



REVIEW

A Survey of Surface Defect Detection in Machine Vision: Addressing Core Challenges, Methodologies, and Dataset Analysis

Langyue Zhao^{1,2}, Yubin Yuan^{3,*} and Yiquan Wu^{2,*}

¹College of Computer Science, Weinan Normal University, Weinan, China

²College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

³College of Information Engineering, Yangzhou University, Yangzhou, China

*Corresponding Authors: Yubin Yuan. Email: yubin.yuan@yzu.edu.cn; Yiquan Wu. Email: nuaatracking@163.com

Received: 05 February 2026; Accepted: 28 April 2026; Published: 15 June 2026

ABSTRACT: This paper presents a systematic survey of machine vision-based surface defect detection technologies, focusing on five core challenges in the field: interference from complex backgrounds, small object detection, class imbalance, dynamic scene modeling, and cross-scenario generalization. It reviews key technical approaches corresponding to these challenges over the past five years. Furthermore, a dataset characterization analysis framework is established around these challenges, summarizing and comparing the characteristics of over 40 publicly available datasets across more than ten scenarios, including PCB, photovoltaic, metal, and pavement surfaces. Quantitative selection metrics (such as the small target coefficient and texture complexity) are proposed for challenges like small target detection and complex backgrounds, offering a methodological guide for aligning research questions with benchmark data. Finally, the paper summarizes current limitations and provides an outlook on new paradigms driven by large-scale models and the construction of high-quality benchmark datasets, aiming to offer valuable references for both research and engineering practices in this field.

KEYWORDS: Surface defect detection; machine vision; complex scenarios; dataset survey; industrial vision

1 Introduction

With the accelerated advancement of intelligent manufacturing, industrial quality control, and smart infrastructure operation and maintenance, surface defect detection, as a core application of intelligent vision systems in industrial scenarios, is increasingly becoming a critical link connecting physical perception and decision-making feedback [1]. Its primary objective is to achieve automated, non-contact, and high-precision identification of minute anomalies on the surfaces of materials, components, or structures through image acquisition, feature modeling, and intelligent analysis. This enables early detection and intervention of defects, effectively ensuring product quality, extending service life, and reducing maintenance costs [2,3].

In the field of basic materials manufacturing, the precision requirements for detecting defects in metallic materials, such as rolling defects and micro-cracks in strip steel, are exceptionally high. Similarly, defects in non-metallic materials, such as bubbles in glass or glaze flaws in ceramics, demand precise identification. The surface quality of these fundamental materials directly impacts subsequent processing techniques and the performance of the final product. Semiconductor manufacturing exhibits particular sensitivity to surface defects. For instance, defects on printed circuit boards (PCBs), such as metal contamination, micro-cracks

in dielectric layers, or cold solder joints, can exponentially accelerate device failure through multi-physics field coupling effects [4].

In the equipment manufacturing sector, solar cells are the core components of photovoltaic power generation. The uniformity of electrode coating and the integrity of separator surfaces directly determine the energy density and safety performance of the battery. Residual micron-scale metal particles can potentially trigger thermal runaway [5]. The surface roughness requirements for precision mechanical components reach sub-micron levels, where any flaw can lead to equipment vibration or loss of accuracy [6].

Furthermore, in the infrastructure domain, such as in road health monitoring, the texture characteristics of asphalt concrete surfaces are closely related to skid resistance, making crack detection crucial for ensuring traffic safety [7]. Common defect detection scenarios are illustrated in Fig. 1.

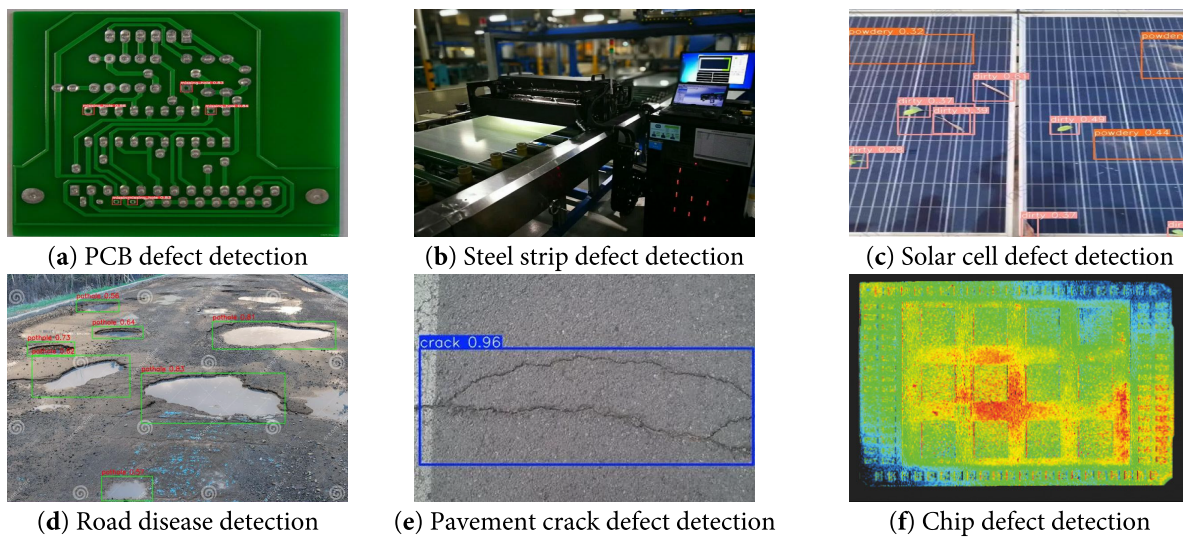


Figure 1: Examples of surface defect detection scenarios.

This survey focuses on five core challenges in surface defect detection: complex background interference, small object detection, class imbalance, dynamic scene modeling, and cross-scenario generalization. Based on these challenges, this survey aims to focus on the following three core aspects: (1) What are the evolutionary trajectories and characteristic strengths and weaknesses of existing technical approaches in addressing the five core challenges of surface defect detection? (2) How can existing public datasets be quantitatively evaluated and selected from the perspective of challenge adaptability? (3) What are the main gaps between current technical systems and industrial application requirements, and what are the promising future research directions?

The literature search was performed across the following academic databases: IEEE Xplore, Web of Science, Scopus, Google Scholar, and CNKI (for Chinese-language publications). The search covered the period from January 2015 to December 2025, with a primary focus on the past five years (2020–2025) to capture the most recent advances, while including seminal earlier works where necessary.

2 Challenges in Machine Vision-Based Surface Defect Detection

Although machine vision-based surface defect detection technology has achieved significant results across various fields, including industrial manufacturing, road inspection, and infrastructure maintenance, its application in real-world complex scenarios still faces numerous challenges. The diversity of defect

types, the unpredictability of object morphology, interference from scene backgrounds, and constraints of deployment environments collectively make it difficult for existing detection systems to fully meet requirements in terms of accuracy, robustness, and generalization. How to construct stable, efficient, and adaptive defect detection models in high-interference environments, with weak-signal targets and dynamic scenes, has become a critical issue constraining the further development of vision systems.

(1) Difficulty in Accurately Identifying Defect Objects under Complex Background Interference.

In industrial and infrastructure imagery, surface defects often exhibit high similarity to background textures, characterized by low contrast, weak edges, and severe occlusion. In tasks such as road crack detection and metal surface flaw identification, factors like light reflection, overlapping patterns, and pseudo-texture interference can significantly degrade the visual system's ability to perceive defect regions. Particularly in reflective materials, complex craft products, or naturally corroded surfaces, traditional static feature extraction networks often fail to distinguish real defects from environmental noise, leading to high false-positive rates. Therefore, there is an urgent need to develop detection architectures with background suppression capabilities and saliency modeling mechanisms to achieve stable perception of foreground defects in complex scenes.

(2) Insufficient Modeling and Representation Capability for Small-Scale and Weakly-Structured Defects.

In high-density integrated manufacturing, photovoltaic module inspection, and microscopic defect detection tasks, defects often manifest as micro-cracks, point-like bubbles, or hairline scratches. Their object regions occupy only a few pixels in images, resulting in extremely sparse feature representation. Existing detection algorithms tend to lose boundary and semantic information of such objects during down sampling, making it difficult for models to learn effective discriminative features. Additionally, issues such as limited positive samples and weak training supervision further constrain system performance. Enhancing model responsiveness in small-object scenarios through foreground enhancement, edge reconstruction, and matching mechanism optimization represents a key technological bottleneck that urgently needs to be addressed.

(3) Challenges Posed by Irregular Geometric Defects and Scale Variations to Feature Structures.

Real-world defect objects often exhibit irregular geometric forms, such as crack bifurcations, rough spalling edges, or fragmented corrosion contours. Their spatial structures are complex, with large scale spans and drastic boundary changes. Particularly in road damage detection, ceramic crack identification, or welding breakpoint inspection, such unstructured targets are easily misclassified or partially ignored by conventional convolutional architectures. Moreover, severe class imbalance and skewed sample distributions in training data cause models to favor high-frequency classes while neglecting long-tail defects. Consequently, constructing feature extraction and decision-making structures with geometric adaptability, multi-scale semantic modeling mechanisms, and sample regulation strategies has become a core pathway to addressing this challenge.

(4) Continuous Modeling and Identity Maintenance of Defect Objects in Dynamic Scenes.

Traditional defect detection primarily focuses on static images. However, in practical applications such as road patrols, pipeline monitoring, or video-based assessment tasks, continuous recognition, tracking, and statistical analysis of defect targets across video sequences are required. Due to factors like viewpoint changes, illumination variations, and occlusion interference, defect targets often experience positional drift, morphological changes, or even temporary disappearance over time. Current models based on frame-independent detection cannot maintain target consistency, leading to duplicate identifications, trajectory

interruptions, and statistical inaccuracies. Therefore, introducing temporal modeling mechanisms and cross-frame identity association strategies to enhance stable recognition capabilities in dynamic scenes is a critical challenge for evolving vision systems toward multi-task perception.

(5) Bottlenecks in Model Adaptability for Cross-Scenario Generalization and Low-Resource Deployment.

Industrial and transportation applications are highly diverse, with significant differences in image distributions due to variations in materials, processes, and acquisition conditions. This makes it difficult for a single trained model to be directly transferred across different domains. Simultaneously, deployment on edge devices often faces constraints such as limited computing power, restricted energy consumption, and stringent real-time requirements, posing severe challenges to model lightweighting and adaptation capabilities. Current methods predominantly rely on static architectures and end-to-end training, lacking dynamic structural adjustments and knowledge transfer mechanisms, which hinders support for multi-source task integration and rapid deployment. Thus, designing a universal detection architecture with structural flexibility, task coordination, and knowledge transferability is essential for realizing the practical implementation of multi-scenario intelligent quality inspection systems.

In summary, as surface defect detection technology progresses toward mature industrial-grade applications, it continues to face profound challenges across multiple dimensions, including feature representation, target matching, temporal reasoning, and cross-domain adaptation. These challenges are interconnected, forming a complex network of research and engineering problems. The subsequent chapters of this paper will systematically review the innovative ideas, methodological frameworks, and technological advancements proposed by researchers worldwide in addressing these five major challenges.

3 A Survey of Surface Defect Detection Methods Based on Machine Vision

To provide a structured overview of the methodological landscape, Fig. 2 presents a taxonomy of surface defect detection methods organized by the five core challenges and their corresponding technical approaches.

3.1 Defect Detection Methods for Small Object Scenarios

Small object detection is a pervasive challenge in surface defect inspection tasks. In typical industrial scenarios such as solder balls and scratches on PCBs, delamination and porosity in composite materials, and micro-cracks in silicon wafers, defects are inherently small in size, have subtle textures, and are easily obscured by background noise. This leads to weak feature representation and ambiguous localization within deep neural networks, significantly impacting the recall rate and precision of detection models. The definition of a small object generally falls into two categories: a relative scale based on the area ratio of the object to the entire image (e.g., an area less than 0.12% or 0.03%), and an absolute scale based on the object's pixel dimensions, the specifics of which depend on the dataset. For instance, the DOTA aerial dataset defines objects with 10–50 pixels as small [8], while the general-purpose MS COCO dataset categorizes objects with resolutions below 32×32 pixels as small [9]. Notably, these definitions primarily originate from the field of general object detection. In high-resolution industrial imaging (e.g., PCBs, wafers), an object spanning hundreds of pixels may still be considered an 'effectively small target' if it occupies an extremely low proportion of the entire image ($<0.01\%$). To address the challenges of small object detection, current research primarily focuses on technological directions such as multi-scale feature fusion, context information enhancement, and super-resolution reconstruction.

- (1) **Multi-Scale Feature Fusion-based Methods.** Critical details of small objects are often lost after multiple downsampling operations in deep networks. Therefore, fusing high-resolution shallow features with semantically rich deep features has become a mainstream approach. Liu et al. proposed

the SSD (Single Shot MultiBox Detector), which first utilized feature maps at multiple scales to detect objects of different sizes, with shallow layers responsible for small objects and deep layers for large ones [10]. Subsequently, building on SSD, Lin introduced the FPN (Feature Pyramid Network) structure, constructing a top-down feature pyramid with lateral connections to effectively fuse shallow details and deep semantics [11].

