tuned to optimize performance [19]. YOLOv8 substitutes the conventional header with a modern decoupled header, transitioning from anchor-based to anchor-free detection method [20]. Regarding the loss function, YOLOv8 utilizes Binary Cross-Entropy (BCE) loss for classification and employs Deep Feature Loss (DFL) in conjunction with Complete Intersection over Union (CIoU) loss for regression tasks. Additionally, YOLOv8 adopts a task-aligned strategy for assigning positive and negative samples, foregoing both Intersection over Union (IoU) allocation and unilateral allocation.

B. DWR module

The DWR module is an expandable residual attention mechanism consisting of two primary components: Regional Residualization (RR) and Semantic Residualization (SR). RR emphasizes regional residualization, while SR is dedicated to semantic residualization. As illustrated in Fig. 1, these components are intricately integrated to enhance the efficient extraction of feature information and the fusion of feature maps from multi-scale sensory fields.

Manuscript received June 2, 2024; revised February 6, 2025.

Shengqiang Cong is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (email: 222085400551@stu.ustl.edu.cn).

Chunna Zhang* is an Associate Professor at University of Science and Technology Liaoning, Anshan 114051, Liaoning China (Corresponding author to provide phone: +86-130-1961-7131; e-mail: lkdzcn@ustl.edu.cn).

Yang Yu is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (email: 222085400542@stu.ustl.edu.cn).

Xiaoping Yue is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (email: 222085400571@stu.ustl.edu.cn).

Jinchi Zhao is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (email: 232085400142@stu.ustl.edu.cn).

Yuming Shen is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (email: 232085400119@stu.ustl.edu.cn).

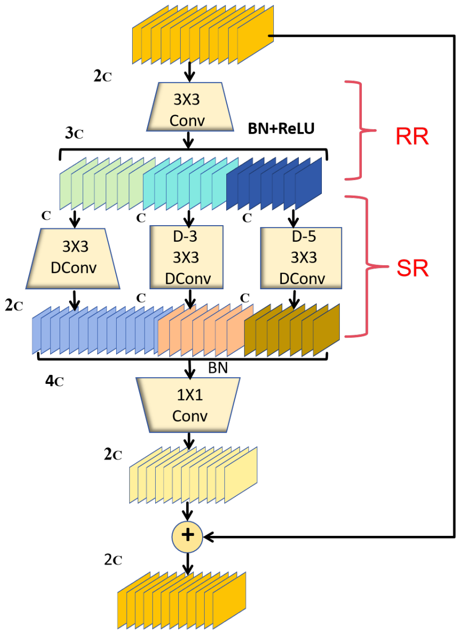


Fig. 1. DWR module diagram

Incorporating RR into deep networks addresses the diverse requirements of various receptive fields within multi-branch architectures. In each branch, a 3x3 convolution is sequentially applied for feature extraction, followed by batch normalization and ReLU activation to produce feature maps. Subsequently, SR utilizes these feature maps as filtering material and inversely aligns the receptive fields to apply an optimal receptive field for each channel. The DWR not only improves feature extraction but also mitigates redundancy in receptive fields. It transforms dilated convolution from merely extracting complex semantic information to executing morphological filtering on each concise feature map, thereby enhancing the acquisition of multi-scale contextual features.

C. SENetv2 Attention Mechanisms

In deep learning, attention mechanisms have widely applications, with the Squeeze-and-Excitation(SE) attention mechanism being particularly distinguished. The SE constructs models through squeezing and excitation operations. SENetv2 is an enhanced version of the SE attention mechanism, integrating the Squeeze aggregated Excitation (SaE) module into its original architecture. The SaE module is inspired by the inception module and incorporates a multi-branch fully connected layer of equivalent size, thereby significantly improving accuracy. The structure of the SaE module is illustrated in Fig. 2.

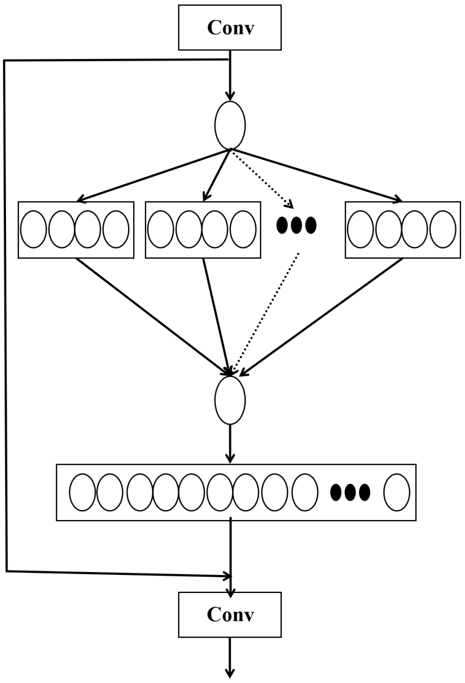


Fig. 2. Structure diagram of the SaE module

The SaE module is a core innovation of SENetV2, which enhances the traditional SE module by incorporating multi-branch fully connected layers, thereby improving the modeling of global features. In the SaE module, the input undergoes a compression operation via multi-branch fully connected layers, followed by an excitation operation that restores the features to their original dimensions. The output is then multiplied channel-wise with the input features to amplify key characteristics. The use of a lower cardinality design optimizes computational efficiency and reduces model complexity. SENetv2 integrates residual connections that transmit the original features to the output, combining them with the features processed by the SaE module. This approach helps mitigate feature loss during backpropagation.

D. WIoU

IoU is a crucial metric for assessing the performance of object detection models in computer vision, especially in tasks like object detection and semantic segmentation. It quantifies the overlap between the predicted bounding box and the ground truth bounding box.

WIoU represents an enhanced variant of IoU that employs the concept of outlier degree to assess the efficacy of anchor boxes and introduces a gradient assignment strategy. WIoU effectively mitigates the effects of low-quality anchor boxes and high-quality anchor boxes, thus shifting attention to medium-quality anchor boxes.

