coatings

MDPI

Review

Deep Learning-Based Steel Surface Defect Detection: A Survey

Xin Wen [1], Zhenhao Yu [1], Yu He [1], Ziteng Wang [1], Xi Chen [1] and Kechen Song [2,*]

[1] School of Software, Shenyang University of Technology, Shenyang 110870, China; wen_xin@sut.edu.cn (X.W.); yuzhenhao@smail.sut.edu.cn (Z.Y.); heyu_142616@sut.edu.cn (Y.H.); ztwang@smail.sut.edu.cn (Z.W.); chenxi0914@sut.edu.cn (X.C.)
[2] School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China
* Correspondence: songkc@me.neu.edu.cn

Abstract

With the rapid development of manufacturing, increasingly stringent requirements for material quality and inspection efficiency have promoted the widespread application of deep learning-based nondestructive testing technologies in industrial quality control. In recent years, steel surface defect detection has expanded from conventional inspection scenarios with controlled imaging conditions, such as steel strips, plates, and welds, to more challenging applications involving the inner surfaces of steel pipes, castings, and complex industrial components. This paper reviews recent advances in deep learning-based steel surface defect detection from an industrial application-oriented perspective, with particular emphasis on developments in application scenarios, datasets, methodological frameworks, and research trends. It analyzes representative reviews and their limitations, summarizes publicly available datasets covering conventional flat steel products, pipes, and castings, examines the defect characteristics of different inspection objects, and reviews the development of deep learning and its industrial applications. Furthermore, defect detection and segmentation methods are systematically organized according to practical requirements, including challenging imaging conditions, limited annotated data, small and low-contrast defects, real-time deployment, and cross-domain generalization. Finally, the key challenges, emerging trends, future research directions, and priorities for the next stage of development are discussed.

Keywords: steel; deep learning; defect segmentation; defect detection; survey

1. Introduction

Steel, as a high-strength material, is widely used in the transportation of media such as hydraulic oil, fuel oil, lubricating oil, and air, covering many important industries including petroleum, natural gas, aerospace, hydropower, and chemicals. In complex industrial manufacturing processes, damage to the material surface is inevitable; common surface defects include cracks, rust spots, and pits. These defects can seriously affect product performance and service life, potentially leading to structural failure and serious safety accidents, as well as wasting resources and reducing revenue. Therefore, identifying and recognizing these defects in the early stages of production is crucial for reducing accident risks and ensuring quality compliance.

With the continuous development of testing technologies, non-destructive testing (NDT) technology plays an increasingly important role in industrial quality control. It can evaluate materials without damaging them, reducing the overall production costs, saving significant time, and improving work efficiency compared to manual inspection.

Check for updates

Academic Editor: Giorgos Skordaris

Received: 2 August 2026
Revised: 27 August 2026
Accepted: 29 August 2026
Published: 2 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Computer vision inspection technology is one of the most commonly used NDT technologies. It is also a non-contact inspection technology, while other NDT technologies generally require special media to achieve good inspection results. This non-contact testing method avoids scratches and contamination on the surfaces of materials requiring high precision, thus ensuring the integrity of the tested samples.

In computer vision inspection, a complete recognition system consists of three parts: image acquisition, defect recognition, and quality control. Industrial cameras are generally used for image acquisition [1], and these cameras can be further divided into two categories: CCD cameras and CMOS cameras. CCD cameras are more frequently used in steel surface defect data acquisition because they can adapt to a wider brightness range, 3–10 times that of a typical CMOS camera, resulting in higher imaging quality. However, CMOS cameras have an advantage in the harsh inspection environment of actual steel mills. Furthermore, a light source is needed for supplementary lighting during image acquisition [2], with LED light sources being the most widely used in steel defect recognition [3]. After acquiring product images through image acquisition equipment, the recognition system calls defect recognition methods to analyze and detect the images.

With the development of deep learning technology and the continuous improvement of image processing, automation, and computing capabilities, deep learning-based detection methods have advantages such as strong automatic feature extraction capabilities, high detection accuracy, and the ability to identify complex defects without the need for manual feature design. This method has become an important tool for achieving efficient recognition and accurate analysis in the field of steel surface defect detection. The application of deep learning in steel surface defect detection has evolved from simple to complex, mainly including the deep learning classification technology stage, the object detection technology stage, the semantic segmentation and attention mechanism stage, and the multi-scale fusion and Transformer stage. Each stage has continuously improved in terms of detection accuracy, localization capability, and feature representation, and its key development milestones are shown in Figure 1.

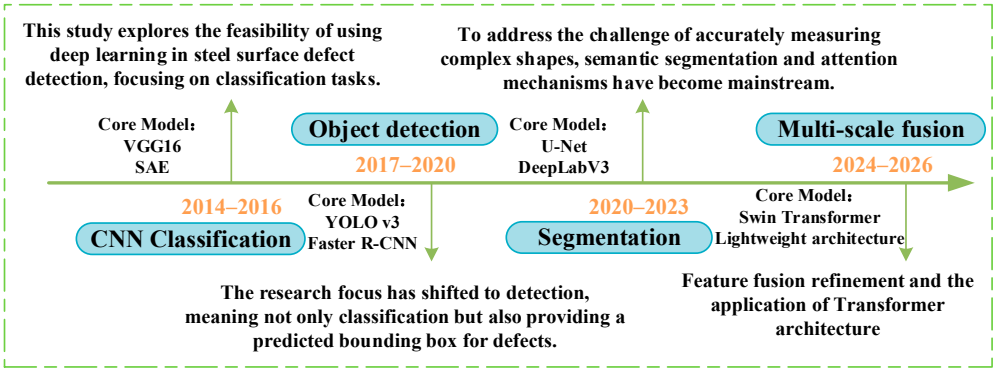


Figure 1. Development history of deep learning-based steel surface defect detection.

In the early stages, research mainly focused on defect classification based on convolutional neural networks. Convolutional layers extracted local visual features such as texture, edges, and grayscale variations, and classifiers were used to determine the defect category. However, classification tasks struggled to provide the specific spatial location of defects within an image, nor could they accurately describe the scale, boundaries, and morphological features of defects. Therefore, as industrial inspection needs shifted from category recognition to defect localization, the limitations of simple classification frameworks became increasingly apparent. Research gradually shifted toward object detection tasks. Object detection methods not only focus on defect category determination, but also predict defect locations through bounding box regression. Compared to classification

methods, object detection frameworks significantly enhance spatial perception capabilities, enabling the simultaneous processing of multiple defect targets in an image and providing their corresponding location ranges. However, steel surface defects often exhibit irregular shapes, blurred edges, and significant scale differences. Rectangular bounding boxes can only roughly describe the defect area, failing to meet the needs of fine measurement and morphological analysis. Therefore, semantic segmentation methods have become an important research direction. Semantic segmentation provides a finer-grained characterization of defect areas through pixel-level classification, enabling models to obtain regional information close to the true defect boundaries. Semantic segmentation methods are particularly suitable for handling complex targets such as long, thin cracks, patchy defects, and defects with irregular boundaries. However, segmentation models typically place higher demands on feature representation quality, annotation accuracy, and computational resources, and still face challenges in scenarios with small target defects, low-contrast defects, and complex background interference. Building upon the methods above-mentioned, multi-scale feature fusion and attention mechanisms have gradually become important technical paths for improving model performance. Steel surface defects often exhibit significant scale diversity in real-world images. Single-scale features struggle to simultaneously capture both global structural information and local detail information. Therefore, multi-scale fusion methods enhance the model's adaptability to defects of different sizes and shapes by integrating shallow detail features with deep semantic features. With the introduction of Transformer and its related attention mechanisms, the model can more effectively capture long-distance dependencies and global contextual information, thereby compensating for the shortcomings of traditional convolutional networks in global modeling. The combination of lightweight structures further addresses the requirements of industrial environments for real-time performance, deployment cost, and computational efficiency.

Table 1 summarizes representative reviews relevant to steel surface defect detection and compares them in terms of subject, dataset perspective, core perspective, research focus, and future trends. Some reviews examined surface defect detection across a broad range of industrial materials, including textiles, ceramics, and steel products. Consequently, their discussions of steel-specific defect characteristics, datasets, and application requirements are relatively limited. Other reviews focused specifically on steel surface defects but organized existing methods primarily according to learning paradigms and supervision strategies, with limited attention to the correspondence between algorithmic capabilities and practical industrial requirements.

Table 1. Overview and summary of steel surface defect detection.

Reference	Subject	Dataset Perspective	Core Perspective	Research Focus	Future Trends
Chen et al. [4]	Steel surface defects	Scarcity of large-scale, high-quality labeled data	Transition from handcrafted to deep features	Detector taxonomy and industrial applicability	Small-defect detection, generalization and lightweight deployment
Ibrahim et al. [5]	Steel strip defects	Need for standardized benchmarks	Unified taxonomy of visual inspection methods	Industrial imaging and supervised inspection	Semi-supervised learning, data augmentation and domain adaptation

Table 1. *Cont.*

Reference	Subject	Dataset Perspective	Core Perspective	Research Focus	Future Trends
Zhang et al. [6]	Automated industrial surface inspection	Limited and unlabeled industrial data	Semi-supervised and unsupervised inspection pipelines	Feature fusion, compression and real-time processing	Efficient processing with subtle-information preservation
Shukla et al. [7]	Mobile industrial anomaly inspection	Industrial anomaly detection datasets	Comparison of supervised, semi-supervised and unsupervised methods	Utilization of unlabeled normal samples	Self-supervised learning, mobile inspection and generalization
Nahar et al. [8]	Multi-material surface defects	Data scarcity and class imbalance	Multi-paradigm framework for defect analysis	Model reliability and edge deployment	Generative, Transformer and foundation models
F Shang et al. [9]	Strip steel surface defects	Data scarcity, class imbalance, and standardized benchmarks	Scenario-specific framework bridging laboratory research and industrial deployment	Small and low-contrast defects, multi-scale feature fusion, lightweight models, and real-time edge deployment	Self-supervised learning, multimodal sensor fusion, and digital twin-driven quality control
Wen X et al. [10]	Steel products and structures	Datasets, acquisition methods and evaluation metrics	Organizes existing hardware and learning-oriented review	Data acquisition, hardware configuration, traditional machine learning and deep learning	Transfer learning, lightweight models and hardware–software co-design
Our Survey	Steel strips; steel pipes; castings	Comparative analysis of 17 public and selected non-public datasets	Needs-oriented organization of detection methods	Complex imaging, limited labels, small defects, real-time deployment and cross-domain generalization	3D inspection, automated inspection and VLM/LLM-enabled multimodal detection

More specifically, Chen et al. [4] described the transition from handcrafted features to deep feature learning and compared the industrial applicability of one-stage and two-stage detectors. Ibrahim et al. [5] established a unified taxonomy covering statistical, spectral, texture-based, and machine-learning methods, with particular attention to industrial image quality and standardized benchmark datasets. Zhang et al. [6] reviewed semi-supervised and unsupervised inspection pipelines, emphasizing multi-source feature fusion, dimensionality reduction, and real-time processing. Shukla et al. [7] extended industrial anomaly detection to drones, robotic arms, and AGVs and highlighted the use of unlabeled normal samples and self-supervised learning. Nahar et al. [8] examined defect detection across multiple industrial materials and discussed multi-paradigm learning, model reliability, edge deployment, and emerging architectures. Collectively, these reviews provide valuable methodological perspectives but differ substantially in material scope, dataset coverage, and attention to practical industrial requirements.

Among these studies, our previous review by Wen X et al. [10], published in 2023, primarily focused on the hardware configurations, data acquisition strategies, and recognition algorithms used in steel surface defect inspection systems. In contrast, the present review places greater emphasis on the systematic analysis and evolution of steel surface defect datasets. It extends dataset coverage beyond conventional steel strips to more challenging inspection targets, including the inner surfaces of steel pipes and castings with complex geometries and imaging conditions. Methodologically, Wen X et al. organized recognition algorithms mainly according to their learning paradigms and supervision strategies,

whereas the present review adopts an application- and needs-oriented framework. This framework organizes detection methods according to practical industrial challenges, including challenging imaging conditions, limited annotations, small and low-contrast defects, real-time deployment, and cross-domain generalization.

A systematic literature search was conducted to identify studies on steel surface defect detection and related industrial visual inspection technologies, with particular emphasis on steel strips, steel pipes, and castings. Relevant publications were retrieved from the Web of Science Core Collection, IEEE Xplore and SpringerLink. Google Scholar was used as a supplementary source to identify potentially relevant studies not indexed in the primary databases and to support backward and forward citation tracking. The search strategy combined terms describing inspection objects with those relating to visual inspection tasks and computational approaches, including steel surface defect detection, steel strip defect, steel pipe defect, casting defect, industrial visual inspection, machine vision, deep learning, object detection, semantic segmentation, and anomaly detection.

The reviewed literature spanned the period from 1992 to 2026, with 112 representative studies ultimately included. Earlier publications were retained primarily to trace the evolution of machine vision and deep learning in industrial defect inspection, whereas the methodological analysis focused predominantly on advances reported between 2024 and 2026. Early studies investigated surface-shape analysis, texture descriptors, and conventional machine-learning techniques. Subsequent research increasingly adopted convolutional neural networks, object detection, semantic segmentation, semi-supervised learning, unsupervised learning, and few-shot learning. More recent studies have explored Transformer-based architectures, lightweight networks, cross-domain generalization, 3D anomaly detection, and the intelligent inspection of geometrically complex objects, particularly pipe inner surfaces and castings.

Studies were considered eligible if they addressed steel products or representative industrial surface defects, investigated defect classification, object detection, segmentation, or anomaly detection, and provided sufficiently documented information on datasets, methodologies, experimental results, or evaluation metrics. Publications with ambiguous or insufficient experimental details were excluded, particularly those that did not clearly specify the source or composition of the datasets, provided inadequate evaluation results, or lacked sufficient experimental evidence to support their conclusions. Duplicate publications, studies unrelated to industrial surface inspection, and studies concerned exclusively with material properties without addressing visual inspection or anomaly detection were also excluded. On this basis, the selected literature was systematically analyzed in terms of inspection objects, dataset characteristics, methodological frameworks, industrial requirements, and emerging research directions.

Based on the above description, this paper introduces the latest advancements in deep learning-based steel surface defect detection. The main contributions of this paper are as follows:

1. This paper details the types and characteristics of surface defects in three of the most commonly used engineering steel materials: strip steel, steel pipe, and castings. It also reviews 17 publicly available datasets and some unavailable datasets, introducing the defect types, data characteristics, and acquisition methods of these datasets.
2. Following the evolution of deep learning detection technology and addressing the practical needs of industrial inspection scenarios, this paper systematically reviews mainstream semantic segmentation algorithms and object detection algorithms. It also summarizes over 20 commonly used experimental evaluation metrics.
3. Based on a systematic summary of existing research results, this paper deeply analyzes the future development trends and potential research directions in the field of steel

surface defect detection, and provides an outlook on the key challenges and scientific problems that need to be addressed in this field.

