Scientific Research Publishing

Open Journal of Applied Sciences, 2026, 16(7), 2604-2631
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

SCD-YOLO: A Detail-Enhanced Network for Strip Steel Surface Defect Detection

Jiacheng Jiang, Yuebin Su*

College of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong, China
Email: *dawnfading@outlook.com, mathsyb@suse.edu.cn

How to cite this paper: Jiang, J.C. and Su, Y.B. (2026) SCD-YOLO: A Detail-Enhanced Network for Strip Steel Surface Defect Detection. *Open Journal of Applied Sciences*, **16**, 2604-2631.
https://doi.org/10.4236/ojapps.2026.167145

Received: July 1, 2026
Accepted: July 27, 2026
Published: July 30, 2026

Abstract

Accurate and real-time surface defect detection is essential for industrial visual inspection. However, strip steel defects often show weak textures, ambiguous boundaries, low contrast, and strong background interference. These characteristics make fine-grained defect representation and accurate localization challenging. Although recent YOLO-based detectors achieve high inference efficiency, their convolution-dominated feature extraction modules and generic prediction heads remain limited in modeling subtle defect cues. To address these issues, this paper proposes SCD-YOLO, a detail-aware real-time detector built upon YOLOv11n. Extending our previous Star-based feature interaction and lightweight shared detection head, we introduce two defect-aware enhancements. First, a Star-Context Enhanced C3k2 structure, termed SC-C3k2, is developed for feature extraction. It embeds context anchor attention into the star-shaped multiplicative interaction path. This design performs contextual recalibration on high-order nonlinear interaction features. As a result, it strengthens weak-texture defect representation and suppresses pseudo-defect activations from complex backgrounds. Second, a Lightweight Shared Detail-Enhanced Convolutional Detection Head, termed LSDECD, is proposed for prediction. It incorporates detail-enhanced convolution into the shared multi-scale prediction transformation. This design reduces prediction-head redundancy and improves sensitivity to local edges, texture discontinuities, and subtle intensity variations. Experiments on the NEU-DET dataset show that SCD-YOLO improves the mAP@0.5 from 75.64% to 78.81% compared with YOLOv11n. The absolute gain is 3.17 percentage points, while the model maintains real-time inference speed. Ablation studies further confirm that SC-C3k2 and LSDECD provide complementary improvements in feature discrimination and detail-aware localization. These results demonstrate that SCD-YOLO achieves a favorable accuracy-efficiency trade-off for real-time strip steel surface defect detection.

Keywords

Strip Steel Surface Defect Detection, YOLOv11n, Star-Context Enhanced C3k2, Detail-Enhanced Detection Head

1. Introduction

Strip steel, known for its high strength, excellent ductility, and versatile formability, is widely used in industrial production, construction, transportation, machinery manufacturing, and various other sectors [1]. However, surface defects such as scratches, inclusions, cracks, patches, pitted surfaces, and rolled-in scales may occur during rolling, cooling, transportation, or storage processes. These defects not only degrade the visual quality of steel products, but may also weaken mechanical properties, reduce service reliability, and even cause severe production accidents such as strip breakage and equipment damage [2]-[4]. Therefore, rapid, accurate, and reliable surface defect detection is of great significance for quality control and intelligent manufacturing in the steel industry.

Historically, surface defect inspection in strip steel production relied heavily on manual visual examination, where experienced inspectors identified abnormal regions according to visual observation and empirical judgment. Although manual inspection is intuitive, it suffers from low efficiency, high labor cost, strong subjectivity, and poor reproducibility, making it unsuitable for modern high-speed and large-scale production lines [5]. To overcome these limitations, traditional machine learning methods were introduced into automated inspection systems. These methods generally consist of handcrafted feature extraction and classical classifiers, such as support vector machines, decision trees, and random forests [6]. Although they improve inspection automation to some extent, their performance highly depends on manually designed texture, shape, or statistical features. As a result, they often exhibit limited adaptability, weak robustness, and insufficient localization capability when facing complex backgrounds, low-contrast defects, and diverse defect morphologies [7].

In recent years, deep learning has greatly advanced industrial defect detection. Convolutional neural network (CNN)-based detectors can automatically learn hierarchical visual representations from data, avoiding the limitations of handcrafted feature design. Existing object detection models can generally be divided into two categories: two-stage detectors and single-stage detectors. Two-stage approaches, such as Faster R-CNN [8] and Mask R-CNN [9], first generate candidate regions and then perform classification and bounding-box refinement. These methods usually provide strong detection accuracy but require relatively high computational cost and inference latency. In contrast, single-stage detectors, including SSD [10], EfficientDet [11], and the YOLO series [12]-[14], directly predict categories and bounding boxes in a single forward pass. Benefiting from their end-to-end design and high inference efficiency, single-stage detectors are more

suitable for real-time industrial inspection scenarios.

Among single-stage detectors, the YOLO family has attracted extensive attention due to its favorable balance between accuracy and speed. Its efficient multi-scale prediction mechanism enables fast localization and classification, making it particularly attractive for online strip steel surface defect detection. With continuous architectural evolution, recent YOLO variants have improved feature extraction, feature fusion, and prediction efficiency. YOLOv11 [15] further provides a compact and deployment-oriented baseline for real-time detection tasks. Nevertheless, directly applying the original YOLOv11n to strip steel surface defect detection still faces several challenges. First, many defects are small, weak-textured, and low-contrast, and their discriminative cues are often hidden in subtle local texture variations rather than salient semantic structures. Second, complex industrial imaging conditions, including uneven illumination, specular reflection, background texture similarity, and pseudo-defect interference, may lead to false detections. Third, the detection head must preserve fine-grained boundary and texture details to accurately localize small or weak-boundary defects. Therefore, improving fine-grained feature representation and detail-aware prediction while maintaining real-time efficiency remains a challenging problem.

To address these limitations, we propose SCD-YOLO, a lightweight and detail-aware detector built upon YOLOv11n for strip steel surface defect detection. Instead of redesigning the entire detection pipeline, SCD-YOLO focuses on two key stages of the detector: feature extraction and prediction. In the feature extraction stage, we propose a Star-Context Enhanced C3k2 structure, termed SC-C3k2, which embeds context anchor attention into the star-shaped multiplicative interaction path. This design enables high-order nonlinear feature composition and contextual recalibration within a unified transformation, improving the representation of weak-texture defects while suppressing pseudo-defect responses. In the prediction stage, we design a Lightweight Shared Detail-Enhanced Convolutional Detection Head, termed LSDECD. By combining shared multi-scale prediction transformation with detail-enhanced convolution, LSDECD reduces redundant prediction parameters and enhances the sensitivity of the detection head to local edges, texture discontinuities, and subtle intensity variations.

The Star module and LSCD head were initially investigated in our previous work [16]. In this paper, we further extend these components by introducing contextual recalibration and detail-enhanced convolution, resulting in Star-CAA and LSDECD, respectively. The main contributions of this work are summarized as follows:

1) Star-CAA enhanced feature extraction. Building upon our previous Star-based feature interaction design, we further propose a Star-CAA enhanced C3k2 structure for strip steel surface defect detection. Different from the original Star module, the proposed structure embeds context anchor attention into the star-shaped multiplicative interaction path. This design enables contextual recalibration to be performed on high-order nonlinear interaction features, thereby im-

proving weak-texture defect representation and suppressing background-induced pseudo-defect activations.

2) Detail-enhanced shared detection head. Based on our previously developed lightweight shared convolutional detection head, we further introduce LSDECD by incorporating detail-enhanced convolution into the shared prediction trunk. The proposed head preserves the compact shared prediction paradigm while strengthening sensitivity to local edges, texture discontinuities, and weak-boundary defect regions.

2. Related Work

2.1. YOLO-Based Real-Time Object Detection

The YOLO series has become one of the most representative families of single-stage object detectors due to its favorable balance between detection accuracy and inference efficiency. Unlike two-stage detectors that first generate region proposals and then perform classification and localization, YOLO formulates object detection as a unified dense prediction problem and directly predicts object categories and bounding boxes in a single forward pass [12]. This end-to-end detection paradigm makes YOLO-based methods particularly suitable for real-time industrial inspection scenarios, where inference latency, deployment cost, and system responsiveness are important considerations.

Recent YOLO variants have continuously improved the accuracy-efficiency trade-off from different perspectives. YOLOv6 introduces an industrial-oriented single-stage detection framework with efficient backbone and neck designs [13]. YOLOv7 improves real-time detection through trainable bag-of-freebies and efficient architectural optimization [14]. YOLOv9 investigates information preservation in deep detection networks through programmable gradient information and efficient feature aggregation [17], while YOLOv10 explores end-to-end real-time detection by reducing post-processing dependency and computational redundancy [18]. These studies show that modern YOLO detectors have evolved from simple convolutional stacking toward more carefully designed feature extraction, feature fusion, assignment, and prediction mechanisms.

In this work, YOLOv11n is adopted as the compact baseline detector due to its streamlined architecture and real-time inference capability [15]. YOLOv11n introduces efficient feature extraction units such as C3k2 and maintains a concise detection pipeline, making it suitable for resource-constrained industrial visual inspection. However, when directly applied to strip steel surface defect detection, the original YOLOv11n still faces task-specific challenges. Many surface defects are weak-textured, low-contrast, and visually similar to background patterns, making it difficult for ordinary convolution-dominated feature extraction to capture subtle and high-order defect cues. In addition, the generic detection head lacks explicit modeling of local structural details, which may limit localization accuracy for ambiguous boundaries, small defect regions, and weak edge discontinuities.

2.2. Strip Steel Surface Defect Detection

Strip steel surface defect detection is more challenging than general object detection because defects often exhibit small spatial extent, irregular morphology, weak texture contrast, and strong similarity to surrounding background structures [19] [20]. In practical production environments, uneven illumination, specular reflection, imaging noise, and pseudo-defect interference further increase the difficulty of robust defect localization. Therefore, an effective strip steel defect detector should simultaneously provide fine-grained representation, strong background discrimination, accurate localization, and real-time inference capability.

