processes

MDPI

Review

Review on Application of Machine Vision-Based Intelligent Algorithms in Gear Defect Detection

Dehai Zhang [1,*], Shengmao Zhou [1], Yujuan Zheng [2] and Xiaoguang Xu [3,*]

[1] Mechanical and Electrical Engineering Institute, Zhengzhou University of Light Industry, Zhengzhou 450002, China; 15110053093@163.com
[2] Xuchang University Library, Xuchang University, Xuchang 461000, China; grass1979@126.com
[3] China Tobacco Henan Industrial Co., Ltd., Xuchang 461000, China
* Correspondence: zhangdehai0318@163.com (D.Z.); xxgjyc@163.com (X.X.); Tel.: +86-139-3745-1186 (D.Z.)

Abstract

Gear defect detection directly affects the operational reliability of critical equipment in fields such as automotive and aerospace. Gear defect detection technology based on machine vision, leveraging the advantages of non-contact measurement, high efficiency, and cost-effectiveness, has become a key support for quality control in intelligent manufacturing. However, it still faces challenges including difficulties in semantic alignment of multimodal data, the imbalance between real-time detection requirements and computational resources, and poor model generalization in few-shot scenarios. This paper takes the paradigm evolution of gear defect detection technology as the main line, systematically reviews its development from traditional image processing to deep learning, and focuses on the innovative application of intelligent algorithms. A research framework of "technical bottleneck-breakthrough path-application verification" is constructed: for the problem of multimodal fusion, the cross-modal feature alignment mechanism based on Transformer network is deeply analyzed, clarifying its technical path of realizing joint embedding of visual and vibration signals by establishing global correlation mapping; for resource constraints, the performance of lightweight models such as MobileNet and ShuffleNet is quantitatively compared, verifying that these models reduce Parameters by 40–60% while maintaining the mean Average Precision essentially unchanged; for small-sample scenarios, few-shot generation models based on contrastive learning are systematically organized, confirming that their accuracy in the 10-shot scenario can reach 90% of that of fully supervised models, thus enhancing generalization ability. Future research can focus on the collaboration between few-shot generation and physical simulation, edge-cloud dynamic scheduling, defect evolution modeling driven by multiphysics fields, and standardization of explainable artificial intelligence. It aims to construct a gear detection system with autonomous perception capabilities, promoting the development of industrial quality inspection toward high-precision, high-robustness, and low-cost intelligence.

Keywords: gear defect detection; traditional image processing; deep learning models; intelligent algorithms; cross-modal feature alignment

check for updates

Academic Editors: Nhat Luong Nhieu, Phan Van-Thanh, Hector Tibo and Chia-Nan Wang

Received: 25 September 2025
Revised: 13 October 2025
Accepted: 16 October 2025
Published: 21 October 2025

Citation: Zhang, D.; Zhou, S.; Zheng, Y.; Xu, X. Review on Application of Machine Vision-Based Intelligent Algorithms in Gear Defect Detection. *Processes* **2025**, *13*, 3370. https://doi.org/10.3390/pr13103370

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Gears, as a core component of industrial mechanical transmission, play a decisive role in the operational reliability and service life of high-end equipment such as automotive, aerospace, and industrial robots, depending on their quality [1]. As shown in Figure 1,

typical gear defects include wear, pitting, spalling, and cracks [2]. These defects not only reduce transmission efficiency and increase vibration and noise but may also cause system failures and catastrophic accidents, leading to significant losses in industrial production [3].

Traditional gear defect detection methods based on human expertise could only perform preliminary inspections. Due to their slow speed and precision variability from human factors, they could not meet the strict quality control requirements of mass production [4]. With technological advances, machine vision-based defect detection emerged. Its non-contact imaging, automated feature extraction, and intelligent decision-making provide an effective solution to overcome the limitations of traditional methods [5].

