Journal of Network Intelligence ©2026 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **11**, Number **1**, February **2026**

An Improved YOLOv11-Based Algorithm for Defect Detection in Radial Tire X-ray Images

Haijun Lian[1]

[1]College of Electronic, Electrical Engineering and Physics,
Fujian University of Technology, Fuzhou 350118, China
lianhjsg@163.com

Jianxing Li[1,2,3,*]

[1]College of Electronic, Electrical Engineering and Physics,
Fujian University of Technology, Fuzhou 350118, China
[2]Technical Development Base of Industrial Integration Automation of Fujian Province,
Fuzhou 350118, China
[3]College of Automation Engineering,
Fujian Polytechnic of Water Conservancy and Electric Power, Sanming 366000, China
linocoo@163.com

Zhenyu Liu[3]

[3]College of Automation Engineering,
Fujian Polytechnic of Water Conservancy and Electric Power, Sanming 366000, China
1434212422@qq.com

Mengfei Chen[3]

[3]College of Automation Engineering,
Fujian Polytechnic of Water Conservancy and Electric Power, Sanming 366000, China
1229754764@qq.com

Rongkun Ye[1]

[1]College of Electronic, Electrical Engineering and Physics,
Fujian University of Technology, Fuzhou 350118, China
2231905031@smail.fjut.edu.cn

Junzo Watada[4]

[4]Graduate School of Information, Production and Systems,
Waseda University, Kitakyushu 808-0135, Japan
junzo.watada@gmail.com

[*]Corresponding author: Jianxing Li

Received May 19, 2025, revised September 13, 2025, accepted December 20, 2025.

Abstract. *To address the challenges in defect detection of radial tire X-ray images, such as significant variation in object scales, blurred boundaries, complex texture backgrounds, and weak feature expression, this paper proposes a defect detection algorithm named TDS-YOLO, based on an improved YOLOv11n architecture. The model incorporates structural enhancements in feature extraction, scale-aware perception, and semantic guidance to improve detection accuracy and robustness. First, a Multi-Scale Edge Information (MSEI) module is designed and incorporated into a newly constructed C3k2-MSEI block to replace the original C3k2 structure in the backbone. This enhancement strengthens texture modeling by integrating multi-scale edge features and improves the localization of blurred defect boundaries. Second, a Fine-grained Spatial Perception Module (FSPM), integrating spatial multi-scale encoding and fine-grained channel attention, replaces the SPPF module in the backbone, thereby enhancing the model's ability to perceive small and low-contrast defects. Third, a Feature-Focused Pyramid Network (FFPN) is introduced in the neck to guide the fusion of multi-source semantic features and reinforce the model's focus on critical regions. Experiments conducted on real-world tire X-ray datasets demonstrate that TDS-YOLO achieves a detection accuracy of 97.3% mAP@0.5 and 65.8% mAP@0.5:0.95 while maintaining a lightweight structure. Both precision and recall outperform YOLOv11n and other mainstream detection models. These results confirm that TDS-YOLO offers high detection performance and strong practical value for multi-scale, low-contrast defect detection in complex industrial environments.*

Keywords: Radial Tire, X-ray Image, Object Detection, YOLOv11, Multi-scale Modeling, Feature Enhancement

1. Introduction.

Tires, as critical components responsible for supporting loads and absorbing shocks in vehicles, directly affect driving stability and road safety through their manufacturing quality and structural integrity. In recent years, the rapid development of the global automotive industry has led to surging demand for tires, placing higher requirements on both performance and manufacturing processes [1, 2]. Among them, radial tires have become the dominant product in the industry due to their excellent wear resistance, low rolling resistance, and superior handling characteristics, and are widely used in passenger cars and transport vehicles [3]. However, the complex manufacturing process of radial tires is highly susceptible to the quality of raw materials, equipment conditions, and production environment, often resulting in internal structural defects. These defects are typically difficult to detect via visual inspection. If left undetected, they may lead to severe safety incidents during actual vehicle operation, posing significant risks to life and property [4, 5].

Currently, the detection of internal tire defects primarily relies on X-ray imaging for nondestructive inspection. However, X-ray images of radial tires are often characterized by complex background textures and low grayscale contrast. Moreover, the diversity of defect types and the presence of blurred boundaries result in ambiguous visual features, which further complicates defect recognition. In practice, many tire manufacturers still depend on manual inspection of X-ray images. This approach is not only inefficient but also heavily dependent on the experience and concentration of inspectors, making it difficult to ensure the accuracy and consistency of the detection results [6–8].

For defect detection in radial tire X-ray images, traditional image processing methods [9–11] often suffer from low detection accuracy and slow processing speed in practical applications. Moreover, these methods typically rely on manually set thresholds, making them difficult to adapt to the demands of modern tire manufacturing. In recent years, deep learning has shown tremendous potential in image recognition and defect detection tasks due to its superior feature extraction capabilities and generalization performance [12]. Li and Peng [13] proposed a defect segmentation method for tire X-ray images based on an improved DeepLabv3+ model. By integrating an encoder-decoder structure, the ASPP module, and a lightweight MobileNetv2 backbone, the model achieved a mean Intersection over Union (MIoU) of 81.42% on a custom dataset. However, its performance was limited when dealing with large-scale defects. Li et al. [14] introduced an industrial-oriented defect classification method named TireNet. This end-to-end framework combines a Siamese architecture with a class-balancing strategy, improving detection performance under complex structures and class imbalance conditions, and achieving high recall rates in experiments. Peng et al. [15] developed an improved YOLO-based model that incorporates an Efficient Layer Aggregation Network (ELAN), a cross-stage feature fusion module (SPPCSPC), and the CBAM attention mechanism, enhancing both feature representation and detection precision. On their custom dataset, the model reached a mAP of 91.3% with an inference time of 9.28 ms per image. Xu et al. [16] proposed another enhanced YOLOv5-based model that integrates Dynamic Snake Convolution (DSConv), a multi-scale feature fusion module, and CBAM. This model significantly improved the detection of elongated and small-scale defects, with mAP@0.5 and mAP@0.5:0.95 increasing by 5.9 and 5.7 percentage points respectively compared to the baseline YOLOv5. However, the study did not evaluate the model's real-time performance.

In recent years, single-stage object detection algorithms have gained widespread adoption in industrial defect recognition tasks due to their high inference efficiency and deployment flexibility. Among them, the YOLO series [17] has demonstrated strong application value across various industrial inspection scenarios, offering a favorable balance between detection speed and accuracy. Recent studies in secure intelligent systems and privacy-aware applications also reflect the growing importance of trustworthy AI in emerging industrial and vehicular environments [18, 19]. YOLOv11n, as a representative lightweight variant of the series, is chosen as the baseline in this study for its excellent performance in small object detection and its favorable balance between accuracy and computational cost. To further enhance its performance in complex scenarios of radial tire X-ray images, this paper proposes an improved detection algorithm named TDS-YOLO, based on the YOLOv11n architecture and enhanced with multi-level feature augmentation mechanisms, aiming to improve both detection performance and practical usability. The main contributions of this work are as follows:

(1) MSEI is embedded into the backbone by constructing a novel C3k2-MSEI block, which improves the network's sensitivity to complex textures and blurred defect boundaries through multi-scale edge feature integration.