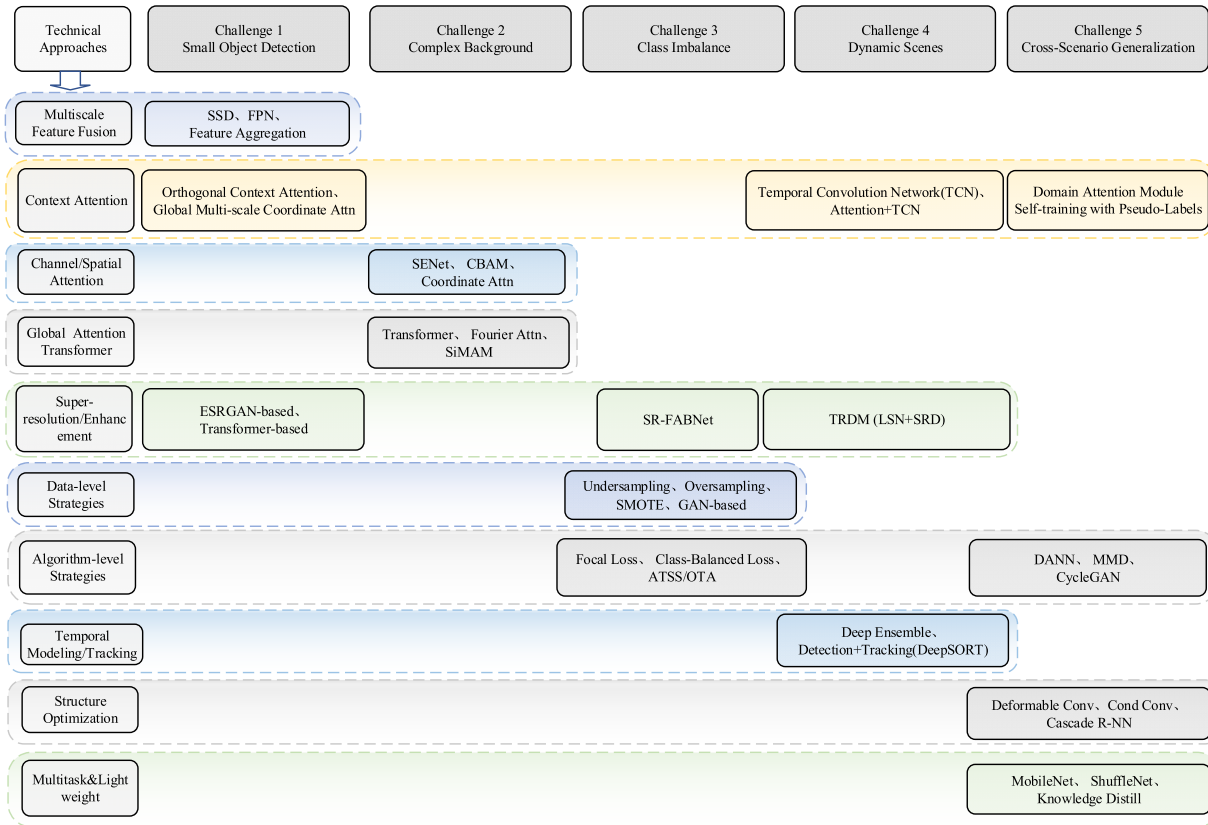


Figure 2: Taxonomy diagram.

These methods effectively mitigate the issue of information loss for small objects during downsampling by constructing semantically enhanced feature pyramids. However, multi-scale fusion models often suffer from high computational complexity and semantic inconsistencies between different feature layers. Shallow features are prone to noise, while deep features are highly abstract; direct concatenation or fusion can lead to redundancy or false detections. Consequently, subsequent research has gradually introduced lightweight structures, attention mechanisms, and local super-resolution enhancement modules to achieve more efficient collaboration between shallow and deep information, providing a solid structural foundation for subsequent context modeling and image enhancement strategies.

- (2) **Context Information-Based Methods.** The limited texture information of small objects makes precise recognition based solely on their intrinsic features difficult. Thus, researchers have introduced context information modeling mechanisms to capture the semantic relationships between a target and its surrounding regions, thereby improving discriminative power. Contextual information includes not only neighboring pixels but also semantic cues from the target’s region, such as scene structure, spatial location, and class priors. Ref. [12] proposed an Orthogonal Context Attention Module, which

simultaneously enhances contextual details from vertical and horizontal directions, strengthening the feature representation of tiny defects in steel strips.

Context modeling methods are particularly effective in scenarios where objects are surrounded by regular structures (e.g., weld seams, steel plate scratches) and are valuable for small object localization in complex textured backgrounds. However, their performance hinges on the accuracy of context modeling; if the background is complex or the contextual structure is ambiguous, they may introduce redundant or even misleading information. Additionally, the computational overhead of context networks requires careful consideration. Therefore, lightweight context modules and adaptive relationship modeling under Transformer architectures (e.g., global attention, dynamic receptive field regulation) will be key future development focuses.

- (3) **Super-Resolution-Based Methods.** Low resolution directly limits the pixel-level information available for small objects, complicating feature extraction. In recent years, a substantial body of research has incorporated super-resolution reconstruction techniques into small object detection. These methods enhance critical details like edges and textures by improving resolution at either the image or feature level. Dwivedi et al. proposed a low-cost, ESRGAN-based driving method for enhancing the spatial resolution of luminescence images, significantly boosting the detection performance of microscopic defects in solar cells [13].

Super-resolution methods are especially effective in low-quality image scenarios, significantly compensating for detail loss in small objects caused by compression, viewing angles, or sampling limitations. However, challenges are evident: on one hand, Generative Adversarial Network (GAN)-based super-resolution methods are computationally intensive and difficult to deploy in real-time; on the other hand, generated images may contain artificial textures or erroneous boundaries, leading to increased false detection rates. Future directions include dynamic sparse computation (activating high-resolution processing only in suspicious regions), end-to-end differentiable upsampling mechanisms, and a paradigm shift from dense prediction towards sparse representations driven by points or contours.

Summary and Comparative Analysis. To facilitate a comparative understanding of the methods discussed above, [Table 1](#) summarizes the key characteristics, advantages, and limitations of representative approaches for small object detection.

Table 1: Comparison of representative small object detection methods.

Method Category	Representative Works	Key Idea	Advantages	Limitations
Multi-Scale Feature Fusion	SSD, FPN	Fusing high-resolution shallow features with semantically rich deep features	Effectively mitigates information loss for small objects during downsampling	High computational complexity; semantic inconsistency across layers

(Continued)

Table 1 (continued)

Method Category	Representative Works	Key Idea	Advantages	Limitations
Context Information-Based	Orthogonal Context Attention	Capturing semantic relationships between target and surrounding regions	Enhances discriminability of targets in structured backgrounds	Performance depends on context accuracy; may introduce noise in complex backgrounds
Super-Resolution-Based	ESRGAN-Based	Reconstructing high-resolution details at image or feature level	Compensates for detail loss in low-quality images	High computational cost; may introduce artificial textures or errors

Real-World Implementation Considerations. In practical industrial deployment, small object detection faces additional challenges beyond algorithmic design. The resolution and optical setup of imaging systems directly determine the upper bound of detectable defect sizes. For applications such as PCB inspection or wafer defect detection, telecentric lenses and high-resolution line-scan cameras are often required to ensure that micron-level defects occupy sufficient pixels. Furthermore, motion blur in conveyor belt scenarios can severely degrade small defect features, necessitating integration with motion compensation hardware or deblurring preprocessing modules. These engineering considerations highlight that effective small defect detection requires a holistic approach combining algorithm design with appropriate imaging system configuration.

3.2 Defect Detection Methods for Complex Backgrounds

In tasks such as PCB inspection, road crack detection, and pipeline corrosion assessment, defects are often characterized by their small size, variable morphology, and high degree of blending with complex, cluttered background textures, resulting in an extremely low signal-to-noise ratio. This poses significant challenges for precise localization and identification. Existing methodologies have largely evolved through two iterative phases: the traditional model-driven stage and the deep learning data-driven stage. Traditional methods, based on handcrafted features and machine learning classifiers, are theoretically sound but susceptible to noise interference [14]. With the rise of CNNs, architectures like U-Net, FPN, DeepLab, and BiFPN have significantly improved accuracy by leveraging large-scale annotated data. However, these models are primarily designed for natural scenes with clear boundaries, whereas industrial defect images feature intricate textures and minuscule targets, creating a “domain shift” [15]. To suppress background noise and accentuate defect regions, researchers have extensively integrated attention mechanisms, which can be broadly categorized into three types: channel attention mechanisms, channel-spatial joint attention mechanisms, and global attention mechanisms.

- (1) **Channel Attention Mechanisms.** This mechanism enhances the response strength of discriminative features by learning the importance of each channel. Zhang et al. integrated SENet into YOLOv5, significantly improving the localization accuracy of insulators in foggy conditions [16]. Hao et al.

embedded Residual Split Attention into YOLOv4, strengthening insulator defect recognition in complex scenes [17]. Zhou et al. proposed IFIFusion, employing a dual-attention Transformer backbone combined with an LWNet and an IFIF multi-scale coupling module to enhance defect perception [18]. Ma et al. designed ELA-YOLO, introducing linear attention within the backbone and constructing a selective feature pyramid to compress computational costs while maintaining high precision [19]. Channel attention can be integrated into existing detectors without significantly altering the network topology, making it a common solution for rapid upgrades in industrial production lines.

- (2) **Channel-Spatial Joint Attention Mechanisms.** Weighting channels alone may still miss crucial positional information. Therefore, researchers began modeling spatial saliency simultaneously. Hao et al. incorporated CBAM into the YOLOv5 backbone to perform explicit saliency enhancement for faulty targets [20]. Zhou et al. similarly utilized CBAM for glass insulator detection [21]. Song et al. placed a Coordinate Attention Module (CAM) in the feature fusion layer, significantly improving the detection rate of bird nests on transmission lines [22]. For multi-force defects and multi-scale scenarios, Wang et al. proposed SDA-PVTDet, designing an SO-Emb and dual-attention blocks to extract blurred boundary features and using FBPN to fuse multi-scale semantics [23]. Pu et al. jointly modeled geometric correlations through RSIR and SCA to accurately distinguish highly similar impeller defects [24]. Sui and Wang designed DMPDD-Net, which utilizes a dual-path parallel attention mechanism (DP-AM), MFFM, and PSPPF to collaboratively suppress noise, achieving significant gains for the challenging “high-similarity background + small target” problem in aluminum profiles [25]. Joint attention mechanisms highlight genuine defects amidst cluttered backgrounds by simultaneously filtering redundant channels and interfering regions.
- (3) **Global Attention Mechanisms.** Local attention mechanisms struggle to capture long-range dependencies. Global attention mechanisms leverage self-attention or deformable convolutions to identify macroscopic structures across a wide field of view. He et al. combined deformable convolutions with a multi-path collaborative attention mechanism at the global level to learn background-target boundaries [26]. Chen et al. proposed a Fourier Attention Network guided by an SR branch, using FAB to analyze frequency domain distributions and cooperating with super-resolution reconstruction to achieve collaborative defect extraction through “global suppression + local enhancement” [27]. Global attention mechanisms significantly improve detection rates for images with multiple coexisting defects or extremely complex textures. However, their computational and memory consumption is substantial, requiring strategies like sparsification or window partitioning for real-time deployment.

Table 2 provides a comparative summary of attention mechanism-based methods for complex background suppression.

Table 2: Comparison of attention mechanisms for complex background suppression.

Attention Type	Representative Works	Key Idea	Advantages	Limitations
Channel Attention	SENet + YOLOv5, ELA-YOLO	Learns channel importance to enhance discriminative features	Easy integration; minimal architecture modification	Lacks spatial awareness

(Continued)

Table 2 (continued)

Attention Type	Representative Works	Key Idea	Advantages	Limitations
Channel-Spatial Joint Attention	CBAM + YOLOv5, Coordinate Attention	Simultaneously models channel and spatial saliency	Effectively filters both redundant channels and interfering regions	Higher computational overhead
Global Attention	Transformer + YOLOv8, Fourier Attention Network	Captures long-range dependencies via self-attention	Handles multi-defect coexistence and complex textures	Substantial memory consumption; requires sparsification for real-time deployment

In summary, in industrial visual inspection, defect regions often exhibit low contrast due to high similarity with background textures or limited imaging conditions, making their saliency difficult to express effectively. Beyond the primary focus on introducing attention mechanisms, current research also employs multi-scale feature fusion strategies (e.g., FPN, U-Net++), frequency-domain analysis techniques (e.g., wavelet or Fourier transforms) to enhance the response to weak abnormal signals, as well as self-supervised or contrastive learning methods for pre-training models on unlabeled data to improve sensitivity to subtle anomalies. Although these methods have improved defect discernibility to some extent, challenges remain, such as attention being easily distracted by strong textures, the high computational overhead of frequency-domain methods, and unstable transfer effects in self-supervised learning. Future trends will place greater emphasis on multimodal information fusion (e.g., combining visible light with polarization/infrared images), generative prior modeling (e.g., diffusion models reconstructing normal sample distributions), and physics-informed network design guided by imaging principles to achieve more robust saliency representation.

3.3 Class-Imbalanced Defect Detection Methods

Addressing the issue of class imbalance caused by uneven distribution of defect samples, current research primarily focuses on two directions: the data level and the algorithm level. This problem is particularly prominent in scenarios such as industrial quality inspection (e.g., rare defects on PCBs, special stains on textiles) and public infrastructure inspection (e.g., road cracks, where the distribution of defects in shape and size is extremely uneven, and severe network cracks are far less frequent compared to numerous fine cracks) [28]. Data-level methods adjust the data distribution through defect sample augmentation (e.g., GAN-based defect generation, data augmentation) and normal sample filtering. Algorithm-level methods enhance the model's ability to identify long-tail and hard-to-detect defect classes by designing improved loss functions (e.g., weighted cross-entropy, Focal Loss), label optimization strategies, and feature enhancement mechanisms (e.g., attention mechanism-guided feature focusing).

Data-level strategies directly adjust the sample composition before training to mitigate distribution bias inherent in the input data. Among these, normal sample filtering methods mainly include undersampling and oversampling. Random undersampling balances classes by randomly removing samples from the majority class but risks discarding useful information and degrading the model's ability to discriminate the majority class [29]. Random oversampling augments the dataset by replicating minority class samples, effectively mitigating information loss caused by undersampling but potentially introducing redundancy

and increasing training burden [30]. In contrast, the SMOTE algorithm and its variants generate synthetic samples via k-nearest neighbor interpolation, balancing class ratios while avoiding redundancy from simple replication [31]. However, they may produce unrealistic samples near feature boundaries, reducing model generalization.