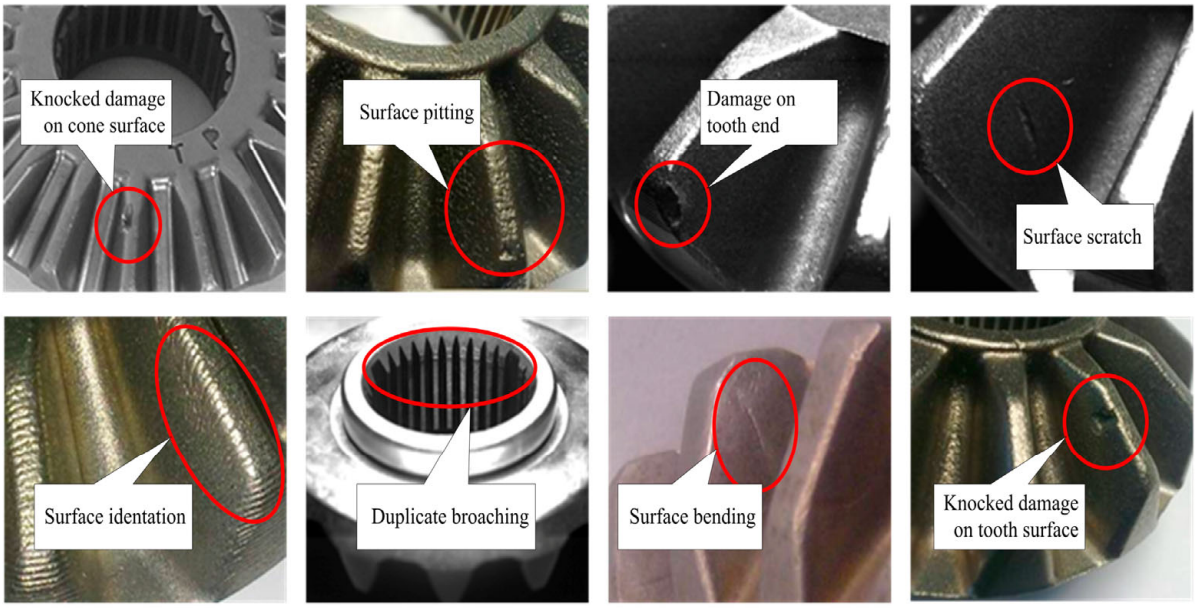


Figure 1. Common types of tooth fault [6].

However, as industrial settings grow more complex, machine vision inspection faces several challenges [7,8]. First, aligning and modeling semantic consistency across multimodal data such as visual, vibration, and thermodynamic signals is difficult due to differences in data types, complicating the fusion process [9]. Second, balancing real-time detection with limited computational resources is challenging because high-speed detection (requiring >20 FPS) requires significant computational power, which is often restricted in industrial environments (e.g., edge devices with only 2 GB memory) [10]. Additionally, models struggle to accurately detect defects with limited training data (e.g., when the number of defective samples is less than 50, the detection accuracy decreases by more than 20% [11]), which restricts the practical application of inspection systems [12]. These technical bottlenecks limit the large-scale application of machine vision inspection systems in smart manufacturing.

In recent years, the continuous breakthroughs in intelligent detection algorithms have provided innovative solutions to the above challenges [13–15]. For instance, cross-modal feature alignment technology, leveraging Transformer networks, has achieved joint embedding of visual and vibration signals, effectively solving the problem of multimodal data fusion [16]. The emergence of lightweight model architectures, such as MobileNet [17] (using depthwise separable convolution to reduce 8–9 times computation), EfficientNet [18] (scaling width/depth/resolution uniformly), and ShuffleNet [19] (adopting channel shuffle to reduce inter-channel dependence), has significantly reduced the computational complexity of models, improved their operational efficiency, and enabled rapid detection under limited computational resources. Few-shot generation models based on contrastive

learning can approximate the performance of fully supervised models with minimal samples (e.g., 10-shot scenario accuracy reaches 90% of fully supervised [20]), enhancing model generalization [21]. These advancements mark a shift in gear detection technology from the traditional "human experience-driven" model to an intelligent paradigm of "data-model collaborative drive."