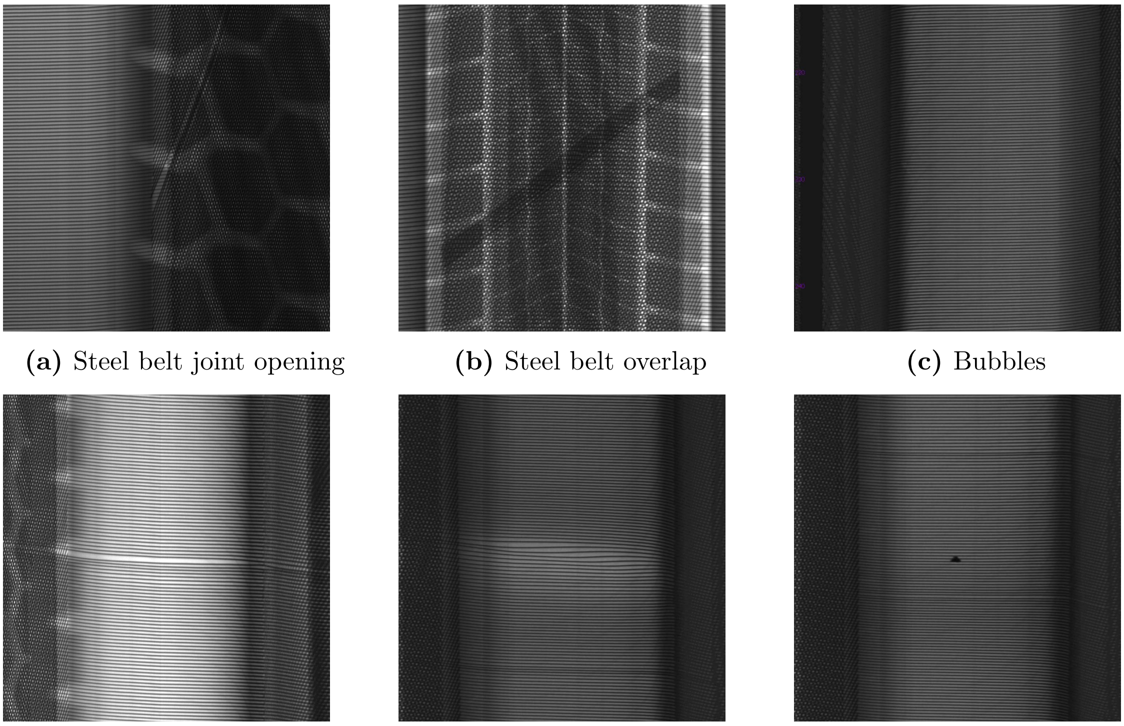
(2) FSPM is developed by combining spatial pyramid pooling with a Fused Stable Grained Channel Attention (FSGCA) mechanism, effectively enhancing the modeling precision for subtle or low-contrast defects.

(3) FFPN is constructed in the neck, leveraging multi-scale feature focusing and semantic-guided fusion to improve localization accuracy in key defect regions.

The proposed TDS-YOLO significantly improves the defect detection accuracy and overall robustness in radial tire X-ray images while maintaining a lightweight architecture.

2. **Dataset.** The dataset used in this study was acquired from real-world X-ray inspection equipment on the production line of Jianxin Tire (Fujian) Co., Ltd. It contains six typical types of internal defects commonly found in radial tires: steel belt joint opening, steel belt overlap, bubbles, steel wire alienation, steel wire sparsity, and internal impurities. Representative samples of each defect category are shown in Figure 1.

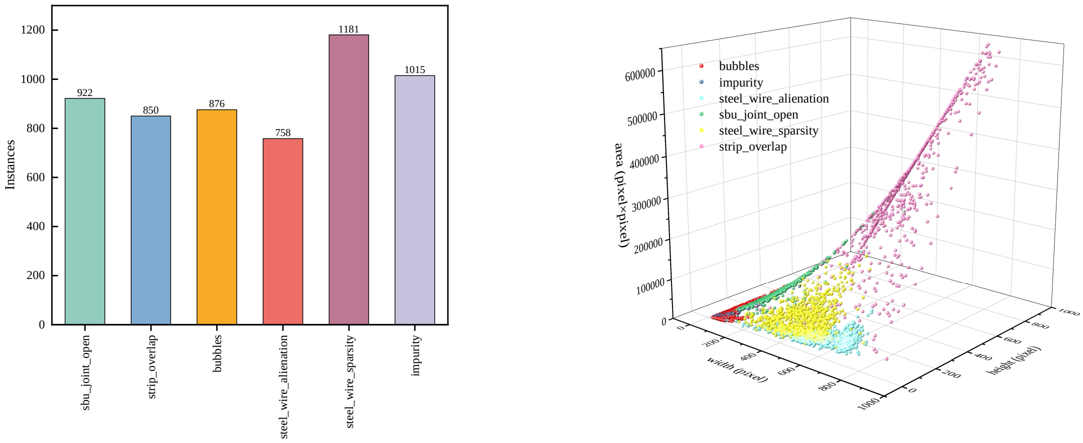
The annotation of defects was conducted using the LabelImg tool. A total of 5,602 annotated X-ray images were collected to form the radial tire defect detection dataset. The dataset was divided into training, validation, and test sets at a ratio of 8:1:1. The distribution of defect labels is presented in Figure 2(a), while Figure 2(b) illustrates the size distribution of various defect categories in terms of width and height. The results reveal substantial variation in target sizes, ranging from small-scale flaws to large structural anomalies, thereby highlighting the significant multi-scale characteristics of the detection task.



(a) Steel belt joint opening **(b)** Steel belt overlap **(c)** Bubbles

(d) Steel wire alienation **(e)** Steel wire sparsity **(f)** Impurities

Figure 1. Representative examples of defect samples.



(a) Label distribution **(b)** Size distribution

Figure 2. Statistical overview of the dataset: (a) number of instances per defect class, and (b) width–height distribution of defect bounding boxes.

3. **The YOLOv11 algorithm.** YOLOv11, as a significant upgrade in the YOLO series, achieves notable improvements in detection accuracy, inference efficiency, and multi-task