Defect sample augmentation is another common data-level solution. Transformation-based data augmentation methods [32], such as image cropping, flipping, rotation, and affine transformations, can generate diverse image variations, thereby improving model robustness. While these methods can alleviate global sample imbalance, they do not directly address intra-image class imbalance (e.g., coexistence of multiple defect types). Synthesis-based augmentation methods like Mixup generate new samples through image interpolation or concatenation, enhancing the distributional diversity and generalization capability of minority class samples to some extent [33]. However, controlling the authenticity of synthetic samples is challenging, and some generated images may deviate from the real distribution. To further enhance sample quality, methods based on generative models (e.g., GANs, diffusion models) are emerging. For instance, the Trans-GAN-Cla developed by Ma et al. is suitable for synthesizing drainage pipe defect images [34]. However, these methods heavily rely on high-quality training data and often struggle to generate sufficiently accurate samples in few-shot, multi-class tasks. Generated images also frequently suffer from issues like content blurring and edge distortion.

Algorithm-level strategies compensate for performance degradation caused by data imbalance by optimizing the training process or model architecture, primarily including label optimization and loss function optimization.

- (1) **Label Optimization Strategies.** Traditional methods often rely on a fixed Intersection over Union (IoU) threshold for positive/negative sample assignment. For example, Faster R-CNN [35] and the YOLO series use 0.5 as the positive sample threshold with an ignore region. Such static labeling methods are simple and straightforward but cannot adapt to targets of varying sizes and complexities, limiting model performance improvement. Consequently, researchers have introduced dynamic label mechanisms. For example, ATSS dynamically sets thresholds based on statistical distributions [36], and OTA even models label assignment as an optimal transport problem, balancing matching quality and training efficiency [37]. Correspondingly, the Noisy Anchor approach guides the model to focus on high-confidence samples by weakening the influence of low-quality samples, thereby enhancing learning capability for minority classes [38].
- (2) **Loss Function Strategies.** Focal Loss dynamically reduces the influence of easy-to-classify samples through a modulating factor, thereby focusing learning on hard samples. It performs well in handling defect class imbalance [39]. Building on this, the Class-Balanced Loss proposed by Cui et al. dynamically adjusts weights based on class frequency [40]. The Balanced Loss proposed by Tan et al. amplifies the loss for long-tail classes through weighting, significantly improving long-tail recognition [41]. Additionally, the AP-Loss ranking loss bridges the gap between training and evaluation by optimizing ranking metrics [42].

Table 3 summarizes the key strategies for addressing class imbalance at both data and algorithm levels.

In summary, research on class imbalance has gradually evolved from single-dimensional techniques toward multi-dimensional integration. Data augmentation, label mechanisms, loss design, and feature modeling now complement and synergistically optimize each other, forming a systematic strategy to address challenges such as difficult learning of minority class samples, sparse distribution, and high false detection rates. This provides more robust technical support for intelligent surface defect detection in complex industrial scenarios.

Table 3: Strategies for class-imbalanced defect detection.

Strategy Level	Approach	Representative Techniques	Applicable Scenarios
Data-level	Resampling	Undersampling, Oversampling, SMOTE	When sample distribution can be adjusted without introducing noise
	Data Augmentation	Transformation-based, Mixup, GAN-based generation	When synthetic samples can approximate real defect distribution
Algorithm-level	Label Optimization	ATSS, OTA, Noisy Anchor	When static IoU thresholds limit learning of minority classes
	Loss Function Optimization	Focal Loss, Class-Balanced Loss, AP-Loss	When reweighting sample contributions during training

3.4 Defect Detection Methods for Dynamic Scenarios

In industrial production, defect detection solutions must be flexibly adjusted according to the characteristics of the application scenario. For static or low-speed moving objects, such as PCBs, circuit boards, and glass panels, single-frame image detection is typically employed. This strategy offers three main advantages: First, complete information can be acquired in a single imaging instance, resulting in high processing efficiency. Second, it eliminates the need for object tracking, thereby reducing system complexity. Third, the model structure is simple and suitable for deployment on edge devices, such as portable inspection systems. However, in dynamic scenarios like high-speed conveyor belts, rotating components, or moving assembly lines, video stream detection technology demonstrates significant advantages [43]. This type of technology addresses four key issues: (1) Mitigating blurring and distortion caused by rapid object motion through motion compensation algorithms; (2) Utilizing temporal information to remove background interference such as metal reflections and dynamic shadows; (3) Employing multi-frame fusion to enhance the recognition capability for minute defects (e.g., micron-level cracks); (4) Enabling defect evolution tracking and monitoring through time series analysis. Typical applications include: full-perimeter detection of surface cracks on rotating equipment, identification of intermittent defects under vibration conditions, multi-angle defect modeling on complex curved surfaces (e.g., blades, gears, aerospace components), and quality tracking of long-running structures such as pipelines and bridges.

Traditional inspection methods and fixed-camera solutions often suffer from response lag and detection blind spots, making it difficult to meet the requirements for fine-grained, real-time monitoring. Traditional image processing methods, such as edge detection and threshold segmentation, rely on handcrafted features and parameter tuning, resulting in poor robustness and susceptibility to lighting variations and noise. To improve performance, some systems have attempted to introduce high-precision industrial cameras, structured light, or multi-spectral equipment. However, stringent environmental adaptability requirements (e.g., explosion-proof, high-temperature resistance) and on-site constraints often lead to high costs and deployment difficulties for such solutions [44].

The rapid development of deep learning has driven breakthroughs in detection capabilities for dynamic scenes. End-to-end detection networks have improved recognition accuracy and adaptability through

automatic feature extraction [45,46]. Liu et al. proposed a domain fluctuation suppression strategy based on deep ensemble learning, which mitigates dynamic data imbalance and cross-domain shifts, making it suitable for real pipeline inspection [47]. Tsung et al. developed a high-performance road health recognition system that combines simulation and measured data to achieve intelligent identification of various road damages, such as longitudinal/transverse cracks and potholes. Deployed on vehicle-mounted edge devices, it effectively improves inspection efficiency [48]. Guo et al. optimized YOLOv5 by incorporating an attention mechanism for complex road defect detection on mobile devices, successfully deploying the system on a portable platform [49].

Beyond video analysis methods, some research has also explored the integration of robotic and Unmanned Aerial Vehicle (UAV) platforms. However, Yang et al. point out that such systems are limited to providing only local positioning and are suited for small-scale scenes. To address this, they proposed the Det-ReconReg framework, which integrates UAV and machine learning technologies to achieve defect detection and localization in large-scale infrastructure, significantly enhancing the system's scalability [50].

To address motion blur and artifacts in dynamic images, Cui et al. proposed a two-stage weld seam detection method called TRDM. It first uses a lightweight network (LSN) for target localization, followed by an SRD network for defect recognition, effectively improving detection accuracy [51]. Liu et al. designed a belt damage detection model combining an attention mechanism with a Temporal Convolutional Network (TCN), processing complex backgrounds and dynamic shadow interference from both spatial and temporal dimensions, significantly improving the robustness of dynamic detection [52].

Regarding industrial implementation, Xu Hao et al. built a cable defect identification system based on YOLOv5, achieving efficient detection and tracking through adaptive parameters and multi-threshold strategies [53]. Yan Hexiang et al. combined YOLOv7-seg with DeepSORT to create a pipeline defect identification framework capable of simultaneous detection and tracking. The model's performance was optimized using a dual-training dataset to meet the demands of pipeline network inspection [54]. These solutions highlight the significant value of the "detection + tracking" technical approach for dynamic defect detection.

It is worth noting that despite considerable progress in existing research, most deep learning models are still based on static distribution assumptions. Real industrial data exhibits dynamic changing characteristics, which easily triggers domain shift problems and leads to performance degradation when models are deployed on production lines. The emergence of these issues reveals that current dynamic scenario defect detection still faces significant challenges in terms of model generalization and cross-domain adaptability, urgently requiring further research and solutions.

3.5 Cross-Scenario Detection Methods for Enhancing Model Generalization Capability

With the increasing demand for high robustness and adaptability in industrial intelligent inspection, traditional surface defect detection methods relying on single models or single features face significant performance bottlenecks in multi-scenario, multi-target, and multi-modal tasks. In recent years, researchers have conducted in-depth exploration around three main research directions: structure optimization-driven generalization enhancement methods, cross-domain adaptation and transfer learning mechanisms, and multi-task fusion with lightweight deployment optimization strategies. These efforts aim to mitigate performance degradation of models across different defect types, complex environments, and heterogeneous data.

- (1) **Structure Optimization-Driven Generalization Enhancement Methods.** Traditional surface defect detection methods often rely on rule-making based on low-level image features like thresholds,

edges, and textures, or combine shallow classifiers (e.g., SVM, K-NN) to achieve defect identification. Although these methods offer strong interpretability in specific scenarios, their generalization ability heavily depends on input image quality, lighting conditions, and background interference, making it difficult for them to adapt to complex and variable industrial environments [55]. In recent years, CNN-dominated deep learning methods have rapidly risen, improving models' ability to represent unstructured data through end-to-end feature extraction and detection strategies. Models like Faster R-CNN, the YOLO series, and SSD represent the mainstream development of two-stage and one-stage methods. Some architectures enhance robustness to scale variations and background interference by incorporating feature pyramids, attention mechanisms (CBAM, SE), and dilated convolutions.

However, most of these structurally improved models still focus on improving accuracy on a single dataset, overlooking robustness across different materials, equipment, and process conditions. To address this problem, researchers have proposed hybrid attention mechanisms (e.g., integrating Transformer self-attention with channel-spatial fusion attention) and dynamic structure regulation strategies, such as deformable convolutions [56] and conditional convolutions (CondConv) [57], thereby enhancing the model's adaptive capability to different defect geometric structures and texture distributions. Additionally, multi-stage architectures like cascade detection heads have been introduced to alleviate instability in detector regression under high IoU thresholds, further improving boundary precision and classification discriminability [58]. These methods provide fundamental support for the generalization of surface defect detection models in typical scenarios involving complex backgrounds, small targets, and strong occlusion.

- (2) **Application of Cross-Domain Adaptation and Transfer Mechanisms in Generalization.** In industrial applications, significant distribution differences often exist across different production batches, manufacturing equipment, and material surfaces. This leads to severe performance degradation—manifested as “domain shift” or “data drift”—when models trained on a source domain are deployed on a target domain. To address this issue, cross-domain learning has become an important means to enhance generalization in recent years, primarily including two categories: Domain Adaptation (DA) and Domain Generalization (DG). In DA methods, typical strategies like Domain-Adversarial Neural Networks (DANN) [59], Maximum Mean Discrepancy (MMD) [60], and adversarial style transfer networks (e.g., CycleGAN) [61] aim to adapt training to unlabeled target domains by minimizing the feature distribution discrepancy between source and target domains. In DG strategies, researchers enhance model performance on unseen target domains through multi-source data joint training, meta-learning optimization, or style perturbation augmentation.

Despite the good results achieved by these methods in classification and segmentation tasks, the specificity of surface defect detection tasks—such as sparse defect instances, large morphological variations, and difficulty in obtaining labels—makes it challenging to apply transfer methods directly. Consequently, some work has proposed self-training mechanisms based on pseudo-labels, dual-decoder and multi-discriminator structures to improve stability and consistency in unlabeled target domains. Other research has introduced domain attention modules to achieve dynamic weighting and adaptive aggregation in feature spaces. Furthermore, for specific defects (e.g., cracks, scratches) with extremely limited representation in the target domain, sample weighting mechanisms and hard example mining strategies are used to enhance the discriminability of feature representations. Overall, cross-domain adaptation mechanisms are gradually becoming a core direction for improving model generalization across different scenarios, particularly suitable for needs like batch updates, equipment migration, and remote deployment in defect detection contexts.

- (3) **Advances in Generalization through Multi-Task Fusion and Lightweight Deployment Strategies.** As task complexity continues to rise in smart manufacturing scenarios, detection models driven by

a single task can no longer meet the comprehensive demands for multi-dimensional objectives like defect classification, localization, segmentation, and counting. Therefore, Multi-Task Learning (MTL) has become an effective means to enhance model generalization and expressive power. By sharing a bottom-level encoder and introducing task-specific branches with optimized objective functions, researchers have constructed models with joint detection-segmentation functions, such as PAD-Net [62] and DefectNet [63]. These models demonstrate stronger robustness and contextual expression capabilities in surface defect scenarios like steel, PCB, and photovoltaic modules. Additionally, designs that integrate auxiliary tasks like classification, feature matching, and defect counting help mitigate overfitting caused by scarce samples and enhance generalized feature expression.

On the other hand, in industrial settings, inspection systems often face constraints in computational power and require deployment flexibility. Therefore, lightweight model design and knowledge distillation strategies have become another core direction for improving practical usability. Researchers have developed various lightweight detectors based on architectures like MobileNet [64] and ShuffleNet [65], incorporating self-attention mechanisms and residual connections for deployment on edge or mobile platforms. Simultaneously, knowledge distillation, as a “large-teaches-small” transfer paradigm, is widely used to transfer the feature distributions, prediction scores, and attention responses from large teacher models to smaller student models. Frameworks like adversarial distillation and multi-teacher collaborative distillation, which integrate multi-task and multi-perspective knowledge from teachers, significantly enhance the generalization capability and compression effectiveness of student models in complex defect detection tasks. For example, Zhou et al. implemented stable detection of various defect types based on knowledge distillation [66], while Hu et al. achieved high-performance PCB defect detection under lightweight conditions using aligned soft-target knowledge distillation [67].

In summary, multi-task fusion and lightweight deployment methods not only improve the synergy and performance of models across multiple sub-tasks in object detection but also greatly promote the deployability of surface defect detection algorithms, establishing an integrated foundation for generalization from perception algorithms to system integration.

4 Evaluation Metrics

In surface defect detection tasks, to comprehensively evaluate model performance across aspects such as detection accuracy, recall capability, overall performance, and adaptability to multi-scale targets, commonly used evaluation metrics primarily include Precision, Recall, Average Precision (AP), mean Average Precision (mAP), and variants of AP at specific thresholds or scales (e.g., AP@0.5, AP@0.75, AP_s, AP_m, AP_l). In industrial applications, the selection and interpretation of these metrics must align with specific production requirements, such as the relative costs of false positives vs. false negatives.