Compared with existing reviews, the innovations of this study focus on three aspects: First, the research scope is more focused. For the first time, intelligent algorithms are clearly divided into three technical directions, namely cross-modal perception, lightweight deployment, and few-shot learning. Second, the classification system is more systematic. It breaks through the traditional "method listing" model and constructs a framework in accordance with the three-stage evolution logic of "traditional image processing-deep learning-intelligent algorithms". Each stage follows a three-level content structure of "technical principle-bottleneck analysis-case verification". Third, the quantitative comprehensive analysis is more sufficient. By supplementing comparative data on parameter scale, mean Average Precision (mAP), GFLOPs (Giga Floating-Point Operations) and Weight Size of different lightweight models, it provides a quantitative basis for technology selection in industrial scenarios and avoids subjective expressions such as "excellent performance".

In the literature screening for this study, priority was given to research results from top international journals in 2020–2024, while pioneering 2010–2019 literature on fundamental models like ResNet and YOLOv3 was retained. The search covered international core databases (IEEE Xplore, Elsevier ScienceDirect, Web of Science) and Chinese authoritative journals (e.g., Journal of Mechanical Engineering), using keyword combinations such as "gear defect detection", "machine vision", and "deep learning".

This paper is structured as outlined in Figure 2. We begin by highlighting the limitations of traditional gear defect detection methods, leading into Section 2, which focuses on "Gear Defect Detection Based on Traditional Image Processing." Next, we transition to Section 3, "Gear Defect Detection Based on Deep Learning Models," in light of the challenges faced by conventional image processing techniques. Then, in response to the bottlenecks and difficulties encountered by these two detection algorithms, we introduce Section 4, "Intelligent Algorithms to Addressing Challenges in Vision Algorithms." Finally, we summarize the preceding content and present four forward-looking outlooks for future research directions.

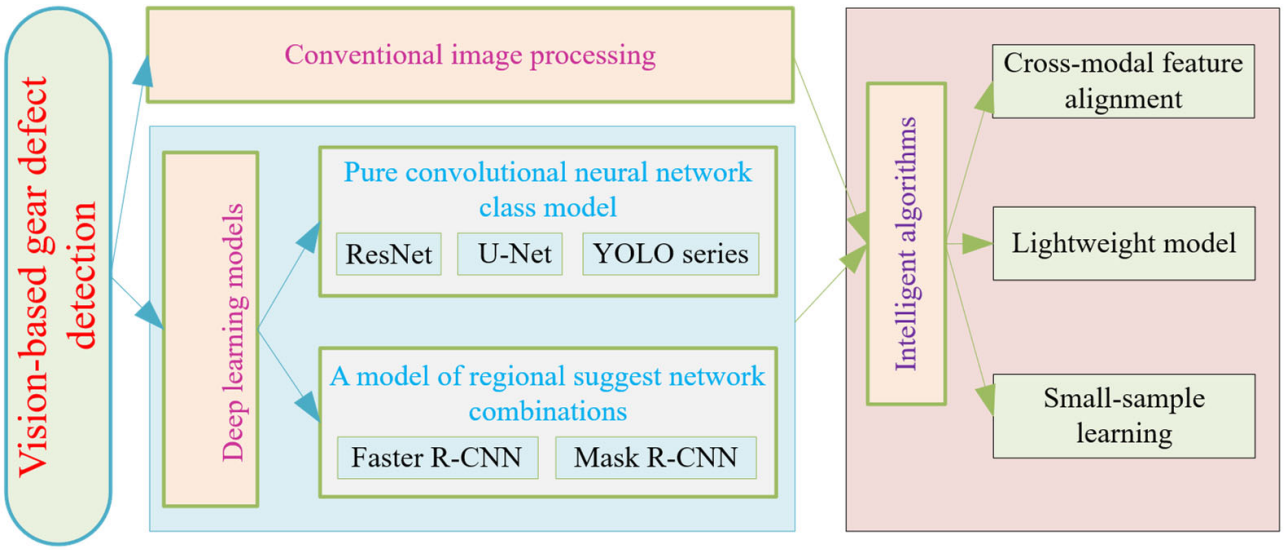


Figure 2. Overall research framework.

