Structural Durability & Health Monitoring

Tech Science Press

https://doi.org/10.32604/sdhm.2026.079118

ARTICLE

An Improved YOLOv11-Based Detection Method for Hidden Void and Loose Defects in Urban Road Ground-Penetrating Radar Images

Bin Chen[1,2,3,*], **Chao Qiu**[1] **and Wanli Cui**[1,2,3]

[1]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China
[2]College of Engineering, Hangzhou City University, Hangzhou, China
[3]Yangtze Delta Institute of Infrastructure, Hangzhou, China

*Corresponding Author: Bin Chen. Email: chenbin@hzcu.edu.cn

Received: 15 January 2026; Accepted: 25 March 2026; Published: 30 June 2026

ABSTRACT: Ground-penetrating radar (GPR) imaging is widely used for detecting hidden defects in urban roads. However, the complex noise environment, large-scale variations in defect features, and the sensitivity of slender defects to annotation errors pose significant challenges to accurate detection. To address these issues, this study proposes an improved object detection framework, termed DFF-MoCA-YOLO, based on YOLOv11 for identifying void and loose defects in GPR images. First, a multi-strategy gated feature fusion module (MSGFF-C3k2) is designed to enhance feature robustness against complex noise and scale variations. Then, a Monte Carlo Attention (MoCAttention) module is introduced to improve defect-feature representation via stochastic sampling and channel recalibration. Subsequently, an adaptive aspect-ratio penalty CIoU loss (CIoU-ARP) is developed to improve bounding box regression accuracy for slender defects. A labeled dataset containing void and loose defects is constructed using multi-source GPR data collected from eight urban roads. Finally, a series of ablation experiments is conducted on the proposed modules. Experimental results demonstrate that the proposed method achieves consistent performance improvements over the baseline YOLOv11 and other mainstream YOLO variants, while maintaining relatively low computational complexity. The results indicate that the proposed framework offers an effective and practical solution for detecting hidden defects in urban roads using GPR images. Moreover, the model's robustness to noise and ability to accurately detect defects at varying scales make it a promising tool for urban infrastructure maintenance. Its efficient performance with minimal computational overhead makes it suitable for real-time defect detection.

KEYWORDS: Road hidden defects; ground-penetrating radar; improvement to YOLOv11

1 Introduction

Urban roads serve as the core carriers of urban transportation networks, and the integrity of their subsurface structures is directly related to public travel safety and overall urban operational efficiency. In recent years, driven by factors such as aging and leakage of underground pipelines, insufficient subgrade compaction, long-term dynamic traffic loading, and rainwater infiltration, hidden subsurface defects—including voids, debonding, and looseness—have become increasingly prevalent. Owing to their strong concealment and high suddenness, these defects pose significant challenges to road safety maintenance and management [1].

Ground-penetrating radar (GPR), owing to its high sensitivity to physical property variations in subsurface media, has become a core nondestructive testing technique for detecting underground defects in

urban roads. In particular, three-dimensional GPR has been widely applied to on-site data acquisition for detecting defects such as voids and looseness, thanks to its superior detection accuracy and high operational efficiency [2]. However, defect identification in three-dimensional GPR images still relies predominantly on manual interpretation. This process is not only labor-intensive but also highly dependent on inspectors' experience, leading to low recognition efficiency and high misjudgment rates. Consequently, it is difficult to meet the engineering demands of large-scale urban road defect screening [3].

The application of GPR in urban infrastructure inspection also relies heavily on efficient survey and data analysis methodologies. Traditional nondestructive testing approaches often suffer from high operational costs, low efficiency of manual inspection, and difficulties in interpreting complex internal structural images [4]. Recent studies and comprehensive reviews indicate that although conventional manual interpretation remains time-consuming and subjective [5], the integration of machine learning—particularly deep learning techniques—can significantly enhance the extraction of complex features from radar waveforms and images, thereby improving defect localization accuracy [5,6]. Moreover, applying these algorithms to GPR data has demonstrated considerable potential for rapidly detecting voids and structural defects in real-world engineering scenarios [7].

1.1 Research Status

Automated defect identification from GPR images has become a prominent research topic in road structural health monitoring. Within the YOLO (You Only Look Once) detection framework, Liu et al. [8] improved YOLOv3 by introducing four-scale detection layers, multi-scale feature fusion, the EIoU loss, and K-means++ clustering, significantly enhancing crack detection accuracy and reducing missed detections of small defects. Wang et al. [9] systematically evaluated multiple YOLOv5 variants with different detection scales for hidden asphalt pavement defect detection in GPR images, demonstrating that increased model complexity does not necessarily yield better performance and that three-scale detection layers are often more suitable than four-scale configurations. Zhang et al. [10] proposed an RFS-YOLO–based framework incorporating receptive field attention and channel attention mechanisms to enhance defect feature focusing while maintaining computational efficiency, and further improved model interpretability using Grad-CAM. Li et al. [11] designed a convolution–attention fusion module for YOLOv11 and combined it with a C2PSA mechanism to strengthen feature extraction from 3D C-scan images, achieving a favorable balance between detection accuracy and real-time performance. Ma et al. [12] applied an EC-YOLOv7 network for defect localization and heatmap generation, showing strong generalization capability under complex interference conditions.

To further address the challenges posed by dense or noisy scenes, the integration of attention mechanisms into the YOLO architecture has emerged as a prominent research direction. Numerous studies have demonstrated that modules such as the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation (SE) networks can significantly enhance feature representation [13,14]. For example, incorporating lightweight SE and CBAM modules into YOLOv7 has been shown to effectively suppress background noise and improve the real-time detection performance for small targets [15]. Similarly, introducing dynamic spatial and channel attention mechanisms into YOLO-based frameworks has successfully reduced missed detections in resource-constrained or complex environments [16,17]. Building on this paradigm, stochastic and dynamic attention mechanisms are now being actively explored to mitigate overfitting to fixed defect patterns and to further improve the robustness of detection models.

Beyond the YOLO series, other deep learning architectures have also been explored. Gao et al. [18] enhanced Faster R-CNN with a deep feature selection network to improve GPR image feature extraction, enabling automatic detection of underground pipelines and uneven settlement. Shin et al. [19] employed

EfficientDet-D3 with compound scaling to achieve efficient pavement void detection, demonstrating a good balance between accuracy and inference speed. Wang et al. [20] proposed GPRI2Net, which integrates DenseUnet and Bi-ConvLSTM to exploit contextual information from long B-scan survey lines, enabling joint dielectric constant inversion and target classification with reduced computational complexity. Hu et al. [21] combined Faster R-CNN with an attention-guided feature pyramid network to accurately identify pipeline locations, while Liu and Zhu [22] developed a Mask R-CNN–based approach for automatic recognition and precise localization of road-void defects in GPR images, achieving performance superior to that of mainstream detection models.

Alongside advances in feature extraction, the optimization of bounding box regression through advanced loss functions has been critical for improving target localization accuracy. The evolution from standard Intersection over Union (IoU) to Distance IoU (DIoU), and Complete IoU (CIoU) has progressively addressed issues such as gradient vanishing and inaccurate shape constraints [23]. Recent comparative analyses indicate that carefully selecting the appropriate IoU variant and re-examining its mathematical relationship with evaluation metrics can directly influence the overall performance of object detectors [24]. Furthermore, researchers have introduced dynamic distance penalties and refined corner-based IoU losses to enhance convergence speed and localization precision [25,26]. Despite these advances, conventional IoU-based loss functions often remain highly sensitive to annotation bias, particularly for elongated targets, underscoring the need for more adaptive aspect-ratio penalty strategies in GPR defect detection.

1.2 Limitations of Existing Studies and Innovations of This Work

Despite these advances, several limitations remain in existing studies, particularly in the targeted discrimination of visually similar defect types (e.g., voids and looseness), the availability of high-quality annotated GPR datasets, and the robustness of detection models under strong noise, scale variation, and elongated-defect annotation sensitivity—issues that motivate the present study:

1. A multi-strategy alternating feature fusion module, termed MSGFF-C3k2, is designed. By integrating ensemble, robust, and consistency-oriented DFF strategies in an alternating manner, the proposed module jointly addresses strong noise interference, large-scale variation, and distribution shift, thereby enhancing the robustness and generalization capability of feature extraction.
2. A Monte Carlo Attention (MoCAttention) module is introduced to collaboratively enhance defect-region feature representation through stochastic sampling and channel recalibration. This design compensates for the limited noise-filtering capability of the feature-extraction stage and jointly optimizes feature focusing and noise suppression.
3. An adaptive aspect-ratio–penalized CIoU loss function (CIoU-ARP) is proposed to specifically address the sensitivity of elongated defect annotations, thereby improving bounding-box regression accuracy.

1.3 Research Objectives and Scope

1.3.1 Research Objectives

To address the challenges encountered in underground defect detection from GPR images—including strong noise interference, large variations in feature scale, sensitivity to annotation bias for elongated defects, and insufficient model generalization—this study aims to develop a DFF-MoCA-YOLO algorithm that jointly balances detection accuracy, generalization capability, and computational efficiency. The proposed method enables fast and accurate identification of underground voids and loose defects in urban roads, providing technical support for road structural health monitoring.