Existing strip steel surface defect detection methods based on YOLO-style frameworks can be roughly categorized into several research directions. One line of work focuses on compact backbone and neck design for efficient deployment. For example, SDD-YOLO emphasizes compact modeling under deployment constraints [21], while AEDN-YOLO enhances feature extraction and multi-scale representation within a one-stage detection framework [22]. TBD-YOLO further adapts detection and feature fusion components for hot-rolled strip surface defect scenarios [23]. These methods improve the practicality of YOLO-style detectors in industrial scenarios, but their feature transformations are still largely based on convolutional aggregation, leaving high-order nonlinear modeling of subtle defect cues less explicitly explored.

Another line of work introduces multi-scale context modeling and receptive-field enhancement. MSC-DNet constructs efficient multi-scale context modules for strip steel defect detection [24], and Trident-LK Net employs a lightweight trident structure with large kernels to enhance multi-scale defect perception [25]. These methods improve contextual aggregation and scale robustness, which are important for defects with varying sizes and irregular shapes. However, they mainly focus on enlarging receptive fields or aggregating multi-scale information, while the interaction among weak local cues and the suppression of pseudo-defect responses remain insufficiently emphasized.

A third line of work attempts to improve defect detection under data scarcity or complex industrial distributions. YOLO-SD integrates simulated feature fusion and generative data augmentation to improve few-shot industrial defect detection [26]. Such data-driven strategies can alleviate sample insufficiency and improve generalization under limited data conditions. Nevertheless, they do not directly address the architectural limitations of feature representation and prediction head design. Overall, existing studies have achieved meaningful progress in real-time strip steel defect detection, yet robustly modeling weak defect patterns and preserving detail-sensitive localization remain important challenges.

2.3. Nonlinear Feature Interaction and Contextual Recalibration

Efficient feature representation is essential for compact visual recognition. Conventional convolutional blocks usually improve representation capacity by increasing depth, width, or kernel size. Although effective, these strategies may in-

troduce additional computational cost and still rely mainly on additive feature aggregation. Recently, StarNet has shown that star-shaped element-wise multiplicative interaction can implicitly map features into high-dimensional nonlinear spaces without significantly widening the network [27]. This property is potentially useful for surface defect detection, because defect patterns are often defined by the joint presence of weak edge deviation, local texture inconsistency, and grayscale perturbation rather than by a single salient response.

However, directly introducing multiplicative feature interaction into industrial defect detection is not always sufficient. In complex strip steel images, background textures, machining traces, illumination artifacts, and imaging noise may produce responses similar to real defects. If nonlinear interaction only amplifies local responses without contextual constraints, it may also strengthen pseudo-defect activations. Therefore, nonlinear feature interaction should be coupled with contextual recalibration to distinguish defect-related responses from background-induced interference.

Contextual attention mechanisms have been widely explored to enhance spatial discrimination in object detection. Context anchor attention, as introduced in PKINet, provides a lightweight mechanism for spatial contextual recalibration and has shown effectiveness in capturing long-range contextual dependencies for detection tasks [28]. Related attention-enhanced detectors further demonstrate that contextual modeling can improve small-object localization and background suppression [29] [30]. Nevertheless, in many existing attention-enhanced designs, attention modules are usually attached as external plug-in components after convolutional feature extraction. Such a design can improve feature weighting, but the attention operation is not necessarily integrated into the internal nonlinear feature composition process.

Different from these designs, this work embeds context anchor attention into the star-shaped multiplicative interaction path of selected C3k2 structures. In this manner, contextual recalibration is performed on the high-order interaction feature rather than on ordinary convolutional features. This allows the module to first construct nonlinear representations for subtle defect cues and then suppress unreliable responses using spatial contextual information. Therefore, the proposed SC-C3k2 can be viewed as a task-oriented adaptation of star-shaped nonlinear interaction for fine-grained strip steel surface defect detection.

2.4. Detail-Aware and Shared Detection Heads

The detection head directly determines the quality of classification confidence and bounding-box localization. In modern one-stage detectors, decoupled and fully convolutional prediction heads have been widely adopted because classification and localization require different feature properties. For example, YOLOX introduces an anchor-free detector with a decoupled head and advanced label assignment strategy, achieving strong detection performance across different model scales [31]. FCOS further demonstrates that object detection can be formulated as a fully con-

volutional dense prediction problem without anchor boxes, providing an effective head design for multi-level feature prediction [32]. In addition, distributional bounding-box regression has been shown to improve localization quality by modeling bounding-box distances as discrete distributions rather than deterministic scalar values. Generalized Focal Loss and Distribution Focal Loss provide effective formulations for learning qualified and distributed bounding boxes in dense object detection [33]. These advances improve general detection performance, but they do not explicitly focus on local detail enhancement for industrial surface defects.

For strip steel surface defects, detail preservation is particularly important. Many defects appear as fine scratches, tiny pits, weak boundary discontinuities, or local grayscale perturbations. Such patterns require the detection head to remain sensitive to local edges, texture changes, and subtle intensity differences. Differential and detail-enhanced convolutional operators provide useful insights for this purpose. Central Difference Convolution enhances fine-grained representation by aggregating intensity and gradient information [34], while detail-enhanced convolution introduces differential priors to strengthen local structural representation [35]. These studies suggest that incorporating explicit detail priors into convolutional feature transformation can improve the modeling of local structural variations.

Another important consideration in real-time detection heads is parameter redundancy across multi-scale prediction levels. Conventional YOLO-style heads often construct separate prediction transformations for different feature levels. Although this design preserves scale-specific flexibility, it may introduce redundant parameters and increase deployment cost. Shared prediction transformations can reduce cross-scale redundancy and improve compactness. However, ordinary shared heads may weaken detail-sensitive perception if local structural cues are not explicitly enhanced. Moreover, normalization is important for stabilizing shared prediction heads, especially when training with limited batch sizes. Group Normalization provides a batch-size-independent normalization strategy and is therefore suitable for small-batch industrial defect detection scenarios [36].

To address these requirements, this work introduces LSDECD, a Lightweight Shared Detail-Enhanced Convolutional Detection Head. LSDECD first aligns multi-scale features into a unified hidden space, then applies a shared detail-enhanced convolutional trunk with group normalization, and finally performs decoupled classification and regression prediction. Compared with conventional independent detection heads, the shared transformation reduces redundant cross-scale prediction transformations. Compared with ordinary shared heads, the detail-enhanced convolution strengthens local edge, texture, and structural responses before final prediction. Therefore, LSDECD is better suited for small, low-contrast, and weak-boundary defects.

3. Methodology

To address the challenges of detecting subtle, irregular, low-contrast, and weak-

boundary defects in strip steel surface images, this work proposes SCD-YOLO, a compact and detail-aware detection framework built upon YOLOv11n. The proposed method retains the general detection pipeline of YOLOv11n, including backbone feature extraction, neck-based multi-scale feature aggregation, and dense prediction. Rather than redesigning the entire detector, SCD-YOLO introduces task-oriented architectural modifications into two critical stages: feature extraction and prediction. The objective is to improve fine-grained defect representation and detail-sensitive localization while maintaining practical real-time inference capability.

As shown in Figure 1, the first improvement is the Star-Context Enhanced C3k2 structure, termed SC-C3k2. Instead of simply increasing network depth or attaching an external attention module, SC-C3k2 enhances selected C3k2 structures by embedding star-shaped multiplicative interaction and contextual recalibration into the internal feature transformation path. The star-shaped interaction strengthens high-order nonlinear feature composition, which is beneficial for modeling subtle defect cues that are difficult to capture using ordinary convolution alone. Meanwhile, the embedded contextual recalibration helps suppress unreliable responses caused by background textures, illumination variations, and pseudo-defect interference.

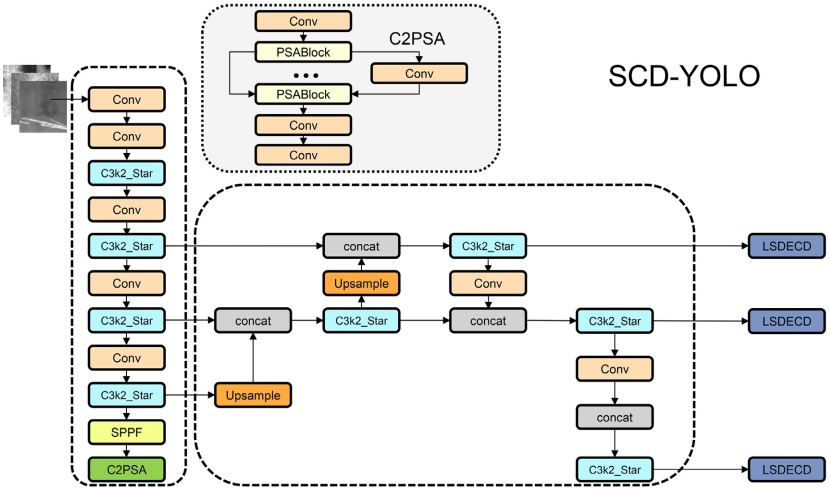

Figure 1. Architecture overview of the proposed SCD-YOLO.

The second improvement is the Lightweight Shared Detail-Enhanced Convolutional Detection Head, termed LSDECD. LSDECD first aligns multi-scale features into a unified hidden space and then applies a shared detail-enhanced convolutional trunk across detection levels. This design reduces redundant cross-scale prediction transformations and encourages scale-consistent defect-sensitive representations. In addition, the detail-enhanced convolution strengthens local structural responses associated with weak edges, texture discontinuities, fine scratches, and subtle intensity variations. The enhanced features are then used for decoupled bounding-box regression and defect classification.