2. Gear Defect Detection Based on Traditional Image Processing

Between 2000 and 2010, traditional image processing techniques dominated gear defect detection research [22,23]. Early studies relied heavily on conventional image processing algorithms, with edge detection being a key technology. Operators like Canny [24] and Sobel [25] were commonly used to extract gear tooth profiles. For instance, the Canny operator could accurately identify gear edge positions, providing a foundation for subsequent analysis. Image registration, using feature matching methods like SIFT [26] and SURF [27], compared gear design drawings with actual images to detect geometric deviations. However, these traditional methods had significant limitations. First, they depended on manually designed features, requiring expertise and experience to select appropriate feature extraction methods and parameters. This increased the difficulty and cost of detection while limiting the methods' generality and adaptability. Second, they were highly sensitive to lighting conditions, leading to unstable detection results in varying environments [28]. Kyriakides et al. [29] demonstrated that the number of light sources is positively correlated with defect visibility. When four UV light sources are activated simultaneously, the false detection rate of cracks on metal surfaces decreases from 12% to 1.5%. Additionally, in complex backgrounds, traditional methods lacked robustness and struggled to accurately identify and extract gear defect features.

During the 2000–2010 traditional image processing era, scholars deeply explored gear defect detection using conventional algorithms and achieved notable results. Shao et al. [30] proposed a vision-based method for gear surface integrity inspection, combining a subpixel edge detector and Canny operator for edge extraction alongside normalized cross-correlation coefficient for target matching, demonstrating high reliability and speed for gear surface inspection. Liu et al. [31] introduced a bidimensional local feature-scale decomposition (BLCD) method, which decomposed images into intrinsic scale components to remove noise and enhance defect visibility, offering a novel solution for gear surface defect detection. Chang et al. [32] combined surface replication with image analysis to study gear tooth wear evolution, using optical and laser scanning confocal microscopy (LSCM) imaging to quantify surface roughness, wear extent, and depth via image processing tools, enabling high-resolution monitoring of gear wear. Wang et al. [33] addressed issues such as low efficiency, poor quality, and instability in gear surface defect detection by proposing a method that converts gear image information into digital image information, which is then subjected to corresponding digital morphological processing to eliminate defective gears. They utilized VS2010 in conjunction with the Halcon library to analyze and process the images, ultimately feeding the processing results back to the control terminal to achieve rapid identification and elimination of defective gears, with detection time errors kept within 0.2 s.

Although traditional image processing methods have made some progress in gear defect detection, their inherent limitations make it difficult to meet the growing industrial demands for higher accuracy and efficiency. For example, the detection accuracy of traditional methods is generally below 90%, and in industrial environments with strong noise and variable illumination, the detection accuracy can drop to 70% or even lower [34].

3. Gear Defect Detection Based on Deep Learning Models

Since 2010, deep learning technology has rapidly emerged, injecting unprecedented innovation into the field of gear defect detection [35–37]. Deep learning has the powerful ability to automatically learn and extract deep, high-discrimination features from massive amounts of data, eliminating the reliance on manually designed features. This characteristic has shown unparalleled advantages in processing complex and variable gear image data [38].

In gear defect detection research and applications, traditional deep learning network models have been widely explored and implemented. Among them, convolutional neural network (CNN) models, such as ResNet [39,40], U-Net [41], and the YOLO series [42–44], leverage their unique convolutional architectures to quickly and efficiently extract both local and global features from gear images, performing exceptionally well in gear defect detection tasks. Another category of models combines region proposal networks, such as Faster R-CNN [45,46] and Mask R-CNN [47–49]. These models use a two-stage detection strategy, first generating potential target regions with region proposal networks and then performing precise classification and localization. This approach not only improves detection accuracy but also allows for more detailed analysis of gear defects [50–53]. Finally, this section systematically reviews and comprehensively summarizes practical application cases of these network models in gear defect detection.

3.1. Gear Defect Detection Based on Pure Convolutional Neural Network Models

3.1.1. Gear Defect Detection Based on ResNet Architecture

ResNet, proposed by Kaiming He et al. [54] in 2015 at Microsoft Research Asia, is a deep neural network model. Its key contribution lies in introducing residual connections, which solve the model degradation and gradient vanishing problems in deep neural networks. This innovation allows the network to increase in depth, enabling it to learn more complex, abstract, and highly discriminative features.