WIoU comprises three iterations, with WIoU v1 primarily addresses geometric measurement issues. These metrics include Euclidean distance and aspect ratio, among others, which can disproportionately influence training and exacerbate penalties for low-quality anchor boxes. WIoU v1 employs a two-layer distance-attention mechanism based on the distance metric. The formulation of WIoU v1 is detailed in Eq. (1), Eq. (2), Eq. (3), and Eq. (4).

$$\mathcal{L}_{WIoUv1} = R_{WIoU} \times \mathcal{L}_{IoU} \tag{1}$$

$$R_{WIoU} = exp(x) \tag{2}$$

$$x = (\frac{(b^{g^t}_{cx} - b_{cx})^2 + (b^{g^t}_{cy} - b_{cy})^2}{(c_w^2 + c_h^2)}) \tag{3}$$

$$\mathcal{L}_{IoU} = 1 - IoU \tag{4}$$

where$\mathcal{L}_{\mathrm{WIoUv1}}$ is the weighted IoU loss, R_{WIoU} is the weight factor, b^t_{cx} is the x-coordinate of the center of the ground truth box, b_{cx} is the x-coordinate of the center of the predicted box, b^t_{cy} is the y-coordinate of the center of the ground truth box, b_{cy} is the y-coordinate of the center of the predicted box, c_w

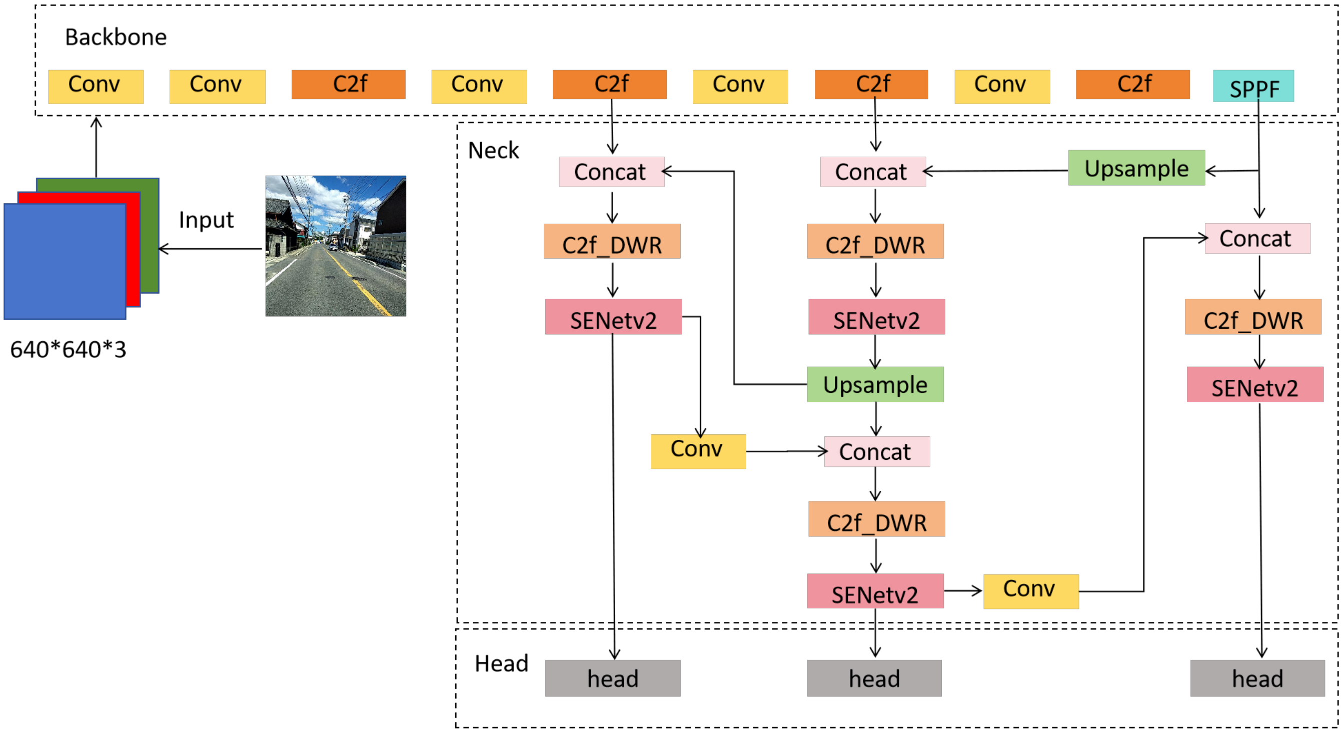


Fig. 3. YOLOv8-DSW

is the width of the smallest enclosing box that covers both the ground truth and predicted boxes, and c_h is the height of the smallest enclosing box that covers both the ground truth and predicted boxes. $\mathcal{L}_{\text{IoU}}$ is the standard IoU loss, defined as 1 minus the IoU.

WIoU v2 is an improved version of the WIoU v1 series, which introduces a static Focusing Mechanism (FM) and a dynamic normalization factor to further optimize the gradient allocation strategy for anchor boxes. WIoU v2 significantly enhances the regression accuracy and training efficiency of the model. In WIoU v2, the focusing factor decreases as IoU decreases, thereby reducing the contribution of low-quality anchor boxes to the loss function and enabling the model to focus more on high-quality and medium-quality anchor boxes. Additionally, WIoU v2 incorporates the running average of IoU (LIoU) as a dynamic normalization factor to adaptively adjust the standardization of IoU values, ensuring that the gradient allocation strategy is dynamically optimized during training. This approach effectively addresses the issue of slowed convergence in the later stages of training. By combining the FM with the dynamic normalization factor, WIoU v2 achieves faster convergence and higher localization accuracy, particularly demonstrating superior performance and generalization ability in scenarios with complex target distributions or inconsistent data quality. The formula for WIoUv2 is as shown in Eq.(5).

$$\mathcal{L}_{WIoUv2} = \mathcal{L}_{IoU}^{\gamma *}\mathcal{L}_{WIoUv}, \gamma > 0 \tag{5}$$

Specifically, $\mathcal{L}_{IoU}^{\gamma *}$ is a monotonically focusing coefficient.