2. Types of Steel Surface Defects and Datasets

2.1. Types of Steel Surface Defects

Throughout the entire steel production process, from raw material smelting and continuous casting to hot and cold rolling, the material is constantly exposed to a complex and ever-changing environment. During this process, the material is subjected to multiple factors, including drastic temperature changes, chemical corrosion, and equipment friction, making it highly susceptible to various types of surface damage. Furthermore, due to differences in processing paths and forming mechanisms, different types of materials exhibit certain differences in the types and manifestations of surface defects.

2.1.1. Surface Defects of Strip Steel

The smelting and continuous rolling process of strip steel is extremely complex. When subjected to harsh conditions such as high-temperature heating, multiple strong extrusions, high-speed friction, and cooling coiling, it is affected by a combination of mechanical and thermodynamic factors such as roll wear, process parameter fluctuations, media pollution, and environmental erosion. As a result, various defects of different shapes are easily generated on its surface.

Among the major defects in strip steel, the most representative are cracks, inclusions, patches, pitting, indentation of iron oxide scale, and scratches. These defects occur far more frequently than other defects, and their characteristics and formation principles are as follows:

Crazing on the surface of strip steel usually appear as a network or irregular polygonal interwoven microcracks. Their formation mechanism is mainly related to the surface condition of the roll and thermal fatigue stress. During hot rolling, the roll is subjected to a repeated rapid cooling and heating alternating stress environment for a long time, and thermal fatigue network cracks are easily generated on the surface. When a roll with such microcracks comes into contact with the high temperature strip steel, the morphology of the cracks will be transferred to the surface of the strip steel, forming crack defects. In addition, the uneven thermal stress caused by the excessive local temperature gradient during the cooling process of the strip steel may also promote the initiation and expansion of surface microcracks [11].

Inclusions appear as dot-shaped, teardrop-shaped or strip-shaped foreign objects extending along the rolling direction on the surface of the strip steel. Their color and reflectivity are usually significantly different from the surrounding metal matrix. The essence of this type of defect is non-metallic impurities or protective slag that have not been fully floated and discharged during the smelting or continuous casting process of the molten steel. During the subsequent rolling deformation process, these non-metallic inclusions exposed on the surface are inconsistent with the plastic deformation capacity of the steel matrix, which not only destroys the continuity of the material surface, but also easily evolves into microcrack sources when subjected to stress, seriously weakening the mechanical properties and fatigue life of the steel [12].

Patches are a discolored area with cloud-like, sheet-like or irregular outlines in a local area of the strip steel, which usually do not have obvious depth characteristics. Its formation mechanism is relatively complex and is generally associated with inadequate control of the fluid medium and temperature during rolling. For instance, nonuniform distribution or the incomplete removal of rolling lubricants or emulsions on the strip surface may lead to localized carbonization or adhesion under high-temperature conditions. In addition, varia-

tions in the local cooling rate during laminar cooling can cause nonuniform surface phase transformation and slight differences in oxidation state. Patch defects primarily degrade the surface quality of steel and may adversely affect its subsequent coating performance [13].

Pitted surface is characterized by dense, tiny pits or holes on the surface of the strip steel. These may be locally clustered or globally diffused, causing the steel surface to lose its proper metallic luster and exhibit rough diffuse reflection characteristics. The causes are mainly divided into two categories: one is mechanical indentation type, where tiny iron oxide scale or broken roll debris is pressed into the strip steel during rolling and then falls off during pickling or leveling processes, leaving tiny pits; the other is corrosion type, where local electrochemical corrosion occurs due to the strip steel being corroded by moisture or acidic atmosphere in the storage or transportation environment, resulting in micro-pitted corrosion [14].

Rolled in scale defects refer to black or grayish-black dot-like, strip-like, or fish-scale-like indentation areas visible on the surface of the strip, with residual oxides often present at the bottom of the pits. The formation of such defects is directly attributed to the failure of the descaling process. If the high-pressure water descaling system fails to completely remove the primary or secondary iron oxide scale from the surface of the billet before hot rolling or during rough rolling, these high-hardness oxides will be forcibly pressed into the soft high-temperature austenitic matrix under the enormous rolling pressure. After cooling, some of the oxide scale falls off to form indentations, while the remaining part is firmly embedded in the steel surface, severely reducing the surface quality of the material and easily leading to strip breakage during subsequent cold working [15].

Scratches are one of the most common linear defects on the surface of steel, usually manifested as long and thin grooves extending in a certain direction. These defects are mostly caused by foreign object scratching, mechanical contact, roll indentation, and friction during transportation and handling. Scratches not only disrupt the continuity of the steel surface and affect its appearance quality, but in more severe cases, they may also form stress concentration areas, thereby inducing early failure of the material [16].

2.1.2. Surface Defects of Steel Pipes

Pipes, as deep-processed products of strip steel and plate steel, or directly hot-rolled/cold-drawn from metal tube blanks, have surface defects that are largely similar to those of materials and processes.

However, due to the unique three-dimensional closed tubular structure of pipelines, the more complex forming and heat treatment processes they undergo, and the harsher service environments they operate in, pipelines have also developed many types of defects with their own distinct characteristics, such as rust, pitting, and white streaks. Their characteristics and formation principles are as follows.

Rust defects typically manifest as reddish-brown, yellowish-brown, or black oxide scale covering the inner and outer surfaces of the pipe wall; in severe stages, the rust layer spalls off, resulting in extensive surface roughness and wall thinning. The primary mechanism underlying its formation is a spontaneous electrochemical reaction between the metal substrate and the surrounding environment. Rusting not only reduces the effective load-bearing cross-section of the pipe wall, but under stress, also readily evolves into corrosion fatigue cracks, ultimately leading to pipe leakage or rupture [17].

Pit defects manifest as localized inward depressions in the pipe wall, resulting in a significant alteration of the pipe cross-sectional curvature; however, they are typically not accompanied by a volumetric loss of the metal material. Mechanical dents severely disrupt the geometric continuity of the pipeline, making the pit edges highly susceptible to severe stress and strain concentrations under cyclic internal pressure loading [18].

White band defects are frequently observed on the surfaces of stainless steel tubes following pickling. Macroscopically, they manifest as light-colored or grayish-white bands extending longitudinally along the steel tube or distributed in a spiral pattern. Compared to the normal metal matrix, the white-band regions exhibit distinct differences in reflectivity and visual color. The underlying mechanism for their formation lies in the continuous casting process, wherein electromagnetic stirring is employed to promote compositional homogeneity within the molten steel. This stirring induces solute redistribution within the mold, creating a segregation zone at the solid–liquid interface; when subsequently subjected to acid pickling, this zone appears as white, ring-like bands [19].

2.1.3. Surface Defects of Castings

In contrast to steel strips and pipes, the formation of castings relies on the physical phase transition processes of liquid metal mold filling and subsequent cooling and solidification. This fundamental difference in manufacturing processes dictates that surface defects in castings exhibit distinct characteristics regarding their physical origins, morphological topology, and the difficulty associated with extracting their visual features. Typical primary defects found on casting surfaces primarily manifest as shrinkage cavities, shrinkage porosity, cold shuts, and surface blowholes [20]. However, currently available datasets for castings are scarce; furthermore, these casting-specific defects—which possess strong domain specificity—often lack fine-grained semantic annotations. Most existing datasets tend to adopt classification paradigms designed for generic steel surface defects, focusing primarily on common anomalies such as scratches, pitting, and patches. Consequently, the core challenge and primary research hurdle for casting defect datasets at this stage lie in overcoming the complexities of their highly intricate non-planar topological structures and the severe interference caused by high-frequency, rough background textures, rather than merely perceiving the diversity of defect categories. This current scarcity of fine-grained, domain-specific annotated data objectively constitutes a practical bottleneck hindering the deep semantic understanding and fine-grained feature representation capabilities of contemporary casting surface defect detection models.

2.2. Steel Surface Defects Dataset

High-quality datasets play a pivotal role in driving the advancement of steel surface defect detection technologies. However, in real-world industrial settings, the data acquisition process is constrained by various physical factors, rendering the procurement of high-quality defect samples a significantly challenging task. Given that modern manufacturing processes widely employ statistical process control mechanisms, product yields typically remain at high levels; consequently, defect samples—occurring as low-probability events—exhibit an inherently highly imbalanced distribution. To construct statistically significant defect datasets, it is often necessary to conduct large-scale screening across massive volumes of products, a requirement that imposes substantial time and cost overheads during the data acquisition phase. We systematically reviewed and summarized the publicly available datasets utilized for steel surface defect detection. Figure 2 shows examples of defects from some of these datasets. Each dataset entry includes its name, access channels, and corresponding references, thereby facilitating further consultation and utilization by researchers and providing reliable data support for subsequent model training, performance evaluation, and comparative analysis of methodologies. A summary of representative datasets in the field of steel defect detection is presented in Table 2.

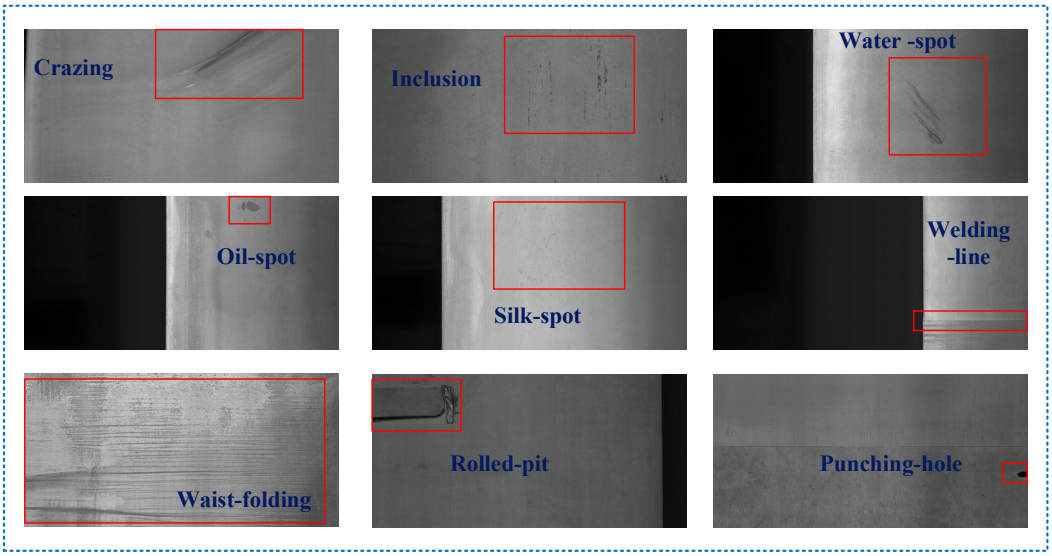


Figure 2. Steel surface defect dataset images [21].

Table 2. Steel surface defects dataset summary.

Dataset	Year	Object	Description	Link
Severstal	2019	Steel strip	The defects are divided into four categories: scratches, inclusions, cracks, and corrosion. A total of 6666 images are included, with a pixel resolution of 256 × 1600, and were taken by a high-frequency camera.	https://www.kaggle.com/competitions/severstal-steel-defect-detection/overview (accessed on 28 August 2026)
NEU-DET	2020	Hot-rolled strip steel	These defects are categorized into six classes: cracks, inclusions, patches, pitted surfaces, rolled-in scale, and scratches. The dataset comprises 300 grayscale images for each category, totaling 1800 images.	https://pan.baidu.com/s/1nBbO-jWDm1_NHDQsc1dRkg (accessed on 28 August 2026)
GC10-DET	2020	Steel strip	This dataset contains ten types of surface defects: punching, welds, crescent-shaped gaps, water stains, oil stains, filamentous spots, inclusions, rolling pits, creases, and waist folds, comprising a total of 3570 grayscale images. All defects were located on the surface of the steel plate.	https://www.mdpi.com/1424-8220/20/6/1562 (accessed on 28 August 2026)
SD-saliency-900	2020	Steel strip	Three categories: inclusion, patches, and scratches consist of total 90,900 steel surface defect detection images and corresponding pixel-wise binary maps.	https://www.kaggle.com/datasets/alex000kim/sdsaliency900 (accessed on 28 August 2026)
NEU-Seg	-	Steel strip	The dataset contains four categories: background, inclusions, and scratches, totaling 4500 images with a resolution of 200 × 200 pixels.	https://github.com/Chris2wzk/UNet-NEU-SEG (accessed on 28 August 2026)
X-SDD	2021	Hot-rolled strip steel	The resolution of each defect image is 128 × 128 pixels, and the image is in 3-channel JPG format. The dataset contains seven types of 1360 defect images, including 238 slag inclusions, 397 red iron sheet, 122 iron sheet ash, 134 surface scratches, 63 oxide scale of plate system, 203 finishing roll printing and 203 oxide scale of temperature system.	https://www.mdpi.com/2073-8994/13/4/706 (accessed on 28 August 2026)

Table 2. *Cont.*

Dataset	Year	Object	Description	Link
KolektorSDD2	2021	General Industrial Products	The dataset consists of images with a resolution of approximately 230 × 630 pixels, including 356 positive samples and 2979 negative samples. The dataset is divided into training and testing sets: the training set consists of 246 positive samples and 2085 negative samples; the testing set contains 110 positive samples and 894 negative samples, with only scratches as the defect type.	https://www.vicos.si/resources/kolektorsdd2/ (accessed on 28 August 2026)
BSData	2021	Casting	This dataset contains three types of defects: pitting, stains, and dents, totaling 1104 3-channel images.	https://github.com/2Obe/BSData (accessed on 28 August 2026)
NEU-FSSD-12	2023	Steel strip	The dataset employed in this study comprises 12 distinct defect categories: welding-line, rolled-in-scale, water-spot, oil-spot, silk-spot, rolled-pit, one-inclusion, crease, warm-folding, punching-hole, crazing, and scratches. It contains a total of 600 images, with a perfectly balanced distribution of 50 images allocated to each defect category.	https://github.com/VDT-2048/CPANet (accessed on 28 August 2026)
NEU-FSC-20	2023	Steel strip	This dataset was collected from three datasets: NEU-CLS, X-SDD, and GC10-DET, with a total of 20 categories. The twenty categories are welding-line, rolled-in-scale, water-spot, oil-spot, silk-spot, rolled-pit, one-inclusion, crease, warm-folding, punching-hole, crazing, crescent-gap, patches, pitted-surface, rolled-in-scale, two-inclusion, iron-sheet-ash, oxide-scale, red-iron, slag-inclusion, and scratches. There are fifty images for each category, for a total of 1000 images.	https://github.com/VDT-2048/FSC-20 (accessed on 28 August 2026)
SSP2000	2024	Steel pipe	This dataset contains nine defect types: cross-striation, peeling, pit, pockmark, rust, sand streak, scale, scratch, and multi-class, totaling 2000 images.	https://ieeexplore.ieee.org/document/10720461 (accessed on 28 August 2026)
TB1500	2025	Steel pipe	This dataset contains six defect types: inclusion, pit, rust, scale, scratch, and white band, totaling 1500 images.	https://ieeexplore.ieee.org/document/11184396 (accessed on 28 August 2026)
CSDD	2025	Casting	This dataset contains 2100 RGB images, each with a resolution of 3648 × 3648. It covers three different types of defects: scratches, spots, and corrosion, and includes 2100 high-resolution images of casting surface defects, totaling 56356 defects.	https://www.ieee-jas.net/en/article/id/cef44e69-fd4f-4da1-baab-0a63f8bcbfa9 (accessed on 28 August 2026)
MSDD	2025	Casting	9332 annotated images covering eight types of defects.	https://www.nature.com/articles/s41597-025-04454-6 (accessed on 28 August 2026)
SDSSP	2025	Steel pipe	Contains 1200 images and four defect categories.	https://link.springer.com/article/10.1007/s11760-025-04805-1 (accessed on 28 August 2026)