Different from generic compact detector designs that mainly focus on reducing parameters or computational cost, SCD-YOLO is designed according to the visual characteristics of strip steel surface defects. SC-C3k2 improves the discriminability of intermediate features before multi-scale aggregation and prediction, whereas LSDECD further enhances the utilization of detail-sensitive cues at the detection head. These two components are complementary and form a progressive enhancement mechanism from feature extraction to final prediction, enabling SCD-YOLO to achieve improved detection accuracy while preserving compact model complexity and real-time applicability.

The following subsections describe the technical details and design rationale of SC-C3k2 and LSDECD.

3.1. SC-C3k2: Star-Context Enhanced C3k2 Structure

Industrial surface defect detection differs substantially from generic object detection. In natural-scene images, target objects usually have relatively complete semantic structures and distinguishable appearance patterns. In contrast, strip steel surface defects are often manifested as weak local perturbations, such as fine scratches, tiny pits, cracks, boundary discontinuities, intensity fluctuations, and subtle texture abnormalities. These defect patterns are usually small in spatial extent, weak in contrast, and highly similar to surrounding background textures. Moreover, practical industrial images are frequently affected by uneven illumination, specular reflection, imaging noise, machining traces, and pseudo-defect interference. Under such conditions, feature extraction based only on stacked local convolutions may be insufficient to distinguish real defects from background structures with similar local responses.

The C3k2 structure in YOLOv11n provides an efficient feature extraction unit by combining compact convolutional transformations with cross-stage feature aggregation. However, its internal transformation is still mainly dominated by local convolutional aggregation. Although local convolution is effective for capturing neighborhood patterns, it has limited ability to explicitly model high-order nonlinear interactions among subtle defect cues. For weak-texture defects, discriminative information often does not originate from a single salient activation, but from the joint response of multiple fine-grained patterns, such as local edge deviation, texture inconsistency, and grayscale fluctuation. Therefore, enhancing the nonlinear representation capability and contextual discrimination of C3k2 is important for robust strip steel surface defect detection.

To address this issue, we propose a Star-Context Enhanced C3k2 structure, termed SC-C3k2. The proposed structure enhances selected C3k2 blocks by embedding a star-context transformation into the internal feature extraction path. Specifically, the StarBlock design from StarNet [27] is introduced to strengthen high-order nonlinear feature interaction, and the context anchor attention (CAA) mechanism from PKINet [28] is incorporated to provide spatial contextual recalibration. As illustrated in **Figure 2**, the proposed Star-CAA transformation first

performs local feature embedding, then constructs high-order nonlinear responses through dual-branch multiplicative interaction, and finally applies contextual recalibration before residual refinement. The key idea is to couple star-shaped multiplicative interaction with embedded contextual recalibration. The former improves nonlinear feature composition for subtle defect cues, while the latter suppresses unreliable activations caused by background textures and pseudo-defect interference.

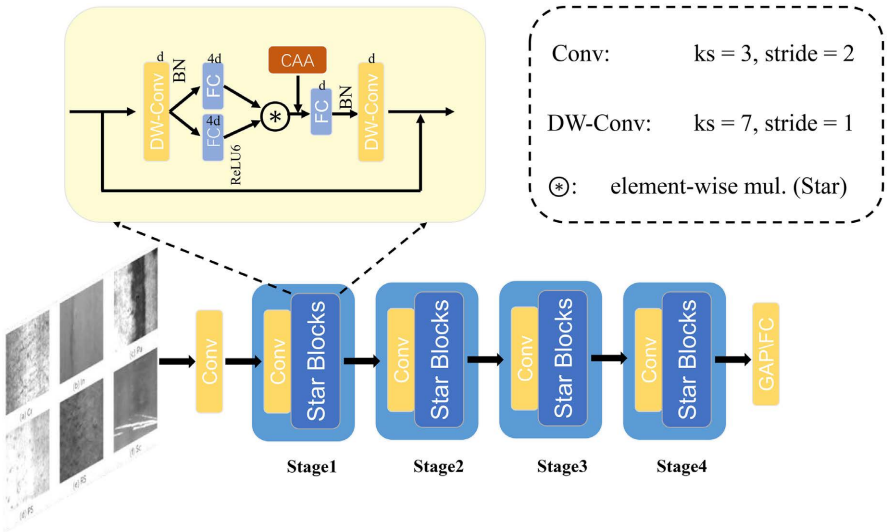

Figure 2. Structure of the proposed Star-CAA transformation.

Given an input feature map

$$X \in \mathbb{R}^{C\times H\times W}, \tag{1}$$

where C, H, and W denote the channel number, height, and width, respectively, the embedded Star-CAA transformation first performs a local feature embedding:

$$X_e = \mathcal{E}(X), \tag{2}$$

where $\mathcal{E}(\cdot)$ denotes the local embedding operation. This step provides a defect-sensitive local representation for subsequent nonlinear interaction.

After local embedding, two parallel feature mappings are applied:

$$X_1 = \phi_1(X_e), \quad X_2 = \phi_2(X_e), \tag{3}$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ denote branch-wise feature transformations. The two branches generate complementary feature responses for multiplicative interaction. One branch produces an activated modulation signal, while the other branch preserves defect-related responses to be modulated.

The star-shaped nonlinear interaction is formulated as

$$X_s = \sigma(X_1)\odot X_2, \tag{4}$$

where $\sigma(\cdot)$ denotes a nonlinear activation function and $\odot$ represents element-wise multiplication. Different from ordinary additive aggregation, the mul-

tiplicative interaction enables high-order feature composition and strengthens the modeling of subtle local cues. This is particularly useful for weak surface defects, whose discriminative patterns are often reflected by the co-occurrence of multiple fine-grained responses rather than a single strong activation.

However, enhancing nonlinear local responses alone may also amplify background-induced activations. To improve contextual discrimination, context anchor attention is further introduced into the star-interaction path:

$$X_c = \mathcal{A}_{\mathrm{CAA}}\left(X_s\right), \tag{5}$$

where $\mathcal{A}_{\mathrm{CAA}}(\cdot)$ denotes the context anchor attention mapping. Unlike conventional attention-enhanced designs that attach attention as an external plug-in after feature extraction, the proposed SC-C3k2 embeds CAA directly into the nonlinear interaction path. Therefore, contextual recalibration is performed on the star-interaction feature X_s, allowing the module to enhance defect-related responses while suppressing unreliable activations caused by pseudo-defects, illumination artifacts, and background texture interference.

Finally, the context-recalibrated feature is projected back to the output space and aggregated with the input through residual learning:

$$Y = X + \mathcal{P}\left(X_c\right), \tag{6}$$

where $\mathcal{P}(\cdot)$ denotes the output projection and refinement operation. The residual connection helps preserve the original feature information and facilitates stable optimization when the transformation is inserted into the detection backbone.

Combining the above operations, the embedded Star-CAA transformation can be summarized as

$$Y = X + \mathcal{P}\left(\mathcal{A}_{\mathrm{CAA}}\left(\sigma\left(\phi_1\left(\mathcal{E}\left(X\right)\right)\right) \odot \phi_2\left(\mathcal{E}\left(X\right)\right)\right)\right). \tag{7}$$

It should be noted that this formulation describes the internal Star-CAA transformation embedded in the selected C3k2 structure, rather than the complete topology of the original C3k2 block. In this way, SC-C3k2 preserves the cross-stage aggregation property of C3k2 while introducing nonlinear feature interaction and contextual recalibration into its internal transformation path.

From the perspective of structural composition, SC-C3k2 contains four functional stages. The first stage is local feature embedding, which captures neighborhood defect patterns. The second stage is dual-branch feature transformation, which constructs complementary representations for nonlinear interaction. The third stage is star-context interaction, where multiplicative feature composition generates high-order nonlinear responses and CAA recalibrates these responses using spatial contextual information. The fourth stage is projection and residual refinement, where the enhanced feature is transformed back to the output space and fused with the original input.

Compared with simply attaching an attention module to a convolutional block, SC-C3k2 integrates contextual recalibration into the nonlinear feature interaction process. Therefore, the attention mapping does not merely reweight ordinary con-

volutional features, but recalibrates the high-order star-interaction representation. This design makes the feature enhancement process more coherent: local patterns are first embedded, subtle cues are then composed through multiplicative interaction, and unreliable responses are subsequently suppressed by contextual recalibration.

In the implementation of SCD-YOLO, the proposed SC-C3k2 is used to replace the original C3k2-style feature transformation modules in both the backbone and the feature-fusion head of YOLOv11n. Specifically, four SC-C3k2 stages are inserted in the backbone after the P2/4, P3/8, P4/16, and P5/32 downsampling layers. In addition, four SC-C3k2 stages are further adopted in the feature-fusion head after multi-scale feature concatenation.

In the overall detector, SC-C3k2 is used to replace selected C3k2 structures in YOLOv11n without changing the general detection pipeline. This replacement enhances the feature extraction stage before multi-scale aggregation and detection prediction. The proposed SC-C3k2 specifically targets two limitations in strip steel surface defect detection: insufficient nonlinear representation for weak-texture defects and inadequate suppression of background-induced interference. By integrating star-shaped high-order interaction with embedded contextual recalibration, SC-C3k2 provides a more discriminative feature extraction unit for fine-grained surface defect detection.

3.2. LSDECD: Lightweight Shared Detail-Enhanced Convolutional Detection Head

The detection head transforms multi-scale features into category confidence scores and bounding-box predictions. For strip steel surface defect detection, this stage is particularly important because many defects are not represented by complete object-level semantics, but by subtle local patterns, such as weak edges, fine scratches, small pits, texture discontinuities, and low-contrast intensity variations. Therefore, an effective detection head should not only preserve semantic discrimination, but also remain sensitive to local structural details for accurate localization. However, conventional YOLO-style detection heads usually employ independent prediction branches for different feature levels, which may introduce redundant parameters across scales. In addition, ordinary convolutional transformations are not explicitly designed to enhance local differential cues, making them less effective for small, weak-boundary, and low-contrast defects.