In contrast, the WIoU v3 version is introduced in this paper incorporates the concept of outlier degree, by the quality of anchor boxes is assessed through the calculation of the outlier degree β. A lower outlier degree corresponds to a higher-quality anchor box. WIoUv3 assigns greater gradient gains to higher-quality anchor boxes to improve the regression performance of the bounding box. Conversely, lower gradient gains are allocated to anchor boxes with higher outlier degrees to mitigate the negative impact of low-quality anchor boxes, thereby enhancing overall model performance. The calculation of WIoU v3 is detailed in Eq. (6), Eq. (7), and Eq. (8).

$$\mathcal{L}_{WIoUv3} = r \times \mathcal{L}_{WIoUv1} \tag{6}$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \tag{7}$$

$$\beta = \frac{\mathcal{L}*_{IoU}}{\overline{\mathcal{L}_{IoU}}} \in [o, +\infty] \tag{8}$$

where α is a hyperparameter and r is the gradient gain.

III. Improved YOLOv8 Road Defect Detection Model

To address the limited feature representation and extraction capabilities in complex scenarios, as well as the slow bounding box regression speed of the current YOLOv8n model, we propose YOLOv8-DSW. First, the integration of a more efficient DWR module for multi-scale feature extraction into the C2f module within the neck of the original YOLOv8n network enhances its ability to extract features in intricate scenes. Second, the inclusion of the SENetv2 attention mechanism in the network's neck improves feature representation. Finally, we replace the CIoU loss function with the WIoU v3 loss function to accelerate regression speed. The architecture of the YOLOv8-DSW model is illustrated in Fig. 3.

A. *Boosts multi-scale detection with DWR*

To enhance multi-scale feature detection capabilities, we introduce the DWR module into the C2f module. The design of the C2f module is inspired by the C3 module in YOLOv5 and consists of two primary components: the upsampling module and the fusion module. The former adjusts low-level feature maps to align with the dimensions of high-level feature maps for subsequent fusion. The latter employs a bottleneck structure for feature fusion, typically consisting of a 1x1 convolution, a 3x3 convolution, and residual connections. The 1x1 convolution reduces dimensionality, then the 3x3 convolution extracts features. Finally, the original features are integrated with the extracted features through residual connections.

The DWR module substitutes the bottleneck module in the original architecture. It first receives the number of channels as input and initializes various convolutional layers. To align the number of channels, the input is processed through a 3x3 convolution, which reduces the channel count to half of its original value. Subsequently, three dilated convolutions with varying dilation rates are applied to the output of the preceding layer. The outputs of these convolutions are concatenated along the channel dimension and then restored to their original dimensions using 1x1 convolutional layers.

The DWR module expands the receptive field without increasing the number of parameters or computational complexity, enabling the network to capture fine-grained details alongside broader contextual information. Furthermore, the residual connections effectively facilitate cross-layer feature propagation, mitigating gradient vanishing issues and enhancing training efficiency and performance. By combining dilated convolutions with a residual connection, the DWR module efficiently aggregates multi-scale features, making it particularly well-suited for detecting targets in complex scenarios. The architecture of C2f_DWR module is shown in Fig.4.

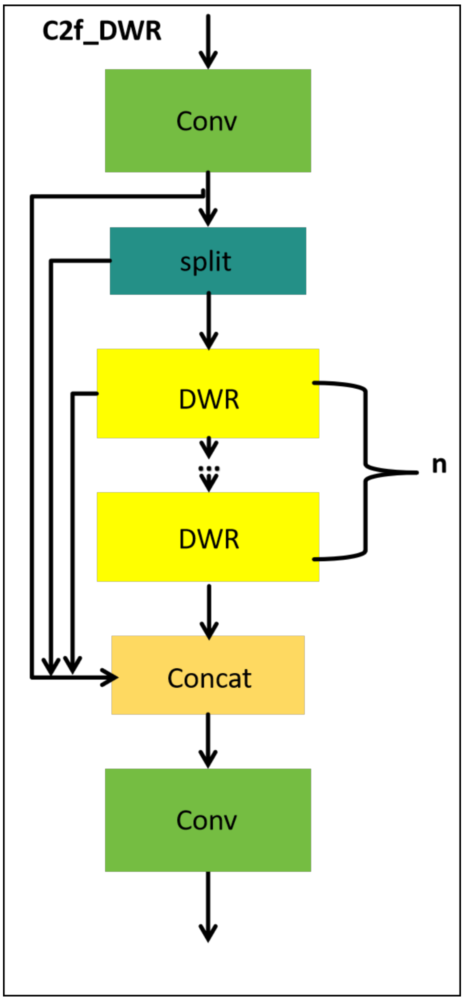


Fig. 4. C2f_DWR

B. Enhances feature expression with SENetv2

We introduce the SENetv2 attention mechanism after the C2f neck network in YOLOv8 to enhance feature representation. The core idea of SENetV2 is to enhance feature representation by applying squeeze-and-excitation operations on channel features and global features, enabling the network to focus more on key features. Additionally, SENetV2 constructs deep networks through residual connections, mitigating feature loss during the backpropagation process. The formulation of the residual connections in SENetv2 is elaborated upon in Eq. (9) and Eq. (10).

$$SENetv2 = x + F(X) \tag{9}$$

$$F(X) = F(X \cdot Ex(\sum Sq(X))) \tag{10}$$

where Sq denotes the squeezing operation and $\sum$Sq signifies the merging of multi-branched features. Ex represents the excitation operation.