Table 2. *Cont.*

Dataset	Year	Object	Description	Link
SCCB-SDD	2025	Steel billet	The dataset comprises 7398 images with resolutions ranging from 64 × 64 to 640 × 640, covering seven representative types of surface defects in steel continuous casting billets, along with background samples. Specifically, the dataset includes background, roll mark, transverse crack, scratches, peeling off, iron oxide scale, slag skin, and longitudinal crack.	https://github.com/Lislttt/STCNet (accessed on 28 August 2026)
Steel Defect X	2026	Mix	A visual-language dataset comprising 7778 images, covering 25 defect categories.	https://arxiv.org/html/2603.21824v1 (accessed on 28 August 2026)
Unpublished	2023	Steel dipper	-	https://ieeexplore.ieee.org/document/10401539 (accessed on 28 August 2026)
Unpublished	2024	Steel pipe	-	https://ieeexplore.ieee.org/document/10702486 (accessed on 28 August 2026)
Unpublished	2024	Steel pipe	-	https://ieeexplore.ieee.org/document/10507010 (accessed on 28 August 2026)

3. Algorithm Classification and Overview

Since 2015, groundbreaking advancements in deep learning have significantly propelled the development of industrial defect recognition technologies. Confronted with the high-throughput, real-time demands of industrial assembly lines—as well as critical challenges such as complex background textures, variable defect scales, and extreme data scarcity—computer vision technology has undergone a profound transformation across two key dimensions: "object detection" and "pixel-level segmentation", evolving from early feature-extraction networks to modern global self-attention architectures.

3.1. Deep Learning-Based Steel Surface Defect Detection

During the early stages of development, the foundational frameworks for detection and segmentation were successively established. In the field of object detection, the R-CNN [22] series pioneered the introduction of convolutional neural networks to this task; by combining region proposal extraction with feature classification, it achieved a significant leap in accuracy. Subsequently, Fast R-CNN [23] and Faster R-CNN [24] introduced region proposal networks, enabling the end-to-end generation of candidate regions. However, their complex multi-stage pipelines struggled to meet the real-time requirements of industrial deployment. Against this backdrop, the YOLO [25] model innovatively reframed the detection task as a global regression problem, achieving single-stage prediction—directly mapping images to bounding boxes—thereby drastically accelerating inference speeds and ushering in a new era of industrial online detection. Concurrently, defect segmentation technology also reached a major milestone. In 2015, the fully convolutional network [26] replaced fully connected layers with convolutional layers, enabling the network to accept inputs of arbitrary size and produce segmentation maps of corresponding dimensions, thereby marking the advent of end-to-end semantic segmentation. U-Net [27], released in the same year, leveraged its symmetrical U-shaped architecture and skip connections to perfectly preserve the edge details of defects; to this day, it remains a core baseline for segmentation tasks.

As research has deepened, the academic community has begun to focus its efforts on addressing the complex problems unique to industrial settings. To tackle the challenges

posed by the vast scale variations of industrial defects—particularly the difficulty in detecting minute flaws—SSD [28] and YOLOv2 [29] pioneered the introduction of multi-scale feature prediction. Subsequently, the development of FPN [30] in 2017 further established the standard paradigm for the fusion of multi-scale defect features. In addressing environments characterized by complex and overlapping defects, Mask R-CNN [31]—proposed in the same year—introduced a mask branch built upon standard detection bounding boxes; this innovation ushered in the era of instance segmentation, successfully resolving the difficult problem of distinguishing between individual instances of the same defect category. Confronted with the extreme and irregular aspect ratios characteristic of industrial defects, the anchor-free paradigm—exemplified by FCOS [32] in 2019—rose rapidly to prominence. By breaking free from the constraints of hyperparameters associated with pre-defined anchor boxes, this approach significantly enhanced the model's generalization capabilities regarding irregularly shaped defects. Furthermore, Focal Loss [33]—introduced by RetinaNet—addressed a critical pain point from the perspective of the loss function: effectively mitigating the severe class imbalance between positive and negative samples frequently encountered in industrial scenarios. At the data level—given the extreme difficulty in collecting real-world samples of specific industrial defects, such as cracks and scratches in steel—generative adversarial networks [34] experienced a surge in application within the field of defect segmentation around 2018. Following DCGAN's [35] resolution of generative stability issues in 2016, GANs have effectively bridged the data gap required for model training by synthetically generating realistic images of industrial defects.

Since 2020, relevant research has shifted significantly toward more challenging real-world scenarios, focusing on issues such as the scarcity of labeled samples and the complexity of background textures. Against this backdrop, the Transformer architecture—originally proposed in the field of natural language processing [36]—began to demonstrate remarkable potential within the computer vision domain. The introduction of DETR [37] broke through the limitations imposed by the local receptive fields of traditional convolutional neural networks, marking the first instance where the Transformer structure was applied to object detection tasks in an end-to-end manner. Leveraging a global self-attention mechanism, this model effectively captures long-range contextual dependencies between defect regions and complex industrial backgrounds while simultaneously completely eliminating cumbersome post-processing steps such as non-maximum suppression. Subsequently, Deformable DETR [38] and DINO [39] addressed the shortcomings of early Transformer models—specifically regarding the detection of small objects and convergence speed—by introducing deformable attention mechanisms and optimizing matching strategies. As these improved models continued to evolve, self-attention mechanisms were widely adopted across a diverse range of specialized tasks. Currently, the Transformer architecture has achieved deep integration between detection and segmentation tasks, emerging as a leading technological direction for achieving high-precision and highly generalizable recognition of industrial defects; its developmental trajectory is illustrated in Figure 3.

3.2. Method for Steel Surface Defect Segmentation

Defect segmentation is primarily categorized into four major types. The most fundamental of these is binary segmentation, which strictly partitions all pixels within an image into two mutually exclusive categories: foreground and background; mathematically, its output typically takes the form of a single mask matrix. Semantic segmentation serves as a natural extension of binary segmentation to multi-class scenarios; its core objective is to assign a predefined class label to every pixel in the image, though it does not distinguish between distinct physical entities belonging to the same category. In contrast, instance segmentation combines the dual characteristics of object detection and pixel-level classi-

fication; it not only requires the identification of an object's class and spatial bounding box, but also necessitates the extraction of the precise contours of specific physical entities, thereby achieving a transition from treating "like objects as an undifferentiated mass" to "distinguishing individual instances". Furthermore, panoptic segmentation aims to unify semantic and instance segmentation by partitioning image content into amorphous, uncountable background "stuff" and discrete, countable "things", thereby assigning both a class label and an instance ID to every pixel.

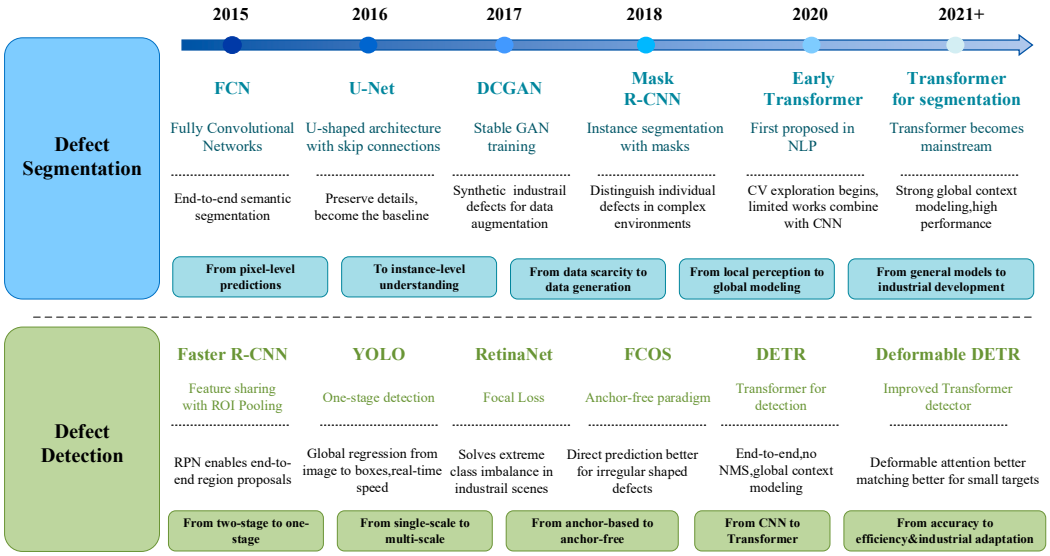


Figure 3. Development chart of steel defect detection technology. FCN [26], U-Net [27], DCGAN [34], Transformer [36], Faster R-CNN [24], YOLO [25], RetinaNet [33], FCOS [32], DETR [37], and Deformable DETR [38].

In the context of steel defect segmentation, binary segmentation stands as the most widely adopted and absolutely dominant paradigm. Although instance segmentation enjoys widespread application in general computer vision scenarios, binary segmentation demonstrates superior adaptability within the specific physical context of industrial metallurgical visual inspection. This advantage stems primarily from the unique morphological characteristics and topological complexities inherent to defects on steel surfaces. Defects such as patches, water marks, embedded mill scale, or diffuse network-like cracks often exhibit extremely high intra-class morphological variability, lacking distinct geometric boundaries or regular topological structures.

In recent years, the field of defect segmentation has exhibited several distinct trends. First, there is a concerted effort to explore lightweight architectures; researchers aim to strike a balance between achieving maximum inference speed and maintaining segmentation accuracy by designing multi-scale perception modules and employing model compression techniques. Second, the field is evolving toward few-shot and weakly supervised learning paradigms; the objective here is to mitigate the high cost associated with acquiring pixel-level annotated data—and to enhance the model's generalization robustness when trained on limited real-world samples—by leveraging mechanisms such as cross-location aggregation and adaptive prototype transformation. Third, by introducing multi-scale contextual modeling and global-local feature fusion mechanisms, researchers are enhancing the model's ability to extract features from—and precisely identify—subtle and complex defects, particularly in environments characterized by high noise levels and low contrast; this approach effectively overcomes issues related to intra-class inconsistency and boundary ambiguity. A detailed classification of these trends is illustrated in Figure 4.

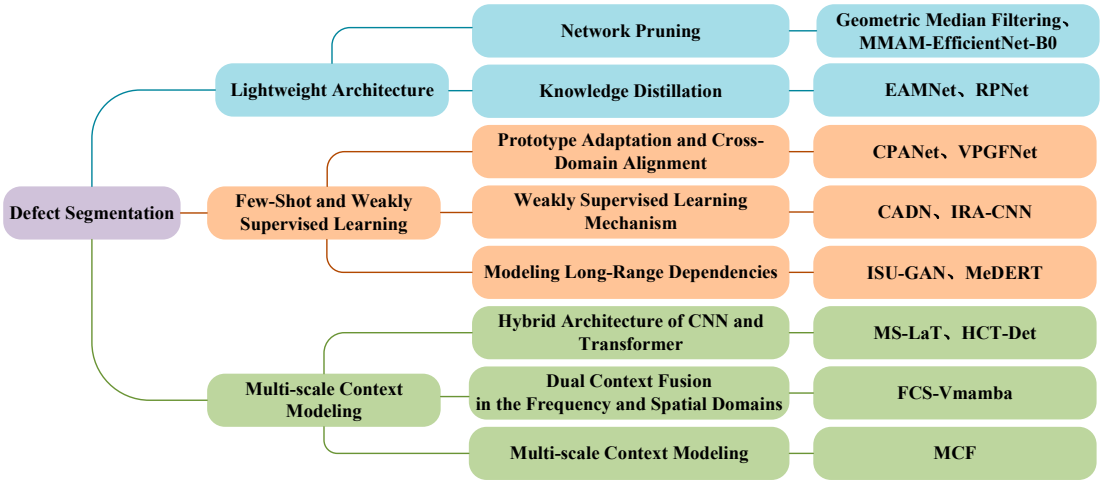


Figure 4. Diagram of a deep learning-based defect segmentation method. Geometric Median Filtering [40], MMAM-EfficientNet-B0 [41], EAMNet [42], RPNet [43], CPANet [44], VPGFNet [45], CADN [46], IRA-CNN [47], ISU-GAN [48], MeDERT [49], MS-LaT [50], HCT-Det [51], FCS-Vmamba [52], MCF [53].

3.2.1. A Lightweight Defect Segmentation Method

In response to the stringent real-time requirements of industrial production lines and the objective reality of limited computational power in edge devices, traditional complex segmentation networks often struggle to meet millisecond-level online detection standards due to their high computational overhead; consequently, researchers have begun to focus their efforts on exploring network lightweighting and inference acceleration techniques. For instance, Chen et al. [54] proposed a local hierarchical information propagation network, which utilizes a dimensional attention propagation module and a vertical semantic propagation module to facilitate local hierarchical information propagation. This network demonstrated strong performance on a custom-built dataset based on a magnetic levitation background. Wang H et al. [55] designed a semi-bilateral skip connection structure and a multi-scale residual context extraction module to compensate for information loss during the encoding process and enhance the network's segmentation capabilities. In their experiments, the model achieved an mIoU of 81.28% and an operating speed of 92.7 FPS on the NEU-Seg steel bar dataset. Zhou et al. [56] introduced a novel lightweight saliency model—the attention-guided dense multi-scale network (ADMNet)—which, by leveraging a multi-scale perception module and a dual attention module, achieved outstanding detection performance with a parameter count of only 0.84 M. Liu et al. [57] proposed a Transformer-based global dual-branch network, termed CGTD-Net; by integrating a multi-channel feature pyramid based on Swin Transformer with an edge detection branch featuring spatial-channel global attention (SCGA), CGTD-Net achieved an mIoU of 75.16% at a high inference speed of 178 FPS, significantly improving the detection performance for tiny and narrow defects. Tan et al. [58] proposed a lightweight detection model that balances both accuracy and real-time performance—the lightweight perception-enhanced network (LPENet). By employing a progressive feature extraction backbone alongside multi-scale context enhancement and boundary enhancement aggregation modules, LPENet achieved an optimal balance between detection accuracy and computational efficiency on the SSP2000 dataset, and was successfully deployed in robots designed for inspecting the internal surfaces of pipes in actual factory environments.