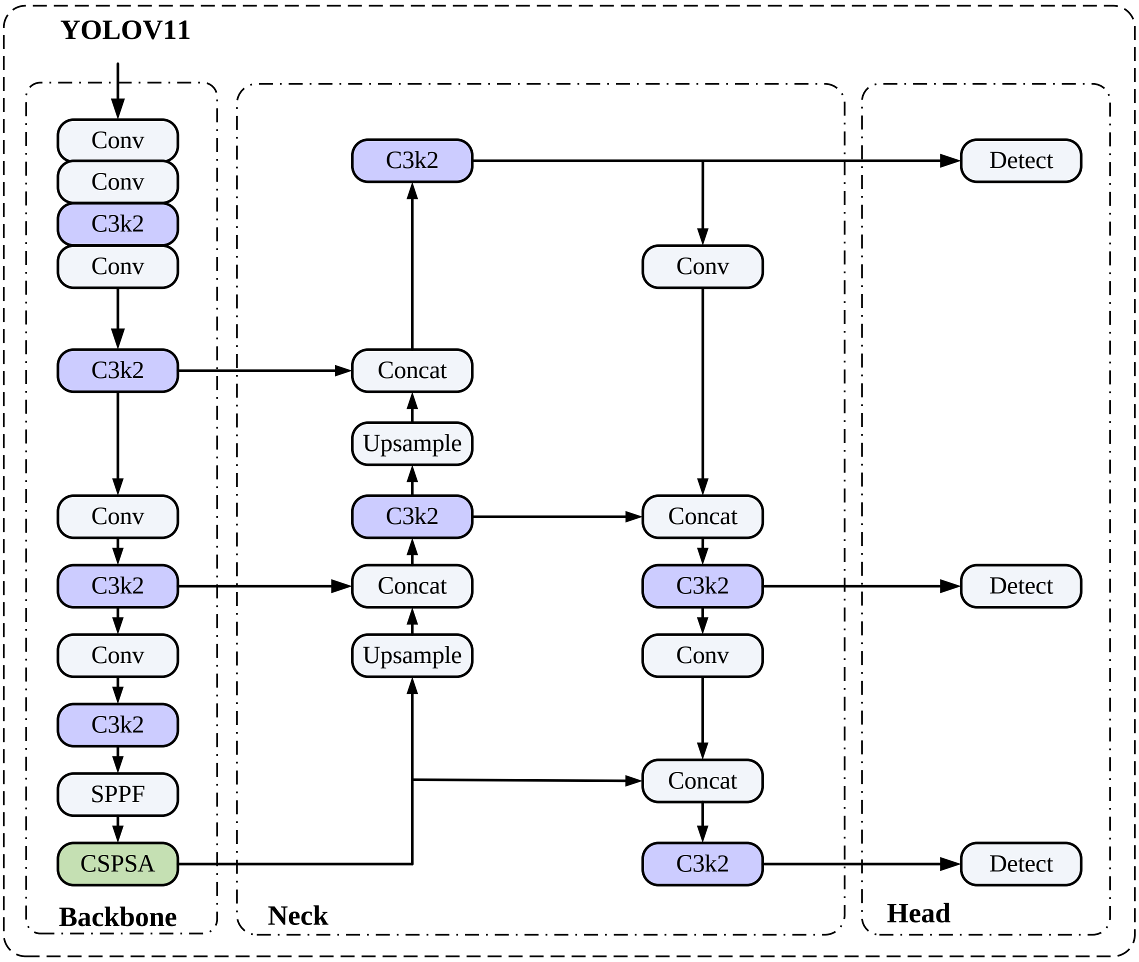


Figure 3. Overall architecture of YOLOv11 model.

adaptability. The model consists of a backbone, a feature fusion neck, and a detection head. It offers multiple scale variants (n, s, m, l, and x) to accommodate varying performance and lightweight deployment requirements across different application scenarios, as shown in Figure 3.

During the feature extraction stage, YOLOv11 is built upon a modified CSPDarknet architecture [20] by introducing the C3k2 module as a replacement for the original C2f structure [21]. This module combines the flexibility of C3k with the computational efficiency of C2f, thereby enhancing the modeling capability for shallow detail features. In addition, the model integrates the Spatial Pyramid Pooling - Fast (SPPF) module to strengthen multi-scale feature representation and incorporates a modified C2PSA module to introduce multi-dimensional attention mechanisms, improving information interaction across both channel and spatial dimensions.

The neck integrates Feature Pyramid Network (FPN) [22] and Path Aggregation Network (PAN) [23] structures to effectively fuse multi-level features and enhance responsiveness to objects of various scales. The detection head adopts a decoupled design that separately predicts class and location, while the inclusion of depthwise separable convolution (DWConv) [24] reduces computational overhead and maintains high inference efficiency. In addition, YOLOv11 exhibits good scalability and modularity, offering potential for extension to multi-task scenarios such as image segmentation and pose estimation.

4. **Improved YOLOv11n Model.** In the analysis of radial tire X-ray images, the presence of diverse defect categories, fuzzy boundaries, and complex background textures significantly increases the difficulty of accurate defect localization and classification. Directly applying the baseline YOLOv11n model often leads to unstable detection performance and reduced robustness under such conditions. Although increasing the model size can enhance feature representation to a certain extent, it inevitably results in higher computational complexity, increased memory usage, and inference latency, which hinders its practicality for real-time industrial deployment. To address these challenges and achieve a better trade-off between detection accuracy and deployment efficiency, this paper proposes an enhanced network architecture based on YOLOv11n, termed TDS-YOLO. The overall structure of the proposed model is illustrated in Figure 4, highlighting its targeted improvements for industrial tire defect detection applications.

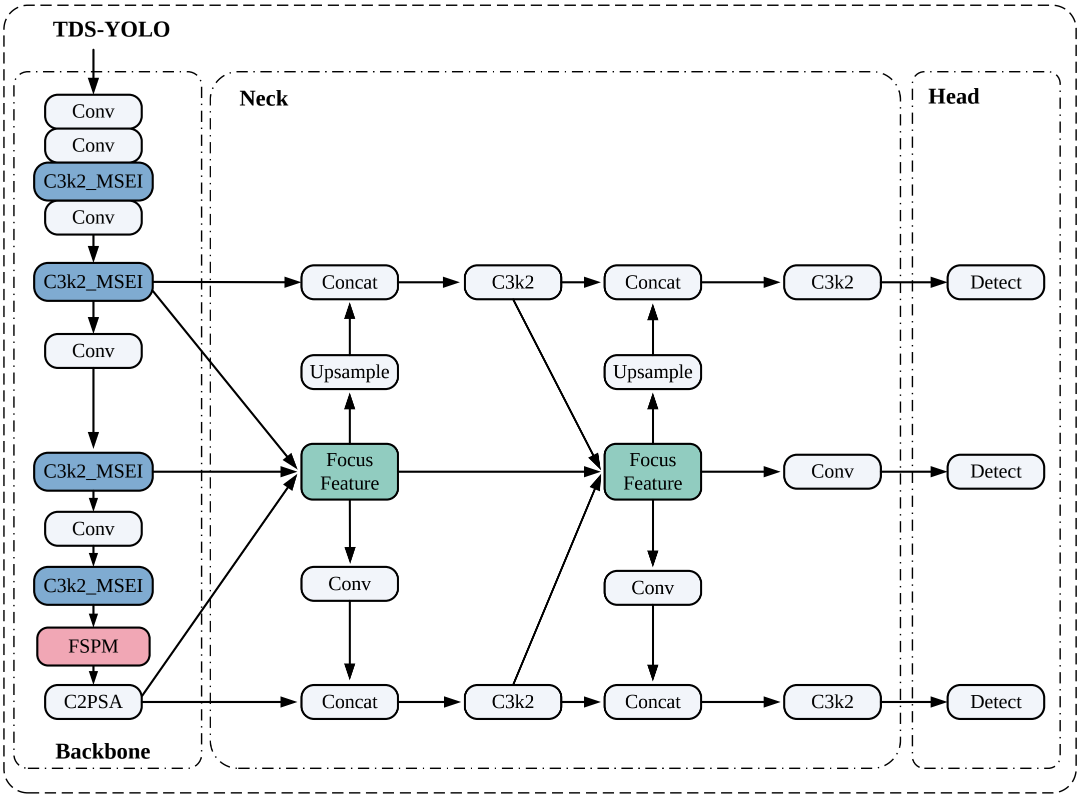


Figure 4. Architecture of the proposed TDS-YOLO model.

4.1. **Improved C3k2 Module.** X-ray images of radial tires often exhibit challenges such as significant variation in defect sizes, blurred boundaries, and complex background textures. Although the C3K2 module adopted in the YOLOv11 backbone offers lightweight advantages, its small convolutional kernel size (2 × 2) limits the overall receptive field. As a result, it struggles to effectively capture large-scale structures and edge details, leading to suboptimal detection performance, especially for small defects and ambiguous boundaries. To enhance the feature representation capability and detection stability of the model under complex defect scenarios, this paper designs a MSEI module, which integrates multi-scale contextual modeling, edge information enhancement, and dynamic feature selection. Based on this, a C3k2-MSEI module is constructed to replace the original C3K2 structure in the backbone network. The architecture is illustrated in Figure 5 and Figure 6.

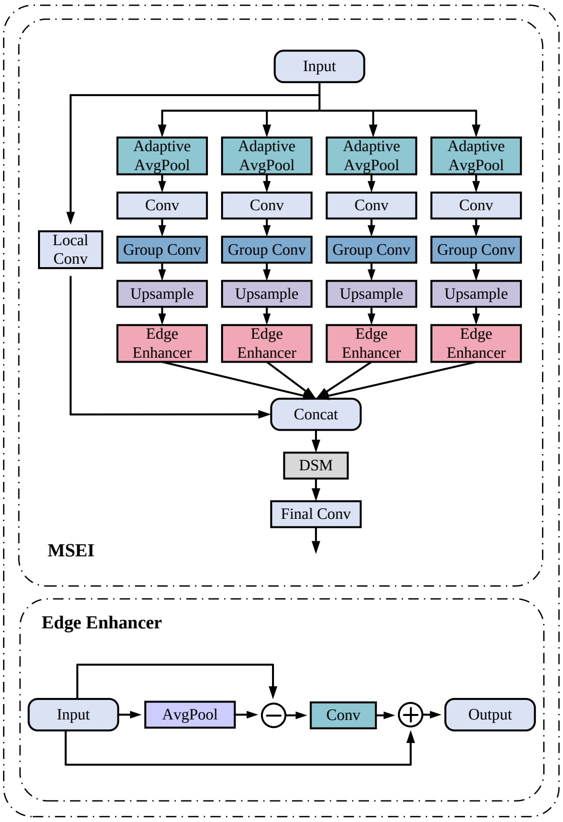


Figure 5. Structure of the MSEI module.