Su et al. [55] developed the SR-ResNetYOLO method for detecting defects on the end faces of metal gears. This method addresses challenges such as diverse gear types, uneven cross-sectional structures, few defect sizes, and multiple scales. It first extracts images of the machined area using a visual saliency-based method, then establishes a multi-scale sampling feature extraction backbone network, ResNet-21, under 16X to obtain high-resolution defect features. Feature maps are fused through a multi-scale fusion module, and defects are classified and located by positioning and classification modules. Experiments show this method achieves good performance in metal gear end face defect detection, with an average precision of 96.66%, a recall rate of 97.07%, and an average detection time of 0.12 s per image, effectively detecting small-sized and multi-scale defects. Han et al. [56] used a Mask R-CNN network based on deep learning algorithms for detecting tiny gear defects, selecting ResNet-101 as the image shared feature extraction network. By removing the unreasonable 3×3 convolution on feature map P5 in the feature pyramid network and setting appropriate anchor sizes and aspect ratios based on sample labeling, the optimized Mask R-CNN achieved a 98.2% gear tooth missing detection rate. Bao et al. [57] proposed a gear defect detection method based on an improved ResNet101 network. By introducing atrous convolution and dense connection operations, the feature extraction ability and stability of the model were significantly enhanced. Experimental results indicate that the improved ResNet101 network performs excellently in gear defect detection, with an accuracy of 96.78% and a recall rate of 96.54%. It offers higher stability and detection precision, making it suitable for quality control in gear production and providing a reliable technical means to ensure gear quality. Relevant results of these models are summarized in Table 1.

3.1.2. Gear Defect Detection Based on U-Net Architecture

In gear defect detection, U-Net fuses high-resolution texture features (such as gear tooth edge and texture) from the shallow layers of the encoder with semantic features from the deep layers of the decoder through skip connections, effectively preserving sub-pixel-level details [58]. This structure is particularly suitable for detecting tiny defects on gear surfaces (such as microcracks and pitting corrosion), enabling accurate segmentation under

complex backgrounds, with the mean Intersection over Union (mIoU) often exceeding 80%. The following are some research findings on gear defect detection using the U-Net network model.

Qin et al. [59] presented a gear pitting measurement method based on multi-scale concatenated attention U-Net. This method leverages multi-scale feature extraction and attention mechanisms to enhance the model's ability to identify pitting areas. From rigorous image preprocessing to detailed feature extraction and precise attention mechanism application, every step is carefully designed and optimized. The experimental results validate that this method achieves high detection accuracy of 93.77% in gear pitting measurement, showing strong robustness and reliability in identifying and segmenting pitting regions, thus offering solid technical support for gear maintenance and upkeep. Wang et al. [60] proposed a gear pitting visual measurement method that integrates DCGAN and U-Net. By first generating high-quality samples with DCGAN and then using U-Net for precise segmentation and measurement of pitting areas, this method demonstrated excellent performance in gear pitting measurement, achieving a low average relative error of 7.83% and an absolute error of 0.18%, which significantly improved the accuracy and robustness of detection. Dong et al. [61] developed a cascaded detection method based on machine vision for gear surface defect detection. This method combines various image processing techniques with deep learning models in a cascaded manner to enhance detection accuracy and efficiency. In gear tooth surface integrity detection, it achieved a Dice coefficient of 84.29% and an IoU of 73.95%, enabling efficient and accurate identification and segmentation of defective areas on gear surfaces. The structure of their GBSU-Net network model is shown in Figure 3. As can be seen from Figure 3, the GBSU-Net network adopts the encoder-decoder structure, and introduces Transformer and skip connection modules. The Transformer module can capture the global feature information of the gear image, and the skip connection module can fuse the shallow and deep features, which is beneficial to improve the detection accuracy of small defects.