The features entering the Squeeze module undergo global average pooling, converting the output into a flattened vector. This operation facilitates subsequent processing and reduces the number of parameters. The features are then passed through a multi-branch fully connected layer, further reducing their dimensions. To mitigate the impact of varying group sizes during compression, the multi-branch fully connected layer adopts a consistent topology to minimize the influence of hyperparameters. Additionally, a convolutional structure using 1x1, 3x3, and 1x1 filters is employed for feature extraction. The excitation operation learns the dependencies between channels using the full connection layer, and then applies activation functions to obtain the weights for each channel. Finally, the weighted multi-branch input is concatenated and the original shape is restored through the fully connected layer. Therefor, the incorporation of the SENetv2 attention mechanism maintains the integrity of the network's original features, enhances the detection capabilities for road defects and improves the ability to adapt across various environments.

C. Improves localization accuracy and speed with WIoU

Bounding box regression is pivotal to target detection, and enhancing its fitting capability can optimize model performance. However, indiscriminately augmenting bounding box regression for low-quality anchor boxes may negatively impact overall performance. YOLOv8 employs CIoU as the loss function for bounding box regression, incorporating an additional penalty term for aspect ratio beyond that of DIoU. While it enhances the accuracy of measuring anchor box similarity, it exerts excessive influence on the training process and does not adequately account for variations among different anchor boxes. To accelerate bounding box regression, the penalty associated with geometric factors is diminished. Consequently, in our YOLOv8-DSW, WIoU is employed in place of the original CIoU.

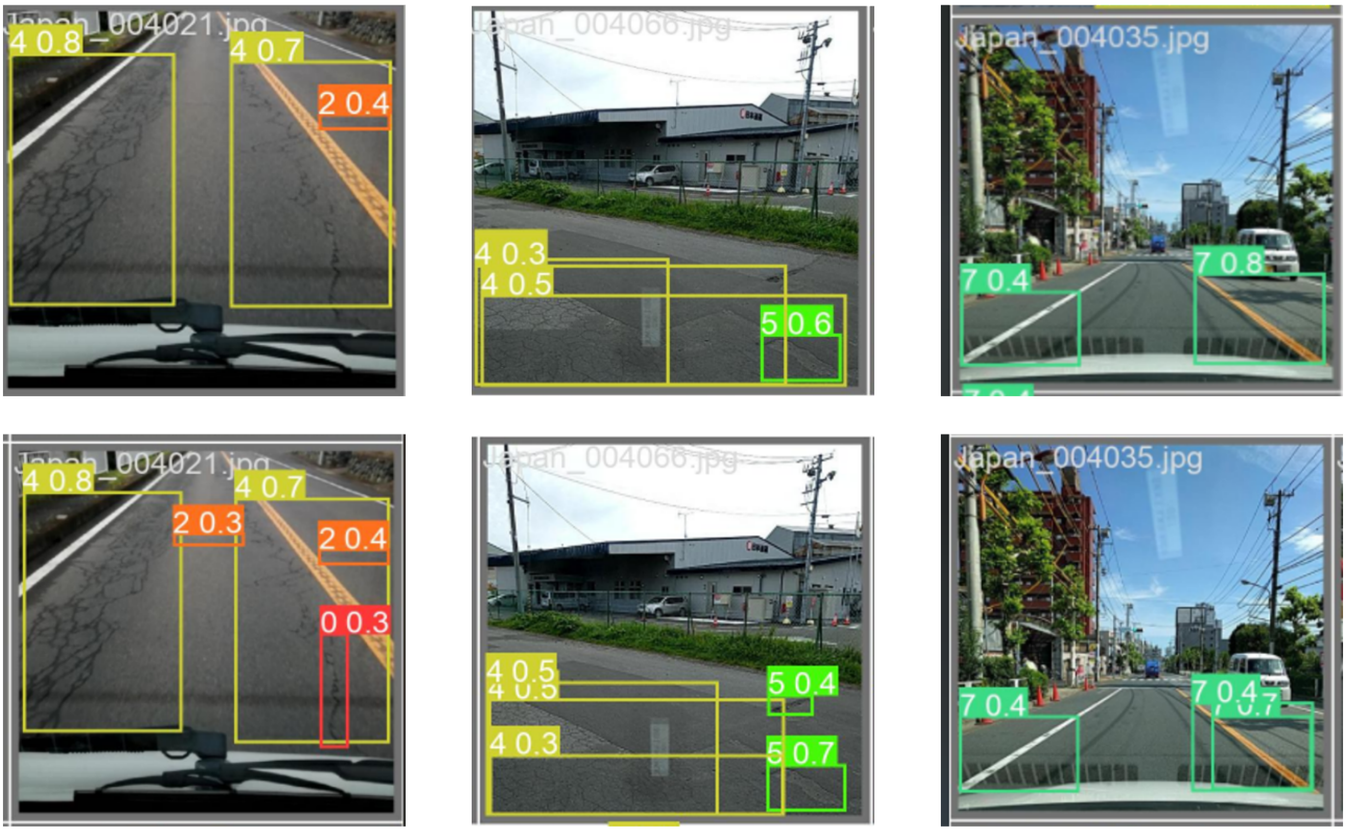


Fig. 5. Comparison of the detection effects of YOLOv8n and YOLOv8-DSW

WIoU v3 is the most advanced version of the Wise-IoU series. WIoU v3 adaptively assigns smaller gradient gains to both low-quality and high-quality anchor boxes, focusing the learning process on medium-quality anchor boxes, which are most critical for improving model performance. This approach effectively reduces the harmful gradients from low-quality anchor boxes while maximize the influence of medium quality anchor frame, resulting in a more balanced and efficient training process. The dynamic focusing mechanism allows WIoU v3 to adjust gradient allocation strategies in real-time based on the current training context. WIoU v3 significantly enhances localization accuracy and convergence speed, outperforming previous versions and other state-of-the-art bounding box regression loss functions.