(1) Precision

Precision represents the proportion of results predicted by the model as defects that are actually defects. It is calculated as the ratio of correctly predicted positive samples (True Positives, TP) to all samples predicted as positive (the sum of True Positives and False Positives, FP).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

(2) Recall

Recall represents the proportion of actual defects that the model successfully detects.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

where FN (False Negative) is the number of actual defects missed by the model.

(3) Average Precision (AP)

AP measures the area under the Precision-Recall curve plotted at varying detection confidence thresholds. A higher value indicates better detection performance.

$$\text{AP} = \int_0^1 p(r) dr \quad (3)$$

where $P(r)$ represents the precision curve at different recall levels. In practice, it is typically computed using interpolation methods or discrete summation, such as the 11-point interpolation (VOC2007) or full-range integration (COCO).

(4) Mean Average Precision (mAP)

mAP is the average of AP values across all classes, serving as a core metric to evaluate the overall detection performance of a model.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

where N is the total number of defect categories, and AP_i is the Average Precision for the i -th defect class.

(5) Intersection over Union (IoU)

IoU measures the overlap between a predicted bounding box and the ground truth box, forming the basis for determining TP and FP.

$$\text{IoU} = \frac{A_{\text{pred}} \cap A_{\text{gt}}}{A_{\text{pred}} \cup A_{\text{gt}}} \quad (5)$$

where B_p is the area of the predicted bounding box and B_g is the area of the ground truth annotation box.

(6) AP at Different IoU Thresholds (e.g., AP@0.5, AP@0.75)

AP@0.5: AP calculated when $\text{IoU} \geq 0.5$, measuring “lenient” matching accuracy.

AP@0.75: AP calculated when $\text{IoU} \geq 0.75$, measuring “strict” matching accuracy.

AP@ [0.5:0.95]: In the COCO standard, the mean AP across 10 IoU thresholds from 0.5 to 0.95 in steps of 0.05, providing a more comprehensive reflection of performance.

(7) AP for Different Object Sizes (AP_s , AP_m , AP_l)

AP_s (Small): Average Precision for small objects (e.g., area $< 32 \times 32$ pixels).

AP_m (Medium): Average Precision for medium-sized objects (area between 32×32 and 96×96 pixels).

AP_l (Large): Average Precision for large objects (area $> 96 \times 96$ pixels).

These metrics are instrumental in analyzing a model’s detection capability for defects of varying scales.

(8) F1-Score (Comprehensive Metric)

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

The F1-score, which balances Precision and Recall, is often used as a comprehensive evaluation metric in practical engineering deployments.

(9) Multi-Object Tracking Metrics (e.g., MOTA, IDF1)

In dynamic scenarios requiring both surface defect detection and tracking, static image-based detection metrics (e.g., mAP, Recall) alone are insufficient for a comprehensive performance evaluation. Therefore, key evaluation metrics from Multi-Object Tracking (MOT) must be introduced. Among these, Multiple Object Tracking Accuracy (MOTA) and the Identity F1-score (IDF1) are the most commonly used.

MOTA comprehensively considers missed detections, false positives, and identity switches. A value closer to 1 indicates a more accurate tracking system. It emphasizes the quality of detection and tracking concerning the overall number of targets and is a core metric for evaluating the overall accuracy of a detection and tracking system.

$$\text{MOTA} = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t GT_t} \quad (7)$$

where FN_t is the number of undetected true targets in frame t ; FP_t is the number of falsely detected pseudo-targets in frame t ; $IDSW_t$ is the number of identity switches in frame t ; and GT_t represents the total number of ground truth targets in frame t .

IDF1 measures the consistency of identity preservation during tracking, i.e., the ability to continuously follow the same target while avoiding frequent ID switches and tracking losses. This metric is particularly suitable for continuous defect monitoring scenarios in dynamic industrial videos. The IDF1 formula is as follows:

$$\text{IDF1} = \frac{2 \cdot \text{IDTP}}{2 \cdot \text{IDTP} + \text{IDFP} + \text{IDFN}} \quad (8)$$

where IDTP (Identity True Positives) is the number of correctly identified and consistently tracked targets; IDFP (Identity False Positives) is the number of targets with incorrectly assigned identities; and IDFN (Identity False Negatives) is the number of targets not successfully tracked continuously.

In industrial quality inspection, the practical implications of evaluation metrics extend beyond numerical performance:

Precision: High precision corresponds to low false positive rates, which is critical for minimizing unnecessary rework and material waste. In applications such as automotive painting inspection, false positives can trigger costly manual verification processes.

Recall: High recall indicates low miss rates, which is paramount for safety-critical components. In aerospace or medical device manufacturing, a missed defect can lead to catastrophic failure, making recall the primary optimization target.

AP_s: This metric specifically evaluates performance on small defects (e.g., micro-cracks, pinholes), which are often the most challenging to detect but can have severe consequences if missed.

F1-score: In engineering practice, the F1-score provides a balanced measure for threshold selection, helping operators find the optimal trade-off between precision and recall based on the cost structure of the specific production line.

For dynamic scenarios requiring both detection and tracking, Multi-Object Tracking (MOT) metrics such as MOTA and IDF1 provide essential performance characterization. MOTA comprehensively accounts for misses, false positives, and identity switches, while IDF1 measures the consistency of identity preservation—critical for applications like defect evolution monitoring where tracking the same defect across frames is required.

5 Datasets

High-quality and diverse datasets are the cornerstone for advancing surface defect detection technology. However, the high confidentiality of industrial environments, the substantial cost of annotation, and the significant heterogeneity of defect morphology collectively result in a scarcity of publicly available data. To facilitate the systematic evaluation of existing methods under various challenges, this section first constructs a dataset classification framework, then proposes dataset selection guidelines, and finally summarizes and analyzes typical datasets.

5.1 Dataset Classification Framework

To systematically understand existing datasets, this paper introduces datasets from the following four dimensions, as detailed in [Table 4](#).

Table 4: Dataset classification framework.

Classification Dimension	Category	Typical Examples	Characteristics & Applications
Data Source	Synthetic	PCB Defect, DAGM	Computer-generated or rendered; precise labeling, strong controllability; but may suffer from domain shift from real-world data
	Real (Lab)	NEU-DET, AITEX	Collected in controlled environments; stable quality; suitable for algorithm benchmarking.
	Real (Industrial Field)	FICS-PCB	Collected in real industrial settings; contain noise and interference; high practical value but costly to annotate.
	Dedicated	MVTec AD, Road Damage Detection Challenge	Designed for specific competitions, usually with standard evaluation protocols.
Defect Type	Structural	Cracks, Fractures, Holes	Geometric shape damage; often requires boundary integrity analysis.
	Textural	Fabric stains, Metal spots	Local texture anomalies; high similarity to background texture.

(Continued)

Table 4 (continued)

Classification Dimension	Category	Typical Examples	Characteristics & Applications
Challenge Characteristics	Material	Oxidation, Corrosion, Delamination	Changes in material properties; may require multi-spectral information.
	Assembly	Missing components, Misalignment, Cold solder joints	Relationship anomalies in multi-component systems.
	Small Objects	NEU-DET, KolektorSDD	Defect pixel area < 1% of total image; features easily lost during downsampling.
	Low Contrast	Concrete Crack Dataset	Grayscale/color difference between defect and background < 30.
	Complex Background	PCB Defect, Fabric Texture Datasets	Background texture is complex; defects are easily obscured.
	Class Imbalance	Strip Steel Defects (some categories are rare)	Sample count for some defect classes < 5% of total samples.
	Multi-Scale	Road Damage Dataset	Defect sizes span multiple orders of magnitude.

5.2 Dataset Selection Methodology

In practical research, an appropriate dataset can be selected based on the specific defect challenges according to the following procedures.

5.2.1 Small Object Dataset Selection

Define ρ_{small} as the small object coefficient:

$$\rho_{\text{small}} = \frac{\text{Defect Pixel Area}}{\text{Total Image Area}} \times 100\% \quad (9)$$

When $\rho_{\text{small}} < 1\%$, the dataset is considered a small object defect dataset. For example, in NEU-DET, the crack-type defects have $\rho_{\text{small}} = 0.3 \pm 0.15\%$.

Application Example. Taking the NEU-DET dataset as an illustration, the crack-type defects have an average $\rho_{\text{small}} = 0.3 \pm 0.15\%$ (based on crack annotations covering approximately 60–120 pixels in 200×200 images). This places NEU-DET firmly in the small object category, explaining why methods employing multi-scale feature fusion or super-resolution techniques consistently achieve significant improvements on this benchmark. Researchers focusing on small object detection should prioritize datasets with $\rho_{\text{small}} < 1\%$ and evaluate performance using AP_s as the primary metric.

The key evaluation metric focuses on AP_s :

$$AP_s = \frac{1}{|S|} \sum_{c \in S} AP(c), S = \{c: \rho_{\text{small}}(c) < 1\%\} \quad (10)$$

where c is the defect category.

5.2.2 Complex Background Dataset Selection

(1) For datasets with repetitive but not strictly periodic background textures (e.g., fabric):

Background texture complexity τ can be determined based on statistical features of the Gray-Level Co-occurrence Matrix (GLCM).

For a grayscale image I , the GLCM P is defined as:

$$P(i, j) = \{(x, y) \mid I(x, y) = i, I(x + dx, y + dy) = j\} \quad (11)$$

where $\mathbf{d} = (dx, dy)$ is the displacement vector, typically $\mathbf{d} \in \{(1, 0), (0, 1), (1, 1), (1, -1)\}$.

First, extract the following features from the GLCM:

Contrast:

$$C = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 P(i, j) \quad (12)$$

Energy:

$$E = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j)^2 \quad (13)$$

Homogeneity:

$$H = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{P(i, j)}{1 + (i, j)^2} \quad (14)$$

Entropy:

$$S = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j) \log P(i, j) \quad (15)$$

Then, the texture complexity τ is defined as a weighted combination of Contrast, Entropy, and the product of inverted Homogeneity and Energy:

$$\tau = \alpha \cdot C + \beta \cdot S + \gamma \cdot \frac{(1 - H)}{(E + \epsilon)} \quad (16)$$

where $\alpha = 0.4$, $\beta = 0.4$, $\gamma = 0.2$ are weighting coefficients, and $\epsilon = 10^{-6}$ prevents division by zero. In practice, use the average value across the four displacement directions.

Thus, the determination of texture complexity can refer to the numerical ranges in [Table 5](#).

Table 5: Ranges of texture complexity τ .

Texture Category	τ Range
Simple Texture (Uniform Surface)	$\tau \approx 1-3$
Moderate Texture (Wood, Stone)	$\tau \approx 4-6$
Complex Texture (Fabric, Hair)	$\tau \approx 7-10$
Highly Complex Texture (Cluttered Background)	$\tau > 10$

(2) For datasets with regular structural textures (e.g., PCB):

Background complexity is judged using the Structural Regularity R . A larger R indicates greater complexity.

$$R = \exp\left(\frac{1}{S + \epsilon}\right) \quad (17)$$

Specifically, regular structures are periodic, so their autocorrelation function will exhibit sharp peaks. The peak sharpness S is:

$$S = \frac{A(0, 0) - A(T_x, T_y)}{T_x^2 + T_y^2} \quad (18)$$

where (T_x, T_y) is the estimated main period, and $A(\Delta x, \Delta y)$ is the autocorrelation function:

$$A(\Delta x, \Delta y) = \frac{\sum_x \sum_y I(x, y)I(x + \Delta x, y + \Delta y)}{\sum_x \sum_y I(x, y)^2} \quad (19)$$

(3) For natural background complexity (e.g., road damage datasets): The judgment can be based on the variety of distracting elements present in the dataset's background, such as shadows, stains, etc.

Examples, Textural Background (AITEX), Fabric texture complexity $\tau = 8.7$; Structural Background (PCB Defect Dataset), Circuit texture regularity $R = 0.92$; Natural Background (RDD2022): Distracting elements (shadows, stains, etc.) > 5 categories.

For evaluation, it is crucial to examine the change in False Positive Rate (FPR) under complex backgrounds:

$$\Delta\text{FPR} = \text{FPR}_{\text{complex}} - \text{FPR}_{\text{simple}} \quad (20)$$

Application Example. For the AITEX fabric defect dataset, the texture complexity τ computed via GLCM features is approximately 8.7, placing it in the “complex texture” category according to Table 5. This high complexity indicates that algorithms must incorporate effective background suppression mechanisms. In contrast, the PCB Defect Dataset exhibits structural regularity with $R = 0.92$, suggesting that methods leveraging periodic structural priors may be particularly effective. When selecting datasets for complex background research, practitioners should first compute τ or R to quantify background complexity, then evaluate model robustness by measuring ΔFPR (increase in false positive rate compared to simple backgrounds).

5.2.3 Class Imbalance Dataset Selection

(1) Long-Tail Distribution Dataset Selection

In the study of long-tail learning and class imbalance problems, accurately quantifying the degree of dataset imbalance is a prerequisite for assessing problem difficulty, designing algorithms, and comparing benchmarks. The academic community currently widely employs three fundamental and complementary statistical indicators: the Imbalance Ratio (IR), the Gini Coefficient, and Entropy-based metrics. They provide formalized measures from three dimensions: extreme differences, distribution inequality, and information uncertainty.

IR measures extreme deviation in data distribution. A larger IR indicates more severe extreme imbalance in the dataset. For example, IR = 100 means the most frequent class has 100 times the samples of the least frequent class [68].

$$IR = \frac{\max\{n_c\}}{\min\{n_c\}}, c \in \{1, 2, \dots, C\} \quad (21)$$

where n_c is the number of samples for class c .

Gini Coefficient [69] quantifies the overall inequality of the class distribution. Its discrete calculation formula based on class sample proportions is:

$$G = \frac{\sum_{i=1}^C \sum_{j=1}^C |n_i - n_j|}{2C \sum_{c=1}^C n_c} \quad (22)$$

where $G \in [0, 1]$, with 0 indicating perfect equality and 1 indicating maximum inequality (all samples belong to one class). The Gini coefficient provides intermediate distribution information that IR lacks; two datasets with the same IR may have significantly different G values.