To address these limitations, we propose a Lightweight Shared Detail-Enhanced Convolutional Detection Head, termed LSDECD. Different from the previously used lightweight shared convolutional head, LSDECD further introduces detail-enhanced convolution into the shared prediction trunk. As shown in **Figure 3**, the proposed head mainly consists of three parts: level-wise channel alignment, a shared detail-enhanced convolutional trunk, and decoupled regression and classification prediction layers. The design aims to retain the compactness of shared prediction while improving the sensitivity of the detection head to local defect details.

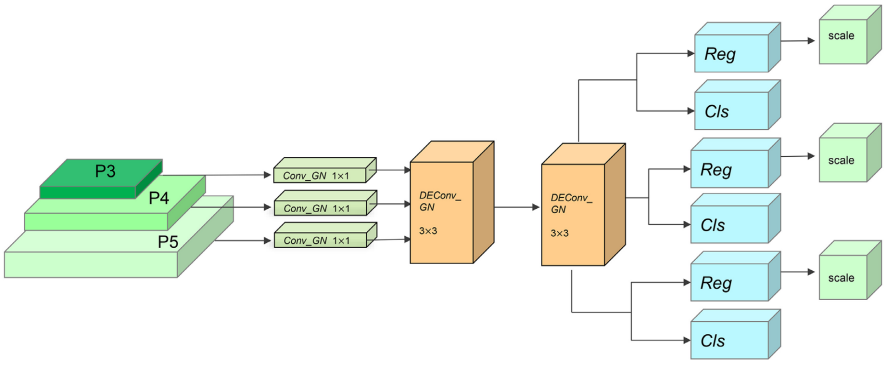

Figure 3. Structure of the proposed LSDECD head.

Given the multi-scale features from the neck,

$$\mathcal{F} = \{F_i\}_{i=1}^{N}, \quad F_i \in \mathbb{R}^{C_i \times H_i \times W_i}. \tag{8}$$

where N denotes the number of detection levels. In our implementation, $N = 3$, corresponding to the three prediction scales of YOLO-style detectors. LSDECD first aligns different feature levels into a unified hidden space by using level-wise 1×1 convolution followed by group normalization:

$$\hat{F}_i = \mathrm{GN}\left(\mathrm{Conv}_{1\times1}^{(i)}\left(F_i\right)\right). \tag{9}$$

where $\mathrm{Conv}_{1\times1}^{(i)}(\cdot)$ denotes the channel alignment layer for the i-th feature level, and $\mathrm{GN}(\cdot)$ denotes group normalization [36]. The hidden channel dimension is set to 256 in all experiments.

After channel alignment, all feature levels are processed by the same shared detail-enhanced convolutional trunk:

$$\tilde{F}_i = \mathcal{T}_{\mathrm{DE}}\left(\hat{F}_i\right). \tag{10}$$

where $\mathcal{T}_{\mathrm{DE}}(\cdot)$ consists of two consecutive DEConv-GN blocks. The same trunk is shared by all detection levels, thereby reducing redundant cross-scale prediction transformations and encouraging scale-consistent defect-sensitive representations.

$$\mathrm{DEConv}(X) = \mathrm{Conv}_{\mathrm{std}}(X) + \sum_{m=1}^{4} \mathrm{Conv}_{\mathrm{diff}}^{(m)}(X). \tag{11}$$

Here, $\mathrm{Conv}_{\mathrm{std}}(\cdot)$ denotes the standard convolution branch, while $\mathrm{Conv}_{\mathrm{diff}}^{(m)}(\cdot)$ denotes the m-th differential detail-prior branch. Specifically, the four differential branches correspond to central-difference, angular-difference, horizontal-difference, and vertical-difference convolutional priors. These branches enhance complementary local structural variations, which is beneficial for modeling weak edges, fine scratches, small pits, texture discontinuities, and ambiguous defect boundaries.

The fusion coefficients of these branches are fixed to 1 rather than implemented as learnable scalar weights. Thus, DEConv can be regarded as a fixed summation of one standard convolution branch and four differential-prior branches, while the convolution kernels inside each branch remain learnable. This design introduces

explicit local detail priors without adding additional branch-weight parameters.

Based on the shared detail-enhanced feature $\tilde{F}_i$, LSDECD adopts decoupled prediction layers for bounding-box regression and defect classification:

$$R_i = \gamma_i \cdot \mathrm{Conv}_{\mathrm{reg}}\left(\tilde{F}_i\right), \quad S_i = \mathrm{Conv}_{\mathrm{cls}}\left(\tilde{F}_i\right), \tag{12}$$

where $R_i \in \mathbb{R}^{4K \times H_i \times W_i}$ and $S_i \in \mathbb{R}^{N_c \times H_i \times W_i}$ denote the regression and classification outputs at the i-th detection level, respectively. Here, $K = 16$ is the number of discrete bins used for distributional bounding-box regression, N_c is the number of defect categories, and γ_i is a learnable scale factor for the i-th detection level. It should be noted that γ_i is only used to calibrate the magnitude of regression outputs across different feature levels and is not a fusion coefficient of DEConv. During training, LSDECD follows the standard YOLO-style loss formulation, including bounding-box regression loss, classification loss, and distribution focal loss. During inference, the regression distribution is decoded into bounding boxes, and the classification logits are converted into category confidence scores by a sigmoid function.

Compared with the original YOLOv11n detection head, LSDECD has two main advantages. First, the shared prediction trunk reduces redundant transformations across multi-scale detection levels, improving the compactness of the prediction head. Second, the introduced detail-enhanced convolution strengthens local differential responses, which is beneficial for detecting small, low-contrast, and weak-boundary defects. In this way, LSDECD preserves the efficiency of shared prediction while improving detail-aware localization and classification. Together with the proposed SC-C3k2 feature extraction structure, LSDECD forms a progressive enhancement mechanism from nonlinear feature representation to detail-sensitive prediction, enabling SCD-YOLO to achieve a better accuracy-efficiency trade-off for strip steel surface defect detection.

4. Experiments and Results

To evaluate the effectiveness of the proposed SCD-YOLO, which integrates the Star-Context Enhanced C3k2 structure and the Lightweight Shared Detail-Enhanced Convolutional Detection Head, we conduct a comprehensive series of experiments on the NEU-DET dataset. These experiments include dataset evaluation, implementation analysis, visualization experiments, comparative studies with representative detectors, and ablation studies. The purpose is to verify whether the proposed method can improve detection accuracy while maintaining a lightweight structure and real-time inference capability.

4.1. Dataset

We evaluate our method on the publicly available NEU-DET dataset, a widely used benchmark for surface defect detection on rolled steel strips. As illustrated in **Figure 4**, NEU-DET contains six common defect categories: crazing (Cr), inclusions (In), patches (Pa), pitted surface (PS), rolled-in scales (RS), and scratches

(Sc). The dataset comprises 1800 grayscale images, each with a resolution of 200 × 200 pixels. We adopt a class-stratified random split with a fixed random seed of 0 to avoid category distribution bias and ensure reproducibility. Specifically, the dataset is divided into training, validation, and testing sets with an 8:1:1 ratio, resulting in 1440 training images, 180 validation images, and 180 testing images. The same split is kept fixed across all baseline comparisons, cross-model comparisons, and ablation experiments.

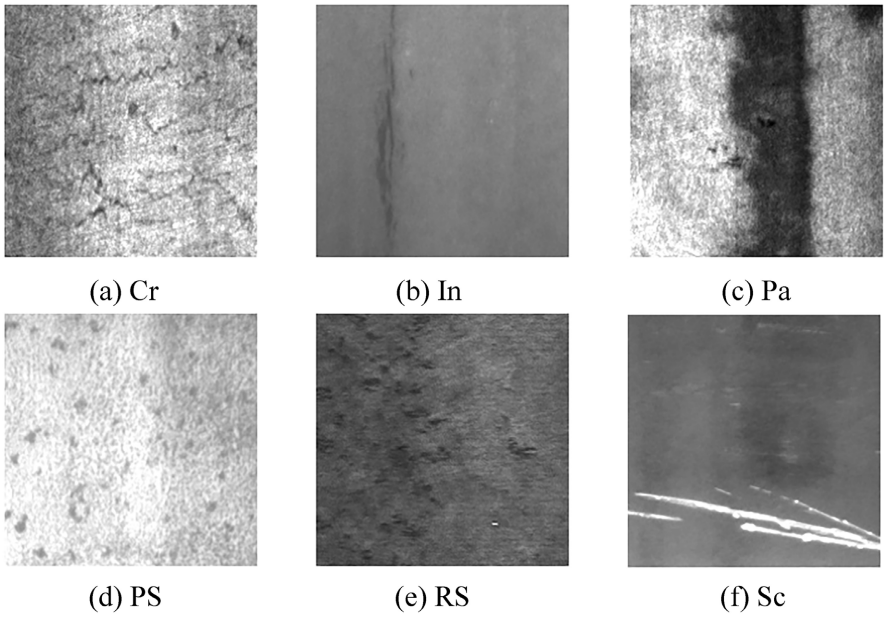

Figure 4. Visualization of representative defect categories in the NEU-DET dataset. (a) crazing; (b) inclusion; (c) patches; (d) pitted surface; (e) rolled-in scales; (f) scratches.

All images are fed into the network at 640 × 640. For inputs with varying aspect ratios, we apply the standard letterbox strategy commonly used in YOLO-style detectors: images are isotropically scaled to fit the target size and then symmetrically padded to 640 × 640. Ground-truth boxes are scaled and shifted with the same transformations. This prevents geometric distortion from anisotropic resizing and maintains consistent spatial geometry.