IV. Experimental Analysis

A. Environment Configuration

We assess our model on the public dataset from the 2020 Global Road Detection Challenge. In this chapter, we provide a comprehensive description of the model parameters, training process, evaluation metrics, ablation experiments, and comparative studies. The hardware configuration comprises a 10GB NVIDIA GeForce RTX 3080 graphics card, PyTorch 2.0.0 as the deep learning framework, Python 3.8, CUDA 11.8, and the Ubuntu 20.04 operating system.

B. Network Training

The images are partitioned into training and validation sets in an 8:2 ratio.The maximum number of training epochs is set to 300. For practical considerations, the input images are normalized to a size of 640 × 640, which represents the largest dimension permitted for deployment on the device. The initial learning rate is established at 0.01, and the SGD optimization strategy is employed to adjust it. To ensure fairness and accuracy, both the training and ablation experiments utilize a consistent set of hyperparameters. The parameter configuration is elaborated in Table 1.

TABLE I
Parameter Table

Parameters	Setup
Epochs	300
Batch Size	16
Imgsize	640
Learning Rate	0.01
Patience	50
Optimizer	SGD
Workers	8
Weight-Decay	0.0005

C. Evaluation indicators

The experiments are conducted to assess the performance of model, utilizing precision (P), recall (R), and mean Average Precision (mAP) as the primary evaluation metrics. Accuracy and recall are computed as illustrated in Eq. (11) and Eq. (12):

$$P = \frac{TP}{TP + FP} \times 100 \tag{11}$$

$$R = \frac{TP}{TP + FN} \times 100 \tag{12}$$

where TP denotes the count of true positives, FP represents the count of false positives, and FN indicates the count of

TABLE III
COMPARISON OF DIFFERENT ATTENTION MECHANISMS

Algorithms	mAP50	mAP50-95	Params/M	GFLOPs
Yolov8n	54.5	29.1	3.0	8.2
Yolov8n+SE	54.3	28.5	3.0	8.1
YOLOv8n+CBAM	55.9	29.0	3.1	8.3
YOLOv8n+LSKA	56.0	29.1	3.3	8.9
YOLOv8n+SEv2(Ours)	56.1	29.2	3.0	8.1

false negatives.the average accuracy across n categories is computed as illustrated in Eq. (13) and Eq. (14):

$$AP = \int_0^1 P(r)d(r) \tag{13}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{n} AP_i \tag{14}$$

where AP denotes the average precision of the model at various recall rates, and N represents the number of categories.

D. Ablation experiments

To evaluate the influence of various components, this paper conducts ablation experiments. These experiments assess how different combinations of modules, such as DWR, SENetv2, and WIoU, impact the performance of road defect detection.

TABLE II
ABLATION TEST RESULTS

Algorithms	P/%	R/%	mAP50	mAP50-95
DWR	56.6	53.7	55.0	29.0
SENetv2	57.7	53.7	56.1	29.2
WIoU	57.8	55.4	55.7	29.2
DWR+SENetv2	58.3	53.3	55.8	29.0
DWR+WIoU	57.0	55.0	56.2	29.7
SENetv2+WIoU	58.7	53.4	56.5	29.3
DWR+SENetv2+WIOU	62.5	52.8	57.0	29.3

As illustrated in Table 2, the integration of DWR, SENetv2, and WIoU into YOLOv8n enhances the mAP50 to varying degrees. The highest mAP50 value is attained when all three modules are incorporated simultaneously. This enhanced model exhibits greater efficiency, with each module contributing positively to overall performance.

E. Comparative Experiments

1) Impact of different attention mechanisms on network performance: To assess the performance of SENetv2, this section compares it with other widely used attention mechanisms. The CBAM attention mechanism consists of two independent sub-modules: the channel attention module and the spatial attention module. This design reduces computational complexity and parameter requirements while enhancing feature extraction efficiency. LSKA, a large separable kernel attention module, decomposes the 2D convolutional kernel in deep convolutional layers into cascaded horizontal and vertical 1D kernels. This approach effectively mitigates the quadratic growth in computation and memory usage without compromising model performance. As shown in Table 3, SENetv2 demonstrates a more substantial improvement in detection accuracy compared to other mechanisms.

2) Classical model comparison experiments: To comprehensively evaluate the YOLOv8-DSW model. This paper conducts comparative experiments using YOLOv7, YOLOv8, and other improved models based on YOLOv8n, aiming to verify the effectiveness of the YOLOv8-DSW model. The comparison of detection effects between YOLOV8-DSW and YOLOv8n is shown in Fig.5. The mAP50 comparison between YOLOv8n model and YOLOv8-DSW model is shown in Fig.6. The comparison of mAP50 and mAP50-95 between YOLOv8-DSW and other models is shown in Fig.7 and Fig.8.

1.We reproduced the YOLOv8-RD road defect detection model by adding the BOT module after the SPPF module in the original YOLOv8 backbone network. Additionally, the model integrates the CA attention mechanism into the head section and incorporates RepGhost into the C2f module. It is specifically designed to detect four types of road defects: longitudinal cracks, transverse cracks, mesh cracks, and potholes. Detection accuracy for these defect types has been improved to varying degrees.

2. We incorporate the MSBlock from YOLO-MS into YOLOv8n for comparative experiments. The MS-Block integrates the CSP module utilized in YOLOv5 with the ELAN module from YOLOv7, focusing on extracting richer multi-scale features. It utilizes the proposed HKS protocol by employing convolutional layers of varying kernel sizes at different stages, thereby achieving an optimal balance between speed and accuracy through large kernel convolutions.