1.3.2 Research Scope

To achieve the above objective, the following research tasks are carried out:

1. High-quality dataset construction: Based on GPR inspection projects conducted on eight urban roads in Binjiang District, Hangzhou, real-world data of void and loose defects are collected. Through annotation and format conversion, a YOLO-formatted dataset is constructed, including key defect information such as location, burial depth, and planar size. The dataset is explicitly divided into training, testing, and validation sets to ensure data authenticity and reliability.
2. Design of the MSGFF-C3k2 module: To cope with strong noise, large-scale variation, and distribution shift in GPR images, a multi-strategy alternating feature fusion module, termed MSGFF-C3k2, is designed to replace the original C3k2 module in YOLOv11. By alternately integrating three DFF strategies, the module enhances the robustness and generalization capability of feature extraction.
3. Integration of the attention mechanism: A Monte Carlo Attention (MoCAttention) module is introduced after the large-object detection layer of YOLOv11n. By combining stochastic sampling and channel recalibration, the defect-region feature representation is strengthened, enabling effective feature focusing and noise suppression.
4. Loss function optimization: An adaptive aspect-ratio–penalized CIoU loss function (CIoU-ARP) is proposed. By dynamically adjusting the aspect-ratio penalty for elongated defects, the method improves bounding-box regression accuracy and alleviates sensitivity to annotation bias.
5. Model performance evaluation: Ablation experiments are conducted to verify the effectiveness of each proposed module. Comparative experiments with mainstream detectors, including YOLOv5–YOLOv13, are performed to comprehensively evaluate the proposed method in terms of precision, recall, mAP50, mAP50-95, and computational cost (GFLOPs and parameter count).

2 Methodology

2.1 YOLOv11 Network Architecture

The YOLOv11 model adopts a four-stage architecture consisting of input processing, backbone feature extraction, neck feature fusion, and detection head prediction. The overall network structure is shown in Fig. 1.

At the input stage, the YOLOv11 input module supports images of varying resolutions. Through preprocessing operations such as image resizing and normalization, the input data are adjusted to meet the model requirements. This flexible input adaptation mechanism not only broadens the model's applicability to multi-scale images but also ensures consistent data processing.

The backbone network is the core component for feature extraction and adopts a multi-scale feature extraction strategy. The C3k2 module is designed around an optimized Cross Stage Partial (CSP) bottleneck structure, in which large convolution kernels are decomposed into two smaller convolutions to improve computational efficiency. In addition, while retaining the Spatial Pyramid Pooling Fast (SPPF) module, the network introduces the C2PSA module to enhance spatial attention modeling of feature maps.

The neck network is responsible for aggregating and transmitting features. YOLOv11 upgrades the C2f module in YOLOv8 to the C3k2 module, significantly enhancing feature aggregation. Meanwhile, the introduced C2PSA module further optimizes the spatial attention distribution, effectively improving the recognition accuracy of target regions.

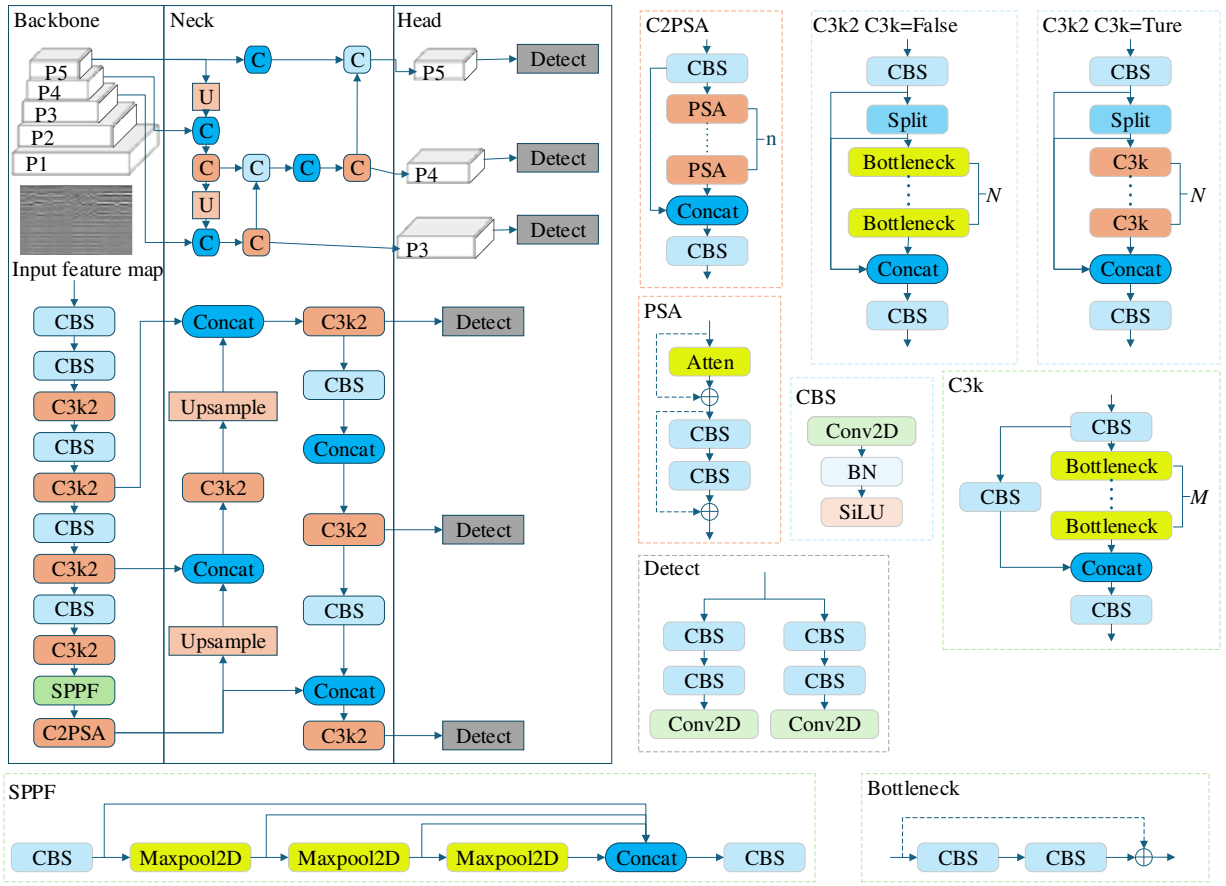

Figure 1: Architecture of the YOLOv11 network.

The detection head converts the fused features from the neck network into final predictions of object locations and categories. This module employs multiple adjustable C3k2 blocks to refine feature representations and incorporates lightweight depthwise separable convolutions (DWConv). While maintaining detection performance, this design significantly improves computational efficiency.

Overall, by systematically optimizing the input processing, feature extraction, feature fusion, and prediction stages, YOLOv11 achieves a favorable balance between detection accuracy and computational efficiency compared with previous YOLO variants.

2.2 Improvements to YOLOv11

Considering the characteristics of GPR images for urban road void and loose defect detection—such as strong noise interference, significant scale variations of defect features, and high sensitivity of slender defects to annotation deviations—the baseline YOLOv11 model exhibits limitations in generalization ability, feature focusing accuracy, and bounding box regression adaptability. To address these issues, this study proposes an improved detection framework, termed DFF-MoCA-YOLO, that systematically enhances YOLOv11 across three aspects: feature extraction, attention mechanisms, and loss function optimization. The overall network architecture is illustrated in Fig. 2.

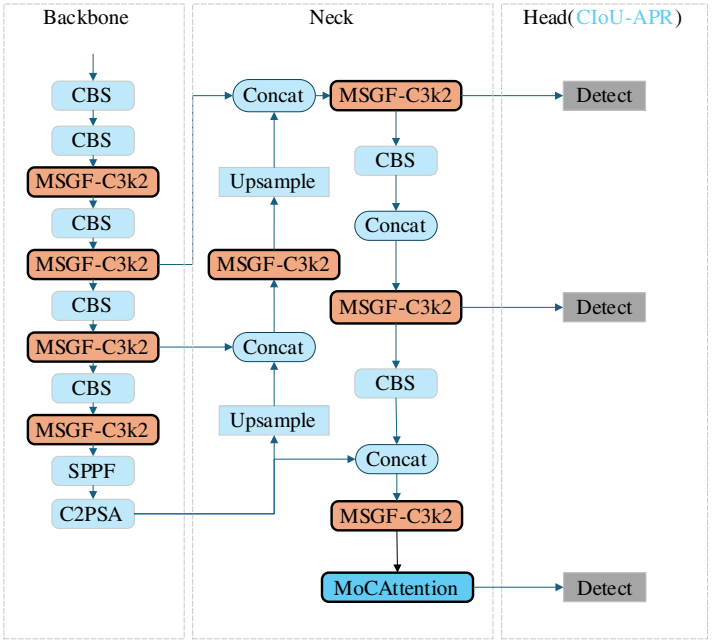

Figure 2: Architecture of the DFF-MoCA-YOLO network.

The specific improvements are summarized as follows:

1. A multi-strategy gated feature fusion module (MSGFF-C3k2) is designed to enhance the robustness of feature extraction under complex GPR image conditions.
2. A Monte Carlo Attention (MoCAttention) module is introduced to strengthen defect-feature representation by combining stochastic sampling and channel recalibration.
3. An adaptive aspect-ratio penalty CIoU loss function (CIoU-ARP) is proposed to improve bounding box regression accuracy for slender defect targets.

These modules collaboratively enhance feature representation, attention allocation, and loss optimization, forming an efficient detection framework tailored for GPR-based urban road defect detection.

2.3 MSGFF-C3k2 Module Design

To address the problems encountered in GPR–based road subsurface distress detection, including strong noise interference in radar images, large variations in defect feature scales, and training–testing distribution shifts, this paper proposes an MSGFF-C3k2 module based on multi-strategy alternating feature fusion (as shown in Fig. 3), which is used to replace the native C3k2 module in the YOLO11 backbone network. The core design objective of this module is to enhance the robustness and generalization capability of feature extraction without compromising the stability of the original model, thereby improving the detection accuracy of voids and looseness in radar images.