Based on information theory, entropy is the core function for measuring the uncertainty of a random variable. The Class Entropy metric is:

$$H_{\text{class}} = - \sum_{c=1}^C p_c \log p_c, p_c = \frac{n_c}{N} \quad (23)$$

where $H_{\text{class}} \in [0, \log C]$, with smaller values indicating greater imbalance.

In addition to the fundamental IR, Gini coefficient, and entropy, the following extended metrics can reveal the structural characteristics of long-tail distributions more deeply.

Tail Index (α): For the class distribution sorted in descending order by sample count, the power-law distribution model is typically used for fitting analysis. Power-law distributions are classic models for describing long-tail phenomena in natural and social systems. Assuming the class distribution follows a power-law pattern: the class rank r and sample count n_r satisfy $P(\text{rank} = r) \propto r^{-\alpha}$. The tail index can be estimated via least squares regression:

$$\log n_{(r)} = \beta - \alpha \log r + \epsilon \quad (24)$$

where $n_{(r)}$ is the sample count of the r -th most frequent class, $\alpha > 0$ is the tail index (a larger value indicates a heavier tail). The tail index is intrinsically related to the Gini coefficient; a larger α typically corresponds to a higher Gini coefficient, but the former focuses more on the decay rate in the tail of the distribution.

To quantify the concentration of sample distribution in a dataset, two key proportional metrics are defined:

Head Class Proportion ($R_{\text{head}}(k)$), The cumulative sample proportion of the top k most frequent classes, reflecting the dominance of head classes.

$$R_{\text{head}}(k) = \frac{\sum_{c=1}^k n(c)}{N}, k \text{ is the number of head classes} \quad (25)$$

This metric can be seen as a specific supplement to the entropy-based metric: when $R_{\text{head}}(k)$ is close to 1, the information entropy will approach 0, indicating a highly concentrated distribution.

Evaluation Metrics can draw on the framework of the LVIS dataset [70], grouping all classes into three sets based on training sample count:

$$AP_r = \frac{1}{|C_r|} \sum_{c \in C_r} AP_c, r \in \{\text{frequent, common, rare}\} \quad (26)$$

where Frequent classes have >100 samples (or are in the top 33%), Common classes have $10 < \text{samples} \leq 100$ (middle 33%), and Rare classes have ≤ 10 samples (bottom 33%). This grouped evaluation method corresponds to the global perspective of the Gini coefficient: the Gini coefficient quantifies the inequality in the training data distribution, while the grouped AP quantifies the manifestation of this inequality in model performance.

Further, define the average precision for head and tail classes:

$$AP_{\text{head}} = \frac{1}{|C_{\text{head}}|} \sum_{c \in C_{\text{head}}} AP_c, AP_{\text{tail}} = \frac{1}{|C_{\text{tail}}|} \sum_{c \in C_{\text{tail}}} AP_c \quad (27)$$

To more sensitively reflect tail performance and incentivize models to achieve a balance between head and tail performance, the Harmonic Mean AP (HM-AP) is introduced:

$$\text{HM-AP} = \frac{2 \cdot AP_{\text{head}} \cdot AP_{\text{tail}}}{AP_{\text{head}} + AP_{\text{tail}}} \quad (28)$$

The harmonic mean is more sensitive to smaller values. When tail class performance is poor, HM-AP will decrease significantly, thereby encouraging models to pay more attention to tail performance during optimization. This metric embodies the fairness principle from information theory: it not only focuses on overall performance but also emphasizes the balance of performance distribution.

Taking the metal surface defect detection dataset GC10-DET as an example, it contains 10 defect classes with typical long-tail distribution characteristics (Table 6). Among the basic metrics, $IR = 12.3$, $Gini \approx 0.68$, and normalized entropy ≈ 0.72 . Among the extended metrics, the top 2 classes (20% of classes) account for approximately 50% of the total samples, and the Pareto ratio $R_{\text{Pareto}} \approx 0.3$. On this dataset, a typical model might achieve $AP_{\text{head}} = 0.85$, but $AP_{\text{tail}} = 0.32$, resulting in $\text{HM-AP} = 0.46$. This analysis shows that basic distribution metrics (high Gini, low entropy) predict significant performance imbalance, and the validation metrics (low HM-AP, tail AP much lower than head AP) confirm this prediction. A complete multi-level indicator analysis provides a solid quantitative foundation for the design and evaluation of long-tail learning algorithms.

Table 7 compares the distribution characteristics of two classic long-tail benchmarks and two road damage datasets. From the tail index α , LVIS has $\alpha \approx 1.0$, conforming to the common Zipf distribution in the real world, which is a typical natural long-tail. In contrast, RDD2022 and UAPD have $\alpha > 2.0$, indicating a steeper tail decay and extreme scarcity of samples in tail classes. This difference not only reflects the

fundamental distinction in distribution shape between artificially constructed long-tails and natural long-tails but also suggests that on RDD2022 and UAPD, models are more prone to overfitting rare tail defect classes (e.g., “potholes”, “folds”), requiring more refined long-tail learning strategies.

Table 6: GC10-DET dataset example.

Class	Sample Count	Class Group
Scratch	800	Head
Punching	650	Head
Weld line	350	Middle
Water spot	320	Middle
Oil spot	280	Middle
Silicon	200	Middle
Oxidation	180	Tail
Rolled pit	120	Tail
Burr	85	Tail
Fold	65	Tail

Table 7: Long-tail characteristics of typical datasets.

Dataset	Classes	IR	Gini	Entropy	Tail Index (α)	Top 20% Proportion	Bottom 40% Proportion
COCO-LT	80	1153	0.812	3.62	1.98	62.3%	5.9%
LVIS	1203	47,024	0.891	6.15	1.02	68.3%	3.9%
RDD2022	4	5.49	0.199	1.79	2.48	41.9%	42%
UAPD	6	14.7	0.335	1.98	2.55	65.9%	8.7%

(2) Multi-Scale Dataset Selection Criteria

In visual recognition tasks, imbalanced scale distribution of objects is a key factor affecting model performance. This phenomenon often intertwines with class long-tail distribution, jointly exacerbating the learning difficulty for models.

Following the COCO dataset evaluation protocol, object scale is divided based on its pixel area, defined as:

$$\text{Object } i \in \begin{cases} \text{Small,} & \text{if } \sqrt{w_i h_i} < 32 \\ \text{Medium,} & \text{if } 32 \leq \sqrt{w_i h_i} \leq 96 \\ \text{Large,} & \text{if } \sqrt{w_i h_i} > 96 \end{cases} \quad (29)$$

where w_i and h_i are the width and height of object i (in pixels), respectively. This division is based on the absolute size of the object in the image, independent of image resolution. Define the proportion of each scale group in the dataset:

$$P_{\text{small}} = \frac{N_{\text{small}}}{N}, P_{\text{medium}} = \frac{N_{\text{medium}}}{N}, P_{\text{large}} = \frac{N_{\text{large}}}{N} \quad (30)$$

where N is the total number of object instances in the dataset, and $N_{\text{small}}, N_{\text{medium}}, N_{\text{large}}$ are the counts of small, medium, and large objects, respectively.

To quantify the balance of the dataset's scale distribution, a Scale Imbalance Index (SII) is defined, drawing inspiration from the concept of variance. For the 3-group case, SII is defined as:

$$\text{SII} = \sqrt{\frac{(P_{\text{small}} - \bar{P})^2 + (P_{\text{medium}} - \bar{P})^2 + (P_{\text{large}} - \bar{P})^2}{3}} \quad (31)$$

where $\bar{P} = 1/3$ is the expected proportion of each group under ideal equilibrium. The range of SII is $[0, \sqrt{2/3}]$, with larger values indicating a more imbalanced scale distribution.

For the more general case of M scale groups:

$$\text{SII}_M = \sqrt{\frac{1}{M} \sum_{m=1}^M \left(P_m - \frac{1}{M}\right)^2} \quad (32)$$

This index shares similar mathematical properties with the Gini coefficient but is specifically designed to quantify the scale dimension, forming a multi-dimensional characterization of data characteristics alongside class-level imbalance metrics.

Evaluation Metrics can introduce the Scale Performance Gap (SPG) metric, defined as the normalized difference in a specific evaluation metric between large-scale and small-scale objects.

$$\text{SPG}_X = \frac{X_{\text{large}} - X_{\text{small}}}{X_{\text{large}}}, X \in \{\text{AP}, \text{Recall}, \text{F1}\} \quad (33)$$

This metric measures, from an algorithmic fairness perspective, whether the model pays equal attention to object instances of different scales. $\text{SPG} \rightarrow 0$ represents ideal scale invariance, while a significant $\text{SPG} > 0$ exposes the inherent limitations of current detection architectures in handling small-scale objects, providing a quantitative basis for subsequent scale-aware network design.

5.3 Dataset Quality and Industrial Applicability Discussion

Beyond quantitative characteristics, the quality of dataset annotations and the alignment with real industrial conditions significantly impact the validity of research findings. This section discusses key considerations for dataset quality assessment.

Annotation Quality. Datasets vary substantially in annotation granularity and consistency. Image-level labels (e.g., classification datasets like ELPV) are suitable for defect classification tasks but insufficient for localization or segmentation. Bounding box annotations (e.g., RDD2022, GC10-DET) enable object detection but may suffer from subjective variations in box boundaries, particularly for irregular defects like cracks. Pixel-level masks (e.g., MVTec AD, CrackForest) provide the highest annotation detail but are costly to produce and may exhibit inconsistencies across annotators. Researchers should consider these factors when selecting datasets for specific tasks; for instance, segmentation models trained on bounding box annotations may achieve inferior boundary precision.

Noise and Diversity. Real industrial environments introduce numerous confounding factors that are often absent in laboratory-collected datasets. Important diversity dimensions include:

Illumination variations: Datasets like DTU-Drone and RDD2022 capture images under diverse lighting conditions, enabling evaluation of illumination robustness.

Background complexity: While datasets like NEU-DET have relatively uniform backgrounds, AITEX and PCB datasets incorporate complex textural backgrounds that challenge detection algorithms.

Sensor variability: Datasets combining visible light, thermal (PTID), and electroluminescence (ELPV) imaging enable multimodal algorithm development.

Industrial Applicability Assessment. Based on the above considerations, Table 8 provides a qualitative assessment of dataset suitability for different industrial application contexts.

Table 8: Qualitative assessment of dataset industrial applicability.

Dataset Category	Representative Datasets	Strengths	Limitations for Industrial Deployment
Laboratory/Controlled	NEU-DET, AITEX	High annotation quality; Controlled conditions for benchmarking	Limited background diversity; May overestimate real-world performance
Industrial Field	FICS-PCB, RDD2022	Realistic noise and variations; High practical relevance	Annotation may be less precise; Class distribution often imbalanced
Synthetic	DAGM, DeepPCB	Perfect annotations; Controllable defect characteristics	Domain gap from real data; May not capture realistic variations
Dedicated Benchmarks	MVTec AD, KolektorSDD	Standardized evaluation protocols; Widely used for comparison	Specific to certain defect types; May not represent full production variability

5.4 Dataset Summaries

5.4.1 PCB Surface Defect Datasets

As the core component of electronic devices, the surface defect detection of PCB is crucial for ensuring the reliability of electronic products. PCB defects typically include solder joint anomalies, circuit shorts, missing holes, etc., characterized by tiny targets, complex background textures, and low contrast between defects and normal areas. Table 9 summarizes mainstream PCB defect detection datasets and provides a comparative analysis across dimensions such as data scale, defect types, annotation quality, and application scenarios.

Table 9: Overview of PCB surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
PKU-Market-PCB [71]	Synthetic	Missing hole, Mouse bite, Open circuit, Short, Spur, Spurious copper	1386 images. Suitable for detection, classification, and registration tasks.	http://robotics.pkusz.edu.cn/resources/dataset/
PCB Defect-Augmented [72]	Synthetic	Missing hole, Mouse bite, Open circuit, Short, Spur, Spurious copper	10,668 images. High resolution. Suitable for segmentation tasks.	https://www.dropbox.com/s/h0f39nyotddibsb/VOC_PCB.zip?dl=0

(Continued)

Table 9 (continued)

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
FICS-PCB [73]	Real	Components (IC, capacitor, resistor, inductor, transistor, diode) across four board colors (green, red, blue, black)	9912 images. Ideal for researching color-invariant feature extraction, component-level defect detection, and cross-color generalization capability validation.	https://www.kaggle.com/datasets/dhruvmak/fics-pcb
PCB DSLR [74]	Real	IC (Integrated Circuit)	748 images. High resolution. Used for chip-level fine-grained inspection and multi-task learning (detection + segmentation + recognition).	http://www.caa.tuwien.ac.at/cvl/research/cvl-databases/pcb-dslr-dataset/
PCB-METAL [75]	Real	IC (5844 instances), Capacitor (3175), Resistor (2670), Inductor (542)	984 images. Exhibits a long-tail distribution. Suitable for component detection and counting tasks.	Collected for academic research.
DEEP PCB	Synthetic	Pin-hole, Mousebite, Open, Short, Spur, Spurious copper	1500 images, 640 × 640 resolution	https://github.com/tangsanli5201/DeepPCB
Micro-PCB	Dedicated	Micro-PCBs	8125 images. Applicable for research on multi-view fusion, rotation invariance, and small object detection.	https://www.kaggle.com/datasets/frettapper/micropcb-images
PCB_ComponentDetection	Dedicated	Components: Capacitor, Resistor, Transformer, Line Filter, etc.	1410 PCB images. Can be used to detect missing components and component orientation errors.	https://www.kaggle.com/datasets/animeshkumarnayak/pcb-fault-detection
PCB	Synthetic	Missing hole, Mouse bite, Open circuit, Short, Spur, Spurious copper	693 PCB images. Contains annotations in both COCO and VOC formats.	https://aistudio.baidu.com/aistudio/datasetdetail/127210

5.4.2 Photovoltaic Cell Surface Defect Datasets

The detection of surface defects on photovoltaic cells is crucial for ensuring the power generation efficiency and long-term reliability of photovoltaic systems. Due to characteristics such as regular texture, fixed structure, and low contrast defects, detection algorithms for photovoltaic cells heavily rely on high-quality datasets. This section collects and organizes nine publicly available datasets, as detailed in Table 10.