3. We integrate the information aggregation-distribution mechanism from the Gold-YOLO model into YOLOv8 for comparative experiments. The GD mechanism effectively addresses the challenge of missing information during the fusion of cross-layer features, a limitation associated with the FPN method employed in YOLO models. This approach entails constructing both low-level and high-level information aggregation-distribution mechanisms to accommodate objects of varying sizes. This integration significantly improves the information fusion capability of the neck network.

The collection and distribution process involves three distinct modules: the Feature Alignment Module (FAM), the Information Fusion Module (IFM), and the Information Injection Module (Inject). The FAM is responsible for collecting and aligning features across all levels, while the IFM integrates these aligned features to generate global information. Subsequently, the Inject module leverages this fused global information to enhance detection performance in each branch by distributing it appropriately.

4. Yolo-JD is an advanced agricultural pest detection model that excels particularly in the detection of jute diseases and represents a significant enhancement of the YOLO architecture. The model's detection accuracy is significantly im-

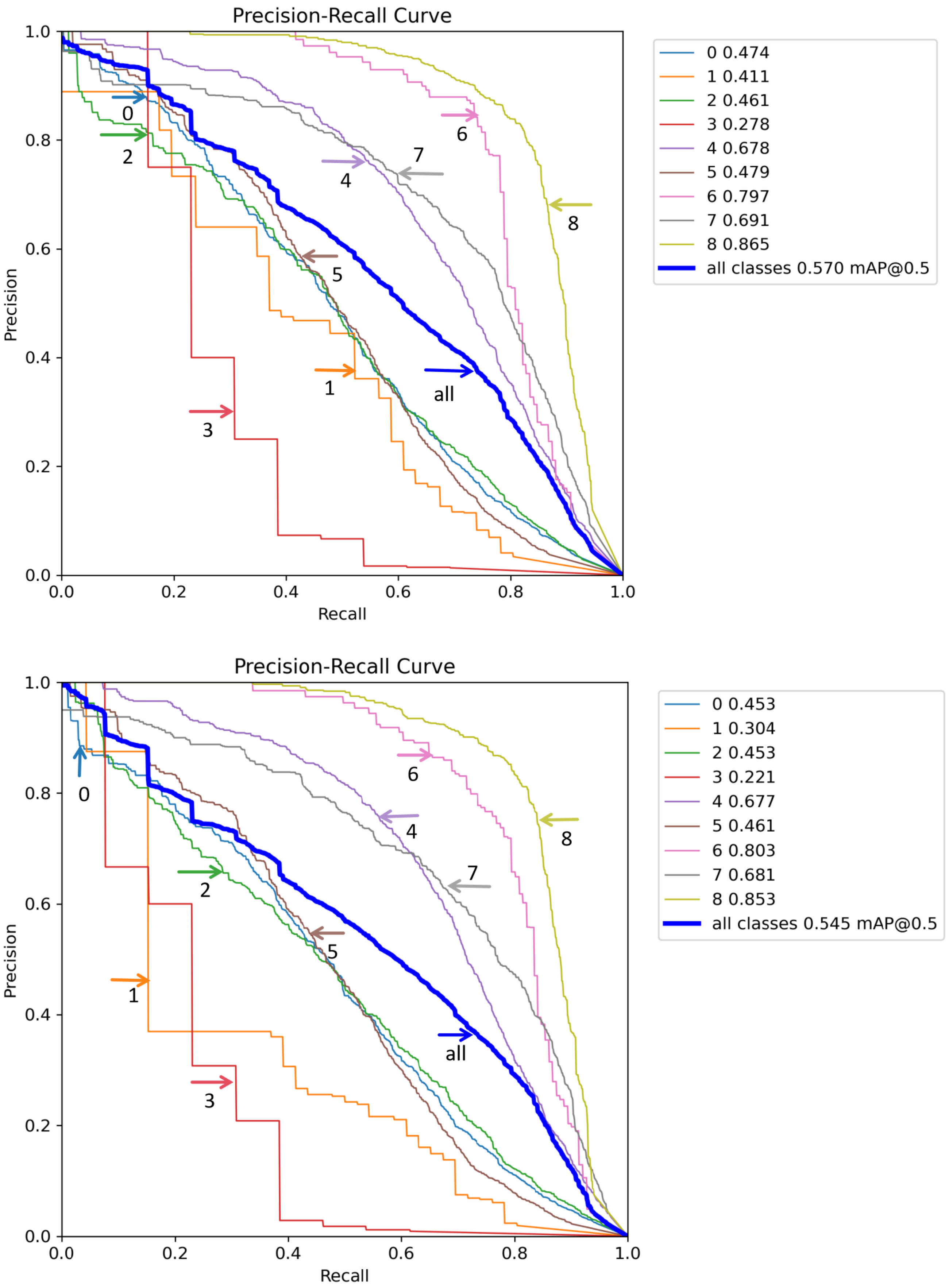
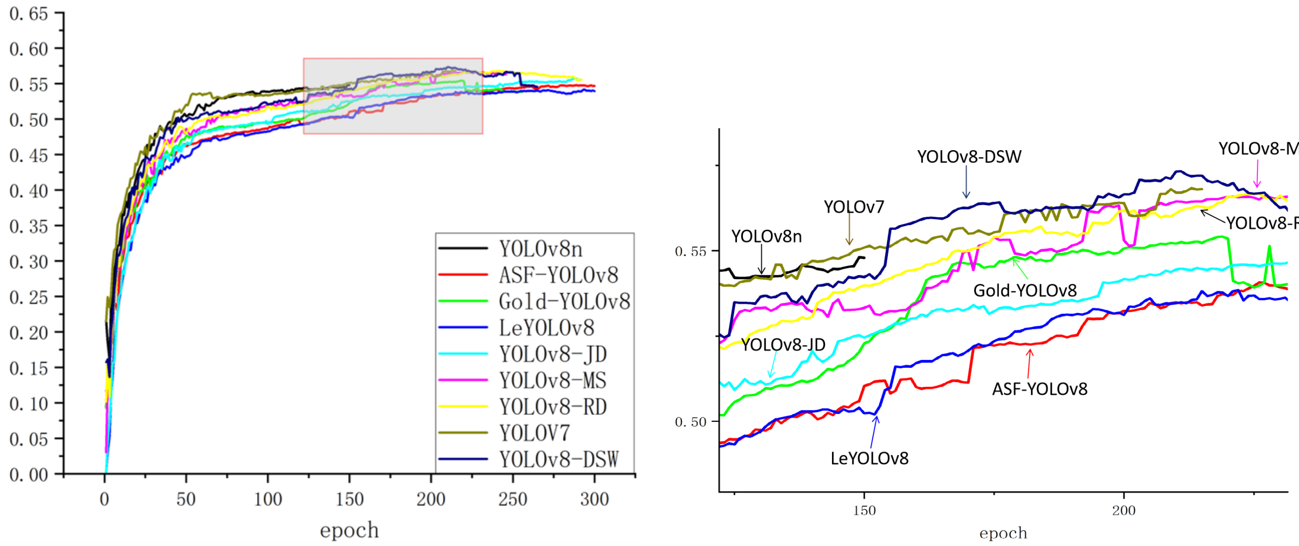