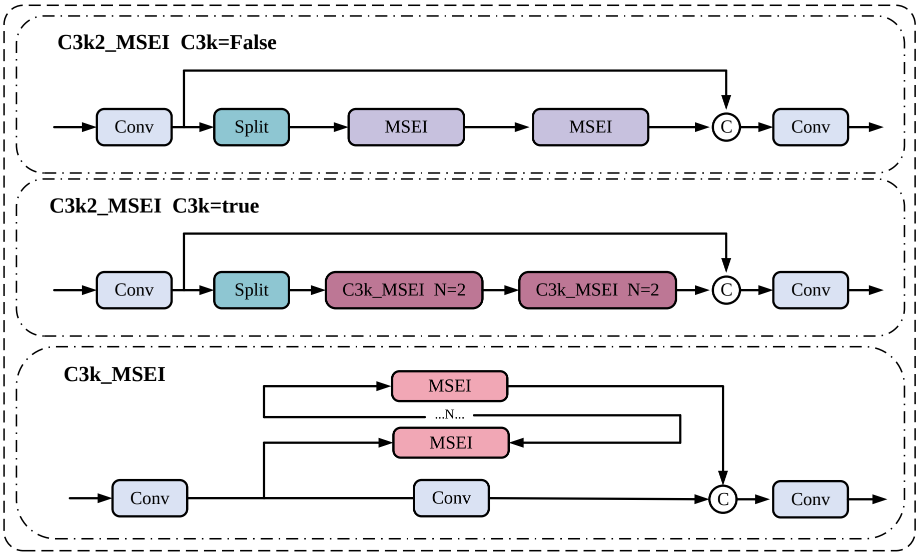


Figure 6. Structural diagram of the C3k2_MSEI module.

The MSEI module incorporates multi-scale feature extraction, edge-aware enhancement, and adaptive feature selection to comprehensively improve the network's ability to model and express features. Specifically, the MSEIS submodule first employs a multi-scale feature extraction unit to obtain hierarchical contextual representations through adaptive pooling operations with varying scales (3, 6, 9, and 12), thereby facilitating comprehensive perception of defects of different sizes. Subsequently, an EdgeEnhancer unit is constructed to strengthen the model's response to blurry boundary regions. This unit performs local average pooling on the input feature map to extract low-frequency components, then subtracts them from the original features to obtain high-frequency residual signals, highlighting edge contours and fine details. These enhanced edge features are further refined and integrated using convolutional operations, and finally fused back with the original features through a residual connection, thereby achieving explicit edge enhancement and effectively mitigating the adverse effects of blurred boundaries on detection performance. The corresponding mathematical formulation is expressed as follows:

$$E_{\text{out}} = X + \sigma\left(X - \text{AvgPool}(X)\right) \tag{1}$$

In the above formulation, X denotes the input feature map, $\text{AvgPool}(\cdot)$ represents the local smoothing information extracted via average pooling, $\sigma(\cdot)$ is the Sigmoid activation function, and E_{out} refers to the output feature after edge enhancement.

To further optimize the fusion and interaction of multi-scale features, all branch features are first aligned through interpolation and then concatenated. Subsequently, a Dual Domain Selection Mechanism (DSM) [25] is introduced to dynamically model the dependencies between features from different domains. This mechanism enables adaptive selection of key features that are highly relevant to the detection task, effectively suppressing background interference and enhancing target-specific information. The overall computation process can be abstracted as follows:

$$F_{\text{out}} = \alpha \cdot \text{LocalAttention}(\text{SpatialGate}(X)) + \beta \cdot X \tag{2}$$

In this expression, X denotes the input feature map; $\text{SpatialGate}(\cdot)$ represents the spatial attention gating module, which selects spatial regions with high significance; $\text{LocalAttention}(\cdot)$ is the local context modeling unit that enhances fine-grained feature representation in the target regions; $\beta, \alpha \in \mathbb{R}^{C\times1\times1}$ is a learnable parameter in the channel dimension used to control the fusion ratio between the enhanced feature and the original input; and F_{out} denotes the optimized feature map output by the DSM module.

Building upon the lightweight characteristics of the original C3K2 structure, the proposed C3k2-MSEIS module further integrates multi-scale contextual modeling, edge information enhancement, and dynamic feature fusion mechanisms. This design significantly improves the model's capability in representing complex textures and detecting defects of varying scales. While effectively enhancing the detection accuracy of radial tire X-ray image defects, the module also strengthens the model's practical deployability and adaptability in real-world industrial environments.

4.2. **Enhancement of the SPPF Module.**

4.2.1. *FSGCA Attention Mechanism.* Attention mechanisms are effective in enhancing a model's ability to focus on target regions and have been widely applied in image defect detection tasks. However, in industrial scenarios such as radial tire X-ray image inspection—where the detection task is highly sensitive to subtle defects and the feature representations are often weak—traditional attention mechanisms typically rely on fully connected layers to model global inter-channel dependencies. This approach lacks effective integration with local channel features, resulting in inaccurate feature weighting.

To address this issue, we propose a dedicated attention mechanism for this domain, named FSGCA, as illustrated in Figure 7. Unlike classical attention modules such as SE [26] and CBAM [27], the core strength of FSGCA lies in its ability to model fine-grained and dynamic dependencies between channels. It further enhances feature stability and improves the adaptability of weight allocation during the feature reweighting and adjustment process.

Specifically, the FSGCA module begins by applying global average pooling (GAP) to the input feature map $F \in \mathbb{R}^{C\times H\times W}$ in order to extract global descriptors for each channel.The corresponding expression is defined as follows:

$$Q_n = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_n(i,j) \tag{3}$$

In the above equation, Q_n denotes the global descriptor of the n-th channel; $F_n(i,j)$ represents the pixel value at spatial position (i,j) of the n-th channel in the input feature map; and H and W are the height and width of the feature map, respectively.

To model both local and global inter-channel dependencies, FSGCA introduces two parallel branches: a local modeling path and a global modeling path. The local branch employs one-dimensional convolution (Conv1D), while the global branch adopts fully connected (FC) layers to extract features from the channel-wise descriptor vector. The corresponding operations are formulated as follows:

$$Q_l = \mathrm{Conv1D}(Q) \tag{4}$$

$$Q_g = \mathrm{FC}(Q) \tag{5}$$

In the above expressions, Q_l denotes the local channel feature, and Q_g represents the global channel feature.Based on this, cross-correlation is applied to compute the fine-grained dependencies between the local and global features, resulting in the channel correlation matrix R.

$$R = Q_g \cdot Q_l^T \tag{6}$$

In the above expression, R denotes the cross-correlation matrix between channels, which captures the fine-grained response relationships between global and local channel features.

To further extract attention cues, FSGCA performs row-wise and column-wise summation on matrix R, yielding the global attention weight ω_g and the local attention weight ω_l. The computation is defined as follows:

$$\omega_g = \sum_j R_{i,j} \tag{7}$$

$$\omega_l = \sum_j R_{j,i} \tag{8}$$

In this equation, ω_g and ω_l represent the global and local channel attention weights, respectively.