These lightweight segmentation methods are particularly suitable for high-speed production lines, robotic inspection platforms, and edge deployment scenarios where latency, memory consumption, and model size are critical constraints. Their main advantage lies in

improving computational efficiency while maintaining acceptable segmentation accuracy. However, lightweight design may also introduce trade-offs between efficiency and fine-grained representation. Excessive parameter compression or simplified feature extraction may impair the ability to segment tiny, low-contrast, or blurry-boundary defects. Furthermore, inference speed is often hardware-dependent, and practical deployment performance may vary across different embedded platforms. Therefore, while lightweight segmentation models are valuable for industrial deployments, their robustness needs further evaluation under different imaging conditions and defect morphologies.

3.2.2. Few-Shot Defect Segmentation Method

Furthermore, the high performance of modern deep networks relies heavily on large-scale, high-quality pixel-level annotated data; however, in real-world industrial settings, samples exhibiting severe defects are often extremely scarce and follow a long-tail distribution, resulting in prohibitively high costs for data acquisition and manual annotation. To overcome this "data-hungry" dilemma and enhance model generalization robustness when encountering unknown scenarios or rare defects, few-shot learning and limited-sample strategies have gradually emerged as another major research hotspot in the field of defect segmentation. For instance, Feng et al. [44] achieved higher segmentation accuracy under limited-sample conditions by integrating cross-location aggregation with few-shot learning. Huang et al. [59] introduced an adaptive prototype Transformer within the context of few-shot semantic segmentation to better localize defects and generate more refined feature representations; by employing hierarchical feature fusion to capture more detailed and contextually relevant information, they enhanced the model's discriminative capability regarding defects across multiple feature scales. Guo et al. [60] proposed an efficient model for real-time, pixel-level detection of metal surface defects—dubbed SPEED (semantic prior and extremely efficient dilated convolution network). By leveraging a semantic prior branch to capture low-level details and an extremely efficient dilated branch to acquire high-level contextual information—combined with an aggregation module for feature fusion—this model effectively addressed the challenges of high intra-class variability among defects and high inter-class similarity. Song et al. [61] proposed a cross-granularity few-shot segmentation model for internal defects in seamless steel pipes—the multi-feature aggregation network (MFANet). By introducing a cross-granularity segmentation paradigm that generalizes feature learning from coarse-grained datasets to fine-grained samples, the model effectively overcame the difficulties associated with sparse defect samples and time-consuming fine-grained annotation, demonstrating state-of-the-art segmentation performance on their self-constructed CGFSDS-9 benchmark. Finally, Ma et al. [62] proposed a multi-context aggregation network (MCA-Net) for few-shot surface defect segmentation; this model incorporates a cross-layer multi-level feature aggregation module to mitigate incomplete segmentation caused by weak feature signals, and features a foreground correction module—comprising background prediction and foreground refinement components—to effectively suppress severe background interference. This method not only achieved state-of-the-art few-shot segmentation performance on the FSSD-12 and FSS-AEB industrial datasets, but was also successfully integrated into a real-world automated defect detection hardware platform, significantly enhancing the model's generalization capabilities and practical applicability regarding unseen defect categories.

Few-shot and limited-sample segmentation methods are highly relevant to industrial inspection because collecting exhaustive pixel-level annotations for all defect categories is often impractical. Their key strength is the ability to improve model adaptability when annotated samples are limited or when new defect categories appear. These methods can reduce annotation costs and enhance the feasibility of deploying segmentation systems in

small-batch or rapidly changing production scenarios. Related studies in industrial fault diagnosis have also shown that interpretable few-shot frameworks can improve diagnostic robustness under limited samples and noisy labels, indicating the broader applicability of few-shot learning to industrial monitoring tasks [63]. Nevertheless, their performance is usually sensitive to support-sample quality, label reliability, category similarity, and domain shifts between training and testing data. In particular, when unseen defects differ significantly from the support samples in texture, scale, or morphology, segmentation stability may decline. Therefore, few-shot methods are promising for rare-defect and new-defect scenarios, but their reliability still depends on representative support samples, reliable annotations, and robust cross-domain feature alignment.

3.2.3. Segmentation Method Based on Multi-Attention and Feature Collaboration

Under typical industrial conditions—characterized by fixed illumination and limited, restricted-view imaging—defects such as minute scratches and shallow pits often exhibit extremely low contrast, strong morphological anisotropy, and severe intra-class inconsistency, making them highly susceptible to being obscured by complex background noise. Given this highly challenging problem of recognizing faint features, a single, localized convolutional operation is no longer sufficient. To address the difficulties associated with feature extraction and the loss of contextual information in industrial surface defect segmentation, Chen et al. [64] proposed a segmentation network incorporating a triple-mechanism architecture: it employs a collaborative perception enhancement module to enrich the feature space and strengthen local perception; introduces a hierarchical attention integration module—utilizing an adaptive weighting mechanism—to achieve the efficient fusion of cross-scale features from both shallow and deep layers; and finally, utilizes a bilateral self-fusion module to thoroughly mine intra-layer features, achieving promising results on the SSP2000 dataset. Pan Y et al. [65] proposed a dual-attention deep learning network; its dual-parallel attention modules facilitate the explicit modeling of rich contextual dependencies within local feature representations across both spatial and channel dimensions, thereby effectively capturing multi-scale contextual relationships. The proposed hybrid model demonstrates enhanced accuracy in detecting defects on steel surfaces. To overcome the issues of intra-class inconsistency and the indistinguishability of inter-class segmentation results inherent in existing defect semantic segmentation methods, Z. Yin et al. [66] introduced a global and local feature fusion module—based on similarity metrics—into their segmentation network to guide the model in further refining and highlighting the fine-grained details of defects. Song et al. [67] designed DSRNet, a model that integrates Transformers with convolutional neural networks (CNNs) and employs a dual-branch encoder to ensure that the extracted features encompass not only rich global semantic information, but also fine-grained local details. Wang et al. [68] proposed an enhanced semantic segmentation framework based on SegFormer; by integrating spatial pyramid pooling with an adaptive feature fusion module—and by incorporating branches for deformable convolution and edge-focusing modules—this framework effectively enhances the model's perception of multi-scale context and the boundaries of irregular defects, achieving an excellent balance between high segmentation accuracy and computational efficiency on the NEU-Seg dataset.

Attention-based, Transformer-based, and multi-scale feature fusion methods are effective for complex segmentation scenarios involving low contrast, irregular shapes, blurred boundaries, and strong background interference. Their advantage lies in enhancing contextual modeling, strengthening boundary perception, and improving the segmentation of defects with large scale variation. However, these methods often introduce additional computational complexity and may require more training data to fully exploit global semantic

modeling. Transformer-based structures, in particular, may be less efficient than compact CNN-based models in strict real-time settings. Moreover, complex attention and fusion modules can improve accuracy on specific datasets but may not always generalize well to unseen production environments.

A summary of the defect segmentation methods used for identifying steel surface defects is presented in Table 3. It should be noted that the experimental results listed in the table are not based on a unified benchmark test and therefore should not be considered as a strictly direct cross-sectional comparison. Due to differences in experimental environment, evaluation metrics, input size, and hardware conditions among different studies, the performance results should be interpreted in conjunction with their respective experimental settings and application scenarios.

Table 3. A Survey of deep learning-based defect segmentation methods.

Dataset	Year	Method	Ref.	Performance
NEU-Seg	2023	SPEED	[60]	mIoU = 92.34%
	2024	DDSNet	[66]	mIoU = 85.12%
	2025	SBSANet	[55]	mIoU = 81.28%; PS = 92.7 Frames/s
	2025	RSM-YOLOv11	[69]	mIoU = 88.6%
	2025	AF-Fuse	[68]	-
	2026	LRTISS	[70]	FPS = 155.17 Frames/s; GFLOPs = 0.92; mIoU = 76.61%
FSSD-12	2022	CPANet	[44]	FB-IoU = 76.3; Mean = 62.6
	2025	MAPTNet	[59]	-
	2025	VPGFNet	[45]	-
	2025	MCANet	[62]	mIoU = 65.6%
	2026	EAMNet	[42]	Mean = 62.52; FBIoU = 75.8
	2026	LRTISS	[70]	mIoU = 76.34%
Severstal	2022	DAN-DeepLabv3+	[65]	mIoU = 89.95%; fwIoU = 97.34%
	2023	SPEED	[60]	mIoU = 92.34%
Self-built dataset	2023	SqueezeNAS	[71]	mIoU = 98.17%
	2024	MFANet	[61]	-
	2024	FASNet	[72]	mIoU = 98.6%
SSP2000	2024	SRPCNet	[64]	MAE = 0.0021; S_α = 0.8441; F_β^w = 0.7698
	2025	LPENet	[58]	-
NEU-DET	2024	CGTD-Net	[57]	mIoU = 75.16%; FPS = 178 frames/s
DUTS-TE	2024	DSRNet	[67]	MAE = 0.029; S_α = 0.903; F_β = 0.895
TB1500	2025	L2IPNet	[54]	MAE = 0.026; S_α = 0.822; F_β^w = 0.8022
Other datasets	2025	MCANet	[62]	mIoU = 57.8%
	2026	ID-MSNet	[73]	mIoU = 86.19%; FPS = 155.17 frames/s
	2023	MLESPNetV2	[74]	-
	2025	MAPTNet	[59]	-
	2024	ADMNet	[56]	-
	2023	GRA-Net	[75]	-
	2025	HMNet	[76]	-

3.3. Methods for Detecting Surface Defects in Steel

Defect detection and defect segmentation are distinct yet complementary tasks. Defect detection focuses primarily on object-level recognition and localization, whereas defect segmentation emphasizes pixel-level delineation and the precise characterization of defect morphology. The developmental trends in defect detection are characterized by three key directions. First, a commitment to balancing end-to-end lightweight design with engineering feasibility—researchers strive to achieve a deep integration of extremely low computational costs and high throughput by designing efficient inference architectures, thereby meeting the stringent real-time requirements of industrial settings. Second, an evolution toward generalized perception and robust recognition under complex and variable operating conditions—this aims to enhance the model's capability to capture features amidst low contrast, high noise, and camouflaged backgrounds, thereby mitigating the adverse effects of environmental fluctuations in industrial field settings on detection performance. Third, an enhancement of the model's localization accuracy for defects exhibiting diverse geometric morphologies, extreme size variations, and irregular shapes—this is achieved by reinforcing global semantic associations and multi-scale adaptive modeling capabilities, thereby overcoming the inherent limitations of traditional detection paradigms when confronted with cross-scale defects and complex spatial distributions. A detailed classification of these trends is illustrated in Figure 5.

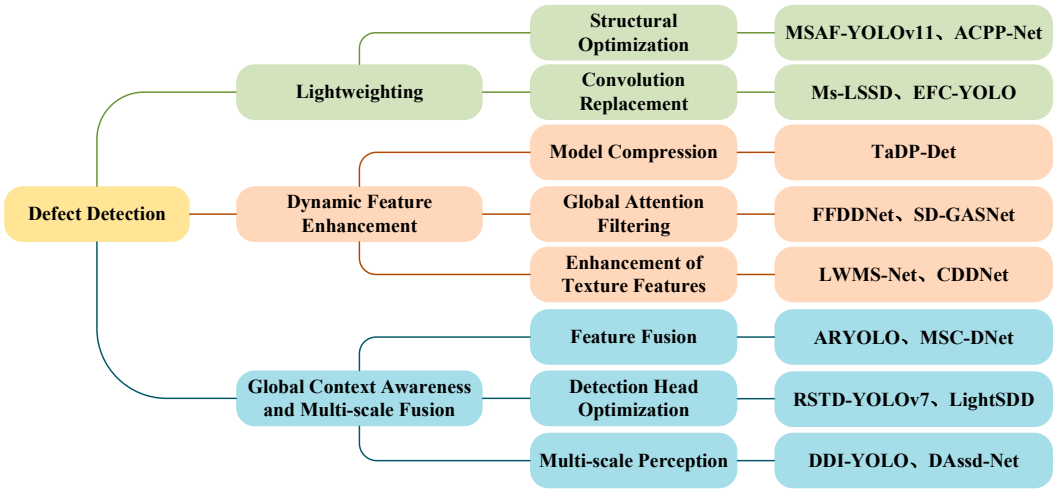


Figure 5. Diagram of a deep learning-based defect detection method. MSAF-YOLOv11 [77], ACPP-Net [78], Ms-LSSD [79], EFC-YOLO [80], TaDP-Det [81], FFDDNet [82], SD-GASNet [83], LWMS-Net [84], CDDNet [85], ARYOLO [86], MSC-DNet [87], RSTD-YOLOv7 [88], LightSDD [89], DDI-YOLO [90], and DAssd-Net [91].

3.3.1. A Lightweight Defect Detection Network

In light of the stringent real-time requirements of industrial production lines and the objective reality of limited computational power in edge devices, traditional complex detection networks often struggle to meet millisecond-level online detection standards due to their high computational overhead; consequently, researchers have begun to focus their efforts on exploring network lightweighting and inference acceleration techniques. By incorporating strategies such as adaptive ghost convolution, SPD convolution, and operator-level pruning, the goal is to drastically compress the number of model parameters while maintaining high-throughput inference performance, thereby enabling the lossless migration and efficient deployment of detection algorithms on low-power embedded terminals. For instance, Wang Z et al. [77] proposed MSAF-YOLO, which achieves multi-scale feature enhancement and adaptive receptive field modeling through the synergistic design

of SPD convolution, triplet attention, EMSC, and CAFM. This approach simultaneously improves the detection accuracy and significantly reduces the number of model parameters; compared to existing benchmark models, MSAF-YOLO utilizes only 1.9 million parameters, making it highly suitable for industrial deployment—particularly on low-power embedded and mobile devices. Duan B et al. [79] designed a lightweight detection framework based on RT-DETR, introducing adaptive partial convolution and an efficient adaptive residual network to enhance feature representation and reduce model complexity. Li et al. [78] addressed the challenges posed by steel surface defects—specifically their small size, diverse morphologies, and uneven distribution—by proposing ACPP-Net. This network integrates adaptive ghost convolution to reduce redundant computations and optimize feature weights, while utilizing a feature enhancement module to bolster its discriminative capabilities regarding complex textures and subtle defects. Li et al. [80] proposed an efficient fusion coordination network, EFC-YOLO. This method incorporates a partial convolution-based fusion-faster module and a lightweight coordinate attention mechanism into the YOLOv7 backbone, and employs a de-weighted BiFPN structure within the network neck to optimize path aggregation, thereby resolving the inherent trade-off between enhancing shallow-layer features and maintaining rapid inference speeds.