The MSGFF-C3k2 module inherits the lightweight "split–process–fusion" architecture of the C3k2 module. Its main innovation lies in replacing the original Bottleneck with a generalization-enhanced bottleneck block that alternately integrates three types of DFF (Diverse Feature Fusion) strategies, namely ensemble, robust, and consistency strategies [27]. The specific structure consists of three key components:

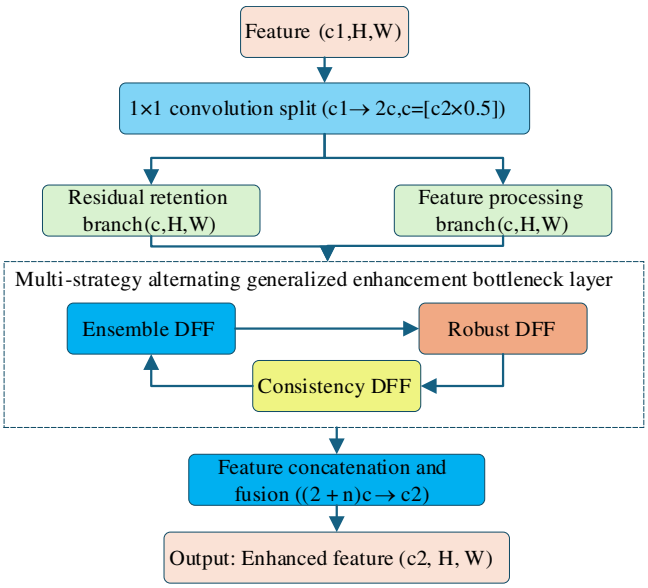

Figure 3: Architecture of the MSGFF-C3k2 module.

Input feature splitting layer: A 1 × 1 convolution is used to divide the input feature channels of the three-dimensional radar image into two branches. One branch is reserved for residual connections, while the other enters the multi-strategy bottleneck layer for feature enhancement, effectively avoiding loss of feature information.

Multi-strategy alternating bottleneck layer: This layer is composed of n generalization-enhanced bottleneck blocks. Inside each bottleneck block, the operations of "1 × 1 convolution for dimensionality reduction—3 × 3 convolution for feature extraction—corresponding strategy-based DFF enhancement—residual fusion" are sequentially performed. In particular, the bottleneck blocks are alternately arranged in a cyclic order of "ensemble DFF → robust DFF → consistency DFF", which not only ensures adaptability to noise and scale variations in radar images but also avoids overfitting caused by a single strategy.

Adaptive feature fusion layer: The residual-preserved branch is concatenated with the output features of all bottleneck blocks, and a 1 × 1 convolution is applied to restore the target number of channels, yielding the final enhanced feature output.

For defect detection in three-dimensional radar images, the three DFF strategies are designed with distinct emphases and form a functionally complementary framework, as detailed below.

The ensemble DFF focuses on comprehensive multi-scale defect-feature capture by constructing parallel branches with different receptive fields. Specifically, a local detail branch employs 3 × 3 convolutions to extract fine-grained information, such as defect boundaries and textures; a global context branch uses adaptive average pooling to capture the overall spatial distribution and contextual relationships of defects; and a medium receptive field branch employs 5 × 5 grouped convolutions to balance local and global representations. These branches are subsequently fused via adaptive weighted aggregation, with the fusion weights dynamically generated from the input features. This design enables accurate adaptation to the wide-scale variations of void and looseness defects in radar images, ranging from centimeter-level localized anomalies to meter-level extended regions, thereby avoiding feature omission caused by a single receptive field.

The robust DFF is specifically designed to address pervasive device noise and environmental interference in radar point-cloud projection images, including surface clutter reflections and underground pipeline

disturbances. It introduces multi-scale dilated convolutions with dilation rates k = 1 and k = 2 to expand the receptive field while preserving feature resolution. In addition, channel shuffling is employed to break rigid inter-channel dependencies, reducing noise accumulation during feature propagation and significantly improving the model's tolerance to complex backgrounds.

The consistency DFF targets mitigating training–testing data distribution shifts. A 3 × 3 smoothing convolution is applied to suppress feature fluctuations caused by high-frequency noise, followed by runtime mean calibration, where feature means are dynamically updated during training and fixed during inference. This strategy alleviates feature distribution discrepancies induced by variations in road materials (e.g., asphalt vs. concrete), climatic conditions (e.g., rainy vs. dry seasons), and acquisition device parameters, thereby ensuring stable detection performance in unseen scenarios.

Meanwhile, a dynamic residual gating mechanism is incorporated into the module. A lightweight convolutional neural network learns weighting coefficients in the range [0, 1], enabling adaptive balancing between enhanced and original features. In regions where defect characteristics are clear, the contribution of enhanced features is emphasized, whereas in complex-background or ambiguous-feature regions, more original features are retained. This adaptive regulation effectively prevents over-enhancement–induced overfitting and ensures stability during both training and inference.

The feature enhancement process of the MSGFF-C3k2 module can be described by Eq. (1):

$$\mathcal{F}_{enhance}(x, s) = \alpha \cdot \sum_{i=1}^{3} w_i \cdot \mathcal{B}_i(x) + (1 - \alpha) \cdot s \tag{1}$$

where x denotes the radar image features after bottleneck processing, and s represents the residual input features; $\mathcal{B}_i(x)$ correspond to the branch feature extraction functions of the ensemble, robust, and consistency DFF strategies, respectively; w_i denotes the adaptive fusion weights $\left(\text{satisfying } \sum_{i=1}^{3} w_i = 1\right)$; α denotes a learnable fusion coefficient (normalized to [0, 1] via a Sigmoid function).

Considering the multi-scale characteristics of subsurface distress features in radar images, the multi-scale feature extraction process of the robust DFF strategy is defined by Eqs. (2) and (3):

$$\mathcal{F}_{multi-scale}(x) = Concat\left(\mathcal{C}_k(x) \,|k \in 1, 2\right) \tag{2}$$

$$\mathcal{C}_k(x) = GN\left(\text{Conv}_{3\times3,dilation=k}(x)\right) \tag{3}$$

where $\text{Conv}_{3\times3,dilation=k}$ denotes a 3 × 3 grouped convolution with dilation rate k; $GN(\cdot)$ denotes the group normalization operation; and $Concat(\cdot)$ denotes the multi-scale feature concatenation operation.

2.4 Design of the MoCAttention Module

In the task of road subsurface distress detection in radar images, large-scale defect features are prone to interference from background noise, while small-scale features are often underrepresented. To address these issues, this paper integrates the Monte Carlo Attention (MoCAttention) module [28] into the large-object detection layer of YOLOv11 and combines it with the MSGFF-C3k2 module to form a feature enhancement network. When MoCAttention is introduced alone, the model's accuracy degrades; however, when it is jointly applied with the MSGFF-C3k2 module, the complementary effects of multi-scale feature fusion and the stochastic sampling–based attention mechanism lead to a significant improvement in road subsurface distress detection accuracy.

The MoCAttention module was not applied to all detection layers primarily due to a joint consideration of detection performance and computational overhead. From the perspective of detection accuracy,

preliminary experiments explored embedding the module separately into the small-, medium-, and large-object detection layers. The results showed that introducing MoCAttention into the small-object detection layer degraded performance, as its random sampling operation disrupted subtle small-scale defect features (e.g., edge textures of minor loose areas), leading to a noticeable decline in key evaluation metrics. When applied to the medium-object detection layer, the model's accuracy failed to improve, instead showing a slight degradation, falling short of the expected optimization effect. In contrast, embedding the module solely in the large-object detection layer enabled effective enhancement of large-scale void and loose-defect representations through random sampling and channel calibration, while suppressing background noise interference, thereby achieving coordinated improvements in both precision and recall.

From a computational cost perspective, embedding the MoCAttention module into all three detection layers would result in a substantial increase in both parameter count and GFLOPs. Compared with the configuration that introduces the module only in the large-object detection layer (2.995M parameters and 6.9 GFLOPs), the full-layer embedding strategy incurs significantly higher computational overhead, which contradicts the core objective of this study—namely, developing a lightweight model suitable for real-time engineering applications. Therefore, by jointly balancing detection performance optimization and computational efficiency, the MoCAttention module was ultimately incorporated only into the large-object detection layer.

The core idea of MoCAttention is to dynamically generate attention masks via a Monte Carlo random sampling strategy, introducing randomness during training to enhance model generalization, while reverting to stable global pooling during inference to ensure detection consistency. The core computational process of this module consists of two steps: first, an attention mask is generated via random sampling; second, channel attention weights are computed using the Squeeze-and-Excitation (SE) mechanism to complete feature reweighting.

During the training stage, the module randomly selects a pooling resolution k from a predefined set of pooling resolutions ($PoolRes = \{1, 2, 3\}$). For an input feature map (X) with dimensions ($B \times C \times H \times W$), a feature-shuffling operation is first applied to enhance randomness (the shuffling strategy can be enabled or disabled via the MoCOrder parameter), followed by adaptive average pooling over the shuffled feature map. If the pooled feature map has a spatial size larger than 1, random index sampling is further performed on the flattened feature dimensions, ultimately yielding a one-dimensional attention mask (A_{sample}), whose mathematical expression is given in Eq. (4):

$$A_{sample} = AdaptiveAvgPool2d\left(Shuffle\left(X\right), k\right) \tag{4}$$

where ($Shuffle\left(\cdot\right)$) denotes the feature shuffling function, which randomly rearranges the spatial dimensions of the feature map. This operation simulates the randomness of subsurface distress distributions in radar images and prevents the model from overfitting to fixed defect location features. After obtaining the attention mask, it is fed into an SE layer composed of two 1 × 1 convolutions to perform channel-wise feature calibration. The first convolution compresses the input channel dimension to ($C_{hid} = \max\left(makeDivisible\left(C_{in}/4, 8\right), 32\right)$) (where ($C_{in}$) denotes the number of input feature channels) and applies ReLU activation. The second convolution restores the channel dimension to the original size, followed by Sigmoid activation to generate the attention weights (W). Finally, element-wise multiplication is applied to reweight the original feature map, thereby enhancing features of the subsurface distress regionand suppressing background noise. Its mathematical expression is given in Eq. (5):

$$W = \sigma\left(Conv2d\left(Conv2d\left(A_{sample}, C_{hid}\right), C_{in}\right)\right),\ X_{out} = X \odot W \tag{5}$$

where ($\sigma(\cdot)$) denotes the Sigmoid activation function, and ($\odot$) denotes element-wise multiplication. During inference, to ensure stable detection results, the module no longer performs random sampling or shuffling. Instead, a 1 × 1 global adaptive average pooling is directly applied to the input feature map to generate the attention mask, thus achieving a balance between generalization during training and reliability during inference.