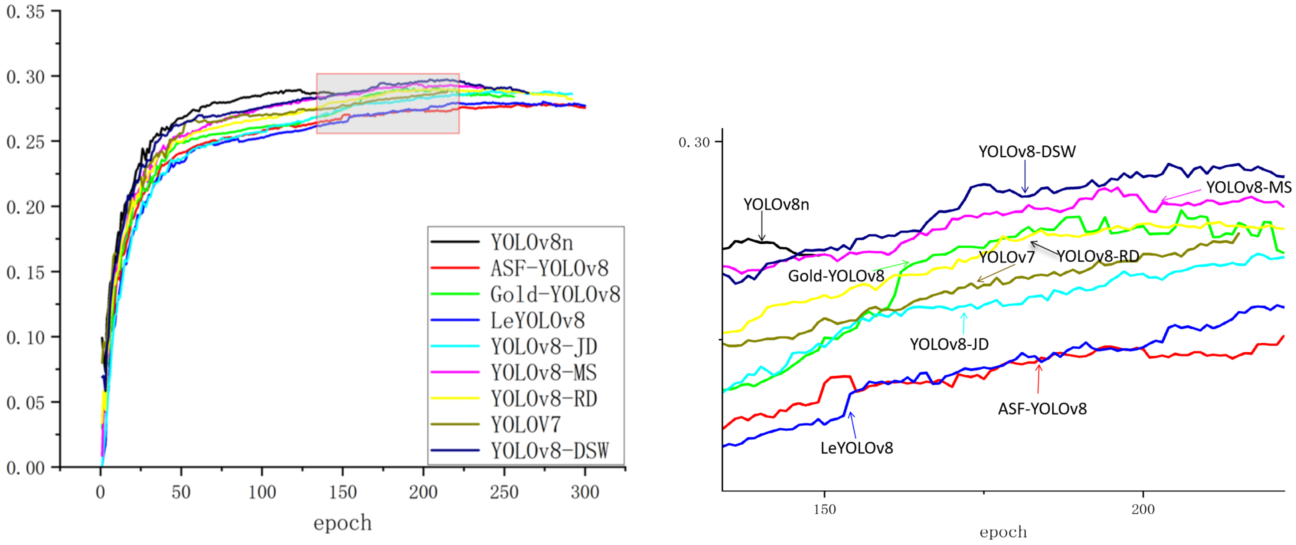
Fig. 6. Comparison of YOLOv8-DSW and YOLOv8nPR curves



(a) comparision chart of mAP@50

(b) partial enlargement of mAP@50

Fig. 7. Comparison of mAP50 models in different models



(c) comparision chart of mAP@50-95

(d) partial enlargement of mAP@50-95

Fig. 8. Comparison of mAP50-95 models in different models

proved through the integration of several custom-developed modules: the SCFEM, DSCFEM, and SPPM modules. Notably, the DSCFEM module is characterized by three SCO sand clock operations. By transforming conventional 3x3 convolutions into a spatially separable format(specifically 3x1 and 1x3 convolutions)SCO not only reduces the number of parameters but also enhances feature extraction capabilities. This paper incorporates the DSCFEM module and the SPPM module into YOLOv8 and conducts comparative experiments to evaluate its performance.

5. The LeYOLO model enhances the efficiency of object detection, particularly for embedded and mobile devices. Through efficient backbone network scaling, the Fast Pyramid Architecture Network (FPAN), and the Decoupled Network-in-Network (DNiN) detection heads, the LeYOLO model significantly reduces computational complexity while maintaining accuracy, offering an effective solution for practical applications with limited resources.

6. The ASF-YOLO model addresses the cell instance segmentation task. Through Scale Sequence Feature Fusion (SSFF) module and Triple Feature Encoder (TFE) module were integrated into the Path Aggregation Network (PANet) to fuse multi-scale feature maps. Specifically, the SSFF module improves the ability to handle objects of different sizes and orientations, while the TFE module focuses on capturing spatial information about small objects. At the same time, the Channel and Position Attention Mechanism (CPAM) enhances the model's attention to important features.