Overall, lightweight detection methods are particularly advantageous for online inspection, embedded deployment, and scenarios where inference speed, model size, and hardware cost are critical constraints. Their main strength lies in reducing computational burden while retaining acceptable detection accuracy. However, these methods also have limitations. Excessive compression may weaken the model's ability to capture small, low-contrast, or irregular defects, especially when defect features are already weak. In addition, the actual acceleration effect may depend heavily on the target hardware platform, operator optimization, and input resolution. Therefore, lightweight models are more suitable for real-time industrial deployment, but their accuracy and robustness should be carefully validated under specific production conditions.

3.3.2. A Defect Detection Method Based on Dynamic Feature Enhancement and Interference Suppression

Faced with the rigorous industrial conditions characterized by drastic lighting fluctuations, extremely low defect contrast, and highly cluttered background textures, traditional discriminative models often encounter bottlenecks involving false positives and missed detections due to insufficient feature representation capabilities. Consequently, the research community has gradually shifted its focus toward the enhancement of salient features and the development of mechanisms to mitigate environmental interference. By integrating techniques such as multi-wavelet domain modeling, dynamic texture enhancement, and constrained deformable convolution, researchers aim to precisely isolate defect signals amidst intense noise interference, thereby significantly boosting the models' feature discriminability and detection robustness regarding camouflaged defects or those with faint boundaries. For instance, Zheng et al. [84] constructed a feature difference enhancement module to accentuate defect edges and detailed information, thereby improving the feature representation capabilities while suppressing redundant interference. Simultaneously, they designed a novel backbone network that effectively models multi-scale contextual and fine-grained features across multiple wavelet domains, achieving superior feature representation and reduced network complexity, and striking a favorable balance between detection accuracy and inference efficiency. Lin et al. [82] proposed a constrained deformable convolutional network designed to efficiently extract defect features; this approach ensures that the receptive field aligns more closely with the actual pixel area of the target object, while a specifically designed constraint function prevents the excessive displacement of sampling points. Furthermore, they introduced global vertical

and horizontal attention mechanisms to effectively minimize interference from irrelevant background information, thereby enabling accurate and efficient detection. Luo et al. [85] proposed CDDNet for the detection of low-contrast and camouflaged surface defects. This network incorporates an initial dynamic texture enhancement module to reinforce weak boundaries and complex texture features; it also employs a lightweight, recurrent, and decoupled fully connected attention mechanism to model long-range spatial dependencies. By integrating an adaptive scale-balancing pyramid convolution module to facilitate cross-scale feature fusion, the network effectively enhances the accuracy and robustness of defect detection within complex background environments. Fu et al. [83] proposed SD-GASNet, a detection network based on a self-distillation model compression strategy. At the feature extraction stage, this method constructs a feature pyramid network that integrates, aligns, enhances, and synchronizes features. By introducing a frequency-domain information collection and distribution mechanism, the network extracts both high- and low-frequency features to effectively counter background interference; additionally, a channel synchronization module is utilized to compensate for information loss during feature fusion. Finally, the adoption of a self-distillation strategy—driven by an enhanced KL divergence loss function—significantly reduces the computational overhead. Luo et al. [81] proposed TaDP-Det, a semi-supervised object detector featuring dual enhancements in feature representation and label filtering. By introducing a texture enhancement module into the shallow layers of the backbone network, the method effectively strengthens the model's feature extraction capabilities in ambiguous regions by amplifying highly discriminative low-level textural cues. Simultaneously, a class-specific dynamic pseudo-label filtering scheme was designed to adaptively determine class-specific thresholds—utilizing a one-dimensional Gaussian mixture model—thereby preserving challenging defects while suppressing erroneous predictions induced by background noise. TaDP-Det effectively mitigates interference arising from complex, low-contrast backgrounds, achieving robust semi-supervised defect detection at an extremely low annotation cost.

These feature enhancement and interference suppression methods are beneficial for complex industrial environments where defects are small, blurred, low-contrast, or easily confused with background textures. Their advantage is that they can strengthen discriminative defect cues and reduce missed detections under challenging imaging conditions. Nevertheless, the introduction of multiple enhancement modules may increase architectural complexity and training difficulty. Some methods also require careful parameter tuning, and their robustness may decline when background texture patterns differ significantly from the training data. For semi-supervised methods such as TaDP-Det, the reduced annotation cost is a clear advantage, but the reliability of pseudo-labels remains an important factor affecting final performance. Therefore, such methods are especially useful when image quality is unstable or defect boundaries are ambiguous, but their generalization ability should be evaluated across different production lines and imaging settings.

3.3.3. A Defect Detection Method Based on Global Context Awareness and Multi-Scale Fusion

Given the highly stochastic spatial distribution of industrial defects and the drastic fluctuations in their size ranges, convolutional operations relying solely on local receptive fields struggle to capture the long-range logical relationships between defects and the topological structures of workpieces; consequently, research is evolving toward a focus on global semantic correlation and multi-scale adaptive feature fusion. By leveraging the global modeling capabilities of Vision Transformers, along with self-supervised contrastive learning and meta-learning paradigms, researchers aim to construct detection architectures endowed with contextual reasoning abilities. These architectures are designed to effectively address the challenges of precise localization and generalized recognition for multi-modal,

cross-scale targets—particularly in scenarios characterized by data scarcity. For instance, Huang J [86] introduced an aggregation-redistribution network to achieve the effective fusion and reconstruction of cross-layer features; by combining global information with attention mechanisms to enhance feature representation capabilities, this approach achieved excellent performance in multi-scale defect detection tasks. Addressing the difficulties associated with detecting surface defects on steel bars—specifically in multi-scale and complex background environments—Liu R et al. [87] proposed the multi-scale contextual detection network (MSC-DNet). This network employs parallel dilated convolutions to capture multi-scale contextual information, incorporates feature enhancement and selection modules to boost feature representation while mitigating information interference, and integrates auxiliary image-level supervision to accelerate model convergence and enhance discriminative power. Song et al. [88] integrated RFBVGG and SimAM attention mechanisms into the YOLOv7 backbone to expand the receptive field and enhance feature representation capabilities; furthermore, they incorporated a Swin Transformer-based STRVGG module into the neck network to improve deep-level semantic modeling, and refined the detection head to boost localization accuracy and convergence speed—thereby achieving superior detection accuracy and efficiency across scenarios involving multi-scale objects, small targets, and complex backgrounds. Finally, addressing the limitations inherent in detecting irregular surface defects on industrial products—particularly in scenarios characterized by sample scarcity and high defect diversity—Huang et al. [92] proposed a novel adaptive cross-Transformer network. By integrating self-supervised contrastive learning, this network effectively mitigates the challenges associated with few-shot learning. Zhang et al. [89] proposed LightSDD, which combines multi-scale dynamic convolution and lightweight cross-level fusion to improve small-object detection while reducing redundant parameters.

Compared with purely convolutional models, Transformer-based and multi-scale fusion methods are better suited for defects with large scale variation, irregular shapes, and complex spatial distributions. Their strength lies in capturing global context, modeling long-range dependencies, and improving the recognition of small or scattered defects. However, these methods often require more training data, higher computational resources, and more careful optimization. Transformer-based structures may also be less efficient than lightweight CNN-based detectors in strict real-time deployment scenarios. In addition, although self-supervised and few-shot learning strategies can reduce dependence on large-scale annotations, their performance may still be sensitive to pretraining data, task design, and domain differences. Thus, these methods are promising for high-accuracy detection and complex defect analysis, but their industrial deployment requires balancing accuracy, speed, memory consumption, and data availability.

Table 4 presents a summary of defect recognition and detection methods utilized for identifying surface defects in steel materials. It should be noted that the experimental results listed in the table are not based on a unified benchmark test and therefore should not be considered as a strictly direct cross-sectional comparison. Due to differences in experimental environment, evaluation metrics, input size, and hardware conditions among different studies, the performance results should be interpreted in conjunction with their respective experimental settings and application scenarios.

Table 4. A Summary of deep learning-based defect detection methods.

Dataset	Year	Method	Ref.	Performance
NEU-DET	2023	DCAM-Net	[90]	mAP = 82.6%; FPS = 100.2 Frames/s
	2023	EFC-YOLO	[80]	mAP = 85.86%; GFLOPs = 39.2; FPS = 73.4 Frames/s
	2023	SCRL-EMD	[91]	-
	2023	MSC-DNet	[87]	mAP = 79.4%
	2023	DAssd-Net	[93]	mAP = 81.97%
	2024	ACViT-SCL	[92]	-
	2024	ACPP-Net	[78]	mAP = 82.1%
	2024	CDDNet	[85]	mAP = 84.2%
	2024	DDI-YOLO	[94]	mAP = 78.3%; Precision = 72.5%; Recall = 71.4%; Parameters = 2.66 M; GFLOPs = 7.5 G; FPS = 158 Frames/s
	2025	FFDDNet	[82]	-
	2025	LWMS-Net	[84]	mAP = 81.6%
	2025	Ms-LSSD	[78]	mAP = 78.7%; FPS = 52.2 Frames/s
	2025	ARYOLO	[86]	mAP = 80.7%
	2025	RSTD-YOLOv7	[87]	-
	2025	MSPro-Net	[95]	mAP = 80.3%
	2025	CE-DETR	[96]	mAP = 78.6%
	2025	GLF-NET	[97]	mAP = 79.2%; FPS = 95 Frames/s
	2025	MC-DDet	[98]	mAP = 80%
	2025	YOLOv8-BSPB	[20]	mAP = 86.9%; FPS = 103.1 Frames/s
	2026	TaDP-Det	[81]	mAP = 71.8%; FPS = 34.3 Frames/s; Parameters = 45.94 M
	2026	LightSDD	[89]	mAP = 76.81%; Precision = 70.18%; Recall = 72.7%; Parameters = 2.42 M
	2026	DGS-YOLO	[99]	mAP = 77.4%
	2026	DEENet	[100]	mAP = 81.4%
	2026	SD-GASNet	[51]	mAP = 81.8%; FPS = 180 Frames/s
	2026	MSAF-YOLOv11	[77]	mAP = 82.5%
GC10-DET	2023	MSC-DNet	[87]	mAP = 71.6%
	2023	SCRL-EMD	[91]	-
	2024	ACPP-Net	[78]	mAP = 71.1%
	2025	LWMS-Net	[84]	mAP = 73.7%
	2025	ARYOLO	[86]	mAP = 71.2%
	2025	MESC-DETR	[101]	mAP = 71.9%; *F*1 = 75.1%; FPS = 33.8 Frames/s
	2025	CE-DETR	[96]	mAP = 71.3%
	2025	STCNet	[102]	-
	2025	MC-DDet	[98]	mAP = 71.8%
	2025	RSTD-YOLOv7	[87]	-
	2026	TaDP-Det	[81]	mAP = 58.2%
	2026	LightSDD	[89]	mAP = 64.92%; Precision = 66.89%; Recall = 63.58%; Parameters = 2.42 M
	2026	MSAF-YOLOv11	[77]	mAP = 70.6%
Other datasets	2023	LSwin Transformer	[103]	mAP = 81.2%
	2024	LMViT	[104]	-
	2024	SSA-YOLO	[105]	-

4. Experimental Evaluation Indicators

In industrial defect recognition, evaluation metrics should be selected according to the output representation and practical requirements of each task. Classification determines whether a sample is defective and identifies its category; it is commonly evaluated using accuracy, precision, recall, F1-score, MCC, and ROC-AUC. For highly imbalanced datasets, recall, F1-score, MCC, and the precision–recall curve are generally more informative than accuracy alone. Object detection evaluates both category prediction and bounding-box localization and is therefore commonly assessed using IoU, AP, and mAP. Segmentation emphasizes pixel-level delineation of defect regions and boundaries and is typically evaluated using pixel accuracy, mIoU, and Dice. For saliency-map-based defect segmentation, MAE, S-measure, E-measure, and F-measure may also be employed. Computational efficiency and industrial deployability are commonly assessed using parameter count, model size, FLOPs or MACs, inference latency, FPS, peak memory consumption, and energy consumption. Although some metrics are applicable to multiple tasks, their calculation granularity differs. Therefore, it is necessary to select the most appropriate evaluation index based on the specific experimental scenario and conditions in order to more comprehensively evaluate accuracy, real-time performance, and industrial feasibility.

4.1. The Precision Class Metrics

Defect classification represents the initial phase of industrial visual inspection, aiming to accurately identify both the presence and type of defects on product surfaces. In practical production settings, a significant imbalance frequently exists between positive and negative samples, necessitating evaluation metrics that demonstrate strong robustness. The foundational classification evaluation metric is derived from a confusion matrix, which encompasses four essential dimensions: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Accuracy serves as the most intuitive metric for assessing the overall proportion of correct predictions made by the algorithm, defined mathematically as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

While accuracy performs well on balanced datasets, it often fails in industrial scenarios due to the "accuracy trap". For example, if the defect rate on a production line is 1%, even if the model predicts all products as acceptable, its accuracy will still be as high as 99%, but it will completely lose its ability to identify defects. Therefore, in academia, it is usually used as an auxiliary indicator rather than a core criterion. To analyze the model's performance on specific categories in more detail, precision and recall are widely used. Precision represents the proportion of samples predicted as defects that are actually defects, while recall represents the proportion of actual defective samples that are successfully detected. In industry, recall is often given higher priority because missed defects (FN) can lead to serious quality accidents or safety risks, while false positives (FP) usually only incur the cost of secondary manual review. The formula is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Based on the trade-off between precision and recall, the F1 score (F1-score), as the harmonic average of the two, is considered one of the authoritative metrics for evaluating the performance of imbalanced datasets. The formula is as follows:

$$F1_{Score} = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

When dealing with extremely imbalanced data, the Matthews correlation coefficient (MCC) is widely recognized as a more statistically significant indicator than the F1 score. The MCC considers all four values in the confusion matrix, and its calculation reflects the correlation between the predicted and actual classifications, as shown in the following formula:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

In practical deployments of industrial algorithms, decision thresholds often need to be adjusted based on process requirements. To evaluate the overall potential of an algorithm at different thresholds, the receiver operating characteristic (ROC) curve and its area under the curve (AUC) provide authoritative metrics. The ROC curve plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis, as shown in the following formula:

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$AUC = \int_0^1 TPR(t) dFPR(t) \tag{8}$$

The AP and mAP metrics are derived from the precision–recall relationship and are widely used to evaluate object detection performance. Average precision summarizes the precision–recall curve for a specific defect category. By varying the confidence threshold, a series of precision–recall pairs can be obtained, with recall plotted on the x-axis and precision on the y-axis. AP is then calculated as the area under the precision–recall curve. Different evaluation protocols may use different interpolation strategies. The 11-point interpolation method is generally used, as shown in the following formula:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots\ldots, 1.0\}} \rho_{interp}(r) \tag{9}$$

Mean average precision (mAP), rather than mean accuracy, is calculated as the arithmetic mean of the AP values across all defect categories:

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{K} \tag{10}$$

where k represents the total number of categories in the object detection task, and i represents the index number of the category.