To enhance the model's sensitivity to local features, FSGCA adopts a dynamic fusion strategy during the weight integration phase. A learnable fusion factor θ is introduced to adaptively control the weighting ratio between local and global features. To mitigate overfitting in the attention distribution and promote diversity in channel selection, a Dropout-based regularization mechanism is applied after the fusion. The corresponding computation is as follows:

$$\omega = \mathrm{Dropout}(\sigma(\theta \cdot \sigma(\omega_l) + (1-\theta) \cdot \sigma(\omega_g))) \tag{9}$$

In this equation, ω denotes the fused channel-wise weighting coefficient, θ is a learnable channel fusion factor, and $\sigma(\cdot)$ refers to the Sigmoid activation function. The Dropout operation is applied to randomly suppress certain channel responses, encouraging feature sparsity and diversity while alleviating overfitting in the attention distribution.

Finally, FSGCA performs channel reweighting and residual fusion by combining the weighted features with the original input features. This process completes the feature recalibration and generates the final output representation:

$$\widetilde{F} = \omega \odot F + F \tag{10}$$

In the above equation, $\widetilde{F}$ denotes the recalibrated output feature map, and $\odot$ represents the element-wise multiplication operation along the channel dimension.

In summary, the FSGCA module adopts a dual-path modeling scheme that captures both local and global channel dependencies with fine granularity, effectively addressing the limited ability of conventional attention mechanisms to perceive local structures. The proposed dynamic fusion strategy not only improves the model's sensitivity to critical channels but also increases the diversity and robustness of channel features. Compared to mainstream attention mechanisms, FSGCA provides more precise channel-wise weighting and stronger feature enhancement, all with maintained computational efficiency. These advantages make it particularly suitable for industrial inspection tasks involving radial tire X-ray images, where sensitivity to subtle defects and boundary details is crucial.

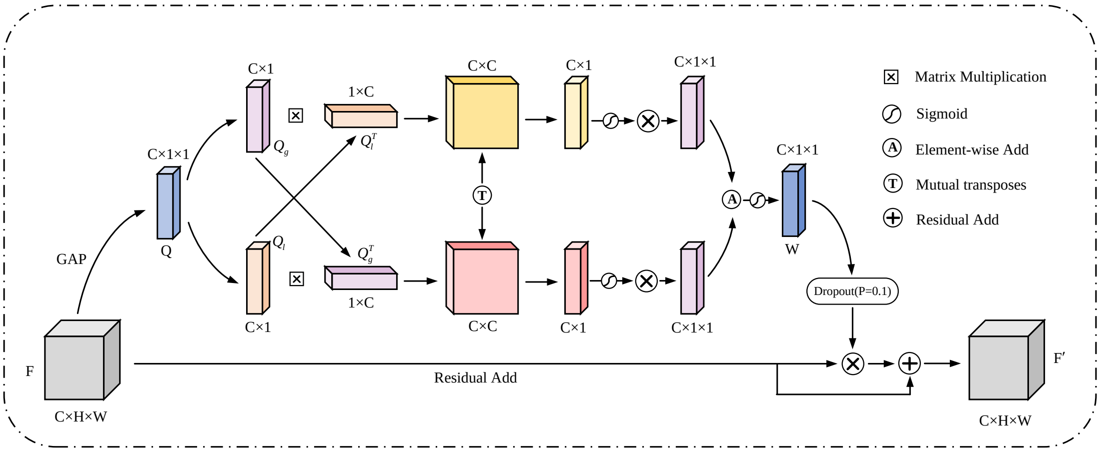


Figure 7. Structure of the proposed FSGCA attention mechanism.

4.2.2. *FSPM Module.* The Spatial Pyramid Pooling Fast (SPPF) module in the YOLOv11n network enables rapid multi-scale feature fusion in the spatial domain, capturing rich information from objects at various scales. However, this design primarily focuses on spatial-level aggregation while lacking detailed feature analysis and adaptive weighting along the channel dimension. To address this limitation and further improve the detection of small defects in radial tire X-ray images, this study integrates the proposed FSGCA attention mechanism into the SPPF structure, resulting in a new module termed the FSPM, as illustrated in Figure 8. This module effectively combines spatial multi-scale fusion with fine-grained dynamic channel weighting, thereby compensating for the original SPPF's limited capacity in channel-wise information modeling.

4.3. **Improved Neck Structure.** To enhance the network's representation capability for multi-scale defect detection—especially in detecting small objects and handling blurry-edge regions—we propose the FFPN as an improved neck module. Building upon the core design of the traditional FPN, FFPN incorporates a feature-focusing mechanism and multi-level guidance strategy to strengthen the representation of key regions and improve cross-scale feature modeling, as illustrated in Figure 9.

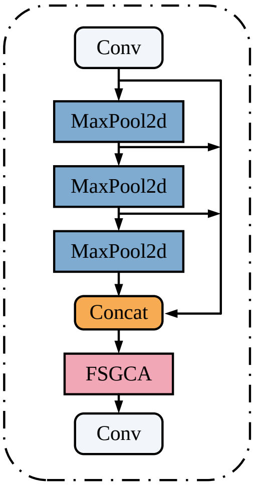


Figure 8. Architecture diagram of the FSPM module.

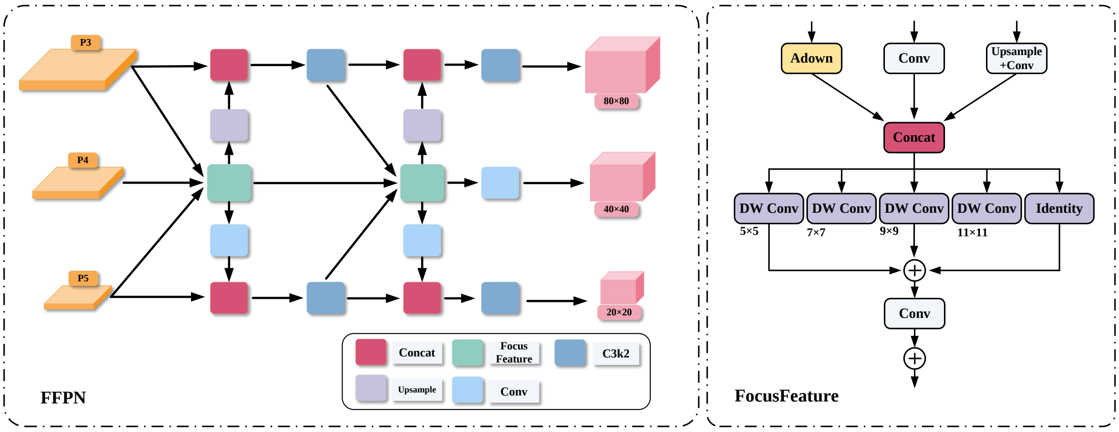


Figure 9. Architecture of the Feature-Focused Pyramid Network (FFPN).

Within the multi-scale fusion pathway, FFPN introduces a Focus Feature module that utilizes spatial attention and local context enhancement to actively identify critical response regions in shallow layers and apply weighted reinforcement to these important locations. Unlike conventional FPNs that rely solely on progressive upsampling, FFPN also facilitates the propagation of deep semantic information to lower layers, promoting effective feature interaction and structural alignment across different levels. This enhances the model's adaptability to small targets and complex structural defects.

The Focus Feature module serves as the core unit for multi-scale feature integration, aiming to enhance the collaborative representation across different feature levels. This module focuses on integrating shallow details, mid-level semantics, and deep contextual information through a structured design, as illustrated in Figure 9. Structurally, the Focus Feature module receives three feature maps from different hierarchical levels, which are processed via upsampling, 1×1 convolution (for channel compression), or downsampling to unify the channel dimensions before being concatenated. The concatenated features are then passed through a set of depthwise separable convolutions (DWConv) with multiple kernel sizes (5, 7, 9, and 11) to capture local contextual information at different receptive fields. These multi-scale responses are fused via channel-wise summation to enhance spatial modeling capability. The fused features are further compressed using pointwise convolution and then adaptively fused with the original concatenated features in a residual manner. This module effectively improves the model's focus on object boundaries, textures, and fine-grained regions while maintaining low computational overhead. Consequently, it enhances detection performance for small and edge-blurred targets.