5.4.3 Transmission Line Insulator Surface Defect Datasets

Surface defect detection for high-voltage transmission line insulators is a critical component of intelligent operation and maintenance in power systems. Influenced by complex field environments, variable

inspection conditions, and diverse defect morphologies, visual inspection of insulator defects faces unique challenges, placing special requirements on dataset construction. The details of relevant datasets are presented in the Table 11.

Table 10: Overview of photovoltaic cell surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
PTID [76]	Real	Thermal anomaly regions	3027 images, thermal imaging. Contains continuous defect evolution sequences. Suitable for semantic segmentation, multimodal fusion, and fine-grained defect localization.	http://vrai.dii.univpm.it/content/photovoltaic-thermal-images-dataset
ELPV [77]	Real	Micro-cracks, Degradation, Short circuits, Solder failures	2624 images, Electroluminescence (EL) imaging, low resolution. Suitable for defect classification and anomaly detection.	https://github.com/zae-bayern/elpv-dataset
ELDDS 1400c5 [78]	Real	Black spots, Micro-cracks, Cracks, Short-circuited cells, Short-circuited strings	1400 images, EL imaging, industrial-grade acquisition equipment. Suitable for object detection and defect classification.	https://universe.roboflow.com/fma04-fayoum-edu-eg/eldds1400c5-dataset
BELI [79]	Real	Typical defects in polycrystalline/monocrystalline cells	593 images, EL imaging. Suitable for classification benchmark testing	https://github.com/TheMakiran/BenchmarkELimages
PVEL-AD [80]	Real	12 classes: Cracks, Broken grid, Black core, Scratches, Fragments, etc.	36,543 images, EL imaging, high resolution. Suitable for object detection, anomaly detection, and classification.	http://aihebut.com/col.jsp?id=118
PVMD	Real	Broken, Bright spots, Black edges, Scratches, Uncharged	1108 images, EL imaging. Suitable for defect classification and anomaly detection.	https://github.com/CCNUZFW/PV-Multi-Defect
ISM [81]	Real	11 types of thermal anomalies + normal class	20,000 images, thermal imaging, very low resolution. Suitable for infrared anomaly detection and classification.	https://github.com/RaptorMaps/InfraredSolarModules
CS [82]	Real	Cracks	1837 images, EL imaging. Suitable for crack segmentation and semantic segmentation.	https://datahub.duramat.org/dataset/crack-segmentation

Table 11: Overview of transmission line insulator surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
CPLID [83]	Real	Normal, Self-detonated, Damaged	2480 images. Includes complex backgrounds (sky, mountains, vegetation). Suitable for object detection and classification.	https://github.com/InsulatorData/InsulatorDataSet
ID	Mixed	Normal, Contaminated, Damaged	1500 images. Includes some video sequences. Suitable for instance segmentation and object detection.	https://ev.po.opole.pl/dataset/

(Continued)

Table 11 (continued)

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
OPDL	Mixed	Normal, Flashover marks, Cracks, Contamination	3200 images. High resolution. Includes infrared and visible light images, multi-scale objects. Suitable for multimodal detection and small object detection.	http://www.dee.eng.ufba.br/dslab/index.php/opdldataset
UPID	Real	Normal, Damaged, Missing, Foreign objects	593 images, EL imaging. Suitable for classification benchmark testing	https://github.com/heitorcfelix/public-insulatordatasets
SFID	Synthetic	Normal, Contamination, Ice accretion, Bird nests	36,543 images, EL imaging, high resolution. Suitable for object detection, anomaly detection, and classification.	https://github.com/zhangzhengde0225/FINet

5.4.4 Wafer Surface Defect Datasets

Wafer defect inspection is a critical step in semiconductor manufacturing. Its defect patterns present unique challenges such as multiple categories, imbalanced distribution, and spatial correlation. Wafer defect data is typically presented in the form of Wafer Maps, where each die is labeled as normal or with a specific defect pattern. The details in the [Table 12](#).

Table 12: Overview of wafer surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
WM-811K [84]	Real	Center, Donut, Edge-Ring, Edge-Loc, Loc, Near-full, None, Random, Scratch	811,457 images. Extreme long-tail distribution. Suitable for classification, anomaly detection, and few-shot learning.	http://mirilab.org/dataSet/public/
MixedWM38 [85]	Mixed	Center, Donut, Edge-Ring, Edge-Loc, Loc, Near-full, Random, Scratch	38,015 images. Class-balanced. Suitable for defect pattern classification.	https://github.com/Junliangwangdhu/WWaferMap
MBWBM	Synthetic	Multi-level defect severity (Levels 0-9)	50,000 images. Supports regression and graded classification tasks.	https://ieee-dataport.org/documents/synthesized-multi-gfa-multi-bin-wafer-bin-map-dataset

5.4.5 Mobile Phone Screen Surface Defect Datasets

The field of mobile phone screen defect detection lacks publicly available benchmark datasets, which is a significant factor constraining its development. The details of relevant datasets are presented in the [Table 13](#).

5.4.6 Wind Turbine Blade Surface Defect Datasets

Wind turbine blade defect detection is a classic challenge in industrial vision, with its core difficulties stemming from the unique application scenarios and defect characteristics. The main challenges are: Blade surface defects (e.g., micro-cracks, leading-edge erosion) occupy an extremely small proportion in drone aerial images captured from a distance, constituting a typical small object detection problem; the inspection background consists of dynamically changing skies and clouds, leading to complex background interference and drastic illumination changes; defect samples are extremely scarce in actual operations, resulting in

extreme class imbalance in datasets; furthermore, algorithms must overcome interference from image blur, low resolution, and adverse weather conditions like rain and fog, while meeting the real-time requirements of drone online inspection. The details of relevant datasets are presented in the Table 14.

Table 13: Overview of mobile phone screen surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
MSD [86]	Mixed	Oil stains, Scratches, Spots	1200 images, high resolution. Suitable for semantic segmentation and defect classification.	https://github.com/jianzhang96/MsD
Glass Cover Defect Dataset	Real	Point-shaped, Line-shaped, Block-shaped	8437 images, class imbalance. Suitable for object detection and defect classification.	https://github.com/lohnushwilVGLassCoverDefectDataset

Table 14: Overview of wind turbine blade surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
DTU-Drone	Real	Cracks, Gel coat damage, Edge erosion, Lightning arrester defects, etc.	Aerial images. Contains illumination variations and complex backgrounds. Suitable for object detection.	https://data.mendeley.com/datasets/hd96prn3nc/1
YOLO Format	Mixed	Cracks, Coating damage, Edge erosion, Lightning arrester defects, Dirt, etc.	13,000 images. Annotated in YOLO format. Suitable for object detection.	https://www.kaggle.com/datasets/ajifoster3/yolo-annotated-wind-turbines586x371
Wind Turbine	Real	Cracks, Coating damage, Lightning strike damage	2240 images. Suitable for defect detection.	https://universe.roboflow.com/sakethh-xv7uo/windturbine-new
Blade30	Real	Cracks, Pinholes, Edge erosion	1302 images. Suitable for blade stitching, segmentation, defect detection and classification, deduplication, contamination detection and classification, and more.	https://github.com/cong-yang/Blade30
Small-WTB Thermal	Real	Cracks, Pinholes, Edge erosion	1000 images. Thermal imaging. Focused on fault detection and classification.	https://github.com/MoShekaramiz/SmallWTB-Thermal1

5.4.7 Road Surface Defect Datasets

A robust road disease detection system must accurately identify distress targets with weak features and variable morphology in complex, dynamic environments with high noise and multiple interferences, while overcoming the impact of data imbalance, imaging variations, and annotation uncertainty. The details of relevant datasets are presented in the [Table 15](#).

Table 15: Overview of road disease datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
CrackForest	Real	Cracks (longitudinal, transverse, network)	118 images. Classic crack segmentation benchmark. Relatively clean background, focused on crack morphology research.	https://github.com/cuilimeng/CrackForest-dataset
CrackTree 200	Real	Cracks	Used for segmentation.	https://github.com/cuilimeng/CrackTree200
DeepCrack	Real	Cracks	Large data volume, high resolution.	https://github.com/yhlleo/DeepCrack
RDD2022 [87]	Real	Longitudinal cracks, ransverse cracks, Alligator cracks, Potholes	47,000+ images. Covers roads from multiple countries (China, Japan, India, etc.). Suitable for detection.	https://www.kaggle.com/datasets/alaagaberh/rdd2022
UAV-RDD2023 [88]	Real	Longitudinal cracks, ransverse cracks, Alligator cracks, Potholes, Oblique cracks, Repair	17,000+ images. Drone aerial imagery, covering 6 countries, highly diverse. Suitable for detection.	https://zenodo.org/records/8429208
UAPD [89]	Real	Longitudinal cracks, ransverse cracks, Alligator cracks, Potholes, Oblique cracks, Repair	3151 images. Large scale variations, class imbalance, noticeable lighting/shadows, small targets, complex background. Can be used for detection and segmentation.	https://github.com/Shuijing725/UAPD
AigleRN	Real	Cracks, Potholes	38 images. Primarily used for performance testing and visual comparison of algorithms on infrared channels.	https://www.irit.fr/~Sylvie.Chambon/AigleRN.html

(Continued)

Table 15 (continued)

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
GAPs384	Real	Cracks, Potholes	196 images. Composed of Google Street View and aerial images, unique perspective (top-down), large scale variations.	https://github.com/Shuijing725/GAPs384
SDNET2018	Real	Cracks (P, M, S three levels)	56,000 images. Ideal for researching long-tail distribution, few-shot defect detection, or training data-hungry deep learning models.	https://digitalcommons.usu.edu/all_datasets/48/

5.4.8 Metallic Surface Defect Datasets

Metallic surface defect detection algorithms need to possess strong capabilities for extracting minute features, discerning complex textures, and demonstrating high robustness to operating conditions, all under extremely low signal-to-noise ratios and severe data imbalance. Therefore, datasets like KolektorSDD (containing micro-cracks) and Severstal (with extreme imbalance characteristics) have become key benchmarks for this challenge. The details of relevant datasets are presented in the [Table 16](#).

Table 16: Overview of metallic surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
NEU	Real	Crazing, Inclusion, Patches, Pitted, Surface, Rolled-in Scale, cratches	1800 images. Balanced distribution. A classic classification benchmark, also usable for segmentation.	https://www.kaggle.com/datasets/fantacher/neu-metal-surface-defects-data
GC10-DET [90]	Real	Punching Hole, Welding Line, Crescent Gap, Water Spot, Oil Spo, Silk Spot, Inclusion, Rolled Pit, Crease, Waist Folding	3570 images. Defect detection and localization in complex industrial scenes.	https://github.com/lvxiaoming2019/GC10-DET-Metallic-Surface-Defect-Datasets
Kolektor SDD/SDD2	Real	Crack, None	399/3347 images. Suitable for fine-grained detection of tiny, low-contrast defects.	https://www.vicos.si/Resources/KolektorSDD
Magnetic Tile	Real	Crack, Fray, Break, Uneven	1344 images. Suitable for detection and classification of surface defect.	https://github.com/abin24/Magnetic-tile-defect-datasets

(Continued)

Table 16 (continued)

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
Severstal Steel	Real	Rolled-in Scale, Inclusion, Crazing, Uneven	12,568 images. Defects occupy a very small proportion of the entire image (long strip-shaped), and the categories are extremely imbalanced.	https://www.kaggle.com/c/severstal-steel-defect-detection/data

5.4.9 Fabric Surface Defect Datasets

The difficulty in fabric defect detection concentrates on the strong interference and adversarial relationship between complex backgrounds and subtle anomalies. Due to the regular or randomly repeating texture structure of fabric itself, subtle defects such as broken ends, holes, or stains are visually highly similar to normal textures, resulting in an extremely low signal-to-noise ratio. Simultaneously, defect morphology, scale, and location are highly random, while defect samples are extremely scarce in actual production, leading to severe data distribution imbalance. Additionally, imaging condition interferences like lighting changes, fabric wrinkles, and deformations further increase the difficulty of stable and accurate detection, requiring algorithms to possess powerful texture representation separation capabilities and few-shot generalization ability. The details of relevant datasets are presented in the [Table 17](#).

Table 17: Overview of fabric surface defect datasets.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
DTU-Drone	Real	Broken_end, Broken_pick, Hole, Kno, Netting_multiple, Slack_end, Stain	245 images. High resolution. Suitable for multi-scale detection and few-shot learning.	https://www.aitex.es/afid/
AITEX	Mixed	Thread, Dirt, Wrinkle, etc.	1000+ images. Complex background, small targets.	https://lmb.informatik.uni-freiburg.de/resources/datasets/tilda.en.html
TILDA	Real	Thread, Dirt, Wrinkle, etc.	1647 images. Sample imbalance.	https://arxiv.org/abs/2207.10414

5.4.10 General Industrial Surface Defect Datasets

In the field of industrial defect detection, public datasets mainly exhibit four characteristics and difficulties. The datasets mostly consist of high-quality images from real scenes and strictly adhere to the setting of training only with normal samples, making them suitable for developing and evaluating unsupervised anomaly detection algorithms. The main detection challenges lie in three aspects: first, tiny defects are difficult to identify against complex backgrounds; second, lighting and angle variations cause difficulties in model generalization; third, logical anomalies (e.g., incorrect assembly, missing parts) are hard to recognize through appearance features, requiring algorithms to possess reasoning capabilities.

These characteristics collectively determine the core problems that current algorithms need to solve: few-shot learning, strong generalization, and reasoning ability. In actual research papers, MVTec AD is now used as the core evaluation dataset in the vast majority of cases, supplemented by other datasets (e.g., the newly released MVTec LOCO, MVTec 3D-AD by MVTec, or real defect datasets like BTAD, VisA, etc.) for additional validation. The details of relevant datasets are presented in the [Table 18](#).

Table 18: Overview of general industrial surface defect dataset.