4.2. Implementation Details

All experiments were conducted on a workstation running Ubuntu 22.04.2 LTS, equipped with an NVIDIA RTX 4090 GPU with 24 GB VRAM. The models were implemented using PyTorch 2.3.0 with CUDA 12.1. We adopt YOLOv11n from the Ultralytics framework [15] as the baseline due to its compact architecture and real-time inference capability. The proposed SCD-YOLO is constructed by integrating SC-C3k2 and LSDECD into the YOLOv11n framework.

For a fair comparison, YOLOv11n, all YOLO-series baselines, and all ablation variants are trained under exactly the same experimental protocol. Specifically, all models use the same training/validation/testing split, optimizer, input resolution,

data augmentation strategy, number of epochs, batch size, and evaluation settings. Unless otherwise specified, stochastic gradient descent (SGD) is adopted as the optimizer, with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 5×10^{-4}. All models are trained for 300 epochs with a batch size of 32 and 4 data loader workers. No early stopping is applied, and the checkpoint with the best validation mAP@0.5 is used for final testing. The input image size is fixed to 640×640 for both training and inference.

To enhance robustness and generalization, we adopt an identical data augmentation pipeline for all compared YOLO-series models and ablation variants. The augmentation strategy includes HSV jitter, random horizontal flipping, translation, scaling, mosaic augmentation, RandAugment, and random erasing. Detailed settings are summarized in Table 1. By keeping the training protocol strictly consistent, the performance differences reported in the ablation study can be attributed to the architectural changes introduced by SC-C3k2 and LSDECD rather than to training-setting variations.

Table 1. Data augmentation configurations.

Category	Augmentation Type (Value)
Photometric	HSV jitter: $h = 0.015$, $s = 0.7$, $v = 0.4$
Geometric	Flip (horizontal: 0.5), translation: 0.1 Scale: 0.5
Composite	Mosaic: 1.0
Auto-augment	Rand Augment
Erasing	Random erasing: 0.4

4.3. Evaluation Metrics

To provide an intuitive assessment of classification performance, we first present the confusion matrix, as shown in Figure 5. This matrix visualizes the prediction distribution across all defect categories. The diagonal elements correspond to correctly classified instances, while the off-diagonal elements indicate misclassifications, which often occur between visually similar defect types.

In this work, we evaluate the proposed detector using widely adopted object detection metrics, including precision (P), recall (R), average precision (AP), and mean average precision (mAP). Precision measures the proportion of correctly identified positive samples among all predicted positives:

$$P = \frac{TP}{TP + FP}, \tag{13}$$

where TP and FP denote true positives and false positives, respectively. Recall quantifies the proportion of correctly identified positive samples among all actual positives:

$$R = \frac{TP}{TP + FN}, \tag{14}$$

where FN is the number of false negatives. Average precision for each defect class is computed by integrating precision over the entire range of recall values:

$$AP = \int_0^1 p(r)\mathrm{d}r, \tag{15}$$

where $p(r)$ is precision as a function of recall r. The mean average precision aggregates AP values across all n classes:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP(i). \tag{16}$$

In addition to detection accuracy, we assess computational efficiency through FLOPs, the number of parameters, model size, and frames per second (FPS). FLOPs measure the computational cost of a single forward pass; the number of parameters reflects memory and storage requirements; and FPS measures end-to-end inference throughput, including preprocessing, model forward inference, and post-processing. Together, these metrics provide a comprehensive evaluation of accuracy, complexity, and deployment potential.

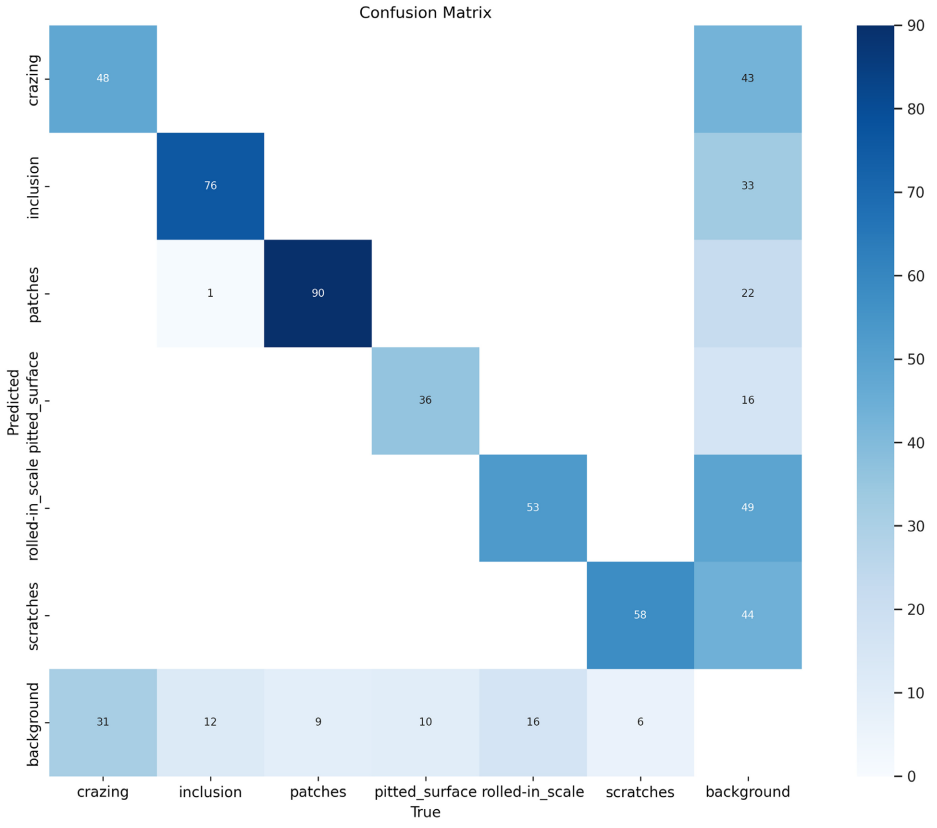

Figure 5. Confusion matrix of the proposed SCD-YOLO.

4.4. Visualization Experiments and Analysis

To further evaluate the effectiveness of the proposed SCD-YOLO, we analyze both training dynamics and qualitative detection performance on the NEU-DET dataset. In addition to quantitative metrics, we provide precision-confidence, recall-

confidence, and F1-confidence curves to investigate the prediction behavior of different models under varying confidence thresholds. These curves offer a more detailed understanding of classification reliability, recall stability, and confidence calibration. Furthermore, qualitative detection results are presented to visually compare the localization capability of YOLOv11n and SCD-YOLO under challenging defect scenarios.

4.4.1. Precision-Confidence Analysis on the NEU-DET Dataset

To evaluate the reliability of positive predictions, we compare the precision-confidence curves of YOLOv11n and SCD-YOLO, as shown in Figure 6. Precision reflects the proportion of correctly detected defects among all predicted defect instances. A higher precision value indicates that the model produces fewer false positives, which is particularly important for industrial inspection systems where false alarms may increase manual rechecking costs and reduce production efficiency.

Compared with YOLOv11n, SCD-YOLO maintains higher precision over a wide range of confidence thresholds. This indicates that the proposed model can better distinguish real defects from background textures and pseudo-defect regions. The improvement can be mainly attributed to the proposed SC-C3k2 structure, which enhances contextual discrimination during feature extraction, and LSDECD, which strengthens detail-aware prediction at the detection head. As a result, SCD-YOLO reduces background-induced false activations and produces more reliable confidence scores for defect regions.

4.4.2. Recall-Confidence Analysis on the NEU-DET Dataset

To assess the capability of detecting actual defect regions, we compare the recall-confidence curves of YOLOv11n and SCD-YOLO, as shown in Figure 7. Recall measures the proportion of correctly detected defects among all ground-truth defect instances. For strip steel surface defect detection, high recall is essential because missed defects may directly affect product quality and downstream manufacturing reliability.

SCD-YOLO exhibits more stable recall performance across different confidence thresholds. This suggests that the proposed method can detect more true defect instances while maintaining robustness to threshold variation. The improvement is especially meaningful for weak-texture, low-contrast, and small-scale defects, which are often difficult to distinguish from background textures. By introducing star-shaped nonlinear interaction and embedded contextual recalibration, SC-C3k2 enhances the representation of subtle defect cues. Meanwhile, LSDECD improves the sensitivity of the prediction head to local edges, texture discontinuities, and fine structural changes. These two components jointly reduce missed detections and improve the completeness of defect coverage.

4.4.3. F1-Confidence Analysis on the NEU-DET Dataset

To further evaluate the confidence calibration and precision-recall trade-off of the

model, we compare the F1-confidence curves of YOLOv11n and SCD-YOLO, as illustrated in Figure 8. The F1 score jointly considers precision and recall, and therefore provides a balanced evaluation of false positives and missed detections under different confidence thresholds.

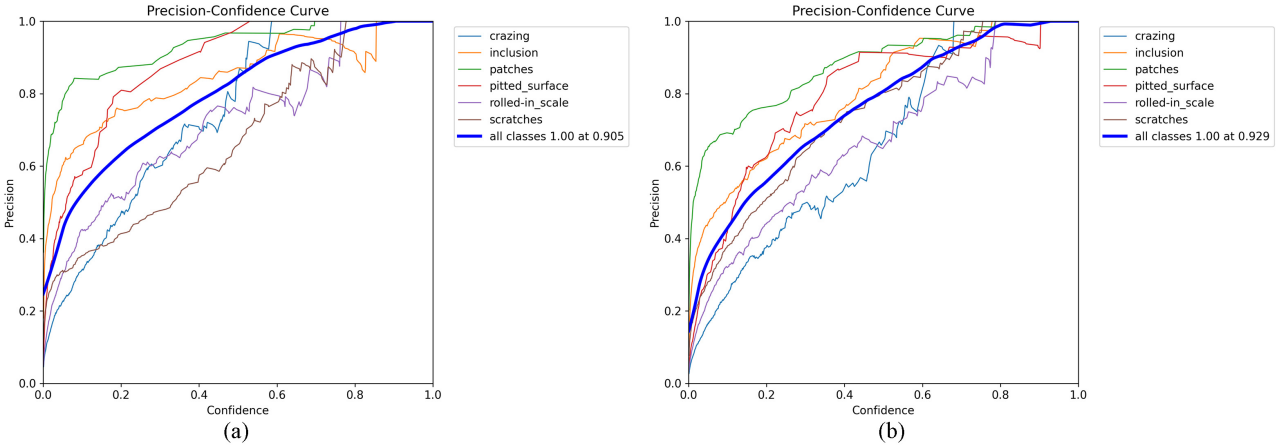

Figure 6. Precision-confidence curves for various defect types. (a) YOLOv11n and (b) SCD-YOLO.