Experimental results demonstrate that the FFPN module significantly improves the model's detection capability under multi-scale scenarios, particularly in recognizing small objects and indistinct boundaries, thereby enhancing the overall robustness and practical applicability of the system.

5. Experimental Results and Analysis.

5.1. Experimental Environment and Parameter Settings. The experiments in this study were conducted on an Ubuntu 20.04 operating system using an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory. The experimental platform was built upon CUDA 12.1, Python 3.8, and the PyTorch 2.2.2 framework. No pre-trained weights were used during training, and the model was optimized using the SGD optimizer. Detailed parameter settings are listed in Table 1.

Table 1. Training Parameter Settings

Parameter Name	Value
Epochs	200
Batch Size	64
Initial Learning Rate	0.01
Momentum	0.937
Weight Decay	0.0005

5.2. Evaluation Metrics. To evaluate the performance of the proposed algorithm on the defect detection task for radial tire X-ray images, several evaluation metrics are employed, including Precision (P), Recall (R), mean Average Precision (mAP), parameter count (Params), computational complexity (GFLOPs), and inference time per image (Inference). Specifically, Params reflect the structural complexity of the model, where larger values imply increased memory and storage consumption. GFLOPs represent the computational cost during inference; higher values indicate greater resource demands. The mAP serves as an indicator of the overall detection accuracy; higher mAP values correspond to better performance. Inference time denotes the average processing duration per image and is indicative of the model's efficiency in real-time applications. The formulas for each metric are defined as follows:

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

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

$$\mathrm{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{13}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}(i) \tag{14}$$

$$\text{Inference Time} = \frac{T_{\text{Inference}}}{N} \tag{15}$$

In the above formulas, TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively. AP represents the average precision, which is computed as the area under the precision–recall (P–R) curve. The mean average precision (mAP) is the mean value of AP over all classes. Specifically, mAP@0.5 is computed at a fixed IoU threshold of 0.5, while mAP@0.5:0.95 averages AP over multiple IoU thresholds from 0.5 to 0.95 in steps of 0.05. $T_{\text{Inference}}$ indicates the total inference time, and N denotes the total number of test images.

5.3. **Comparison of Attention Mechanisms.** To evaluate the effectiveness of the attention mechanism integrated into the proposed FSPM module, a series of comparative experiments were conducted. In these experiments, the original attention mechanism in FSPM was individually replaced with several widely adopted attention modules, including Coordinate Attention (CA) [28], Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation (SE), and Simple Attention Module (SimAM) [29]. All experiments were performed under identical network structures and training configurations to ensure fair comparison. The variations in detection performance for different attention mechanisms were evaluated using mAP@0.5 and mAP@0.5:0.95, and the corresponding performance curves are illustrated in Figure 10.

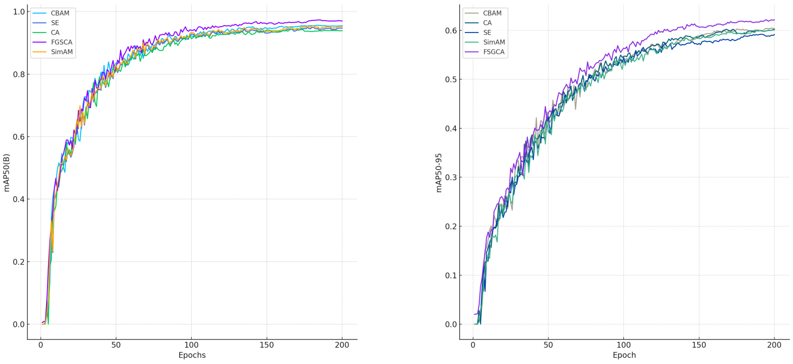


Figure 10. Comparison of mAP values for integrated attention mechanisms.

Table 2. The performance comparison of the FSPM module with different integrated attention mechanisms.

Attention Mechanism	P/%	R/%	mAP_{50}/%	Params/M
SE	92.1	88.2	93.4	2.77
CBAM	92.4	90.1	94.1	2.89
CA	92.9	91.2	92.8	2.79
SimAM	92.2	91.6	93.8	2.65
FSGCA	93.3	92.4	95.1	2.87

As shown in Figure 10, the proposed FSGCA outperforms other attention mechanisms in terms of detection performance. This improvement is attributed to its fine-grained inter-channel dependency modeling and dynamic weight adjustment strategy, which enable more effective differentiation between subtle defect regions and complex backgrounds.

Furthermore, Table 2 presents the detailed evaluation metrics for each attention mechanism. Experimental results demonstrate that FSGCA achieves 95.1% mAP@0.5, 92.4% recall, and 93.3% precision on the radial tire X-ray defect detection task, while maintaining a lightweight design with only 2.87M parameters. This balance between accuracy and model efficiency highlights the practical value of FSGCA in industrial inspection scenarios.

5.4. **Ablation Study.** To comprehensively verify the effectiveness of the proposed improvements, six sets of ablation experiments were conducted. All experiments were performed under identical environmental configurations and parameter settings. The results are summarized in Table 3, where "✓" indicates that the corresponding module is included.

Table 3. Ablation Experiment Results

Baseline	C3k2 -MSEI	FSPM	FFPN	P/%	R/%	mAP $_{50}$/%	mAP $_{50\text{-}95}$/%	Params (M)
✓				92.6	90.2	93.7	61.2	2.62
✓	✓			92.3	92.6	95.6	63.2	2.61
✓		✓		93.3	92.4	95.1	63.1	2.87
✓		✓	✓	94.4	94.3	96.5	64.9	2.76
✓	✓		✓	92.8	93.6	96.1	63.6	2.86
✓	✓	✓	✓	95.2	95.6	97.3	65.8	3.01

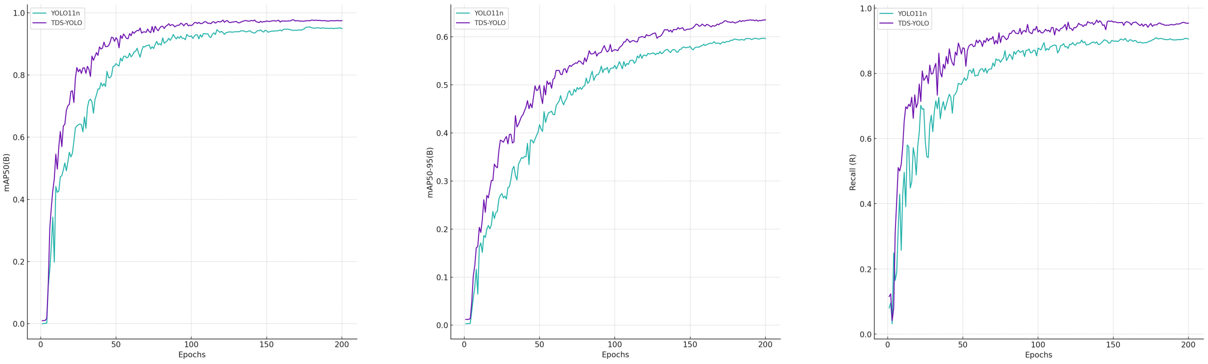


Figure 11. Comparison of TDS-YOLO and the baseline model in terms of performance metrics (mAP@0.5, mAP@0.5:0.95, and Recall).

As shown in Table 3, introducing individual modules such as C3k2-MSEI, FSPM, or FFPN into the YOLOv11n baseline model results in noticeable improvements in detection performance, while maintaining a reasonable level of model complexity and computational cost. Further evaluation was conducted through incremental integration of the proposed modules.