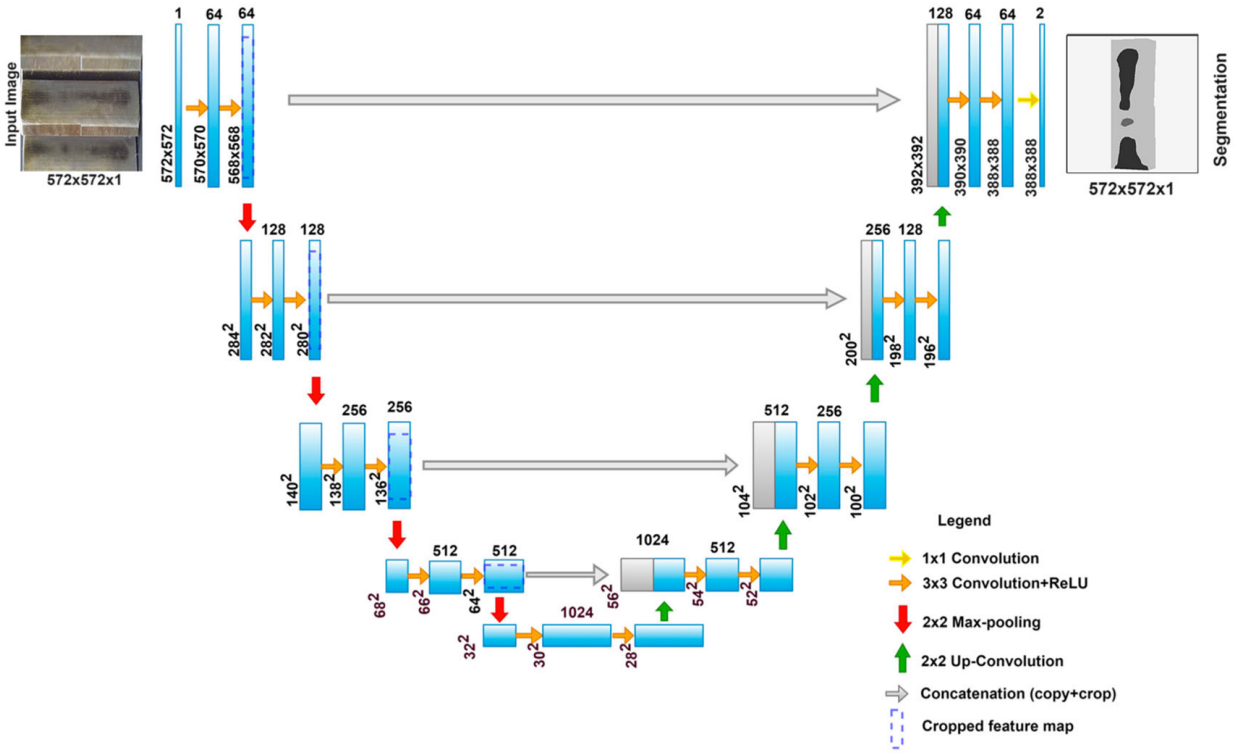


Figure 3. Basic U-Net structure for gear defect segmentation.

3.1.3. Gear Defect Detection Using YOLO Series Architecture

The YOLO series has consistently led in real-time object detection with its unique single-stage detection paradigm and efficient network architecture [62]. Through continuous iterative optimization, it has established significant advantages in detection accuracy (mAP), inference speed (FPS), and cross-scenario generalization, evolving from YOLOv1 to YOLOv8 [63–66]. To address performance bottlenecks, researchers have explored several key areas. In terms of network architecture, they introduced CSPNet for feature reuse and optimized multi-scale feature fusion with BiFPN [67,68]. For loss function design, SIoU Loss improved bounding box regression by incorporating angle penalties, while Focal-EIoU Loss enhanced model convergence by dynamically adjusting weights for hard samples [69,70]. In terms of training strategies, the combination of Mosaic data augmentation and Self-Adversarial Training (SAT) [71,72], along with Exponential Moving Average (EMA) model updates, effectively improved the generalization ability of the model [73,74].