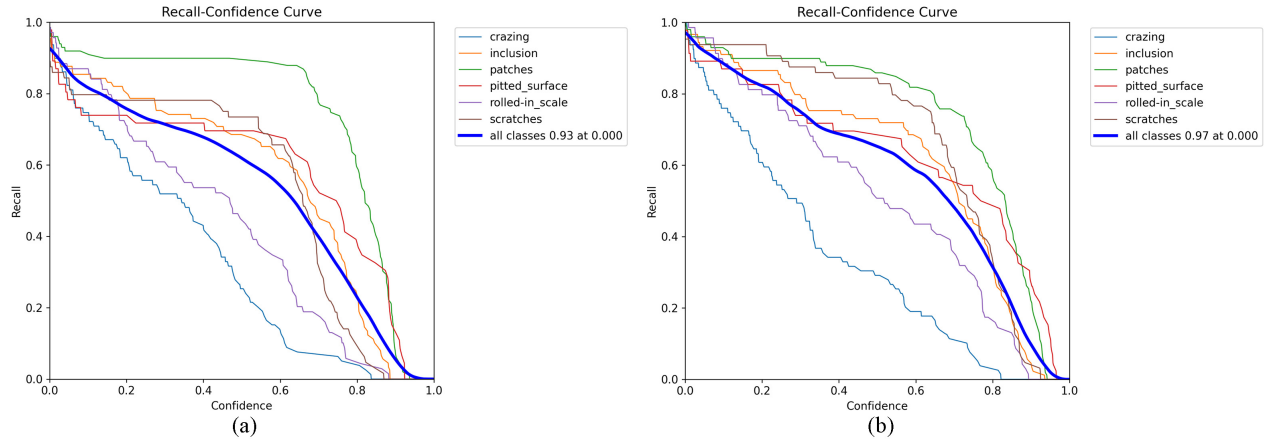

Figure 7. Recall-confidence curves for various defect types. (a) YOLOv11n and (b) SCD-YOLO.

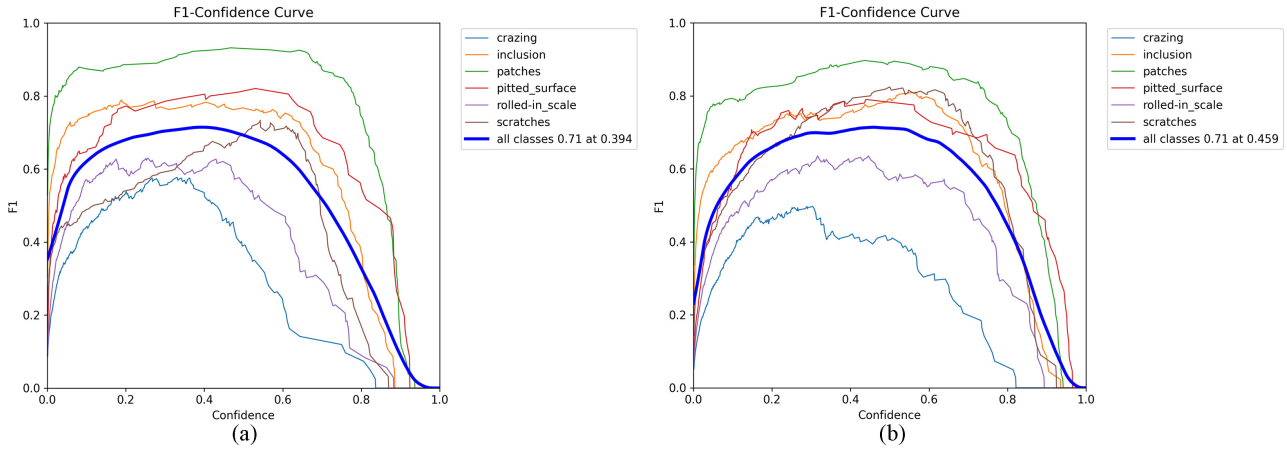

Figure 8. F1-confidence curves for various defect types. (a) YOLOv11n and (b) SCD-YOLO.

SCD-YOLO achieves a more favorable precision-recall balance and maintains stable F1 scores across a broader range of confidence thresholds. This indicates that the proposed detector produces more reliable predictions and is less sensitive to threshold variation. The improved confidence calibration can be attributed to the complementary effects of SC-C3k2 and LSDECD. SC-C3k2 suppresses background-induced pseudo-defect activations during feature extraction, while LSDECD enhances detail-aware prediction at the detection head. Together, they reduce uncertain responses and improve the reliability of defect classification and localization.

4.4.4. Enhanced Localization and Coverage in Qualitative Results

As shown in **Figure 9**, the proposed SCD-YOLO exhibits more robust detection performance than YOLOv11 under complex surface-texture conditions. For large-area defects such as pitted surface, SCD-YOLO produces more complete bounding boxes that better cover the defective regions, whereas YOLOv11n tends to generate relatively fragmented or incomplete predictions in several samples. This indicates that the proposed model has a stronger capability in capturing spatially distributed defect patterns with blurred boundaries and non-uniform grayscale variations.

For weak-texture and low-contrast defects such as rolled-in scale, SCD-YOLO detects more potential defective regions and provides more confident responses. In contrast, YOLOv11 shows weaker sensitivity to subtle local texture changes, especially when the defect appearance is similar to the surrounding background. This improvement can be attributed to the enhanced feature representation introduced by the SC-C3k2 module, which strengthens local texture modeling and suppresses background interference during feature extraction. Meanwhile, the LSDECD detection head further improves the utilization of detail-sensitive features, enabling the model to better localize small and irregular defects.

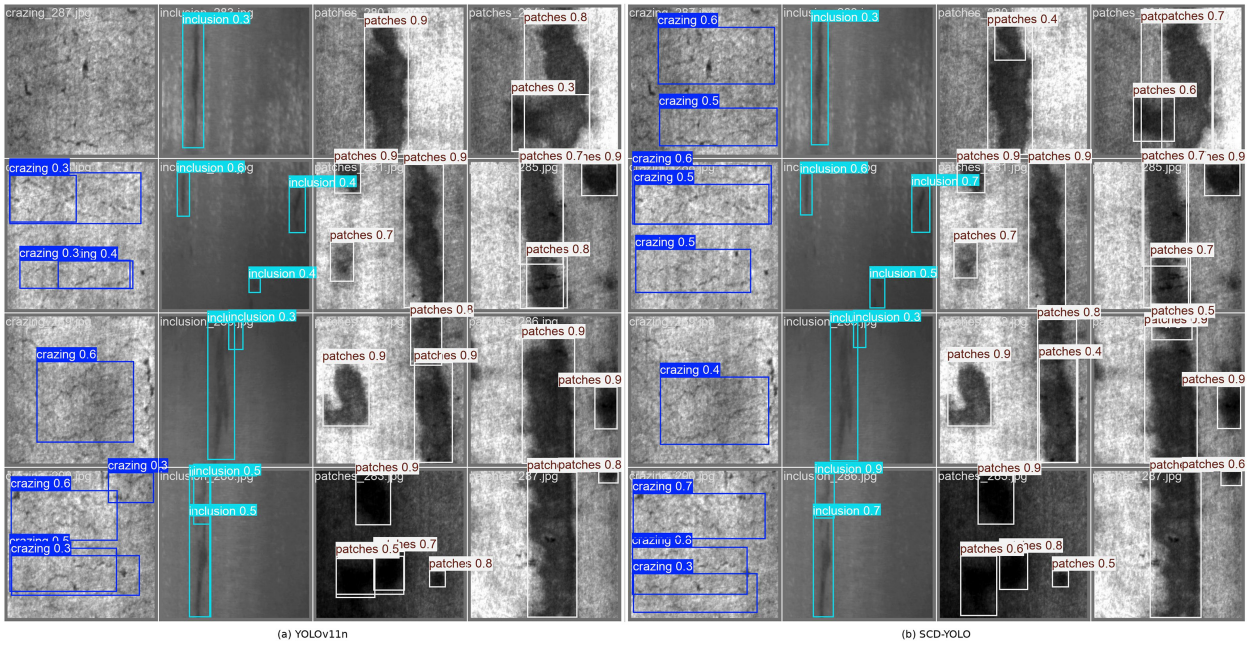

Figure 9. Qualitative detection comparison between the proposed SCD-YOLO and YOLOv11n on the NEU-DET dataset.

In addition, SCD-YOLO maintains better detection consistency across different defect categories, including crazing, pitted surface, patches, and rolled-in scale. The visual comparison demonstrates that the proposed method achieves improved defect completeness, stronger response to low-contrast regions, and better robustness in cluttered industrial backgrounds. These qualitative results are consistent with the quantitative improvements reported in Table 2, further verifying the effectiveness of the proposed SCD-YOLO for steel surface defect detection.

Table 2. Comparison of detection results among YOLO-series detectors on the NEU-DET dataset.