Specifically, replacing all C3k2 blocks in the backbone with C3k2-MSEI led to a 1.9% increase in mAP@0.5 and a 2.0% gain in mAP@0.5:0.95, demonstrating its effectiveness in multi-scale feature modeling and edge information representation. Subsequently, substituting the original SPPF module with the proposed FSPM resulted in additional improvements of 0.5% in mAP@0.5 and 0.4% in mAP@0.5:0.95. This highlights the enhanced ability of FSPM to capture fine-grained defect signals and distinguish them from complex backgrounds by combining spatial multi-scale fusion with channel-wise attention. Building upon this, the introduction of the FFPN neck structure led to further gains of 2.4% in Precision, 1.2% in mAP@0.5, and 2.2% in mAP@0.5:0.95, reflecting the advantage of multi-scale feature focusing and guidance in detecting complex structural defects.

Ultimately, the integrated TDS-YOLO model incorporating all proposed modules achieved the best overall performance, reaching 95.2% Precision, 95.6% Recall, 97.3% mAP@0.5, and 65.8% mAP@0.5:0.95. Notably, these improvements were obtained without significantly increasing the parameter count, validating both the effectiveness of each module and the coherence of the overall design. Furthermore, Figure 11 presents a comparative analysis between TDS-YOLO and the baseline model in terms of mAP@0.5, mAP@0.5:0.95, and Recall, clearly demonstrating the superior detection performance of the proposed approach.

5.5. **Comparative Experiments.** To comprehensively evaluate the overall performance of the proposed TDS-YOLO model, several representative object detection algorithms were selected for comparison. These include the two-stage method Faster R-CNN [30], the classic lightweight single-stage SSD [31], and multiple versions from the YOLO series, such as YOLOv5n, YOLOv8n, YOLOv10n, YOLOv11n, YOLOv11s, and YOLOv12n [32]. All models were trained and tested under the same dataset and experimental conditions. Evaluation metrics include Precision (P), Recall (R), mAP@0.5, number of parameters (Params), computational complexity (GFLOPs), and inference time per image (Inference/ms).

Table 4. Comparison of Detection Algorithms

Model	P/%	R/%	mAP_{50}/%	Params (M)	GFLOPs	Inference (ms)
Faster R-CNN	88.3	79.4	86.9	43.39	221	24.8
SSD	83.9	80.8	84.2	22.4	65.3	27.1
YOLOv5n	89.4	87.6	91.4	2.51	7.1	2.1
YOLOv8n	91.9	89.8	91.8	3.16	8.9	2.4
YOLOv10n	89.3	85.9	89.9	2.76	8.4	3.2
YOLOv11n	92.6	90.2	93.7	2.62	6.6	1.8
YOLOv11s	93.1	92.7	94.2	9.45	21.7	4.1
YOLOv12n	90.5	91.2	93.9	2.61	6.7	2.2
TDS-YOLO	95.2	95.6	97.3	3.01	9.8	3.0

As shown in Table 4, Faster R-CNN achieves only 86.9% mAP@0.5 with an inference latency of 24.8 ms, which remains inadequate for real-time industrial inspection tasks. While SSD consumes slightly fewer resources, its detection accuracy is only 84.2%, with a relatively low recall, indicating limited overall performance.

In contrast, the YOLO series demonstrates a better trade-off between accuracy and efficiency. Models like YOLOv5n, YOLOv8n, and YOLOv12n exceed 90% mAP@0.5 with inference times controlled within 2–3 ms, showing practical deployment potential. Among them, YOLOv11n achieves the best performance in lightweight settings, with 93.7% mAP@0.5 and a fast inference time of only 1.8 ms, indicating strong practicality. Although YOLOv11s delivers higher precision, its significantly increased parameter count and complexity limit its suitability for resource-constrained environments.

Building on these, the proposed TDS-YOLO model achieves further improvements while maintaining a relatively low parameter count (3.01M) and reasonable computational complexity (9.8 GFLOPs). It reaches 97.3% mAP@0.5, with 95.2% Precision and 95.6% Recall. Although the inference time is slightly higher than that of YOLOv11n, the proposed model demonstrates the most balanced performance overall, confirming its effectiveness and practical value for radial tire X-ray defect detection tasks in industrial settings.

6. **Visualization Analysis.** To provide an intuitive evaluation of TDS-YOLO's performance across different defect types, a set of visualization analyses was conducted using samples from six defect categories. These visual results were compared with those of the YOLOv11n model, as shown in Figure 12. The comparison includes original X-ray images, detection output, and heatmaps generated using the HiResCAM [33] method. HiResCAM is adopted in this study for heatmap generation, as it provides more accurate and higher-resolution feature activation visualizations, making it more suitable than traditional Grad-CAM methods for precisely analyzing the model's attention to defect regions.

From the detection results, TDS-YOLO consistently delivers higher confidence predictions across all defect categories, particularly demonstrating greater stability and accuracy in low-contrast and edge-blurred regions. In the heatmap comparison, TDS-YOLO generates more focused and precise high-response regions for localized defects such as "internal impurities," "bubbles," and "steel belt joint opening." The red-highlighted areas closely align with actual defect locations, whereas YOLOv11n exhibits more scattered and inaccurate activation responses. For structurally complex and spatially extensive defects, such as "steel wire alienation," "steel wire sparsity," and "steel belt overlap," TDS-YOLO is also capable of producing more complete and spatially continuous heatmaps, which better capture the overall defect contours.

In summary, the visualization results demonstrate TDS-YOLO's advantages in multi-scale object perception, spatial feature extraction, and accurate defect localization, underscoring its practical value for deployment in complex industrial inspection scenarios.

7. **Conclusion.** To enhance the detection performance of multi-scale defects in radial tire X-ray images, this paper proposes TDS-YOLO, a detection algorithm built upon YOLOv11n with multiple structural enhancements. To address limitations in feature extraction and scale modeling, a C3k2-MSEI module is designed to replace the original C3k2 in the backbone, thereby improving the network's ability to model complex edges and multi-scale targets. Additionally, the SPPF is replaced with the proposed FSPM, which integrates multi-scale modeling with fine-grained channel attention to improve the perception of small defects. Furthermore, FFPN is introduced to aggregate multi-source semantic features and guide the model toward critical defect regions, enhancing spatial localization accuracy.

Experimental results demonstrate that TDS-YOLO surpasses mainstream detection models with 97.3% mAP@0.5, 65.8% mAP@0.5:0.95, 95.2% Precision, and 95.6% Recall. It improves the detection of small and blurred defects while maintaining efficient inference. Ablation and visualization studies confirm the effectiveness of the proposed modules and the model's suitability for complex industrial scenarios.

Future work will focus on designing more efficient lightweight attention mechanisms and optimizing contextual modeling strategies to further improve detection robustness under complex backgrounds.