Intersection over union (IoU), also known as the Jaccard index, is a core metric for evaluating the degree of overlap between the predicted and ground truth regions. Its mathematical expression is:

$$IoU = \frac{A \cap B}{A \cup B} = \frac{TP}{FP + TP + FN} \tag{11}$$

For pixel-level segmentation tasks, the Dice coefficient is another authoritative metric equivalent to the F1 score. It quantifies the similarity between the predicted mask and the ground truth mask, and its formula is defined as follows:

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{12}$$

where $|A|$ is the total number of target pixels predicted by the model, and $|B|$ is the total number of target pixels in the true labels.

MAE measures the average absolute difference between the predicted defect probability map and the true label across all pixels, and its formula is as follows:

$$MAE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} |P(i,j) - G(i,j)| \tag{13}$$

where H represents the height of the feature map, W represents the width of the feature map, and i,j represent the spatial coordinates of the image matrix, with i representing the row coordinate and j representing the column coordinate. $P(i,j)$ is the predicted value output by the model at the coordinates, and $G(i,j)$ is the true value of the model at the coordinates.

S-measure, E-measure, and F-measure are the core standards for evaluating the quality of saliency prediction maps, particularly suitable for industrial defects with low contrast and blurred boundaries. S-measure is used to evaluate the spatial structural similarity between the predicted map and the ground truth label, and the formula is as follows:

$$S_\alpha = \alpha \cdot S_0 + (1 - \alpha) \cdot S_r \tag{14}$$

where α is a hyperparameter ranging from 0 to 1, S_0 denotes the object-aware structural similarity, and S_r denotes the region-aware structural similarity between the predicted map and the ground-truth mask.

The E-measure is an alignment-based evaluation metric that combines image-level statistical information and local pixel-level matching information. The formula is as follows:

$$E = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \varnothing(i,j) \tag{15}$$

where H represents the height of the feature map, W represents the width of the feature map, and i and j represent the spatial coordinates of the image matrix, i being the row coordinate and j being the column coordinate. $\varnothing(i,j)$ is the value of the augmented alignment matrix at the coordinates. First, the global mean of the predicted image and the true label are calculated separately. After centering the image, the similarity between local pixels is calculated, which reflects the alignment accuracy of the pixel after incorporating global cognition.

F-measure is a harmonic mean of precision and recall, used to comprehensively evaluate the model's classification performance at the pixel level. The formula is as follows:

$$F_\beta = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{16}$$

In the fields of saliency detection and fine segmentation, β^2 is usually set to 0.3 to give precision a higher weight, because in practical applications, incorrectly marking the background as a defect is often more unacceptable than missing some edges.

4.2. The Efficiency Class Metrics

Efficiency metrics are primarily used to measure the performance of defect detection algorithms. Their core purpose is to evaluate whether the algorithm possesses real-time detection capabilities in actual industrial production environments, while also indirectly reflecting the complexity of the algorithm's structure. Common efficiency evaluation metrics in practical applications include training duration, testing duration, parameters, inference time, FPS, and FLOPs.

Among these metrics, training duration and testing duration are the most intuitive and readily available, directly reflecting the time cost of an algorithm during the experimental phase. However, these two metrics are highly dependent on the specific hardware environment, such as GPU model and memory size. Even with the same algorithm, results may vary significantly on different devices, making fair comparisons difficult if relying solely on time metrics.

In contrast, parameter provides a more general metric, representing the total number of parameters that a model needs to learn, typically in the millions. This metric reflects not only the model's storage overhead but also, to some extent, its complexity. A larger parameter count means that the model consumes more disk space and static memory. The size of the parameter count determines whether the model can be deployed on resource-constrained edge devices, as shown in the formula below:

$$Params = (k_h \times k_w \times C_{in} \times C_{out}) + C_{out} \tag{17}$$

where k_h and k_w are the kernel sizes, and C_{in} and C_{out} are the input and output channels, respectively.

The related inference time and FPS are used to evaluate the model's operational efficiency in the actual deployment phase: inference time refers to the time required to process a single image, while FPS represents the number of images the model can process per second, and is an important indicator of real-time performance. The FPS formula is as follows:

$$FPS = \frac{Total\ Frames}{Total\ Time} \tag{18}$$

Additionally, FLOPs represents the total number of floating-point multiply-add operations required for a single forward pass of a model, which is utilized to characterize its computational complexity and is typically measured in billions or higher orders of magnitude. This metric reflects the model's demand for computational resources. Theoretically, a lower FLOPs count indicates a more lightweight model execution. However, it should be noted that a low FLOPs value does not necessarily guarantee a strictly faster actual inference speed, as memory access speeds and the level of operator optimization also significantly impact the final latency. By reflecting the model's hardware computational requirements, this metric serves as a crucial basis for evaluating the algorithm's lightweight nature and its deployment feasibility. The FLOPs formula is as follows:

$$FlOPs = 2 \times (k_h \times k_w \times C_{in} \times C_{out} \times H_{out} \times W_{out}) \tag{19}$$

where H_{out} is the height of the output feature map and W_{out} is the width of the output feature map.

GPU load refers to the percentage of GPU computing cores or memory used during model execution, usually expressed as a percentage. This metric is primarily used to evaluate the model's efficiency in utilizing hardware resources. GPU time, on the other hand, focuses on measuring the actual time cost of the model performing computational tasks on the GPU, only counting the time consumed by the neural network's forward

propagation and related operator operations. This metric excludes additional overhead such as data interaction between the CPU and GPU, thus more accurately reflecting the computational efficiency of the model architecture itself.

It is important to emphasize that hardware configuration has a significant impact on model inference latency and frame rate. When the same model is evaluated, differences in GPU or CPU architecture, computational throughput, memory capacity and bandwidth, and hardware acceleration support may lead to substantially different runtime performance. In addition, input resolution, batch size, numerical precision, software framework, compiler, and computational library versions can influence execution efficiency. Therefore, the FPS and inference latency values reported in different studies should not be considered directly comparable when the hardware platforms or evaluation protocols differ. When these conditions cannot be standardized, parameter count and FLOPs can provide supplementary measures of model complexity, but they cannot replace performance measurements conducted on the actual target device.

5. Current Challenges and Future Research Directions

5.1. Current Challenges

Despite substantial progress on public benchmarks, the performance of steel surface defect detection models does not always transfer reliably to production environments. Variations in illumination, metallic reflection, mechanical vibration, surface contamination, production speed, and camera configuration can degrade image quality and produce distribution shifts between training and deployment data. These effects are particularly severe for small, low-contrast, and irregular defects on steel strips, pipe inner surfaces, and castings. Image enhancement, multi-scale feature fusion, deformable convolution, and attention mechanisms can improve defect representation; however, their robustness across production lines and changing operating conditions has not yet been consistently demonstrated. A practical solution therefore requires the combination of acquisition-side control, including stable illumination, optical filtering, synchronized imaging, and periodic calibration, with data augmentation, domain adaptation, and multimodal sensing when RGB imaging alone is insufficient.

Data limitations remain another major obstacle. Industrial defect datasets are generally small, long-tailed, and dominated by normal samples, while annotations may contain inaccurate categories or boundaries. Recent studies have investigated self-supervised, semi-supervised, and few-shot learning to exploit unlabeled data and reduce annotation dependence. Nevertheless, pseudo-label errors, limited defect diversity, and distribution differences between the pretraining and target data may still restrict performance. Future work should combine self-supervised pretraining, uncertainty-filtered pseudo-labeling, active learning, and physically realistic defect synthesis. Synthetic samples should be evaluated not only by visual appearance, but also by whether they reproduce plausible defect morphology and improve performance on independently collected production data.

Model architecture introduces a further trade-off between detection accuracy and deployment efficiency. Multi-scale networks and recent Transformer-, CNN–Transformer-, and Mamba-based models strengthen contextual and long-range feature representation but often increase memory consumption and inference latency. Conversely, pruning, quantization, knowledge distillation, and lightweight detection heads reduce computational cost but may remove features required to identify subtle defects. Future studies should therefore adopt hardware–software co-design and report end-to-end latency, memory consumption, throughput, and energy usage on the actual target device, rather than relying solely on parameter counts or floating-point operations. Uncertainty estimation, confidence calibration, out-of-distribution detection, and human review mechanisms should also be

incorporated into safety-critical inspection systems. In addition, 3D inspection, multimodal sensor fusion, and digital twin-based validation represent promising directions for geometrically complex components, although their industrial effectiveness still requires evaluation using standardized, cross-line datasets; the specific details are shown in Table 5.

Table 5. Current challenges and potential solutions.

Challenge	Industrial Manifestation	Feasible Solutions
Unstable imaging conditions	Illumination variation, specular reflection, vibration, contamination, and motion blur	Controlled illumination, optical filtering, synchronized acquisition, image normalization, and targeted augmentation
Scarce and imbalanced data	Few rare-defect samples and dominance of normal or common classes	Self-supervised pretraining, semi-supervised learning, few-shot learning, active learning, and class-balanced loss functions
Small and low-contrast defects	Weak features are lost during down sampling or compression	High-resolution feature branches, multi-scale fusion, deformable convolution, attention mechanisms, and region-of-interest-based inference.
Limited cross-domain generalization	Performance degradation across products, cameras, and production lines	Domain adaptation, domain generalization, feature normalization, and test-time adaptation.
Real-time edge deployment	Limited memory and computing capacity on industrial devices	Lightweight backbones, pruning, quantization, knowledge distillation, and hardware acceleration
Complex component geometry	Occlusion and incomplete representation in 2D images of pipe interiors and castings	Multi-view imaging, 3D reconstruction, depth sensing, and multimodal sensor fusion
Model reliability	Overconfident predictions and undetected out-of-distribution samples	Confidence calibration, uncertainty estimation, out-of-distribution detection, and human-in-the-loop verification

5.2. Future Research Directions

5.2.1. 3D Defect Detection

In the future evolution path of steel surface defect detection technology, the deep coupling from three-dimensional modeling perception to the automation of the detection process is becoming the core line driving the leap of the industrial inspection paradigm and the reconstruction of the operation and maintenance system. 3D defect detection has shown great potential in the field of defect and anomaly detection of high-precision industrial products [106]. Traditional two-dimensional vision has inherent bottlenecks in terms of illumination dependence, complex background interference and lack of depth information, while multimodal fusion and high-precision reconstruction technology are gradually breaking through these limitations. By introducing micro lidar, dynamic structured light, polarization vision and high-frequency acoustic sensors for multi-source heterogeneous data collaboration, the system can achieve a high-fidelity three-dimensional reconstruction of pipeline topology in extremely complex pipe environments. Further combining cutting-edge algorithms such as neural radiation field and three-dimensional point cloud deep learning, it can not only accurately depict the spatial morphology of cracks, corrosion pits and small deformations, but also effectively suppress environmental noise and multipath scattering, and achieve the high-density mapping of material surface geometric features and textures. This cross-dimensional representation enables defect assessment to leap from planar image-level discrimination to spatial-level quantitative analysis. By directly calculating material thinning rates, stress concentration gradients, and volumetric losses, and deeply interconnecting with finite element analysis, it provides rigorous physical quantitative support for evaluating the integrity of pipeline structures.

However, in the process of fully moving toward true 3D perception and digital twin systems, the current scarcity of high-quality spatial prior data is a real bottleneck restricting algorithm evolution. The high reflectivity and non-planar topology of complex metal surfaces make it difficult for traditional 2D images to meet the requirements for high-precision geometric feature extraction. Against this backdrop, the multi-light imaging paradigm based on active photometric stereochemistry becomes an important transitional bridge. Through multi-source, multi-viewpoint physical illumination modulation, this method can effectively reconstruct the normal vectors and depth cues of the workpiece surface, forming a 2.5D perception mechanism based on enhanced image features, thus achieving a good balance between the needs of 3D representation and current computing power and data constraints.

Nevertheless, this 2.5D perception based on enhanced image features is merely a transitional node in technological evolution. Looking to the future, with a complete upgrade in the dimension of perception, holographic digital twin models of pipelines constructed using multi-source historical inspection data will dominate the inspection system. By introducing the time dimension, the system will achieve a fundamental leap from static three-dimensional reconstruction to dynamic four-dimensional spatiotemporal evolution analysis. This holographic mapping can not only achieve precise three-dimensional localization of current defects, but also predictively extrapolate the propagation trend of pipeline fatigue cracks or corrosion evolution based on the spatiotemporal evolution patterns of multi-period historical inspection data. Ultimately, this will drive a fundamental upgrade in pipeline operation and maintenance models from passive post-event detection to proactive predictive maintenance.

5.2.2. Automation of the Detection Process

In the development of automated inspection processes, the traditional intermittent operation mode of "manual dominance and offline verification" will accelerate its evolution toward a fully autonomous, continuous and closed-loop unmanned paradigm. Automated defect detection systems are crucial for the timely detection of defects and for repair or updates [107]. Future inspection terminals will have high mobility and autonomous operation capabilities in extremely restricted physical environments. At the same time, lightweight perception models and domain adaptive algorithms deployed at the edge enable the system to complete real-time online reasoning from multi-scale feature extraction to defect semantic analysis within a very short time window of receiving massive sensor data streams. This low-latency processing mechanism of "perception as diagnosis" not only endows the system with the ability to continuously learn and self-evolve under complex and unknown working conditions, but also promotes the extension of the inspection network toward "integrated inspection and maintenance", thereby building a closed-loop intervention system of "precise perception—intelligent decision-making—in-situ repair".