Tu et al. [75] proposed the YOLO-GEAR algorithm for detecting metal gear defects. By designing a lightweight C2f-Faster module, they reduced the model's parameters and computations, boosting its efficiency. The integration of an EMA attention module and a BiFPN structure further enhanced both detection speed and accuracy, meeting industrial standards for efficiency (FPS > 30). Yang et al. [76] used visual saliency-based image extraction to eliminate irrelevant features and reduce complexity. They replaced the neck network's feature pyramid with a weighted BiFPN to enhance multi-scale adaptability and efficiency, combined CBAM with C3 modules to form the CBAM-C3 attention module for better detection of small defects, and optimized hyperparameters with an improved sparrow algorithm. Experiments showed that the SF-YOLO model achieved 98.01% accuracy and an F1 score of 0.99 on metal gear end-face defect test sets, with an average detection time of 0.025 s per image, making it highly effective for real-time online detection. Zhang et al. [77] introduced an industrial gear surface defect detection method based on an improved YOLOv5 network. By incorporating CBAMC3 attention modules and analyzing the BiFPN_concat module, the network efficiently extracted and fused defect features. Using a cosine annealing function to adjust the learning rate improved the network's learning ability. Compared to the original YOLOv5, the improved version showed a 13.1% increase in recall and a 12% increase in mAP 0.5, processing 25 frames per second. Ma et al. [78] proposed the YOLO-CHD model by enhancing the YOLOv8 framework. Adding a C2f layer after the initial convolution in the backbone network preserved small-target feature details, improving detection capability. The feature fusion network integrated shallow, large-feature maps and used ASFF-4H for detecting tiny targets, significantly boosting efficiency. Experimental results demonstrated that the improved model achieved an mAP of 73.3%, a 3.3% increase from the original model, with an inference speed of 42 frames per second, effectively meeting the demands of rough machining quality inspection for gear surface defects and providing a more efficient and precise solution for gear quality detection in production.

3.2. Gear Defect Detection Based on Region Proposal Network Integrated Models

3.2.1. Gear Detection Based on Faster R-CNN Architecture

In gear defect detection, Faster R-CNN addresses the low efficiency of candidate region generation in traditional R-CNN by introducing a Region Proposal Network (RPN)—it abandons traditional selective search, generates high-quality candidate regions directly on convolutional feature maps, significantly reducing computational costs and improving detection speed [79]. and further optimizes overall performance when combined with end–to–end training. This feature enables it to accurately adapt to the needs of gear detection scenarios: when faced with gear datasets containing hundreds to thousands of

images (covering samples of different models and wear levels), it can efficiently complete batch detection tasks; even if there are light fluctuations such as strong light reflection and shadow occlusion in the production line, it can still reduce the missed detection of tiny defects like 0.1 mm microcracks through accurate candidate region localization [80]. Ultimately, it achieves dual improvements in detection speed and accuracy, meeting the practical needs of industrial batch detection.

In practical applications, Indasyah et al. [81] used a Faster RCNN-based method to detect various gear surface defects. Under static conditions (single defect, stationary conveyor), accuracy reached 86%; for multiple defects (two types), it was 76%. Even with the conveyor moving at 2–8.3 RPM, accuracy remained at 83.32%, showing strong robustness. Allam et al. [82] integrated domain knowledge with Faster R-CNN and trained the model with detailed boundary annotations. On a test set of 30 gears, the model achieved 88% precision and 86% recall (see Figure 4 for the model structure), proving highly effective for gear defect detection, reducing manual inspection efforts. Miltenovic et al. [83] built a deep-learning model using Faster R-CNN, trained with high-resolution images and precise labels, achieving high-precision gear defect detection and classification, especially in pitting detection, providing key support for gear maintenance. Girshick et al. [84] proposed R-CNN, combining region proposals and CNN. Using high-capacity CNNs to process bottom-up region proposals, it enables accurate object localization and segmentation. With supervised pre-training on auxiliary tasks and fine-tuning for specific domains, model performance improved significantly. On the PASCAL VOC 2012 dataset, mAP exceeded previous best results by over 30%, reaching 53.3%, highlighting R-CNN's excellence in object detection and semantic segmentation.