In this paper, we compare seven models with YOLOv8-DSW under the same experimental conditions. The statistics from the comparative experiments are presented in Table 4. The experimental results demonstrate that the YOLOv8-DSW model exhibits varying degrees of improvements in both mAP50 and mAP50-95 average detection accuracies when compared to the YOLOv8n model. Notably, despite integrating three distinct modules, YOLOv8-DSW does not increase the parameter count and computational complexity compared to the original YOLOv8n model. When com-

TABLE IV
COMPARISON OF DATA FROM DIFFERENT MODELS

Algorithms	mAP50	mAP50-95	Params/M	GFLOPs
LeYOLOv8	54.1	27.8	3.1	8.0
YOLOv8n	54.5	29.1	3.0	8.2
YOLOv8-DSCFEM	54.7	27.8	3.0	8.2
ASF-YOLOv8	54.8	27.8	3.0	8.6
Gold-YOLOv8	55.5	29.0	6.0	11.9
YOLOv8-RD	56.5	29.0	2.8	7.5
YOLOv8-MS	56.6	29.3	3.3	8.0
YOLOv7	56.8	29.0	37.62	105.39
YOLOv8-DSW(Ours)	57.0	29.3	2.9	8.1

pared to the road defect detection model YOLOv8-RD, YOLOv8-DSW improves mAP50 by 0.5%. Furthermore, in comparison to other YOLO enhancements, YOLOv8-DSW outperforms Gold-YOLOv8, YOLOv8-MS, YOLOv8-DSCFEM, LeYOLOv8, ASF-YOLOv8 and YOLOv7 by 1.5%, 0.4%, 2.3%, 2.9%, 2.2% and 0.2%, respectively, in mAP50, highlighting the effectiveness of YOLOv8-DSW.

V. CONCLUSION

In this paper, we propose YOLOv8-DSW, which incorporates the DWR module into the C2f module of the original YOLOv8 network. This enhancement enables more effective extraction of multi-scale contextual information. Furthermore, we integrate the SENetv2 attention mechanism to improve feature characterization. Finally, we replace the original CIoU regression loss function in YOLOv8 with the more accurate WIoU v3 loss function to accelerate bounding box regression. Comprehensive experimental results demonstrate that YOLOv8-DSW outperforms state-of-the-art models in key metrics, including precision, recall, mAP50, and mAP50-95. Our YOLOv8-DSW provides a robust approach for advancing road defect detection.

REFERENCES

[1] W. Cao, Q. Liu, and Z. He, “Review of pavement defect detection methods,” *IEEE Access*, vol. 8, pp. 14531–14544, 2020.

[2] Z. Ren, F. Fang, N. Yan, and Y. Wu, “State of the art in defect detection based on machine vision,” *Int. J. Precision Eng. Manuf.-Green Technol.*, vol. 9, no. 2, pp. 661–691, 2022.

[3] S. D. Nguyen, T. S. Tran, V. P. Tran, H. J. Lee, M. J. Piran, and V. P. Le, “Deep learning-based crack detection: A survey,” *Int. J. Pavement Res. Technol.*, vol. 16, no. 4, pp. 943–967, 2023.

[4] H. Alzraiee, A. Leal Ruiz, and R. Sprotte, “Detecting of pavement marking defects using faster R-CNN,” *J. Performance Construct. Facilities*, vol. 35, no. 4, pp. 04021035, 2021.

[5] Y. Xu, D. Li, Q. Xie, Q. Wu, and J. Wang, “Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN,” *Measurement*, vol. 178, p. 109316, 2021.

[6] M. S. Arman, M. H. Hasan, F. Sadia, A. S. Shakir, K. Sarker, and F. A. Himu, “Detection and classification of road damage using R-CNN and faster R-CNN: A deep learning approach,” in *Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2*, Springer, 2020, pp. 730–741.

[7] F. J. Du and S. J. Jiao, “Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection,” *Sensors*, vol. 22, no. 9, p. 3537, 2022.

[8] C. Yi, J. Liu, T. Huang, H. Xiao, and H. Guan, “An efficient method of pavement distress detection based on improved YOLOv7,” *Measurement Science and Technology*, vol. 34, no. 11, p. 115402, 2023.

[9] K. Yan and Z. Zhang, “Automated asphalt highway pavement crack detection based on deformable single shot multi-box detector under a complex environment,” *IEEE Access*, vol. 9, pp. 150925–150938, 2021.

[10] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, “A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[11] Z. Lin, L. Zhu, J. Zhang, Y. Zhang, and X. Liu, “Research on Improving YOLOv5s Algorithm for Defect Detection in Cylindrical Coated Lithium-ion Batteries,” *Engineering Letters*, vol. 32, no. 7, pp. 1521–1528, 2024.

[12] Y. Zhang, M. Ma, Z. Wang, J. Li, and Y. Sun, “POD-YOLO Object Detection Model Based on Bi-directional Dynamic Cross-level Pyramid Network,” *Engineering Letters*, vol. 32, no. 5, pp. 995–1003, 2024.

[13] Z. Zhang, W. Cui, Y. Tao, and T. Shi, “Road Damage Detection Algorithm Based on Multi-scale Feature Extraction,” *Engineering Letters*, vol. 32, no. 1, pp. 151–159, 2024.

[14] R. Shan, X. Zhang, and S. Li, “A Method of Pneumonia Detection Based on an Improved YOLOv5s,” *Engineering Letters*, vol. 32, no. 6, pp. 1243–1254, 2024.

[15] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.

[16] M. Hussain, “YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection,” *Machines*, vol. 11, no. 7, p. 677, 2023.

[17] X. Wang, H. Gao, Z. Jia, and Z. Li, “BL-YOLOv8: An improved road defect detection model based on YOLOv8,” *Sensors*, vol. 23, no. 20, p. 8361, 2023.

[18] H. Wang, C. Liu, Y. Cai, L. Chen, and Y. Li, “YOLOv8-QSD: An improved small object detection algorithm for autonomous vehicles based on YOLOv8,” *IEEE Transactions on Instrumentation and Measurement*, vol. 2024, p. 2513916.

[19] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, “UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios,” *Sensors*, vol. 23, no. 16, p. 7190, 2023.

[20] X. Zhai, Z. Huang, T. Li, H. Liu, and S. Wang, “YOLO-Drone: an optimized YOLOv8 network for tiny UAV object detection,” *Electronics*, vol. 12, no. 17, p. 3664, 2023.