In the context of the deep integration of automation and intelligence, the robustness and engineering feasibility of underlying visual algorithms increasingly rely on a data-driven paradigm that closely adheres to real physical constraints. Traditional idealized, heavily manually selected balanced datasets are no longer sufficient to meet the generalization requirements of models under extreme conditions within edge computing frameworks. In future fully autonomous detection architectures, sensor data from various front-end intelligent nodes will break through the limitations of manual pre-screening and directly drive models to make dynamic decisions in an end-to-end manner. Therefore, a dataset construction paradigm that conforms to the original, long-tail distribution characteristics of industrial sites can not only provide a high-confidence stress test benchmark for the embedded deployment of lightweight edge models, but also help to promote a fundamental leap

from "offline static data fitting" to "online continuous data flow modeling" of underlying algorithms, thereby laying an indispensable underlying logic and data foundation for building a new intelligent pipeline operation and maintenance ecosystem with full-domain autonomy and closed-loop management.

5.2.3. Multimodal Defect Detection with VLMs and LLMs

As industrial visual inspection gradually evolves from single-modality visual perception toward multimodal semantic understanding, the integration of vision-language models (VLMs) and large language models (LLMs) into steel surface defect detection has emerged as a promising research direction. Existing studies have demonstrated that deep learning-based methods have achieved substantial progress in industrial visual anomaly detection and steel surface defect recognition; however, challenges remain in modeling complex defect morphologies, achieving robust cross-scene generalization, recognizing defects under limited-sample conditions, and improving the interpretability of detection results [108,109]. Compared with conventional vision models, VLMs can align visual features with textual semantics, enabling defect detection systems not only to identify defect locations and categories but also to incorporate defect descriptions, process knowledge, and quality assessment criteria for higher-level semantic representation of defect morphology, severity, and potential impact. Therefore, introducing VLMs into steel surface defect detection may enhance the recognition and understanding of small, low-contrast, blurred-boundary, and rare defects while improving the interpretability and traceability of the detection results.

Furthermore, recent studies have highlighted the potential of large models in machine monitoring and fault diagnostics, particularly in industrial knowledge representation, cross-modal information integration, and intelligent decision support [110]. LLMs can serve as important components for industrial knowledge modeling, semantic reasoning, and decision support, thereby working together with visual inspection models to construct multimodal defect analysis frameworks. In such frameworks, visual models are primarily responsible for extracting the spatial locations, texture structures, and morphological characteristics of defect regions, whereas LLMs can integrate detection outputs with production process parameters, historical quality records, and industrial standards to support the comprehensive analysis of defect causes, risk levels, and potential treatment strategies. For common challenges in industrial scenarios, such as insufficient labeled samples, imbalanced defect categories, and the difficulty of collecting representative samples for emerging defect types, semi-supervised learning and related approaches have shown considerable application potential [111]. Building on this basis, the use of LLMs for textual priors, prompt learning, and automatic annotation may further reduce manual labeling costs and improve model adaptability across different material types, imaging conditions, and production environments. Therefore, future research on steel surface defect detection may further explore the deep integration of VLMs, LLMs, and industrial prior knowledge, thereby promoting the transition of defect detection systems from visual recognition toward semantic understanding, knowledge reasoning, and intelligent decision-making.

6. Summary

Steel is a core component of global industrial production, and the development of steel surface defect detection technology has driven high-quality steel production. Therefore, research on steel surface defect detection technology is of great significance worldwide. The various defect detection methods summarized in Section 3 of this paper were almost all published within the last five years, with most being methods from the last three years. These represent the most cutting-edge technological developments in the field and also

represent the best-performing methods. This paper first summarizes representative reviews in the field, comparing their characteristics and challenges with this paper. It also introduces the characteristics of surface defects in three common engineering steel materials: strip steel, steel pipe, and castings, as well as commonly used, authoritative, and newly released datasets in the field. Second, this paper systematically integrates detection methods based on the needs of real-world industrial scenarios and summarizes more than 20 performance evaluation indicators. Finally, it summarizes the future development trends of steel surface defect detection technology.

Future research should focus on lightweight model design, anti-interference generalization, and global multi-scale localization to overcome deployment challenges in industrial scenarios and improve the model's anti-interference ability and accurate defect morphology capture capabilities. The purpose of this study is to provide researchers and practitioners with detailed and clear references to promote continuous innovation in steel surface defect detection technology in the industrial field.

Author Contributions: Conceptualization, X.W.; Methodology, Y.H.; Validation, Z.Y., Y.H. and X.W.; Formal analysis, Z.Y. and Z.W.; Investigation, X.C. and X.W.; Resources, K.S.; Data curation, K.S.; Writing—original draft preparation, X.W., Z.Y., Y.H., Z.W., X.C. and K.S.; Writing—review and editing, X.W., Z.Y., Y.H., Z.W., X.C. and K.S.; Visualization, Z.Y. and Z.W.; Supervision, K.S.; Project administration, X.W.; Funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Steger, C.; Ulrich, M.; Wiedemann, C. *Machine Vision Algorithms and Applications*; Wiley-VCH: Weinheim, Germany, 2018.
2. Shi, T.; Kong, J.-Y.; Wang, X.-D.; Liu, Z.; Zheng, G. Improved Sobel Algorithm for Defect Detection of Rail Surfaces with Enhanced Efficiency and Accuracy. *J. Cent. South Univ.* **2016**, *23*, 2867–2875. [CrossRef]
3. Lin, H.-I.; Wibowo, F.S. Image Data Assessment Approach for Deep Learning-Based Metal Surface Defect-Detection Systems. *IEEE Access* **2021**, *9*, 47621–47638. [CrossRef]
4. Chen, F.; Fu, L.; Zhang, Y.; Li, J.; Zhang, Q.; Bi, S. A Review of Deep Learning-Based Steel Surface Defect Detection. *Acad. J. Sci. Technol.* **2025**, *15*, 198–202. [CrossRef]
5. Ibrahim, A.A.M.S.; Tapamo, J.-R. A Survey of Vision-Based Methods for Surface Defects' Detection and Classification in Steel Products. *Informatics* **2024**, *11*, 25. [CrossRef]
6. Zhang, J. A Survey of Semi-supervised and Unsupervised Learning Methods for Industrial Defect Anomaly Detection. *J. Comput. Sci. Front. Trends* **2025**, *1*, 1–15. [CrossRef]
7. Shukla, V.; Shukla, A.; SK, S.P.; Shukla, S. A systematic survey: Role of deep learning-based image anomaly detection in industrial inspection contexts. *Front. Robot. AI* **2025**, *12*, 1554196. [CrossRef] [PubMed]
8. Nahar, L.; Awrangjeb, M.; Islam, M.S. AI-enabled defect detection in industrial products: A comprehensive survey, key insights and future research challenges. *Adv. Eng. Inform.* **2026**, *69*, 104067. [CrossRef]
9. Shang, F.; Chen, H.; Sun, B.; Miao, R.; Zhou, X.; Xing, C.; Zhang, P. A Review of Deep Learning-Based Surface Defect Detection Techniques for Strip Steel. *Measurement* **2026**, *281*, 121905. [CrossRef]
10. Wen, X.; Shan, J.; He, Y.; Song, K. Steel Surface Defect Recognition: A Survey. *Coatings* **2023**, *13*, 17. [CrossRef]
11. Song, K.; Yan, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **2013**, *285*, 858–864. [CrossRef]
12. Luo, Q.; Fang, X.; Liu, L.; Yang, C.; Sun, Y. Automated Visual Defect Detection for Flat Steel Surface: A Survey. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 2629–2644. [CrossRef]

13. Wen, X.; Li, K.; He, Y.; Yin, H.; Song, K. MCF-Det: A Cross-Domain Fusion Framework for Few-Shot Blade Defect Detection. *IEEE Trans. Ind. Inform.* **2026**, *22*, 4872–4883. [CrossRef]
14. Neogi, N.; Mohanta, D.K.; Dutta, P.K. Review of vision-based steel surface inspection systems. *EURASIP J. Image Video Process.* **2014**, *2014*, 50. [CrossRef]
15. Gaiser, G.; Presoly, P.; Bernhard, C. High-Temperature Oxidation of Steel under Linear Flow Rates of Air and Water Vapor—An Experimental Determined Set of Data. *Metals* **2023**, *13*, 892. [CrossRef]
16. Ming, L.; Wangqi, S.; Ying, Z.; Yating, C.; Zuguo, C. WSS-YOLO: An Improved Industrial Defect Detection Network for Steel Surface Defects. *Measurement* **2024**, *236*, 115060. [CrossRef]
17. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2405–2415. [CrossRef]
18. Woodham, R.J. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* **1992**, *19*, 139–144. [CrossRef]
19. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [CrossRef]
20. Zheng, P.; Wang, H.; Sang, Z.; Zhong, R.Y.; Liu, Y.; Liu, C.; Mubarok, K.; Yu, S.; Xu, X. Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. *Front. Mech. Eng.* **2018**, *13*, 137–150. [CrossRef]
21. Lv, X.; Duan, F.; Jiang, J.-j.; Fu, X.; Gan, L. Deep Metallic Surface Defect Detection: The New Benchmark and Detection Network. *Sensors* **2020**, *20*, 1562. [CrossRef] [PubMed]
22. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
23. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
26. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]
27. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
28. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
29. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
30. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [CrossRef]
31. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969. [CrossRef]
32. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9626–9635. [CrossRef]
33. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007. [CrossRef]
34. Li, D.; Yang, J.; Kreis, K.; Torralba, A.; Fidler, S. Semantic Segmentation with Generative Models: Semi-Supervised Learning and Strong Out-of-Domain Generalization. *arXiv* **2021**, arXiv:2104.05833. [CrossRef]
35. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In Proceedings of the 4th International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.

37. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In *Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020*; Springer: Cham, Switzerland, 2020; pp. 213–229. [CrossRef]
38. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *arXiv* **2020**, arXiv:2010.04159. [CrossRef]
39. Zhang, H.; Li, F.; Liu, S.; Zhang, L.; Su, H.; Zhu, J.; Ni, L.M.; Shum, H.-Y. DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection. *arXiv* **2022**, arXiv:2203.03605. [CrossRef]
40. Hao, Z.; Wang, Z.; Bai, D.; Tong, X. Surface defect segmentation algorithm of steel plate based on geometric median filter pruning. *Front. Bioeng. Biotechnol.* **2022**, *10*, 945248. [CrossRef] [PubMed]
41. Guo, N.; Fan, H.; Dong, Q.; Gu, R.; Xu, S. A Lightweight Model for Hot-Rolled Steel Strip Surface Defect Recognition. *Sensors* **2026**, *26*, 1618. [CrossRef] [PubMed]
42. Zhang, J.; Ding, H.; Peng, M.; Tu, S.; Chen, G.; Liu, Y. Few-Shot Cross-Episode Adaptive Memory for Metal Surface Defect Semantic Segmentation. *Sci. Rep.* **2026**, *16*, 5660. [CrossRef] [PubMed]
43. Zhao, J.; Qian, X.; Zhang, Y.; Shan, D.; Liu, X.; Coleman, S.; Kerr, D. A Knowledge Distillation-Based Multi-Scale Relation-Prototypical Network for Cross-Domain Few-Shot Defect Classification. *J. Intell. Manuf.* **2024**, *35*, 841–857. [CrossRef]
44. Feng, H.; Song, K.; Cui, W.; Zhang, Y.; Yan, Y. Cross Position Aggregation Network for Few-Shot Strip Steel Surface Defect Segmentation. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5007410. [CrossRef]
45. Zhao, Y.; Chen, Q.; Du, Y.; Gong, W. VPGFNet: Cross-domain few-shot visual prompt graph fusion network for industrial defect segmentation. *J. Comput. Des. Eng.* **2025**, *12*, qwafi05. [CrossRef]
46. Zhang, J.; Su, H.; Zou, W.; Gong, X.; Zhang, Z.; Shen, F. CADN: A weakly supervised learning-based category-aware object detection network for surface defect detection. *Pattern Recognit.* **2021**, *109*, 107571. [CrossRef]
47. Meng, X.; Li, J.; Zhang, Y.; Ma, S. Improved residual attention convolutional neural network for rotating machinery fault diagnosis in the presence of strong noise. *Signal Image Video Process.* **2025**, *19*, 506. [CrossRef]
48. Wang, C.; Xie, H. MeDERT: A Metal Surface Defect Detection Model. *IEEE Access* **2023**, *11*, 35469–35478. [CrossRef]
49. Guo, Y.; Zhong, L.; Qiu, Y.; Wang, H.; Gao, F.; Wen, Z.; Zhan, C. Using ISU-GAN for unsupervised small sample defect detection. *Sci. Rep.* **2022**, *12*, 11604. [CrossRef] [PubMed]
50. Zhao, L.; Zheng, Y.; Peng, T.; Zheng, E. Metal Surface Defect Detection Based on a Transformer with Multi-Scale Mask Feature Fusion. *Sensors* **2023**, *23*, 9381. [CrossRef] [PubMed]
51. Chen, X.; Zhang, X.; Shi, Y.; Pang, J. HCT-Det: A High-Accuracy End-to-End Model for Steel Defect Detection Based on Hierarchical CNN–Transformer Features. *Sensors* **2025**, *25*, 1333. [CrossRef] [PubMed]
52. Zhang, Y.; Ma, Z.; Cui, T.; Zhao, T.; Wang, Q.; Wang, J. Wafer Defect Recognition for Industrial Inspection: FCS-VMamba Model and Experimental Validation. *J. Imaging* **2026**, *12*, 142. [CrossRef] [PubMed]
53. Hu, J.; Chen, W.; Zhao, L. MCF: Multi-scale Context Fusion for Strip Steel Surface Detection. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; pp. 1–8. [CrossRef]
54. Chen, H.; Song, K.; Tan, Y.; Xue, X.; Yan, Y.; Tao, X.; Shen, F. Float in the Air: Defect Detection in Pipelines With Electromagnetic Levitation Technology. *IEEE/ASME Trans. Mechatron.* **2026**, *31*, 1675–1686. [CrossRef]
55. Wang, H.; Zhang, C.; Wang, Y.; Ni, P.; Wang, Y. Segmentation of Inner Surface Defects of Stainless Steel Pipes Based on Semi-bilateral Efficient Self-Attention Network. *J. Nondestruct. Eval.* **2025**, *44*, 39. [CrossRef]
56. Zhou, X.; Shen, K.; Liu, Z. ADMNet: Attention-Guided Densely Multi-Scale Network for Lightweight Salient Object Detection. *IEEE Trans. Multimed.* **2024**, *26*, 10828–10841. [CrossRef]
57. Liu, H.; Chen, C.; Hu, R.; Bin, J.; Dong, H.; Liu, Z. CGTD-Net: Channel-Wise Global Transformer-Based Dual-Branch Network for Industrial Strip Steel Surface Defect Detection. *IEEE Sens. J.* **2024**, *24*, 4863–4873. [CrossRef]
58. Tan, Y.; Song, K.; Chen, H.; Zhang, Y.; Yan, Y. A Lightweight Perception Enhancement Network for Real-Time and Accurate Internal Surface Defect Detection of Cold-Drawn Steel Pipes. *IEEE Sens. J.* **2026**, *26*, 1383–1394. [CrossRef]
59. Huang, J.; Wu, Y.; Zhou, X.; Lin, J.; Chen, Z.; Zhang, G.; Xia, L.; Zhang, J. Multiscale Adaptive Prototype Transformer Network for Few-Shot Strip Steel Surface Defect Segmentation. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 5016514. [CrossRef]
60. Guo, B.; Wang, Y.; Zhen, S.; Yu, R.; Su, Z. SPEED:Semantic Prior and Extremely Efficient Dilated Convolution Network for Real-Time Metal Surface Defects Detection. *IEEE Trans. Ind. Inform.* **2023**, *19*, 11380–11390. [CrossRef]
61. Song, K.; Feng, H.; Cao, T.; Cui, W.; Yan, Y. MFANet: Multifeature Aggregation Network for Cross-Granularity Few-Shot Seamless Steel Tubes Surface Defect Segmentation. *IEEE Trans. Ind. Inform.* **2024**, *20*, 9725–9735. [CrossRef]
62. Ma, Y.; Liu, M.; Jiang, S.; Wang, X.; Bian, Y.; Wang, Y. Multi-Context Aggregation Network With Foreground Correction for Automated Few-Shot Defect Segmentation. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 13777–13787. [CrossRef]
63. Qiu, H.; Wang, B.; Qin, Y.; Ding, A.; He, Z.; Liu, J.; Huang, X. An Interpretable Few-Shot Framework for Fault Diagnosis of Train Transmission Systems with Noisy Labels. *J. Dyn. Monit. Diagn.* **2025**, *4*, 65–75. [CrossRef]