2.5 Improvement of the CIoU Loss Function (CIoU-ARP)

To address the problems that the traditional CIoU loss is sensitive to annotation bias for elongated defects and exhibits weak generalization in radar image–based defect detection tasks, this paper proposes an improved CIoU strategy with an adaptive aspect-ratio penalty (CIoU with Adaptive Ratio Penalty, CIoU-ARP). The core idea of this strategy is to dynamically adjust the penalty strength for aspect ratio based on the morphological characteristics of road subsurface defects in radar images. In this way, the original CIoU's capability to constrain bounding box position and shape is preserved, while adapting to the large annotation bias commonly observed for elongated road defects in radar images. The traditional CIoU is calculated as shown in Eq. (6):

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - v \cdot \alpha_{ciou} \tag{6}$$

where ($\rho^2(b, b^{gt})$) denotes the squared Euclidean distance between the centers of the predicted box b and the ground-truth box b^{gt}, c is the diagonal length of the minimum enclosing convex box covering both boxes, v is the penalty term measuring aspect ratio consistency, and (α_{ciou}) is the weighting coefficient. The conventional formulations of v and α_{ciou} are given in Eqs. (7) and (8), respectively:

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right) \tag{7}$$

$$\alpha_{ciou} = \frac{v}{(1 - IoU) + v} \tag{8}$$

The proposed improvement to the traditional CIoU comprises three parts. First, elongated defects in radar images are defined by an aspect ratio threshold ($\gamma = 5$); that is, when the aspect ratio of either the predicted box or the ground-truth box satisfies ($max(w/h, h/w) > \gamma$), the defect is regarded as elongated. Second, for elongated defects, the penalty weighting coefficient is dynamically adjusted to mitigate accuracy degradation from excessive penalization. The improved weighting coefficient is formulated as shown in Eq. (9):

$$\alpha_{arp} = \begin{cases} 0.6 \cdot \alpha_{ciou}, & max(w_1/h_1, h_1/w_1) > 5 \; or \; max(w_2/h_2, h_2/w_2) > 5 \\ \alpha_{ciou}, & others \end{cases} \tag{9}$$

Finally, to improve numerical stability and avoid overfitting, the improved CIoU value is normalized and clamped to the interval [−1, 1], as shown in Eq. (10):

$$IoU_{ARP} = clamp\left(IoU - \frac{\rho^2(b, b^{gt})}{c^2} - v \cdot \alpha_{arp}, -1, 1\right) \tag{10}$$

This improvement strategy is implemented solely through lightweight conditional judgments and coefficient adjustments, introducing no additional computational overhead, and remains fully compatible with the original CIoU calling logic. Experimental results demonstrate that this improvement exhibits strong

coupling with the other two improvements proposed in this paper. When all three proposed modules are jointly integrated into the baseline model, the overall mAP50 increases by approximately 2.0 percentage points compared with the original YOLOv11n. In contrast, when any two improvements are combined, except for combinations that include the attention module, the performance degrades significantly. This phenomenon arises because different improvements optimize the loss function in partially conflicting directions, and a globally optimal loss can only be achieved through their joint action. Specifically, the dynamic penalty of CIoU-ARP complements the feature enhancement and anchor optimization introduced by the other two improvements, thereby not only improving the detection accuracy of elongated defects but also ensuring stable detection performance for regular defects.

3 Experimental Setup

3.1 Experimental Environment and Hyperparameter Configuration

To verify the effectiveness of the DFF-MoCA-YOLO algorithm, all experiments were conducted on Ubuntu. The deep learning framework used was PyTorch 2.0.0, and the programming language was Python 3.10.16. In terms of hardware, an RTX 4090 GPU and an Xeon Platinum 8352V CPU were employed to ensure efficient experimental execution. The core hyperparameter settings are as follows: the number of training epochs was set to 100, the batch size was 32, the input image size was uniformly resized to 640 × 640, the initial learning rate (lr0) was set to 0.01, and the stochastic gradient descent (SGD) optimizer was adopted. The detailed configurations are listed in Table 1.

Table 1: Experimental environment and hyperparameter configuration.

Environment Item	Configuration	Parameter	Configuration
Operating System	Ubuntu	Epoch	100
Deep Learning Framework	PyTorch2.0.0	Batch Size	32
Programming Language	Python3.10.16	Image Size	640 × 640
GPU	RTX4090	lr0	0.01
CPU	Xeon Platinum 8352V	Optimizer	SGD

In addition to the core configurations, this study also adopts the following key hyperparameter settings based on the algorithm's characteristics and training stability requirements. The weight decay coefficient is set to 0.0005 to alleviate overfitting. The warm-up strategy includes 3.0 warm-up epochs, a warm-up momentum of 0.8, and a warm-up bias learning rate of 0.1, ensuring stable gradient updates during the early training stage. To address the limited dataset size, data augmentation is employed with a Mosaic augmentation probability of 1.0 and a Mixup probability of 0.15, enabling precise control over augmentation intensity. The Mosaic augmentation is disabled after 10 epochs to balance data diversity in the early training phase and convergence stability in the later stage. The bounding box loss weight (box) is set to 7.5 and the classification loss weight (cls) to 0.5, aligning with the loss optimization priorities of the object detection task. The number of training workers is set to 16 to match the available hardware computing capability. Automatic mixed precision (AMP) is disabled, along with single-class training (single_cls = False), checkpoint resuming (resume = False), and data caching (cache = False), to ensure experimental reproducibility and consistent results.

3.2 Evaluation Metrics and Dataset

3.2.1 Evaluation Metrics

To comprehensively and objectively evaluate the performance of target detection algorithms, this study adopts Precision (p), Recall (r), and mean Average Precision (mAP50 and mAP50-95) as the primary evaluation metrics. In addition, the computational complexity (GFLOPs) and number of parameters are reported to assess detection performance and computational cost.

Precision reflects the proportion of true positive samples among all samples the model predicts as positive, while Recall measures the proportion of true positive samples correctly detected among all actual positive samples. Their definitions are given as follows:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

where *TP* (True Positive) denotes the number of underground defect samples correctly detected by the model, *FP* (False Positive) represents the number of non-defect regions incorrectly classified as defects, and *FN* (False Negative) indicates the number of actual defect samples that are missed by the model. A higher Precision implies a lower false alarm rate, while a higher Recall indicates a lower miss detection rate.

The metrics mAP50 and mAP50–95 denote the mean Average Precision at an Intersection over Union (*IoU*) threshold of 0.5, averaged over *IoU* thresholdsfrom 0.5 to 0.95 with a step size of 0.05, respectively. Both metrics are computed based on Intersection over Union (*IoU*) and Average Precision (*AP*). *IoU* measures the spatial overlap between the predicted bounding box and the ground-truth box and is defined as:

$$IoU = \frac{Area\left(Prediction\right) \cap Area\left(GroundTruth\right)}{Area\left(Prediction\right) \cup Area\left(GroundTruth\right)} \tag{12}$$

The Average Precision (*AP*) is calculated as the area under the Precision–Recall (P–R) curve:

$$AP = \int_0^1 P\left(R\right) dR \tag{13}$$

Specifically, mAP50 reflects the overall detection capability of the model under a conventional localization criterion, whereas mAP50–95 imposes stricter constraints on bounding-box localization accuracy and thus provides a more comprehensive evaluation of model robustness.

GFLOPs (Giga Floating-Point Operations) represent the number of floating-point operations during forward inference and directly characterize the model's computational complexity. The number of parameters indicates the total count of trainable parameters, reflecting the storage cost. Together, these two metrics assess the algorithm's practicality: lower GFLOPs indicate higher computational efficiency, and fewer parameters imply reduced memory requirements, making the model more suitable for real-time underground defect detection in GPR imaging applications.

3.2.2 Dataset

The experimental dataset was constructed based on a GPR inspection project commissioned by the Comprehensive Administrative Law Enforcement Bureau of Binjiang District, Hangzhou. The project was jointly carried out by three institutions to which the research team belongs and covered eight urban roads in Binjiang District, Hangzhou, with a total survey length of 10,498.11 m. The inspection scope included typical road scenarios, such as motor vehicle lanes, non-motor vehicle lanes, and sidewalks. Data acquisition strictly followed industry standards, including the Technical Specification for Comprehensive

Detection of Urban Road Underground Defects. Vehicle-mounted three-dimensional GPR and multi-frequency two-dimensional radar were used to collect data. During this eight-month inspection campaign, more than 1200 radar images of suspected looseness and void defects were collected. After excluding locations affected by nearby underground utilities, borehole verification was conducted at selected defect sites using a Mingzhuo MZ-58 impact drill, a Zhuochi air blower, and an Airuipu F408B industrial endoscope. As a result, 677 borehole verification images were obtained, each paired with its corresponding GPR defect image. The distribution of these verified defects is shown in Fig. 4. The radar types and parameters used in the experiments are listed in Table 2, and representative examples of defect-related GPR images and corresponding borehole verification are illustrated in Fig. 5.