Dataset	Source	Defect Types	Characteristics Analysis	Download Link
DAGM 2007	Synthetic	10 classes of artificial grayscale anomalies	10,000+ images. Regular texture, clear contrast between background and defect. Used for algorithm prototyping and teaching.	https://conferences.mpi-inf.mpg.de/dagm/2007/competition.html
MVTec AD [91]	Real	15 classes (e.g., scratches, dents, contamination)	5350 images. The “gold standard” benchmark for 2D anomaly detection. Training set contains only normal samples. Used for unsupervised learning.	https://www.mvtec.com/company/research/datasets/mvtec-ad
MVTec 3D-AD [92]	Real	10 classes (e.g., scratches, dents, contamination)	4000+ images. High resolution. Benchmark for 3D anomaly detection.	https://www.mvtec.com/company/research/datasets/mvtec-3d-ad
MVTec AD 2	Real	8 classes, more challenging scenes	8000+ images. Focuses on complex, challenging scenes.	https://www.mvtec.com/company/research/datasets/mvtec-ad-2
MVTec LOCO AD	Real	5 classes of structural anomalies (scratches, dents) and logical anomalies (object misplacement, missing)	3644 images. Specialized dataset for logical anomaly detection.	https://www.mvtec.com/company/research/datasets/mvtec-loco
BTAD	Real	3 types of industrial defects	2540 images. High resolution. Used for anomaly detection.	https://datasetninja.com/btad
VisA	Real	12 classes of surface and structural defects	10,821 images. Multi-category general-purpose dataset.	https://opendatalab.org.cn/OpenDataLab/VisA

5.4.11 Summary of Dataset Challenge Analysis

Although current public datasets provide important benchmarks for research in this field, their composition and settings still have certain limitations, creating a gap with the complex practical demands of industrial inspection. Existing datasets mainly exhibit the following characteristics: Samples are mostly static single-frame images, lacking temporal information on defect formation and propagation during the production process; the learning paradigm heavily relies on fully supervised settings, typically providing complete pixel-level annotations, which does not fully align with the reality of scarce defect samples and high labeling costs in industrial settings; data modalities and scenarios are relatively singular, with most datasets targeting specific materials (e.g., steel, fabric) and collected under controlled conditions, lacking cross-modal benchmarks that integrate multi-source signals (e.g., visible light, thermal imaging, 3D profiles) and systematic coverage of complex backgrounds and variable operating conditions. Therefore, future research needs to focus on constructing video datasets containing temporal information, exploring benchmarks for weakly supervised and unsupervised learning with very few or even zero labeled samples, and establishing comprehensive datasets covering multimodal information and complex scenarios, in order to promote algorithm validation and improvement under conditions closer to real industrial environments.

6 Conclusion and Prospect

6.1 Conclusion

This paper provides a systematic review of five core challenges in machine vision-based surface defect detection: complex background interference, small object detection, class imbalance, dynamic scene modeling, and cross-scenario generalization. By combining the technical paths of traditional image processing, CNN-based deep learning methods, and the emerging Transformer architecture, the advantages and limitations of various methods in dealing with specific challenges are revealed. Research shows that attention mechanism, multi-scale feature fusion, context modeling, dynamic label assignment and loss function design, temporal reasoning and domain adaptation technologies have become key means to improve the robustness, accuracy and adaptability of detection models in complex industrial environments.

Concurrently, this paper comprehensively summarizes and analyzes the characteristics of publicly available datasets covering a wide range of industrial scenarios. While these datasets have propelled algorithmic research, they also expose limitations misaligned with real industrial needs, such as dominance by static single-frame images, heavy reliance on full supervision, and singularity in modalities and scenarios. Constructing benchmark datasets that better reflect reality (e.g., containing temporal information, supporting weak/unsupervised learning, fusing multimodal data) will be a crucial foundation for advancing technology deployment.

Overall, although deep learning methods have made significant progress, the current technological system still exhibits a fragmented, “challenge-driven, scenario-specific” characteristic. A “generalized” defect visual inspection system capable of stable operation, efficient generalization, and easy deployment in variable and demanding real industrial environments still faces dual theoretical and engineering challenges.

6.2 Prospect

Looking ahead, surface defect detection technology will evolve towards greater intelligence, generalizability, lightweight design, and trustworthiness. In conjunction with advancements in cutting-edge technologies like large models, the following directions warrant in-depth exploration:

Revolution of Visual Large Models and Foundation Models: The perceptual capabilities of current mainstream models based on CNNs and ViTs remain constrained by specific tasks and datasets. The emergence

of Vision Foundation Models and Vision-Language Models (VLMs) brings the potential for a paradigm shift. Their potential is evident in: 1) Few-shot generalization ability. Leveraging the powerful visual-language alignment and in-context learning capabilities of large models, it may be possible to identify unseen defect types using only a few examples or textual descriptions (e.g., “micron-level linear scratch”), alleviating detection challenges for few-shot and new defect categories. 2) Cognitive ability for complex scenes. Large models possess stronger scene understanding, relational reasoning, and commonsense knowledge. This aids in distinguishing real defects from background pseudo-textures and understanding “logical anomalies” (e.g., incorrect/missing parts), enabling semantic-level defect diagnosis closer to that of human experts. 3) Unified multimodal representation and generative augmentation. Generative large models like diffusion models can synthesize rare defect samples with higher quality or generate realistic anomalous samples based on normal samples to address data imbalance. Simultaneously, large models can serve as unified encoders to fuse multimodal information (e.g., visible light, infrared, X-ray, 3D point clouds), enhancing the comprehensiveness and reliability of defect detection.

Integrated “Detection-Tracking-Diagnosis-Decision” Intelligent Systems: Future systems should evolve beyond mere defect locators into closed-loop intelligent agents integrating real-time perception, tracking, evolution analysis, and predictive maintenance decision-making. This requires deep integration of temporal modeling, knowledge graphs, physical mechanism models, and reinforcement learning to achieve the leap from “seeing defects” to “understanding defect evolution, predicting failure risks, and guiding maintenance strategies.”

Lightweight and Adaptive Architectures for Edge Computing: To meet the real-time, low-power, and privacy-preservation demands of industrial sites, models must be extremely lightweight. Future research needs to combine Neural Architecture Search (NAS), dynamic networks, adaptive computation, and advanced model compression and distillation techniques to develop “agile” models that can dynamically adjust computational resources based on input complexity, enabling high-performance deployment on resource-constrained edge devices.

Building an Open, Collaborative Industrial Vision Ecosystem: Promote the establishment of large-scale open-source benchmark datasets encompassing richer scenarios, multiple modalities, and incorporating temporal and physical information. Simultaneously, explore distributed collaborative training frameworks based on federated learning and privacy-computing technologies to aggregate industry knowledge and jointly train more powerful, general-purpose defect detection foundation models while ensuring the data privacy of participating enterprises.

Acknowledgement: The authors acknowledge the valuable comments and suggestions that helped improve this paper.

Funding Statement: This work was supported in part by the Natural Science Foundation of Shaanxi Province of China under Grant 2024JC-YBQN-0695.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Yiquan Wu and Langyue Zhao; methodology, Langyue Zhao and Yubin Yuan; investigation, Langyue Zhao; writing—original draft preparation, Langyue Zhao; writing—review and editing, Yiquan Wu and Yubin Yuan; supervision, Yiquan Wu; funding acquisition, Yiquan Wu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data sharing is not applicable.

Ethics Approval: This study did not involve human participants or animal subjects. Ethical approval is not applicable.

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

References

1. Zhao LY, Wu YQ. Research progress of surface defect detection methods based on machine vision. *Chin J Sci Instrum.* 2022;43(1):198–219. (In Chinese). doi:10.4028/www.scientific.net/amr.403-408.1356.
2. Gao Y, Li X, Wang XV, Wang L, Gao L. A review on recent advances in vision-based defect recognition towards industrial intelligence. *J Manuf Syst.* 2022;62:753–66. doi:10.1016/j.jmsy.2021.05.008.
3. Czimmermann T, Ciuti G, Milazzo M, Chiurazzi M, Roccella S, Oddo CM, et al. Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors.* 2020;20(5):1459. doi:10.3390/s20051459.
4. Wu YQ, Zhao LY, Yuan YB, Yang J. Research status and the prospect of PCB defect detection algorithm based on machine vision. *Chin J Sci Instrum.* 2022;43(8):1–17. (In Chinese). doi:10.23977/autml.2024.050112.
5. Liu YQ, Wu YQ. Review of defect detection algorithms for solar cells based on machine vision. *Opt Precis Eng.* 2024;32(6):868–900. (In Chinese). doi:10.37188/ope.20243206.0868.
6. Lin SY, Wu YQ. Vision-based LCD/OLED defect detection methods: a critical summary. *J Image Graph.* 2024;29(5):1321–45. (In Chinese). doi:10.11834/jig.230518.
7. Amirkhani D, Allili MS, Hebbache L, Hammouche N, Lapointe JF. Visual concrete bridge defect classification and detection using deep learning: a systematic review. *IEEE Trans Intell Transp Syst.* 2024;25(9):10483–505. doi:10.1109/TITS.2024.3365296.
8. Xia GS, Bai X, Ding J, Zhu Z, Belongie S, Luo J, et al. DOTA: a large-scale dataset for object detection in aerial images. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA.* doi:10.1109/CVPR.2018.00418.
9. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in context. In: *Proceedings of the 13th European Conference on Computer Vision ECCV 2014; 2014 Sep 6–12; Zurich, Switzerland.*
10. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: single shot MultiBox detector. In: *Proceedings of the 14th European Conference on Computer Vision ECCV 2016; 2016 Oct 11–14; Amsterdam, The Netherlands.* doi:10.1007/978-3-319-46448-0_2.
11. Lin TY. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–22; Salt Lake City, UT, USA.*
12. Wang JM, Wang MX, Zhang J, Lu JQ, Pi QX, Zhang JX, inventors. Central South University, assignee. Strip steel surface micro defect detection network and method. China patent CN117350968A. 2024 Jan 5.
13. Dwivedi P, Weber JW, Lee Chin R, Trupke T, Hameiri Z. Deep learning method for enhancing luminescence image resolution. *Sol Energy Mater Sol Cells.* 2023;257(9):112357. doi:10.1016/j.solmat.2023.112357.
14. Ding L, Wang X, Li D. Visual saliency detection in high-resolution remote sensing images using object-oriented random walk model. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2022;15:4698–707. doi:10.1109/JSTARS.2022.3179461.
15. Nair V, Radhakrishnan A, Chithra R, James A. Memristive pixel-CNN loop generate for CNN generalisations. *IEEE Trans Nanotechnol.* 2023;22:120–5. doi:10.1109/TNANO.2023.3248108.
16. Zhang ZD, Zhang B, Lan ZC, Liu HC, Li DY, Pei L, et al. FINet: an insulator dataset and detection benchmark based on synthetic fog and improved YOLOv5. *IEEE Trans Instrum Meas.* 2022;71(8):1–8. doi:10.1109/tim.2022.3194909.
17. Hao K, Chen G, Zhao L, Li Z, Liu Y, Wang C. An insulator defect detection model in aerial images based on multiscale feature pyramid network. *IEEE Trans Instrum Meas.* 2022;71:1–12. doi:10.1109/tim.2022.3200861.
18. Zhou X, Zhang Y, Liu Z, Jiang Z, Ren Z, Mi T, et al. IFIFusion: a independent feature information fusion model for surface defect detection. *Inf Fusion.* 2025;120(9):103039. doi:10.1016/j.inffus.2025.103039.
19. Ma R, Chen J, Feng Y, Zhou Z, Xie J. ELA-YOLO: an efficient method with linear attention for steel surface defect detection during manufacturing. *Adv Eng Inform.* 2025;65(2):103377. doi:10.1016/j.aei.2025.103377.
20. Hao S, Yang L, Ma X, Ma RZ, Wen H. YOLOv5 transmission line fault detection based on attention mechanism and cross-scale feature fusion. *Proc CSEE.* 2023;43(6):2319–30. (In Chinese).
21. Zhou M, Li B, Wang J, He S. Fault detection method of glass insulator aerial image based on the improved YOLOv5. *IEEE Trans Instrum Meas.* 2023;72:5012910. doi:10.1109/TIM.2023.3269099.