Model	Params (M)	GFLOPs	Model Size (MB)	*P* (%)	*R* (%)	mAP@0.5 (%)	FPS
YOLOv8n	3.01	8.1	6.0	71.10	70.98	75.66	134.23
YOLOv9t	1.97	7.6	4.5	68.91	68.45	75.03	153.16
YOLOv10n	2.27	6.5	5.5	72.51	66.03	73.52	165.73
YOLOv11n	2.58	6.3	5.2	73.84	68.40	75.64	130.49
YOLOv12n	2.56	6.3	5.3	65.46	68.48	74.50	146.22
SCD-YOLO	2.71	7.8	6.0	77.62	70.30	78.81	119.63

4.5. Comparative Experiments

To comprehensively assess the performance of the proposed SCD-YOLO framework, we conduct two groups of comparative experiments on the NEU-DET dataset. The first group compares SCD-YOLO with representative YOLO-series detectors, including YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and YOLOv12n [37], to evaluate its accuracy-efficiency trade-off within the YOLO family. The second group compares SCD-YOLO with several classical and state-of-the-art surface defect detection methods, including SGS-YOLO [16], Faster R-CNN [8], YOLO-SDS [3], and RT-DETR [38], to further verify its competitiveness in industrial defect inspection.

4.5.1. Comparison with YOLO-Series Detectors

Table 2 reports the comparison results among different YOLO-series detectors. Compared with the YOLOv11n baseline, SCD-YOLO improves the mAP@0.5 from 75.64% to 78.81%, yielding an absolute gain of 3.17 percentage points. Meanwhile, SCD-YOLO achieves 77.62% precision and 70.30% recall with 2.71M parameters and 7.8 GFLOPs. Although its end-to-end FPS is lower than that of YOLOv11n, SCD-YOLO still reaches 119.63 FPS, satisfying the real-time requirement of industrial inspection. These results indicate that the proposed SC-C3k2 and LSDECD modules improve fine-grained defect representation and detail-aware prediction while preserving practical real-time performance.

As shown in Table 2, SCD-YOLO achieves the highest mAP@0.5 among the compared YOLO-series detectors. Compared with YOLOv8n and YOLOv11n, SCD-YOLO improves the overall detection accuracy while maintaining a compact

model size of 6.0 MB. Although YOLOv10n and YOLOv12n achieve higher FPS, their mAP@0.5 values are lower than that of SCD-YOLO. This indicates that SCD-YOLO provides a more favorable balance between detection accuracy and inference efficiency, which is important for industrial inspection scenarios where missed detections and false alarms may lead to quality risks.

4.5.2. Comparison with Representative Defect Detection Methods

To further evaluate the competitiveness of SCD-YOLO, we compare it with several representative surface defect detection methods on the NEU-DET dataset. As reported in Table 3, both overall mAP@0.5 and per-class AP values are evaluated across six defect categories, including crazing (Cr), inclusion (In), patches (Pa), pitted surface (Ps), rolled-in scale (Rs), and scratches (Sc). In addition, the number of parameters and GFLOPs are reported to provide a comprehensive comparison of detection accuracy and computational complexity.

Table 3. Performance comparison of different methods on the NEU-DET dataset.

Metrics	SGS-YOLO	Faster R-CNN	YOLO-SDS	RT-DETR	**SCD-YOLO**
Params (M)	1.5	111.4	1.97	19.8	2.7
GFLOPs	4.7	60.5	5.3	57.0	7.8
mAP@0.5 (%)	78.5	75.1	77.7	62.4	78.8
Cr (%)	58.5	38.9	47.1	28.6	56.0
In (%)	85.5	77.9	79.8	77.86	86.3
Pa (%)	94.9	85.3	95.4	85.76	93.9
Ps (%)	82.4	89.9	82.6	63.4	81.5
Rs (%)	66.4	67.5	66.9	51.8	68.5
Sc (%)	83.4	90.9	94.6	66.8	86.8

As shown in Table 3, SCD-YOLO achieves an overall mAP@0.5 of 78.8%, which is the highest among the compared methods. Compared with Faster R-CNN, YOLO-SDS, and RT-DETR, SCD-YOLO obtains higher overall detection accuracy. Compared with SGS-YOLO, SCD-YOLO also achieves a slightly higher mAP@0.5, while maintaining a compact model scale of 2.7 M parameters. Although its GFLOPs are higher than those of several lightweight methods, SCD-YOLO remains much more efficient than heavy detectors such as Faster R-CNN and RT-DETR. These results indicate that the proposed method provides a favorable balance between detection accuracy and computational complexity.

From the perspective of category-wise performance, SCD-YOLO achieves the best AP on rolled-in scale, with AP values 68.5%, respectively. These results suggest that the proposed SC-C3k2 and LSDECD components are beneficial for detecting weak-texture and background-confusable defects. For inclusion and scratches, SCD-YOLO obtains competitive AP values of 86.3% and 86.8%, respectively, although YOLO-SDS show higher performance on these categories. For patches and pitted

surface, SCD-YOLO also maintains competitive results, but its AP remains lower than the best-performing methods in the corresponding categories. This indicates that extremely irregular or category-specific defect patterns remain challenging and may require further targeted enhancement.

Overall, SCD-YOLO achieves the highest overall mAP@0.5 among the compared methods while maintaining a compact parameter scale and practical computational cost. In addition, the proposed model achieves an end-to-end inference speed of 119.63 FPS and a model file size of 6.0 MB under the adopted experimental setting. These results demonstrate that SCD-YOLO improves overall detection performance while retaining practical real-time applicability for strip steel surface defect inspection.

4.6. Error Analysis

Although SCD-YOLO improves the overall detection performance, several defect categories remain challenging. As reported in Table 3, crazing still obtains a relatively low AP compared with other categories. This is mainly because crazing defects usually appear as thin, fragmented, and low-contrast crack-like patterns, whose boundaries are easily confused with background texture variations. The proposed SC-C3k2 improves nonlinear feature interaction and contextual recalibration, which helps suppress background interference; however, extremely weak or discontinuous defect responses may still be partially missed. Pitted surface and scratches also show room for improvement. Pitted surface defects often exhibit dense and irregular distributions with blurred regional boundaries, which may lead to incomplete localization. Scratches are typically slender and direction-sensitive, requiring stronger long-range structural continuity modeling than local detail enhancement alone. In addition, patches and rolled-in scales may present large intra-class variation and strong similarity to surrounding steel textures, resulting in occasional category confusion. These observations indicate that SCD-YOLO is effective in enhancing weak-texture representation and detail-aware prediction, but the limited image resolution, small dataset scale, and high inter-class similarity of NEU-DET still make extremely subtle, elongated, and background-confusable defects difficult to detect reliably.

4.7. Ablation Study

To clarify the contribution of the newly proposed components, we conduct ablation experiments by distinguishing previously developed components from the extensions introduced in this work. The Star module and LSCD are inherited from our previous work and serve as prior architectural bases. In contrast, Star-CAA and LSDECD are the main improvements proposed in this paper. To ensure a fair comparison, all ablation variants are trained and evaluated under the same experimental settings, including the optimizer, data augmentation pipeline, number of training epochs, input image size, and stopping criterion. Therefore, the ablation study is designed to answer three questions: whether contextual recalibration im-

proves the original Star interaction, whether detail-enhanced convolution improves the original LSCD head, and whether the two newly proposed extensions are complementary when integrated into YOLOv11n.

As shown in Table 4, directly transferring the previous Star module to the YOLOv11n baseline decreases the mAP@0.5 from 75.64% to 72.43%, indicating that star-shaped multiplicative interaction alone may disturb the feature distribution when contextual regulation is absent. After introducing context anchor attention into the Star interaction path, the Star-CAA variant improves the mAP@0.5 to 75.41%, demonstrating that contextual recalibration helps stabilize nonlinear feature interaction and suppress unreliable background responses.

Table 4. Ablation study results on the NEU-DET dataset.

Star	Star-CAA	LSCD	LSDECD	Params (M)	GFLOPs	mAP@0.5 (%)
-	-	-	-	2.6	6.3	75.64
√	-	-	-	2.5	6.4	72.43
-	√	-	-	3.0	8.1	75.41
-	-	√	-	1.7	4.3	75.09
-	-	-	√	2.3	6.0	77.0
√	-	√	-	1.4	4.5	76.88
-	√	-	√	**2.7**	**7.8**	**78.81**

For the detection head, the previous LSCD design reduces the model complexity to 1.7 M parameters and 4.3 GFLOPs while achieving an mAP@0.5 of 75.09%. This confirms the compactness of the shared prediction design. However, LSCD still lacks explicit detail enhancement for weak edges and local texture discontinuities. By further incorporating detail-enhanced convolution, LSDECD improves the mAP@0.5 to 77.00%, exceeding the YOLOv11n baseline by 1.36 percentage points. This result verifies that the detail-enhanced prediction trunk is beneficial for localizing small, low-contrast, and weak-boundary defects.

The complete SCD-YOLO, which integrates the newly proposed Star-CAA and LSDECD, achieves the best mAP@0.5 of 78.81%. Compared with the previous Star + LSCD configuration, the proposed Star-CAA + LSDECD design improves the mAP@0.5 by 1.93 percentage points. These results demonstrate that the improvement of SCD-YOLO mainly comes from the two extensions introduced in this work, namely contextual recalibration within star-shaped interaction and detail-enhanced convolution within the shared detection head.

5. Conclusion

In this work, we propose SCD-YOLO, a compact and detail-aware detection framework for strip steel surface defect inspection. Built upon the YOLOv11n baseline, SCD-YOLO introduces two complementary architectural components: a Star-Context Enhanced C3k2 structure for fine-grained feature extraction and a Lightweight

Shared Detail-Enhanced Convolutional Detection Head for detail-aware prediction.

The proposed SC-C3k2 enhances the feature extraction stage by embedding context anchor attention into the star-shaped multiplicative interaction path. This design improves high-order nonlinear feature representation and strengthens contextual discrimination, enabling the detector to capture weak defect cues while suppressing pseudo-defect activations caused by complex surface backgrounds. The proposed LSDECD improves the prediction stage by combining shared multi-scale prediction transformation with detail-enhanced convolution. By reducing redundant cross-scale prediction transformations and enhancing local structural responses, LSDECD strengthens the localization and recognition of small, weak-boundary, and low-contrast defects.