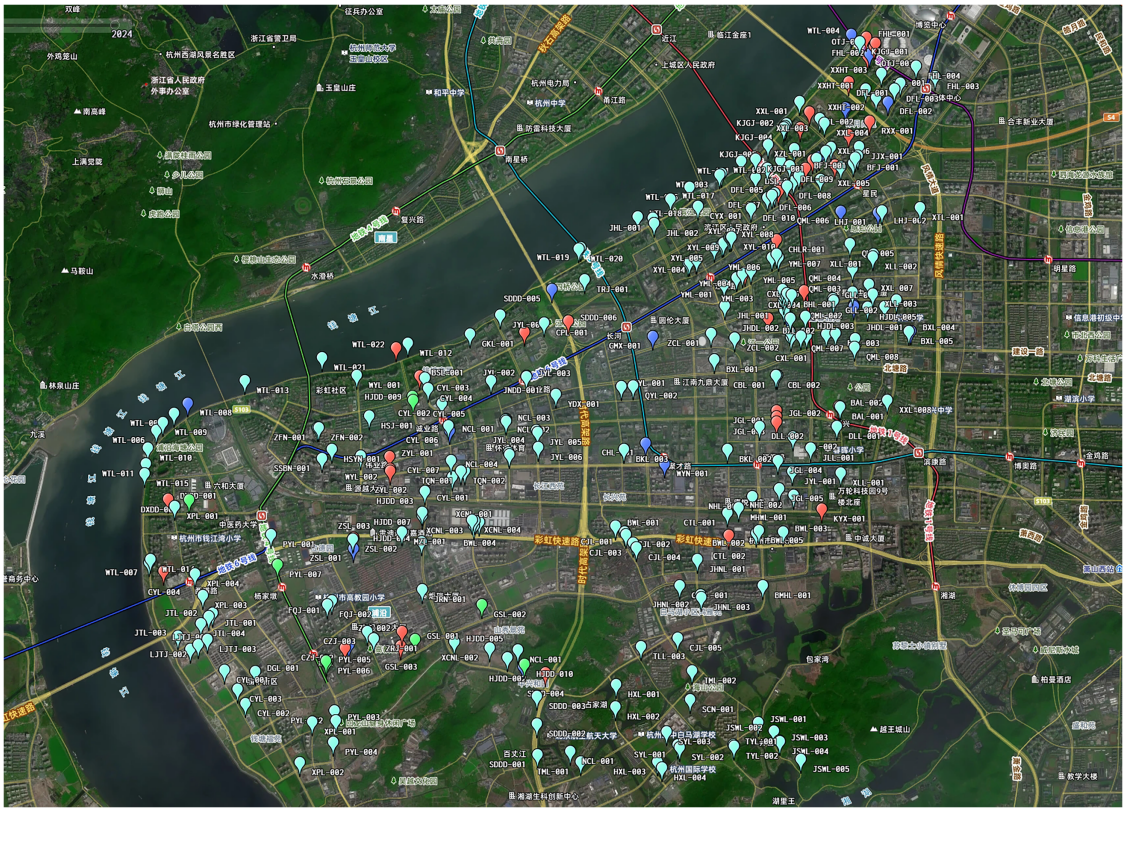

Figure 4: Spatial distribution of borehole-verified defects.

Table 2: Radar types and parameters.

Radar Type	Antenna Frequency	Sampling Interval	Time Window	Channel Configuration	Survey Speed
CrossOverCO1760 2D Radar	170,600 MHz	/	1050 ns	2	5 km/h
GeoScope 3D Radar	200–3000 MHz	7.5 cm	50 ns	20	15 km/h
Mobyscan-200V 3D Radar	450 MHz	50 mm	62 ns	15	15 km/h

Figure 5: Representative examples of defect-related GPR images and corresponding borehole verification: (**a**) example of defect-related GPR images with looseness defects; (**b**) example of defect-related GPR images with void defects; (**c**) example of borehole verification for looseness defects; (**d**) example of borehole verification for void defects.

To meet the training and validation requirements of YOLO-based object detection models, the collected raw GPR data were preprocessed by annotating them with LabelMe and converting them into the required format. A YOLO-format dataset containing two types of road subsurface defects-voids and looseness-was ultimately constructed. The dataset comprises 677 valid annotated samples. Following common dataset splitting ratios for training, testing, and validation, 588 samples were allocated to the training set, 76 to the test set, and 73 to the validation set. This dataset effectively supports the training, tuning, and performance evaluation of road subsurface defect detection models, providing high-quality data support for research on intelligent identification of road subsurface defects.

4 Experimental Results and Analysis

4.1 Ablation Experiments

To investigate the independent and synergistic effects of the MSGFF-C3k2 module (Module 1), the MoCAttention module (Module 2), and the CIoU-ARP loss (Module 3) on the performance of YOLOv11, a series of ablation experiments was conducted. The experimental results and radar comparison charts are presented in Table 3 and Fig. 6.

Taking YOLOv11n (Algorithm A) as the baseline, the model achieves a precision of 0.379, a recall of 0.583, an mAP50 of 0.501, and an mAP50–95 of 0.285.

When only the MSGFF-C3k2 module is introduced (Algorithm B), the recall increases markedly to 0.686, and the mAP50 slightly improves to 0.507, indicating that this module effectively enhances the model's target recall capability. However, both precision and mAP50–95 exhibit a slight decline. This is mainly because MSGFF-C3k2 expands the receptive field through multi-strategy feature fusion and strengthens multi-scale defect feature extraction, while simultaneously amplifying background noise and device-induced interference in GPR images. As a result, the false-positive rate increases, and the performance on the more localization-sensitive mAP50–95 metric deteriorates.

Table 3: Ablation experiments.

Algorithm	Module 1	Module 2	Module 3	Precision (p)	Recall (r)	mAP50	mAP50-95	GFLOPs	Parameters
A	×	×	×	0.379	0.583	0.501	0.285	6.3	2,582,542
B	✓	×	×	0.390	0.686	0.507	0.266	6.4	2,469,670
C	×	✓	×	0.412	0.554	0.443	0.223	6.7	3,108,110
D	×	×	✓	0.399	0.432	0.430	0.227	6.3	2,582,542
E	✓	✓	×	0.593	0.578	0.589	0.320	6.9	2,995,238

(Continued)

Table 3 (continued)

Algorithm	Module 1	Module 2	Module 3	Precision (p)	Recall (r)	mAP50	mAP50-95	GFLOPs	Parameters
F	✓	×	✓	0.454	0.508	0.437	0.239	6.4	2,469,670
G	×	✓	✓	0.492	0.591	0.558	0.312	6.7	3,078,042
H	✓	✓	✓	0.586	0.655	0.608	0.341	6.9	2,995,238

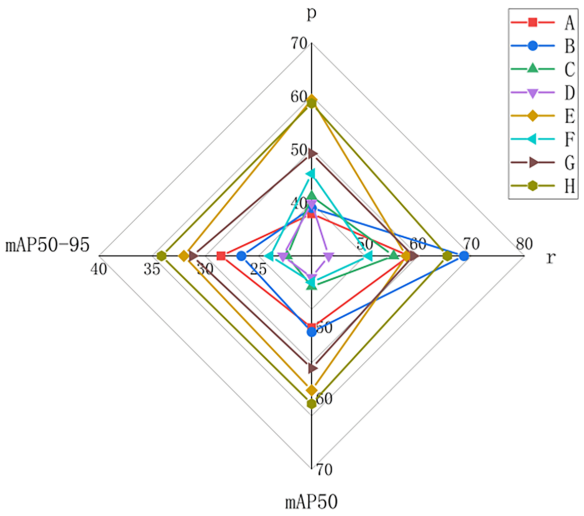

Figure 6: Radar chart of core evaluation metrics for the ablation experiments.

When only the MoCAttention module is incorporated (Algorithm C), precision improves to 0.412, demonstrating the attention mechanism's ability to focus on discriminative defect features. Nevertheless, recall and average precision decrease, suggesting that a single attention module cannot ensure comprehensive detection performance. Although the random sampling and channel calibration mechanisms of MoCAttention can precisely emphasize core defect features, they do not adequately address the strong noise and the vulnerability of small-scale defect features in radar images. Moreover, the stochastic sampling process may disrupt fine-grained defect structures, ultimately leading to an increased miss rate.

When only the CIoU-ARP loss function is used (Algorithm D), precision improves marginally, while recall and average precision decline significantly. This indicates that the proposed loss function must be coupled with effective feature-extraction modules to work properly. CIoU-ARP optimizes only the aspect-ratio penalty for elongated defects; without sufficiently discriminative defect features, the adaptive penalty may lead to overly conservative bounding-box regression, resulting in a large number of missed detections.

From the perspective of module coupling, the MSGFF-C3k2 module performs comprehensive feature extraction, addressing the strong noise and large-scale variations inherent in GPR images; the MoCAttention module performs feature selection and focusing, accurately localizing core defect regions while suppressing irrelevant noise features; and the CIoU-ARP module performs loss optimization and calibration, improving bounding-box regression accuracy for elongated defects. Together, these modules form a closed-loop synergy of feature extraction—feature refinement—loss optimization, in which the limitations of any single module are compensated by the others.

When MSGFF-C3k2 and MoCAttention are combined (Algorithm E), precision increases substantially to 0.593, with mAP50 and mAP50–95 reaching 0.589 and 0.320, respectively, clearly demonstrating the

synergistic effect of feature enhancement and attention mechanisms. When all proposed modules are integrated (Algorithm H), the model achieves optimal performance, with a precision of 0.586, a recall of 0.655, an mAP50 of 0.608, and an mAP50–95 of 0.341. Compared with the baseline model, mAP50 and mAP50–95 increase by 21.4% and 19.6%, respectively, while the computational cost remains modest, with GFLOPs of only 6.9 and a parameter count of 2,995,238. These results verify the effectiveness and rationality of the proposed multi-module fusion strategy.