64. Chen, H.; Song, K.; Cui, W.; Zhang, T.; Yan, Y.; Li, J. SRPCNet: Self-Reinforcing Perception Coordination Network for Seamless Steel Pipes Internal Surface Defect Detection. *IEEE Trans. Ind. Inform.* **2025**, *21*, 950–959. [CrossRef]
65. Pan, Y.; Zhang, L. Dual Attention Deep Learning Network for Automatic Steel Surface Defect Segmentation. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 1468–1487. [CrossRef]
66. Yin, Z.; Qin, L.; Han, G.; Shi, X.; Zhang, F.; Xu, G.; Bi, Y. DDSNet: Deep Dual-Branch Networks for Surface Defect Segmentation. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2525316. [CrossRef]
67. Song, X.; Guo, F.; Zhang, L.; Lu, X.; Hei, X. Salient Object Detection With Dual-Branch Stepwise Feature Fusion and Edge Refinement. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 2832–2844. [CrossRef]
68. Wang, J.; Long, Z.; Lee, M.J.; Feng, Y.; Xu, Y.; Zhang, Y.; Alexander, I.; Tang, X.; Fan, R. Edge-Aware and Deformable Feature Fusion for Steel Surface Defect Segmentation. In Proceedings of the 2025 IEEE International Conference on Real-time Computing and Robotics (RCAR), Toyama, Japan, 14–19 July 2025; pp. 1016–1021. [CrossRef]
69. Shan, Z.; Haoyan, H.; Zhu, C.; Du, S.; Jing, H.; Haibin, W. RSM-YOLOv11: Lightweight Steel Surface Defect Segmentation Algorithm Research Based on YOLOv11 Improvement. *IEEE Access* **2025**, *13*, 111681–111698. [CrossRef]
70. He, Z.; Chen, Y.; Zhang, Z.; Huang, Z. LRTISS: A Lightweight Real-Time Network for Industrial Surface Defect Segmentation. *Measurement* **2026**, *275*, 121410. [CrossRef]
71. Huang, S.; Xiong, L. Rapid Recognition and Semantic Segmentation of Molten Steel Distribution Images on a Steel Ladle Surface based on the Improved SqueezeNAS Network. In Proceedings of the 2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC), Hefei, China, 26–28 August 2023; pp. 762–767. [CrossRef]
72. Feng, H.; Song, K.; Cui, W.; Zhou, Z.; Yan, Y. A Suspected Defect Screening-Guided Lightweight Network for Few-Shot Aviation Steel Tube Surface Defect Segmentation. *IEEE Sens. J.* **2024**, *24*, 38113–38124. [CrossRef]
73. Saberironaghi, M.; Ren, J.; Saberironaghi, A. ID-MSNet: An Enhanced Multi-Scale Network with Convolutional Attention for Pixel-Level Steel Defect Segmentation. *Algorithms* **2026**, *19*, 294. [CrossRef]
74. Peng, Y.; Wang, C.; Hao, Y.; Zhen, L.; Chen, G.; Kumar, N.; Yu, K. High-Precision Surface Crack Detection for Rolling Steel Production Equipment in ICPS. *IEEE Internet Things J.* **2024**, *11*, 4586–4599. [CrossRef]
75. Xiao, M.; Yang, B.; Wang, S.; Mo, F.; He, Y.; Gao, Y. GRA-Net: Global Receptive Attention Network for Surface Defect Detection. *Knowl.-Based Syst.* **2023**, *280*, 111066. [CrossRef]
76. Xiao, H.; Ren, Z.; Cai, S. HMNet: High-Resolution Mamba Network for Semantic Segmentation of Surface Defects. In Proceedings of the 2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Compiegne, France, 7–9 May 2025; pp. 13–18. [CrossRef]
77. Wang, Z.; Zhou, W.; Li, Y. MSAF-YOLO: An Efficient Multi-Scale Attention Fusion Network for High-Precision Steel Surface Defect Detection. *Measurement* **2026**, *257*, 118640. [CrossRef]
78. Li, R.; Hou, K.; Zhu, M.; Dai, Q.; Ni, J.; Liu, X.; Li, X. ACPP-Net: Enhancing Strip Steel Surface Defect Detection With Efficient Adaptive Convolution and Channel-Spatial Pyramid Pooling. *IEEE Access* **2024**, *12*, 152072–152086. [CrossRef]
79. Duan, B.; Wang, D.; Ma, Y.; Wang, G.; Liu, H. Multisource Data-Driven Intelligent Method for Detecting Surface Defects in Cold-Rolled Copper Strips. *Eng. Appl. Artif. Intell.* **2025**, *152*, 110730. [CrossRef]
80. Li, Y.; Xu, S.; Zhu, Z.; Wang, P.; Li, K.; He, Q.; Zheng, Q. EFC-YOLO: An Efficient Surface-Defect-Detection Algorithm for Steel Strips. *Sensors* **2023**, *23*, 7619. [CrossRef] [PubMed]
81. Luo, Q.; Zhan, W.; Su, J. TaDP-Det: Semi-Supervised Texture-Aware Dynamic Pseudo-Labeling Detector for Industrial Surface Defect Detection. *Sensors* **2026**, *26*, 1085. [CrossRef] [PubMed]
82. Lin, Z.; Li, Z.; Yu, J.; Hu, M.; Wang, X. FFDDNet: Flexible Focused Defect Detection Network. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 5019812. [CrossRef]
83. Fu, J.; Zhang, Z.; Peng, T.; Hu, X.; Zhang, J. SD-GASNet: Efficient Dual-Domain Multi-Scale Fusion Network with Self-Distillation for Surface Defect Detection. *Sensors* **2026**, *26*, 23. [CrossRef] [PubMed]
84. Zheng, X.; Liu, W.; Huang, Y. LWMS-Net: A Novel Defect Detection Network Based on Multi-Wavelet Multi-Scale for Steel Surface Defects. *Measurement* **2025**, *252*, 117393. [CrossRef]
85. Luo, Q.; Li, B.; Su, J.; Yang, C.; Gui, W.; Silven, O.; Liu, L. CDDNet: Camouflaged Defect Detection Network for Steel Surface. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5000313. [CrossRef]
86. Huang, J.; Zhang, X.; Jia, L.; Zhou, Y. An Improved You Only Look Once Model for the Multi-Scale Steel Surface Defect Detection with Multi-Level Alignment and Cross-Layer Redistribution Features. *Eng. Appl. Artif. Intell.* **2025**, *145*, 110214. [CrossRef]
87. Liu, R.; Huang, M.; Gao, Z.; Cao, Z.; Cao, P. MSC-DNet: An Efficient Detector with Multi-Scale Context for Defect Detection on Strip Steel Surface. *Measurement* **2023**, *209*, 112467. [CrossRef]
88. Song, H. RSTD-YOLOv7: A steel surface defect detection based on improved YOLOv7. *Sci. Rep.* **2025**, *15*, 19649. [CrossRef] [PubMed]
89. Zhang, Y.; Xu, B.; Yang, G. Steel surface defect detection based on multi-scale dynamic convolution and lightweight cross-stage fusion. *PeerJ Comput. Sci.* **2026**, *12*, e3485. [CrossRef]

90. Chen, H.; Du, Y.; Fu, Y.; Zhu, J.; Zeng, H. DCAM-Net: A Rapid Detection Network for Strip Steel Surface Defects Based on Deformable Convolution and Attention Mechanism. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5005312. [CrossRef]
91. Hu, X.; Yang, J.; Jiang, F.; Hussain, A.; Dashtipour, K.; Gogate, M. Steel Surface Defect Detection Based on Self-Supervised Contrastive Representation Learning with Matching Metric. *Appl. Soft Comput.* **2023**, *145*, 110578. [CrossRef]
92. Huang, X.; Li, Y.; Bao, Y.; Zheng, W. Adaptive Cross Transformer With Contrastive Learning for Surface Defect Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2534717. [CrossRef]
93. Wang, J.; Xu, P.; Li, L.; Zhang, F. DAssd-Net: A Lightweight Steel Surface Defect Detection Model Based on Multi-Branch Dilated Convolution Aggregation and Multi-Domain Perception Detection Head. *Sensors* **2023**, *23*, 5488. [CrossRef] [PubMed]
94. Zhang, T.; Pan, P.; Zhang, J.; Zhang, X. Steel Surface Defect Detection Algorithm Based on Improved YOLOv8n. *Appl. Sci.* **2024**, *14*, 5325. [CrossRef]
95. Tastimur, C. Mean shift based prototypical network for steel surface anomaly recognition. *Sci. Rep.* **2025**, *15*, 45097. [CrossRef] [PubMed]
96. Wang, Y.; Cheng, H.; Shi, J.; Du, X.; Wang, P. Co-training encoder-based transformer method for steel surface defect detection. *J. Electron. Imaging* **2025**, *34*, 023058. [CrossRef]
97. Ma, Y.; Zhang, Z.; Ma, S.; Shi, K.; Fan, C. GLF-NET: Global and Local Dynamic Feature Fusion Network for Real-Time Steel Strip Surface Defect Detection. *IEEE Access* **2025**, *13*, 26063–26078. [CrossRef]
98. Guan, S.; Hou, L.; Shen, J.; Song, Y. MC-DDet: Lightweight Mini-Channel Detector for Steel Strip Surface Defect Detection. *IEEE Sens. J.* **2025**, *25*, 40214–40222. [CrossRef]
99. Zhang, X.; Jian, M.; Zhou, S.; Cong, S.; Sheng, Z. DGS-YOLO: A Lightweight YOLOv8-Based Method for Steel Surface Defect Detection. *IEEE Sens. J.* **2026**, *26*, 5902–5913. [CrossRef]
100. Pan, W.; Zhong, R.; Huang, J.; Li, Y.; Zhang, W.; Liu, T.; Liu, Y. DEENet: An edge-enhanced CNN–Transformer dual-encoder model for steel surface defect detection. *Sci. Rep.* **2026**, *16*, 6692. [CrossRef] [PubMed]
101. Zhou, S.; Cai, Y.; Zhang, Z.; Yin, J. MESC-DETR: An Improved RT-DETR Algorithm for Steel Surface Defect Detection. *Electronics* **2025**, *14*, 2232. [CrossRef]
102. Li, S.; Chen, Z.; Zhang, X.; Yang, T.; Gong, D. STCNet: A hybrid Transformer and CNN network for surface defect detection of steel continuous casting billets. *Meas. Sci. Technol.* **2025**, *36*, 105602. [CrossRef]
103. Zhu, W.; Zhang, H.; Zhang, C.; Zhu, X.; Guan, Z.; Jia, J. Surface Defect Detection and Classification of Steel Using an Efficient Swin Transformer. *Adv. Eng. Inform.* **2023**, *57*, 102061. [CrossRef]
104. Karim, S.T.; Siraj, F.M.; Uddin, J. Steel Surface Defect Detection Using Learnable Memory Vision Transformer. *Comput. Mater. Contin.* **2025**, *82*, 499–520. [CrossRef]
105. Huang, X.; Zhu, J.; Huo, Y. SSA-YOLO: An Improved YOLO for Hot-Rolled Strip Steel Surface Defect Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5040017. [CrossRef]
106. Lu, H.; Liang, H.; Zhang, J.; Hu, C.; Wang, J.; Gao, C. C3D-AD: Toward Continual 3D Anomaly Detection via Kernel Attention with Learnable Advisor. *arXiv* **2025**, arXiv:2508.01311. [CrossRef]
107. Sun, L.; Zhu, J.; Tan, J.; Li, X.; Li, R.; Deng, H.; Zhang, X.; Liu, B.; Zhu, X. Deep Learning-Assisted Automated Sewage Pipe Defect Detection for Urban Water Environment Management. *Sci. Total Environ.* **2023**, *882*, 163562. [CrossRef] [PubMed]
108. Li, Z.; Yan, Y.; Wang, X.; Ge, Y.; Meng, L. A Survey of Deep Learning for Industrial Visual Anomaly Detection. *Artif. Intell. Rev.* **2025**, *58*, 279. [CrossRef]
109. Zhou, Y.; Zhao, Z. MPA-YOLO: Steel Surface Defect Detection Based on Improved YOLOv8 Framework. *Pattern Recognit.* **2025**, *168*, 111897. [CrossRef]
110. Chen, X.; Lei, Y.; Li, Y.; Parkinson, S.; Li, X.; Liu, J.; Lu, F.; Wang, H.; Wang, Z.; Yang, B.; et al. Large Models for Machine Monitoring and Fault Diagnostics: Opportunities, Challenges, and Future Direction. *J. Dyn. Monit. Diagn.* **2025**, *4*, 76–90. [CrossRef]
111. Shi, C.; Wang, K.; Zhang, G.; Li, Z.; Zhu, C. Efficient and Accurate Semi-Supervised Semantic Segmentation for Industrial Surface Defects. *Sci. Rep.* **2024**, *14*, 21874. [CrossRef] [PubMed]

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