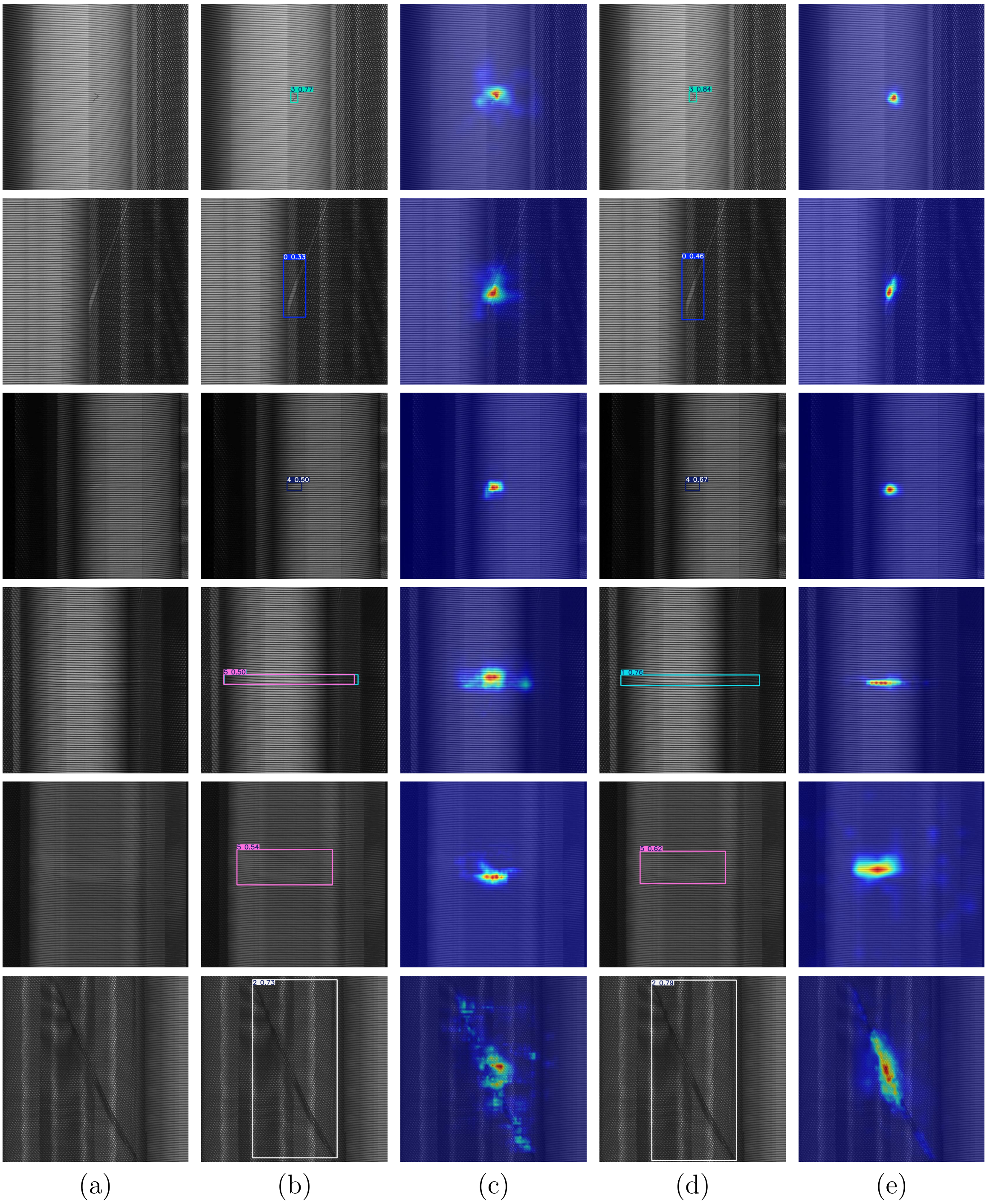


Figure 12. Visualization results of different defect types using YOLOv11n and the proposed TDS-YOLO across six defect categories. Each group of five images represents: (a) original X-ray image with defect; (b) detection result of YOLOv11n; (c) HiResCAM heatmap of YOLOv11n; (d) detection result of TDS-YOLO; and (e) HiResCAM heatmap of TDS-YOLO.

Acknowledgment. This work is supported by the Natural Science Foundation of Fujian Province of China (No. 2020J01876), and the Education and Scientific Research Project for Young and Middle-aged Teachers in Fujian Province (No. JAT231218).

REFERENCES

[1] B. Shoul, Y. Marfavi, B. Sadeghi, E. Kowsari, P. Sadeghi, and S. Ramakrishna, "Investigating the potential of sustainable use of green silica in the green tire industry: A review," *Environmental Science and Pollution Research*, vol. 29, no. 34, pp. 51 298–51 317, 2022.

[2] J. Gaidhane, I. Ullah, and A. Khalatkar, "Tyre remanufacturing: A brief review," *Materials Today: Proceedings*, vol. 60, pp. 2257–2261, 2022.

[3] R. Wang, Q. Guo, S. Lu, and C. Zhang, "Tire defect detection using fully convolutional network," *IEEE Access*, vol. 7, pp. 43 502–43 510, 2019.

[4] Z. Zheng, H. Yang, L. Zhou, B. Yu, and Y. Zhang, "Hlu 2-net: a residual u-structure embedded u-net with hybrid loss for tire defect inspection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[5] D. Jeong, S. Kim, J. Lee, S. B. Choi, M. Kim, and H. Lee, "Estimation of tire load and vehicle parameters using intelligent tires combined with vehicle dynamics," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2020.

[6] Y. Zhang, N. Gu, X. Zhang, and C. Lin, "Tire x-ray image defects detection based on adaptive thresholding method," in *Parallel Architectures, Algorithms and Programming: 10th International Symposium, PAAP 2019, Guangzhou, China, December 12–14, 2019, Revised Selected Papers 10*. Springer, 2020, pp. 118–129.

[7] M. Rafiei, J. Raitoharju, and A. Iosifidis, "Computer vision on x-ray data in industrial production and security applications: A comprehensive survey," *IEEE Access*, vol. 11, pp. 2445–2477, 2023.

[8] M. Tamborski, I. Rojek, and D. Mikołajewski, "Revolutionizing tire quality control: Ai's impact on research, development, and real-life applications," *Applied Sciences*, vol. 13, no. 14, p. 8406, 2023.

[9] G. Zhao and S. Qin, "High-precision detection of defects of tire texture through x-ray imaging based on local inverse difference moment features," *Sensors*, vol. 18, no. 8, p. 2524, 2018.

[10] Q. Guo, C. Zhang, H. Liu, and X. Zhang, "Defect detection in tire x-ray images using weighted texture dissimilarity," *Journal of Sensors*, vol. 2016, no. 1, p. 4140175, 2016.

[11] Y. Zhang, D. Lefebvre, and Q. Li, "Automatic detection of defects in tire radiographic images," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 3, pp. 1378–1386, 2015.

[12] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897–3904, 2023.

[13] X. Li and C. Peng, "Tire defect detection based on improved deeplabv3+," in *2023 42nd Chinese Control Conference (CCC)*. IEEE, 2023, pp. 7465–7470.

[14] Y. Li, B. Fan, W. Zhang, and Z. Jiang, "Tirenet: A high recall rate method for practical application of tire defect type classification," *Future Generation Computer Systems*, vol. 125, pp. 1–9, 2021.

[15] C. Peng, X. Li, and Y. Wang, "Td-yoloa: An efficient yolo network with attention mechanism for tire defect detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.

[16] G. Xu, A. Li, X. Wang, C. Xu, J. Chen, and F. Zheng, "Research on x-ray nondestructive defect detection method of tire based on dynamic snake convolution yolo model," *Scientific Reports*, vol. 14, no. 1, p. 29587, 2024.

[17] 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.

[18] T.-Y. Wu, H. Wu, M. Tang, S. Kumari, and C.-M. Chen, "Unleashing the potential of metaverse in social iov: An authentication protocol based on blockchain," *CMC-Computers, Materials & Continua*, vol. 84, no. 2, pp. 3175–3192, 2025.

[19] C.-M. Chen, Q. Miao, S. Kumar, and T.-Y. Wu, "Privacy-preserving authentication scheme for digital twin-enabled autonomous vehicle environments," *Transactions on Emerging Telecommunications Technologies*, vol. 34, no. 1, p. e4751, 2023. [Online]. Available: https://doi.org/10.1002/ett.4751

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[21] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with yolov8," *arXiv preprint arXiv:2305.09972*, 2023.

[22] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2117–2125.

[23] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8759–8768.

[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[25] Y. Cui, W. Ren, X. Cao, and A. Knoll, "Focal network for image restoration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 001–13 011.

[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[27] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 3–19.

[28] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 713–13 722.

[29] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 863–11 874.

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[31] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[32] Y. Tian, Q. Ye, and D. Doermann, "Yolov12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, 2025.

[33] R. L. Draelos and L. Carin, "Use hirescam instead of grad-cam for faithful explanations of convolutional neural networks," *arXiv preprint arXiv:2011.08891*, 2020.