22. Song LY, Liu S, Wang K, Yang JD. Identification method of power grid components and defects based on improved EfficientDet. *Trans China Electrotech Soc.* 2022;37(9):2241–51. (In Chinese).
23. Wang S, Cheng DJ, Fang XF, Zhang CY. SDA-PVTDet: a spatial-cross dual attention pyramid vision transformer detector for casting defect detection in radiography images. *Expert Syst Appl.* 2025;269(1):126385. doi:10.1016/j.eswa.2025.126385.
24. Pu C, Wang J, Zhang Y, Niu M, Wu Q, Lin Z. Geometric spatial constraints network for slender and tiny surface defect detection. *Adv Eng Inform.* 2025;65(7):103138. doi:10.1016/j.aei.2025.103138.
25. Sui T, Wang J. DMPDD-net: an effective defect detection method for aluminum profiles surface defect. *IEEE Trans Instrum Meas.* 2025;74:3500313. doi:10.1109/TIM.2024.3497168.
26. He M, Qin L, Wang Y, Deng X, Liu Q, Zhang Y, et al. A weakly supervised contrastive learning pretraining method for visual defect detection of transmission lines. *IEEE Trans Instrum Meas.* 2025;74(6):1–15. doi:10.1109/tim.2025.3577837.
27. Chen L, Meng K, Zhang H, Zhou J, Lou P. SR-FABNet: super-resolution branch guided Fourier attention detection network for efficient optical inspection of nanoscale wafer defects. *Adv Eng Inform.* 2025;65(3):103200. doi:10.1016/j.aei.2025.103200.
28. He M, Qin L, Deng X, Liu Q, Liu K. Bolt-YOLO: research on an algorithm framework for detecting bolt defects in transmission lines. *IEEE Trans Power Deliv.* 2025;40(3):1718–29. doi:10.1109/TPWRD.2025.3559034.
29. Xie Y, Huang X, Qin F, Li F, Ding X. A majority affiliation based under-sampling method for class imbalance problem. *Inf Sci.* 2024;662(2):120263. doi:10.1016/j.ins.2024.120263.
30. Zhang R, Lu S, Yan B, Yu P, Tang X. A density-based oversampling approach for class imbalance and data overlap. *Comput Ind Eng.* 2023;186(5):109747. doi:10.1016/j.cie.2023.109747.
31. França Aires L, Oliveira Schmidt J, Hübner GR, Menine Schaf F, Moro Franchi C, Pinheiro H, et al. Convolutional neural networks using the SMOTE algorithm and features fusion for wind turbine fault prediction. *IEEE Lat Am Trans.* 2025;23(3):191–7. doi:10.1109/tla.2025.10879178.
32. Chao X, Zhang L. Few-shot imbalanced classification based on data augmentation. *Multimed Syst.* 2023;29(5):2843–51. doi:10.1007/s00530-021-00827-0.
33. Kim TH, Cho JH, Kim Y, Chang JH. Deep-learning-based prediction algorithm for fuel-cell electric vehicle energy with shift mixup. *IEEE Sens J.* 2024;24(9):14529–38. doi:10.1109/JSEN.2024.3373078.
34. Ma D, Fang H, Wang N, Lu H, Matthews J, Zhang C. Transformer-optimized generation, detection, and tracking network for images with drainage pipeline defects. *Comput Aided Civ Infrastruct Eng.* 2023;38(15):2109–27. doi:10.1111/mice.12970.
35. 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(6):1137–49. doi:10.1109/TPAMI.2016.2577031.
36. Zhang S, Chi C, Yao Y, Lei Z, Li SZ. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA.* doi:10.1109/cvpr42600.2020.00978.
37. Ge Z, Liu S, Li Z, Yoshie O, Sun J. OTA: optimal transport assignment for object detection. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA.* doi:10.1109/CVPR46437.2021.00037.
38. Li H, Wu Z, Zhu C, Xiong C, Socher R, Davis LS. Learning from noisy anchors for one-stage object detection. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA.* doi:10.1109/CVPR42600.2020.01060.
39. Jiao R, Fu Z, Liu Y, Zhang Y, Song Y. A defective bolt detection model with attention-based RoI fusion and cascaded classification network. *IEEE Trans Instrum Meas.* 2023;72:1–11. doi:10.1109/tim.2023.3318688.
40. Cui Y, Jia M, Lin TY, Song Y, Belongie S. Class-balanced loss based on effective number of samples. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA.* doi:10.1109/cvpr.2019.00949.

41. Tan J, Wang C, Li B, Li Q, Ouyang W, Yin C, et al. Equalization loss for long-tailed object recognition. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. doi:10.1109/cvpr42600.2020.01168.
42. Chen K, Lin W, Li J, See J, Wang J, Zou J. AP-loss for accurate one-stage object detection. *IEEE Trans Pattern Anal Mach Intell.* 2021;43(11):3782–98. doi:10.1109/tpami.2020.2991457.
43. Zhu H, Yang H, Yin ZP. Review of machine vision detection methods for texture surface defects. *Mech Sci Technol Aersp Eng.* 2023;42(8):1293–315. (In Chinese).
44. Zhang M, Cao Y, Jiang K, Li M, Liu L, Yu Y, et al. Proactive measures to prevent conveyor belt failures: deep learning-based faster foreign object detection. *Eng Fail Anal.* 2022;141(10):106653. doi:10.1016/j.engfailanal.2022.106653.
45. Yang L, Fan J, Huo B, Li E, Liu Y. A nondestructive automatic defect detection method with pixelwise segmentation. *Knowl Based Syst.* 2022;242(12):108338. doi:10.1016/j.knosys.2022.108338.
46. 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(7):9725–35. doi:10.1109/TII.2024.3383513.
47. Liu Y, Gao H, Guo L, Qin A, Cai C, You Z. A data-flow oriented deep ensemble learning method for real-time surface defect inspection. *IEEE Trans Instrum Meas.* 2020;69(7):4681–91. doi:10.1109/tim.2019.2957849.
48. Tsung CK, Kristiani E, Chiu CK, Liu JC, Yang CT. HPPH: computer-vision-based service for high-performance pavement health recognition. *IEEE Internet Things J.* 2025;12(11):15987–96. doi:10.1109/JIOT.2025.3530253.
49. Guo K, He C, Yang M, Wang S. A pavement distresses identification method optimized for YOLOv5s. *Sci Rep.* 2022;12(1):3542. doi:10.1038/s41598-022-07527-3.
50. Yang G, Zhao B, Zhang J, Wen J, Li Q, Lei L, et al. Det-Recon-Reg: an intelligent framework toward automated UAV-based large-scale infrastructure inspection. *IEEE Trans Instrum Meas.* 2025;74:3539516. doi:10.1109/TIM.2025.3571118.
51. Cui W, Song K, Zhang Y, Jia X, Liu X, Yan Y. TRDM: a two-stage real-time discrimination method for spiral weld defects under dynamic distorted imaging. *IEEE Trans Automat Sci Eng.* 2025;22:15420–34. doi:10.1109/tase.2025.3570251.
52. Liu M, Zhu Q, Yin Y, Fan Y, Su Z, Zhang S. Damage detection method of mining conveyor belt based on deep learning. *IEEE Sens J.* 2022;22(11):10870–9. doi:10.1109/JSEN.2022.3170971.
53. Xu H, Zhu JB, Kang DX, inventors. Inspur Software Group Co., Ltd., assignee. A defect visual detection and tracking method for semi-conductive band cables. Chinese patent CN119273616A. 2025 Jan 7.
54. Tongji University, assignee. A pipeline defect detection and tracking method and device. Chinese patent CN202310506434.8. 2023 Sep 5.
55. Lai Z, Liang G, Zhou J, Kong H, Lu Y. A joint learning framework for optimal feature extraction and multi-class SVM. *Inf Sci.* 2024;671(3):120656. doi:10.1016/j.ins.2024.120656.
56. Wang D, Zhang J, Du B, Zhang L, Tao D. DCN-T: dual context network with transformer for hyperspectral image classification. *IEEE Trans Image Process.* 2023;32:2536–51. doi:10.1109/TIP.2023.3270104.
57. Yang B, Bender G, Le QV, Ngiam J. Condconv: conditionally parameterized convolutions for efficient inference. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019 Dec 8–14; Vancouver, BC, Canada.
58. Wang J, Peng X, Qiao Y. Cascade multi-head attention networks for action recognition. *Comput Vis Image Underst.* 2020;192(8):102898. doi:10.1016/j.cviu.2019.102898.
59. Liu X, Yoo C, Xing F, Oh H, El Fakhri G, Kang JW, et al. Deep unsupervised domain adaptation: a review of recent advances and perspectives. *APSIPA Trans Signal Inf Process.* 2022;11(1):1–51. doi:10.1561/116.00000192.
60. Wang W, Li H, Ding Z, Nie F, Chen J, Dong X, et al. Rethinking maximum mean discrepancy for visual domain adaptation. *IEEE Trans Neural Netw Learning Syst.* 2023;34(1):264–77. doi:10.1109/tnnls.2021.3093468.
61. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep.* 2019;9(1):16884. doi:10.1038/s41598-019-52737-x.

62. He S, Ding L, Dong D, Liu B, Yu F, Tao D. PAD-net: an efficient framework for dynamic networks. arXiv:2211.05528. 2022.
63. Li F, Li F, Xi Q. DefectNet: toward fast and effective defect detection. *IEEE Trans Instrum Meas.* 2021;70:1–9. doi:10.1109/tim.2021.3067221.
64. Sinha D, El-Sharkawy M. Thin MobileNet: an enhanced MobileNet architecture. In: *Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*; 2019 Oct 10–12; New York, NY, USA. doi:10.1109/UEMCON47517.2019.8993089.
65. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. doi:10.1109/CVPR.2018.00716.
66. Zhou Q, Wang H, Tang Y, Wang Y. Defect detection method based on knowledge distillation. *IEEE Access.* 2023;11(41):35866–73. doi:10.1109/ACCESS.2023.3252910.
67. Hu Z, Zhang Z, Liu S, Zhao D, Zheng L, Tang L. ASTKD-PCB-LDD: high-performance PCB defect detection model with align soft-target knowledge distillation and lightweight network design. *J Supercomput.* 2025;81(4):531. doi:10.1007/s11227-025-07045-9.
68. Liu Z, Miao Z, Zhan X, Wang J, Gong B, Yu SX. Large-scale long-tailed recognition in an open world. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2019 Jun 15–20; Long Beach, CA, USA. doi:10.1109/CVPR.2019.00264.
69. Zhou B, Cui Q, Wei XS, Chen ZM. BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020 Jun 13–19; Seattle, WA, USA. doi:10.1109/cvpr42600.2020.00974.
70. Gupta A, Dollár P, Girshick R. LVIS: a dataset for large vocabulary instance segmentation. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2019 Jun 15–20; Long Beach, CA, USA. doi:10.1109/CVPR.2019.00550.
71. Huang W, Wei P. A PCB dataset for defects detection and classification. *J Eng.* 2018;14(8):1–9. doi:10.1049/joe.2019.1183.
72. Tang S, He F, Huang X, Yang J. Online PCB defect detector on a new PCB defect dataset. arXiv:1902.06197. 2019.
73. Lu H, Mehta D, Paradis O. FICS-PCB: a multi modal image dataset for automated printed circuit board visual inspection. *Cryptology EPrint Archive*. [cited 2026 Jan 1]. Available from: https://www.researchgate.net/publication/344475848_FICS-PCB_A_Multi-Modal_Image_Dataset_for_Automated_Printed_Circuit_Board_Visual_Inspection.
74. Pramerdorfer C, Kampel M. A dataset for computer-vision-based PCB analysis. In: *Proceedings of the 2015 14th IAPR International Conference on Machine Vision Applications (MVA)*; 2015 May 18–22; Tokyo, Japan. doi:10.1109/MVA.2015.7153209.
75. Mahalingam G, Gay KM, Ricanek K. PCB-METAL: a PCB image dataset for advanced computer vision machine learning component analysis. In: *Proceedings of the 2019 16th International Conference on Machine Vision Applications (MVA)*; 2019 May 27–31; Tokyo, Japan. doi:10.23919/MVA.2019.8757928.
76. Pierdicca R, Paolanti M, Felicetti A, Piccinini F, Zingaretti P. Automatic faults detection of photovoltaic farms: solAIr, a deep learning-based system for thermal images. *Energies.* 2020;13(24):6496. doi:10.3390/en13246496.
77. Buerhop-Lutz C, Deitsch S, Maiera A, Gallwitz F, Berger S, Doll B, et al. A benchmark for visual identification of defective solar cells in electroluminescence imagery. In: *Proceedings of the 35th European PV Solar Energy Conference and Exhibition*. Brussels, Belgium: The European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC); 2018. p. 1287–9.
78. Rodriguez AR, Holicza B, Nagy AM, Vörösházi Z, Bereczky G, Czúni L. Segmentation and error detection of PV modules. In: *Proceedings of the 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*; 2022 Sep 6–9; Stuttgart, Germany. doi:10.1109/ETFA52439.2022.9921572.
79. Pratt L, Mattheus J, Klein R. A benchmark dataset for defect detection and classification in electroluminescence images of PV modules using semantic segmentation. *Syst Soft Comput.* 2023;5(1):200048. doi:10.1016/j.sasc.2023.200048.

80. Su B, Zhou Z, Chen H. PVEL-AD: a large-scale open-world dataset for photovoltaic cell anomaly detection. *IEEE Trans Ind Inform.* 2023;19(1):404–13. doi:10.1109/TII.2022.3162846.
81. Millendorf M, Obropta E, Vad-Havkar N. Infrared solar module dataset for anomaly detection. In: *Proceedings of the 8th International Conference on Learning Representations*; 2020 Apr 26–30; Addis Ababa, Ethiopia.
82. Chen X, Karin T, Libby C, Deceglie M, Hacke P, Silverman TJ, et al. Automatic crack segmentation and feature extraction in electroluminescence images of solar modules. *IEEE J Photovoltaics.* 2023;13(3):334–42. doi:10.1109/jphotov.2023.3249970.
83. Tao X, Zhang D, Wang Z, Liu X, Zhang H, Xu D. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Trans Syst Man Cybern Syst.* 2020;50(4):1486–98. doi:10.1109/TSMC.2018.2871750.
84. Wu MJ, Jang JR, Chen JL. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Trans Semicond Manuf.* 2015;28(1):1–12. doi:10.1109/TSM.2014.2364237.
85. Wang Y, Ni D. Multi-bin wafer maps defect patterns classification. In: *Proceedings of the 2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE)*; 2019 Apr 20–21; Hangzhou, China. doi:10.1109/SMILE45626.2019.8965299.
86. Zhang J, Ding R, Ban M, Guo T. FDSNeT: an accurate real-time surface defect segmentation network. In: *Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2022 May 23–27; Singapore. doi:10.1109/ICASSP43922.2022.9747311.
87. Arya D, Maeda H, Ghosh SK, Toshniwal D, Sekimoto Y. RDD2022: a multi-national image dataset for automatic road damage detection. *Geosci Data J.* 2024;gdj3.260. doi:10.1002/gdj3.260.
88. Yan H, Zhang J. UAV-PDD2023: a benchmark dataset for pavement distress detection based on UAV images. *Data Brief.* 2023;51(12):109692. doi:10.1016/j.dib.2023.109692.
89. Zhu J, Zhong J, Ma T, Huang X, Zhang W, Zhou Y. Pavement distress detection using convolutional neural networks with images captured via UAV. *Autom Constr.* 2022;133(2):103991. doi:10.1016/j.autcon.2021.103991.
90. Lv X, Duan F, Jiang JJ, Fu X, Gan L. Deep metallic surface defect detection: the new benchmark and detection network. *Sensors.* 2020;20(6):1562. doi:10.3390/s20061562.
91. Bergmann P, Fauser M, Sattlegger D. MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2019 Jun 15–20; Long Beach, CA, USA.
92. Bergmann P, Jin X, Sattlegger D, Steger C. The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization. *arXiv:2112.09045.* 2021.