To more intuitively illustrate the performance improvements brought by the proposed modules, Fig. 7 presents the Precision–Recall (PR) curves of the baseline YOLOv11 (Algorithm A, mAP50 = 0.501) and the DFF-MoCA-YOLO model integrating all improvement modules (Algorithm H, mAP50 = 0.608). It can be observed that the PR curve of DFF-MoCA-YOLO lies entirely above that of the baseline YOLOv11, and the coverage areas of the curves for both defect categories are noticeably larger. Specifically, the precision of the looseness category increases from 0.645 to 0.656, while the precision of the void category rises from 0.357 to 0.560. Moreover, the overall mAP50 across all categories improves from 0.501 to 0.608. These results indicate that the improved model maintains higher precision across different recall levels, and that both the stability and overall detection performance are significantly enhanced.

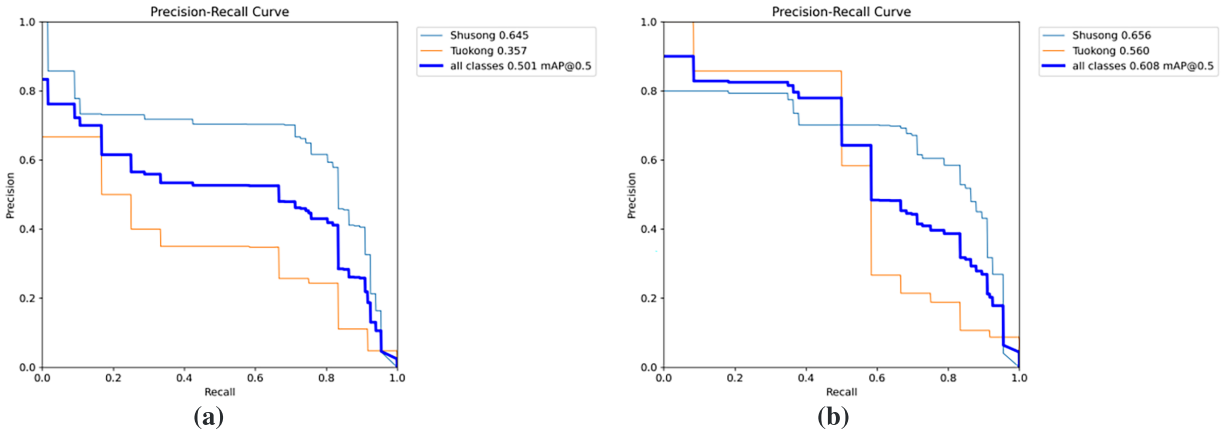

Figure 7: Precision–recall (PR) curves: (**a**) YOLOv11;(**b**) DFF-MoCA-YOLO.

Further analysis of the trade-off between detection accuracy improvement and computational overhead indicates that, compared with the baseline YOLOv11n model (GFLOPs = 6.3, parameters = 2,582,542), the proposed DFF-MoCA-YOLO model, incorporating all improvement modules, increases GFLOPs to 6.9, corresponding to only a 9.5% rise, while the number of parameters increases to 2,995,238, an increase of 15.9%. In contrast, the core performance metrics show substantially larger gains: mAP50 improves by 21.4%, mAP50-95 by 19.6%, and precision and recall increase by 0.207 and 0.072, respectively. These results demonstrate that the performance improvement significantly outweighs the growth in computational cost.

From an engineering application perspective, the proposed model maintains GFLOPs below 7.0 and a parameter count under 3.0M, enabling real-time inference on standard GPUs and even on embedded devices. This computational efficiency satisfies the real-time requirements of vehicle-mounted mobile detection systems for urban road subsurface defects. Further comparisons with mainstream algorithms, such as YOLOv6 (GFLOPs = 11.8, mAP50 = 0.502) and YOLOv9 (GFLOPs = 26.7, mAP50 = 0.499), show that DFF-MoCA-YOLO achieves higher detection accuracy with substantially lower computational overhead, thereby validating the efficiency of the proposed improvement strategies.

In summary, the proposed method achieves significant accuracy gains with only a modest increase in computational cost, effectively balancing detection performance and engineering practicality. This makes it a lightweight and efficient solution for subsurface defect detection in GPR images.

4.2 Comparison with Mainstream Algorithms

To further verify the overall performance of the DFF-MoCA-YOLO algorithm, it is compared with mainstream object detection algorithms from YOLOv5 to YOLOv13 (excluding YOLOv7). The partial prediction results of two types of defects for different models are illustrated in Fig. 8.

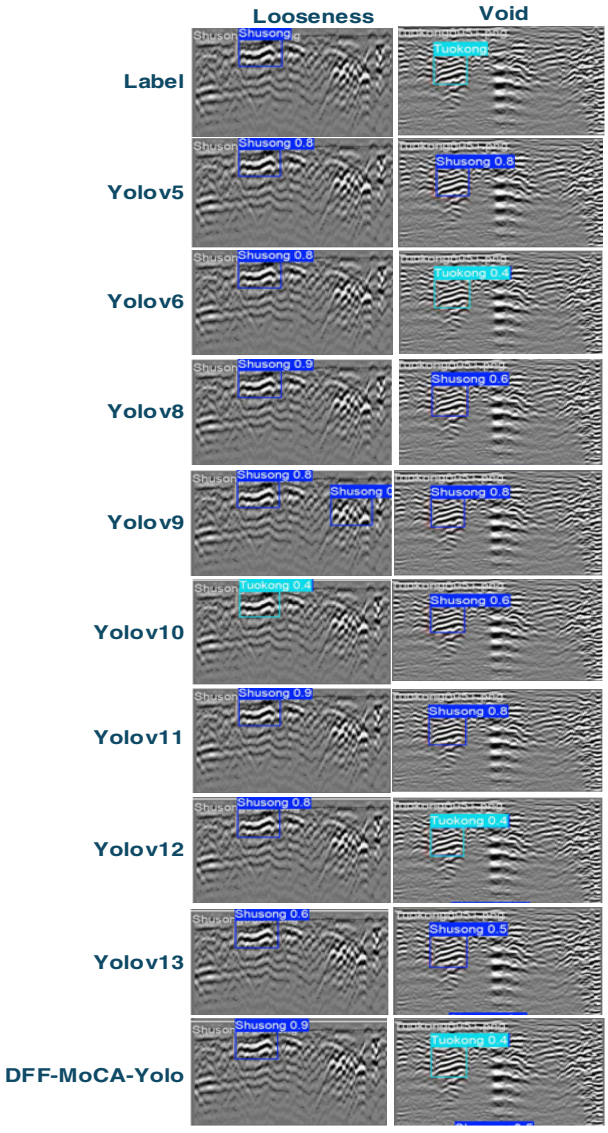

Figure 8: Partial prediction results of two types of defects for different models.

The core considerations for selecting this comparison scope are threefold. First, the YOLO series is designed with an explicit emphasis on balancing real-time performance and detection accuracy, which closely aligns with the engineering objective of this study—namely, achieving high detection accuracy at low computational cost. Although two-stage detectors such as Faster R-CNN and Transformer-based

models (e.g., DETR) exhibit competitive performance in certain scenarios, their computational overhead is substantially higher, with GFLOPs typically exceeding 30, making them unsuitable for vehicle-mounted real-time detection applications.

Second, since the proposed method is developed based on YOLOv11, comparisons within the same algorithmic family can effectively minimize the influence of architectural differences, thereby more accurately highlighting the contributions of the proposed improvements, including the MSGFF-C3k2 module and the MoCAttention mechanism.

Third, YOLO-based detectors have been widely adopted for underground defect detection and related engineering applications, thereby rendering these comparisons more representative and practically meaningful for the target domain.

It should be noted that this study does not include lightweight specialized models (e.g., MobileNet-YOLO variants) or the latest Transformer-based detection frameworks. As a result, the performance positioning of the proposed method across all categories of detection architectures cannot be fully assessed. Expanding the comparison scope to include these models will be considered in future work.

The experimental results, radar comparison charts, and prediction visualizations for the two defect categories for each model are shown in Table 4 and Fig. 9.

The comparison results indicate that DFF-MoCA-YOLO outperforms existing YOLO-series algorithms across all core evaluation metrics. Compared with the baseline YOLOv11, the improved model achieves increases of 0.207 in precision, 0.072 in recall, 0.107 in mAP50, and 0.056 in mAP50-95, representing a significant improvement in key detection performance. Compared with YOLOv6, although the recall of DFF-MoCA-YOLO is slightly lower (0.705 for YOLOv6 vs. 0.655 for DFF-MoCA-YOLO), its precision increases by 0.158, mAP50 by 0.106, and mAP50-95 by 0.062. Meanwhile, the GFLOPs are reduced from 11.8 to 6.9, and the number of parameters decreases from 4,233,942 to 2,995,238, achieving a favorable balance between detection accuracy and computational efficiency. Although YOLOv9 achieves a recall of 0.677, its mAP50 and mAP50-95 are lower than those of the improved YOLOv11, and its computational cost is substantially higher, with GFLOPs reaching 26.7 and the number of parameters totaling 7,167,862. The remaining algorithms also fall significantly behind the improved YOLOv11 in terms of mAP50 and mAP50-95.

Table 4: Comparison results of mainstream algorithms.