Experimental results on the NEU-DET dataset demonstrate the effectiveness of the proposed method. Compared with the YOLOv11n baseline, SCD-YOLO improves the mAP@0.5 from 75.64% to 78.81%, achieving an absolute gain of 3.17 percentage points. Meanwhile, the proposed model maintains a compact model size of 6.0 MB and an end-to-end inference speed of 119.63 FPS, indicating practical real-time applicability for industrial inspection scenarios. Ablation experiments further confirm that SC-C3k2 and LSDECD provide complementary improvements in defect representation and detail-aware prediction.

Despite these improvements, several limitations remain. First, the current speed evaluation is conducted on a desktop GPU, and deployment performance on edge or embedded devices still requires further validation. Second, extremely small or low-contrast defects remain challenging under cluttered backgrounds. Third, redundant or overlapping predictions may still occur in complex surface-texture scenarios. Future work will therefore focus on edge-device deployment, lightweight small-defect enhancement, and improved localization strategies to further improve the robustness and deployment reliability of SCD-YOLO.

Funding

Research Project on the Reform of Graduate Education and Teaching in Sichuan University of Science and Engineering (No. JG202440).

Author Contributions

Conceptualization, J.J.; methodology, J.J.; software, J.J.; validation, J.J.; formal analysis, J.J.; investigation, J.J.; data curation, J.J.; visualization, J.J.; writing—original draft preparation, J.J.; writing—review and editing, J.J. and Y.S.; supervision, Y.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Wang, J., Chen, T., Xu, X., Zhao, L., Yuan, D., Du, Y., *et al.* (2024) An Improved

YOLOv8 Model for Strip Steel Surface Defect Detection. *Applied Sciences*, **15**, Article 52. https://doi.org/10.3390/app15010052

[2] Kou, X., Liu, S., Cheng, K. and Qian, Y. (2021) Development of a YOLO-V3-Based Model for Detecting Defects on Steel Strip Surface. *Measurement*, **182**, Article ID: 109454. https://doi.org/10.1016/j.measurement.2021.109454

[3] Chu, Y., Yu, X. and Rong, X. (2024) A Lightweight Strip Steel Surface Defect Detection Network Based on Improved YOLOv8. *Sensors*, **24**, Article 6495. https://doi.org/10.3390/s24196495

[4] Saklakoglu, N., Bolouri, A., Irizalp, S.G., Baris, F. and Elmas, A. (2021) Effects of Shot Peening and Artificial Surface Defects on Fatigue Properties of 50CrV4 Steel. *The International Journal of Advanced Manufacturing Technology*, **112**, 2961-2970. https://doi.org/10.1007/s00170-020-06532-y

[5] Zheng, X., Zheng, S., Kong, Y. and Chen, J. (2021) Recent Advances in Surface Defect Inspection of Industrial Products Using Deep Learning Techniques. *The International Journal of Advanced Manufacturing Technology*, **113**, 35-58. https://doi.org/10.1007/s00170-021-06592-8

[6] Wen, X., Shan, J., He, Y. and Song, K. (2022) Steel Surface Defect Recognition: A Survey. *Coatings*, **13**, Article 17. https://doi.org/10.3390/coatings13010017

[7] Pan, Y. and Zhang, L. (2022) Dual Attention Deep Learning Network for Automatic Steel Surface Defect Segmentation. *Computer-Aided Civil and Infrastructure Engineering*, **37**, 1468-1487. https://doi.org/10.1111/mice.12792

[8] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[9] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 201, 2980-2988. https://doi.org/10.1109/iccv.2017.322

[10] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al.* (2016) SSD: Single Shot Multibox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV* 2016, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[11] Tan, M., Pang, R. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 10778-10787. https://doi.org/10.1109/cvpr42600.2020.01079

[12] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[13] Li, C.Y., Li, L.L., Jiang, H.L., Weng, K.H., Geng, Y.F., Li, L., *et al.* (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv: 2209.02976.

[14] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 7464-7475. https://doi.org/10.1109/cvpr52729.2023.00721

[15] Jocher, G., Chaurasia, J., Qiu, A. and Stoken, K. (2024) Ultralytics YOLOv11. GitHub Repository. https://github.com/ultralytics/ultralytics

[16] Jiang, J., Su, Y. and Song, J. (2026) A YOLOv8-Based Network for Surface Defect Detection in Strip Steel. *Journal of Computer and Communications*, **14**, 106-129. https://doi.org/10.4236/jcc.2026.142006

[17] Wang, C., Yeh, I. and Mark Liao, H. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Computer Vision—ECCV* 2024, Springer, 1-21. https://doi.org/10.1007/978-3-031-72751-1_1

[18] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., *et al.* (2024) YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems* 37, Vancouver, 10-15 December 2024, 107984-108011. https://doi.org/10.52202/079017-3429

[19] He, Y., Song, K., Meng, Q. and Yan, Y. (2020) An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features. *IEEE Transactions on Instrumentation and Measurement*, **69**, 1493-1504. https://doi.org/10.1109/tim.2019.2915404

[20] Jain, S., Seth, G., Paruthi, A., Soni, U. and Kumar, G. (2020) Synthetic Data Augmentation for Surface Defect Detection and Classification Using Deep Learning. *Journal of Intelligent Manufacturing*, **33**, 1007-1020. https://doi.org/10.1007/s10845-020-01710-x

[21] Wu, Y., Chen, R., Li, Z., Ye, M. and Dai, M. (2024) SDD-YOLO: A Lightweight, High-Generalization Methodology for Real-Time Detection of Strip Surface Defects. *Metals*, **14**, Article 650. https://doi.org/10.3390/met14060650

[22] Wei, M., Chen, B., Liu, J., Yuan, N., Liu, J. and Ji, Z. (2024) AEDN-YOLO: An Efficient One-Stage Detection Network for Strip Steel Surface Defects. *Engineering Research Express*, **6**, Article ID: 035415. https://doi.org/10.1088/2631-8695/ad681d

[23] Kong, S., Kong, Y., Chi, X., Feng, X. and Ma, L. (2025) A TBD-YOLO-Based Surface Defect Detection Method for Hot Rolled Steel Strips. *Russian Journal of Nondestructive Testing*, **61**, 137-149. https://doi.org/10.1134/s1061830924603192

[24] Liu, R., Huang, M., Gao, Z., Cao, Z. and Cao, P. (2023) MSC-DNet: An Efficient Detector with Multi-Scale Context for Defect Detection on Strip Steel Surface. *Measurement*, **209**, Article ID: 112467. https://doi.org/10.1016/j.measurement.2023.112467

[25] Yang, S., Wang, X., Qian, X., Yu, Y. and Jin, W. (2023) Trident-LK Net: A Lightweight Trident Structure Network with Large Kernel for Muti-Scale Defect Detection. *IEEE Access*, **11**, 131073-131080. https://doi.org/10.1109/access.2023.3333918

[26] Wen, Y. and Wang, L. (2024) YOLO-SD: Simulated Feature Fusion for Few-Shot Industrial Defect Detection Based on YOLOv8 and Stable Diffusion. *International Journal of Machine Learning and Cybernetics*, **15**, 4589-4601. https://doi.org/10.1007/s13042-024-02175-7

[27] Ma, X., Dai, X., Bai, Y., Wang, Y.Z. and Fu, Y. (2024) Rewrite the Stars. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 16-22 June 2024, 5694-5703. https://doi.org/10.1109/cvpr52733.2024.00544

[28] Cai, X., Lai, Q., Wang, Y., Wang, W., Sun, Z. and Yao, Y. (2024) Poly Kernel Inception Network for Remote Sensing Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 16-22 June 2024, 27706-27716. https://doi.org/10.1109/cvpr52733.2024.02617

[29] Quan, Z. and Sun, J. (2025) A Feature-Enhanced Small Object Detection Algorithm Based on Attention Mechanism. *Sensors*, **25**, Article 589. https://doi.org/10.3390/s25020589

[30] Chen, W., Liu, J., Liu, T. and Zhuang, Y. (2025) PCPE-YOLO with a Lightweight

Dynamically Reconfigurable Backbone for Small Object Detection. *Scientific Reports*, **15**, Artice No. 29988. https://doi.org/10.1038/s41598-025-15975-w

[31] Ge, Z., Liu, S.T., Wang, F., Li, Z.M. and Sun, J. (2021) YOLOX: Exceeding YOLO Series in 2021. arXiv: 2107.08430.

[32] Tian, Z., Shen, C., Chen, H. and He, T. (2019) FCOS: Fully Convolutional One-Stage Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 9626-9635. https://doi.org/10.1109/iccv.2019.00972

[33] Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J. and Yang, J. (2020) Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Proceedings of the* 34*th International Conference on Neural Information Processing Systems*, Vancouver, 6-12 December2020, 21002-21012.

[34] Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., *et al.* (2020) Searching Central Difference Convolutional Networks for Face Anti-Spoofing. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 5294-5304. https://doi.org/10.1109/cvpr42600.2020.00534

[35] Chen, Z., He, Z. and Lu, Z. (2024) Dea-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention. *IEEE Transactions on Image Processing*, **33**, 1002-1015. https://doi.org/10.1109/tip.2024.3354108

[36] Wu, Y. and He, K. (2018) Group Normalization. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, 3-19. https://doi.org/10.1007/978-3-030-01261-8_1

[37] Tian, Y., Ye, Q. and Doermann, D. (2025) YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv: 2502.12524.

[38] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., *et al.* (2024) DETRs Beat Yolos on Real-Time Object Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 16-22 June 2024, 16965-16974. https://doi.org/10.1109/cvpr52733.2024.01605