Algorithm	P	r	mAP50	mAP50-95	GFLOPs	Parameters
YOLOv5	0.467	0.492	0.403	0.22	7.1	2,503,334
YOLOv6	0.428	0.705	0.502	0.279	11.8	4,233,942
YOLOv8	0.358	0.682	0.48	0.233	8.1	3,006,038
YOLOv9	0.402	0.677	0.499	0.272	26.7	7,167,862
YOLOv10	0.375	0.549	0.396	0.209	8.2	2,695,196
YOLOv11	0.379	0.583	0.501	0.285	6.3	2,582,542
YOLOv12	0.461	0.454	0.447	0.239	5.8	2,508,734
YOLOv13	0.295	0.627	0.432	0.226	6.9	2,448,285
DFF-MoCA-YOLO	0.586	0.655	0.608	0.341	6.9	2,995,238

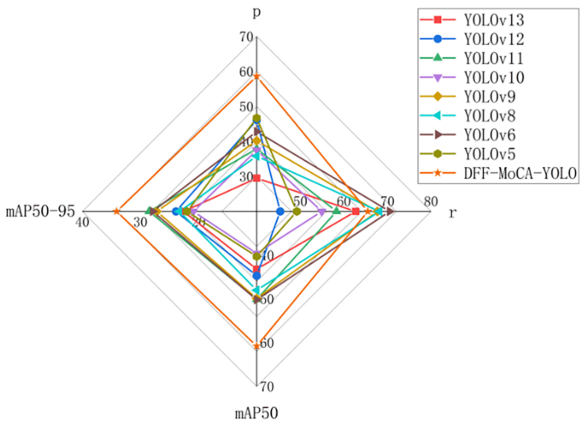

Figure 9: Radar chart of core evaluation metrics for comparison with mainstream algorithms.

Overall, while maintaining low computational overhead (GFLOPs = 6.9) and a moderate parameter count (2,995,238), DFF-MoCA-YOLO achieves comprehensive improvements in precision, recall, and average precision, thereby validating the effectiveness and practical applicability of the proposed improvement strategies.

4.3 Hyperparameter Sensitivity Analysis

To further validate the overall performance, stability, and rationality of hyperparameter selection for the proposed DFF-MoCA-YOLO algorithm, this section presents additional sensitivity comparison experiments focusing on the input image resolution and the elongated-defect decision threshold γ in the CIoU-ARP loss function. By quantitatively analyzing the impact of different parameter settings on model performance, these experiments provide a basis for optimal hyperparameter selection.

4.3.1 Sensitivity Analysis of Input Image Resolution

To investigate the sensitivity of model performance to input image size, experiments were conducted using four typical resolutions: 320 × 320, 480 × 480, 640 × 640, and 800 × 800. All other hyperparameters were kept unchanged. Detection performance and computational cost under different resolutions were compared, and the results are reported in Table 5.

Table 5: Sensitivity comparison results under different input image resolutions.

Resolution	p	r	mAP50	mAP50-95	GFLOPs	Parameters
320	0.451	0.496	0.496	0.246	6.9	2,995,238
480	0.775	0.375	0.515	0.249	6.9	2,995,238
800	0.336	0.659	0.473	0.266	6.9	2,995,238
DFF-MoCA-YOLO (640)	0.586	0.655	0.608	0.341	6.9	2,995,238

The experimental results indicate that model performance is highly sensitive to input image resolution.

At low resolution (320 × 320), fine-grained defect features in GPR images—such as the edge textures of elongated voids and the point-distribution details of loose zones—are severely degraded. As a result, both

precision and recall remain relatively low, with mAP50 at 0.496, failing to meet the accuracy requirements for defect detection.

At medium resolution (480 × 480), although a high precision of 0.775 is achieved, recall drops sharply to 0.375, resulting in a pronounced imbalance characterized by "high precision but low recall". This is mainly because this resolution cannot simultaneously capture the global structure of large-scale defects and the local details of small-scale defects, leading to many small, loose defects being missed.

At high resolution (800 × 800), background noise, such as surface-clutter reflections and device-induced interference, is excessively amplified. Precision decreases to 0.336, and mAP50 drops to 0.473, failing to achieve the expected performance improvement. Moreover, the increased image size results in higher memory consumption during inference.

In contrast, the 640 × 640 resolution achieves the best balance between precision and recall, with mAP50 reaching 0.608 and mAP50-95 reaching 0.341, yielding the optimal overall performance. It is noteworthy that GFLOPs and parameter counts remain unchanged across different resolutions (6.9 GFLOPs and 2,995,238 parameters), indicating that input resolution affects only inference-stage feature extraction efficiency rather than the core model architecture or trainable parameters.

Consequently, 640 × 640 is selected as the input resolution, as it aligns with the five-stage downsampling structure of the YOLOv11 backbone (640/32 = 20), maximizes the preservation of defect features in GPR images, and achieves a favorable balance between detection accuracy and runtime efficiency, meeting the requirements of vehicle-mounted mobile inspection scenarios.

4.3.2 Sensitivity Analysis of the Elongated-Defect Threshold

The CIoU-ARP loss function is designed to address annotation bias associated with elongated defects in GPR images. Its core mechanism involves computing the conventional CIoU penalty and weight coefficient, identifying elongated defects whose aspect ratio exceeds a predefined threshold using torch.max(w/h, h/w), reducing the penalty strength to 60% of the original value, and finally normalizing the loss to the range [−1, 1] to improve numerical stability. The choice of $\gamma = 5$ is motivated by dataset characteristics and model compatibility: most elongated defects in the dataset (e.g., extremely narrow voids) exhibit aspect ratios exceeding this threshold, making it a reasonable boundary between elongated and regular defects, while also being well matched to the 640 × 640 resolution and the YOLOv11n regression mechanism.

To verify the rationality of the elongated-defect threshold γ in the CIoU-ARP loss function, five representative values (γ = 3, 4, 5, 6, 7) were evaluated under identical hyperparameter settings. The corresponding performance results are summarized in Table 6.

Table 6: The corresponding performance results.

CIoU-ARP	p	r	mAP50	mAP50-95	GFLOPs	Parameters
$\gamma = 3$	0.406	0.663	0.458	0.240	6.9	2,995,238
$\gamma = 4$	0.406	0.663	0.458	0.240	6.9	2,995,238
$\gamma = 6$	0.406	0.663	0.458	0.240	6.9	2,995,238
$\gamma = 7$	0.406	0.663	0.458	0.240	6.9	2,995,238
DFF-MoCA-YOLO ($\gamma = 5$)	0.586	0.655	0.608	0.341	6.9	2,995,238

The sensitivity analysis shows that when $\gamma < 5$ (i.e., γ = 3 or 4), a large number of non-elongated defects are incorrectly classified as elongated, causing the CIoU-ARP aspect-ratio penalty to be excessively activated.

This leads to increased bounding-box regression bias and a significant degradation in core accuracy metrics (mAP50 and mAP50-95).

When $\gamma > 5$ (i.e., $\gamma = 6$ or 7), some genuinely elongated defects fail to meet the threshold and are therefore excluded from the penalty adjustment mechanism. As a result, the loss function cannot effectively target elongated defects, and detection accuracy remains low.

Only when $\gamma = 5$ does the model accurately distinguish between elongated and non-elongated defects, enabling the penalty mechanism to align closely with the morphological characteristics of defects in GPR images. Under this setting, the model achieves optimal overall performance, with a precision of 0.586, a recall of 0.655, mAP50 of 0.608, and mAP50-95 of 0.341.

Additional validation confirms that GFLOPs and parameter counts remain unchanged across different γ values (6.9 GFLOPs and 2,995,238 parameters), indicating that γ functions purely as a loss-function hyperparameter and does not affect the core model structure or computational overhead. This further confirms the rationality and lightweight nature of the chosen threshold.

5 Conclusion and Limitations

5.1 Conclusion

In this study, a DFF-MoCA-YOLO detection framework for road subsurface defects in GPR images is proposed. By jointly optimizing the feature extraction module, attention mechanism, and bounding box regression loss, the proposed method achieves consistent improvements in detection robustness, accuracy, and generalization under complex GPR imaging conditions. The main conclusions can be summarized as follows:

(1) The MSGFF-C3k2 module enhances the model's robustness to noise interference and scale variations through multi-strategy alternating feature fusion. The MoCAttention module strengthens feature representation of defect regions via stochastic sampling and channel-wise calibration. The CIoU-ARP loss function improves bounding-box regression accuracy for elongated targets by introducing an adaptive aspect-ratio penalty.

(2) Experimental results demonstrate that DFF-MoCA-YOLO achieves significant performance improvements on the self-constructed dataset. Compared with the baseline YOLOv11, mAP50 increases by 21.4% and mAP50-95 by 19.6%, while precision and recall reach 0.586 and 0.655, respectively. In terms of computational cost, the model has 2.99M parameters and only 6.9 GFLOPs, maintaining low complexity while substantially improving detection accuracy.

(3) Compared with mainstream algorithms such as YOLOv5–YOLOv13, DFF-MoCA-YOLO achieves the best overall performance across four core metrics, namely precision, recall, mAP50, and mAP50-95. In particular, it achieves comprehensive performance gains while maintaining a lightweight design (GFLOPs = 6.9), demonstrating strong engineering practicality and deployment potential.

Furthermore, the proposed method not only enhances detection accuracy but also provides critical insights for structural durability assessment. The identified void and looseness defects directly compromise pavement structural integrity: voids, as subsurface cavities, reduce the load-bearing capacity of the roadbed and can lead to sudden collapse under traffic loading; looseness, characterized by decreased material density, indicates early-stage deterioration that may accelerate the development of more severe distresses such as cracking and settlement. By enabling accurate localization and classification of these defects, DFF-MoCA-YOLO supports the prioritization of maintenance interventions—allowing engineers to distinguish between urgent threats (e.g., large voids requiring immediate grouting) and areas requiring routine monitoring

(e.g., localized looseness). This capability facilitates data-driven durability assessments and cost-effective maintenance scheduling, thereby extending pavement service life and enhancing urban road safety.

5.2 Limitations

Despite the promising results, several limitations remain and should be addressed in future work.

First, the proposed model shows a strong dependence on soil type. The experimental dataset primarily consists of silty clay soil, and model performance may be affected by variations in soil moisture content, compaction degree, and medium homogeneity. In regions with sandy or gravelly soils, differences in radar-wave propagation characteristics and defect responses may degrade detection accuracy, and model adaptability remains to be validated.

Second, the dataset's geographical coverage is limited. All data were collected from road scenes in Binjiang District, Hangzhou, where climatic conditions, construction standards, and underground utility distributions are region-specific. The model's generalization to other climatic zones (e.g., arid or cold regions) or to different urban road systems has not yet been verified, limiting its direct nationwide applicability.

Third, the dataset's scale and diversity remain insufficient. Although the 677 samples support model training, the dataset remains small compared with large-scale public benchmarks. Moreover, only two defect types—voids and looseness—are included, while other common underground defects, such as water-rich zones and cracks, are not. These limitations may restrict model robustness in complex, multi-defect scenarios.

Fourth, external independent validation is lacking. Model performance was evaluated only using internally split test and validation sets, without cross-platform or cross-region validation on third-party datasets or real engineering deployments. Consequently, the reported results may not fully reflect model behavior in complex real-world environments.

5.3 Future Work

To address the above limitations, future research will focus on several directions. First, the dataset scale and diversity will be expanded by collecting GPR data across different soil types and geographical regions, and by incorporating additional defect categories, such as water-rich zones and cracks, to enhance generalization. Second, cross-region and cross-soil adaptability will be explored by introducing environmental factors (e.g., soil type and moisture content) as auxiliary inputs and developing adaptive adjustment mechanisms. Third, external validation and engineering pilot applications will be conducted across multiple cities and road scenarios, with model refinement guided by practical feedback. Finally, multimodal fusion detection schemes integrating technologies such as LiDAR and infrared sensing will be investigated to overcome the limitations of single-modal GPR detection and further improve defect recognition accuracy in complex environments.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm their contribution to the study as follows: Conceptualization: Bin Chen; methodology: Chao Qiu; software: Chao Qiu; validation: Bin Chen, Chao Qiu, Wanli Cui; formal analysis: Bin Chen; investigation: Chao Qiu; data curation: Bin Chen; writing—original draft: Chao Qiu; writing—review & editing: Bin Chen; visualization: Chao Qiu; supervision: Bin Chen, Wanli Cui. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript

ARP	Adaptive Aspect-Ratio Penalty
CIoU	Complete Intersection over Union
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DFF	Diverse Feature Fusion
GPR	Ground-Penetrating Radar
MoCA	Monte Carlo Attention
MSGFF	Multi-Strategy Gated Feature Fusion
SPPF	Spatial Pyramid Pooling Fusion
YOLO	You Only Look Once

References

1. Wang D, Lyu H, Tang F, Ye C, Zhang F, Wang S, et al. Road Structural defects detection and digitalization based on 3D ground penetrating radar technology: a state-of-the-art review. China J Highw Transp. 2023;36(3):1–19. (In Chinese). doi:10.19721/j.cnki.1001-7372.2023.03.001.
2. Luo X, He J, Zhang D, Zhu J, Li M, Zhang B, et al. Evaluating subsurface cavities detection using innovative laser dynamic deflectometer for efficient and large-scale urban road network inspections. Tunn Undergr Space Technol. 2025;159:106471. doi:10.1016/j.tust.2025.106471.
3. Pham MT, Lefèvre S. Buried object detection from B-scan ground penetrating radar data using faster-RCNN. In: 2018 IEEE International Geoscience and Remote Sensing Symposium; 2018 Jul 22–27; Valencia, Spain. p. 6804–7. doi:10.1109/IGARSS.2018.8517683.
4. Wang L, Liu Z, Gu X, Wang D. Three-dimensional reconstruction of road structural defects using GPR investigation and back-projection algorithm. Sensors. 2024;25(1):162. doi:10.3390/s25010162.
5. Peng C, Yang B, Li M, Zhang G, Sun H, Jiang Z. Automatic road subsurface distress recognition from ground penetrating radar images using deep learning-based cross-verification. arXiv:2507.11081. 2025. doi:10.48550/arXiv.2507.11081.
6. Sui X, Leng Z, Wang S. Machine learning-based detection of transportation infrastructure internal defects using ground-penetrating radar: a state-of-the-art review. Intell Transp Infrastruct. 2023;2:liad004. doi:10.1093/iti/liad004.
7. Liu P, Ding Z, Zhang W, Ren Z, Yang X. Using ground-penetrating radar and deep learning to rapidly detect voids and rebar defects in linings. Sustainability. 2023;15(15):11855. doi:10.3390/su151511855.
8. Liu Z, Gu X, Chen J, Wang D, Chen Y, Wang L. Automatic recognition of pavement cracks from combined GPR B-scan and C-scan images using multiscale feature fusion deep neural networks. Autom Constr. 2023;146(4):104698. doi:10.1016/j.autcon.2022.104698.
9. Wang P, Zhang L, Xing C, Tan Y, Leng Z, Sui X. Intelligent identification of hidden defects in asphalt roads using GPR based on loss function and anchor box optimization. J Perform Constr Facil. 2025;39(6):04025057. doi:10.1061/jpcfev.cfeng-5008.
10. Zhang Y, Lv H, Ni Y, Ye C, Wang D, Tang F. Automatic recognition of hidden road defects from GPR images using an enhanced CNN approach. J Transp Eng Part B Pavements. 2025;151(2):04025021. doi:10.1061/jpeodx.pveng-1699.

11. Li Y, Zhang W, Lv S, Yu J, Ge D, Guo J, et al. YOLOv11-CAFM model in ground penetrating radar image for pavement distress detection and optimization study. Constr Build Mater. 2025;485:141907. doi:10.1016/j.conbuildmat.2025.141907.
12. Ma Y, Lei W, Pang Z, Zheng Z, Tan X. Rebar clutter suppression and road defects localization in GPR B-scan images based on SuppRebar-GAN and EC-Yolov7 networks. IEEE Trans Geosci Remote Sens. 2024;62(1):1–14. doi:10.1109/TGRS.2024.3373025.
13. Chen Y, Deng J, Wang Z, Li M, Pan Z, Yu H. Efficient defect detection method for YOLOv5 circuit board based on RepVGG and SE attention mechanism. In: Proceedings of the 2025 3rd Asia Conference on Computer Vision, Image Processing and Pattern Recognition; 2025 May 23–25; Xiamen, China. p. 154–60. doi:10.1117/12.3076246.
14. Zhou L, Ma B, Dong Y, Yin Z, Lu F. DCFE-YOLO: a novel fabric defect detection method. PLoS One. 2025;20(1):e0314525. doi:10.1371/journal.pone.0314525.
15. Kang Z, Liao Y, Du S, Li H, Li Z. SE-CBAM-YOLOv7: an improved lightweight attention mechanism-based YOLOv7 for real-time detection of small aircraft targets in microsatellite remote sensing imaging. Aerospace. 2024;11(8):605. doi:10.3390/aerospace11080605.
16. Di R, Fan H, Feng H, Lv Z, Shu L, Xie R, et al. MFE-YOLO: a multi-scale feature enhanced network for PCB defect detection with cross-group attention and FIoU loss. Entropy. 2026;28(2):174. doi:10.3390/e28020174.
17. Li A, Hamzah R, Khatijah Nor Abdul Rahim S, Gao Y. YOLO algorithm with hybrid attention feature pyramid network for solder joint defect detection. IEEE Trans Compon Packag Manuf Technol. 2024;14(8):1493–500. doi:10.1109/TCPMT.2024.3409773.
18. Gao Y, Pei L, Wang S, Li W. Intelligent detection of urban road underground targets by using ground penetrating radar based on deep learning. J Phys Conf Ser. 2021;1757(1):012081. doi:10.1088/1742-6596/1757/1/012081.
19. Shin SP, Lee SY, Le THM. Feasibility of efficientDet-D3 for accurate and efficient void detection in GPR images. Infrastructures. 2025;10(6):140. doi:10.3390/infrastructures10060140.
20. Wang J, Liu H, Jiang P, Wang Z, Sui Q, Zhang F. GPRI2Net: a deep-neural-network-based ground penetrating radar data inversion and object identification framework for consecutive and long survey lines. IEEE Trans Geosci Remote Sens. 2022;60(4):5106320. doi:10.1109/TGRS.2021.3111445.
21. Hu H, Fang H, Wang N, Liu H, Lei J, Ma D, et al. A study of automatic recognition and localization of pipeline for ground penetrating radar based on deep learning. IEEE Geosci Remote Sens Lett. 2022;19(6):4026405. doi:10.1109/LGRS.2022.3198439.
22. Liu Q, Zhu C. Automatic detection for road voids from GPR images using deep learning method. In: Proceedings of the 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL); 2023 May 12–14; Zhuhai, China. p. 617–20. doi:10.1109/CVIDL58838.2023.10167036.
23. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: faster and better learning for bounding box regression. Proc AAAI Conf Artif Intell. 2020;34(7):12993–3000. doi:10.1609/aaai.v34i07.6999.
24. Zhai H, Cheng J, Wang M. Rethink the IoU-based loss functions for bounding box regression. In: Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC); 2020 Dec 11–13; Chongqing, China. p. 1522–8. doi:10.1109/itaic49862.2020.9339070.
25. Zhang H, Zhang S. Focaler-IoU: more focused intersection over union loss. arXiv:2401.10525. 2024. doi:10.48550/arXiv.2401.10525.
26. Wang Q, Cheng J. LCornerIoU: an improved IoU-based loss function for accurate bounding box regression. In: Proceedings of the 2021 International Conference on Intelligent Computing, Automation and Systems (ICICAS); 2021 Dec 29–31; Chongqing, China. p. 377–83. doi:10.1109/icicas53977.2021.00085.
27. Yang J, Qiu P, Zhang Y, Marcus DS, Sotiras A. D-net: dynamic large kernel with dynamic feature fusion for volumetric medical image segmentation. arXiv:2403.10674. 2024. doi:10.48550/arXiv.2403.10674.
28. Dai W, Liu R, Wu Z, Wu T, Wang M, Zhou J, et al. Exploiting scale-variant attention for segmenting small medical objects. IEEE Trans Neural Netw Learning Syst. 2026;2026:1–18. doi:10.1109/tnnls.2025.3645355.