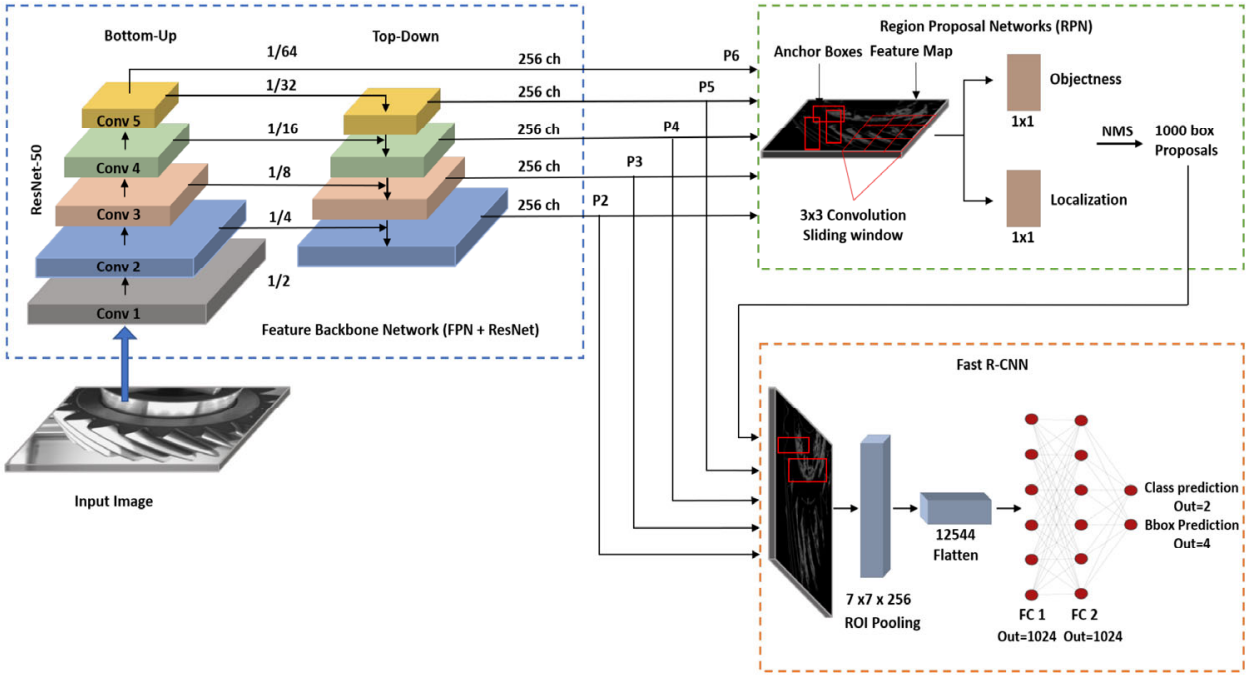


Figure 4. Structure of the Faster R-CNN deep learning network for defect detection [82].

3.2.2. Gear Defect Detection Based on Mask R-CNN Architecture

Mask R-CNN, a benchmark two-stage instance segmentation framework, offers unique advantages for the fine detection of gear surface defects. It innovates by decoupling instance segmentation into combined optimization for object detection and pixel-level mask prediction, outputting defect categories, locations, and geometric contours via a parallel branch architecture [85]. Unlike traditional semantic segmentation methods, its

use of the RoIAlign layer instead of the RoI Pooling layer effectively eliminates spatial quantization errors in feature maps (reducing the error by more than 50%) [86]. This allows it to precisely capture sub-pixel edge features of gear surface defects like pitting and wear. Compared with Faster R-CNN, Mask R-CNN adds a mask prediction branch on the basis of the existing classification and bounding box regression branches. This mask branch uses a fully convolutional network (FCN) to generate a binary mask for each region proposal, which can accurately segment the defect area at the pixel level. Therefore, Mask R-CNN can not only detect the position and category of gear defects but also obtain the detailed shape and size information of defects, which is more conducive to the quantitative analysis of gear defects.

Xi et al. [87] proposed a Mask R-CNN method for gear pitting detection, integrating multi-path feature fusion and a dual attention mechanism. As shown in Figure 5, this approach enhances feature representation by combining features from different levels through a well-designed strategy. The dual attention mechanism (including channel attention and spatial attention) can focus on the key features of pitting areas, and its recognition accuracy is improved by 22.7% compared with traditional segmentation methods. The implementation involves meticulous image preprocessing, efficient multi-path feature extraction, precise dual attention mechanism application, accurate pitting area segmentation, and comprehensive defect evaluation. Experimental results show that this method (DAMF Mask R-CNN) outperforms traditional models like Mask R-CNN, Mask Scoring R-CNN, U-Net, and Deeplabv3+.

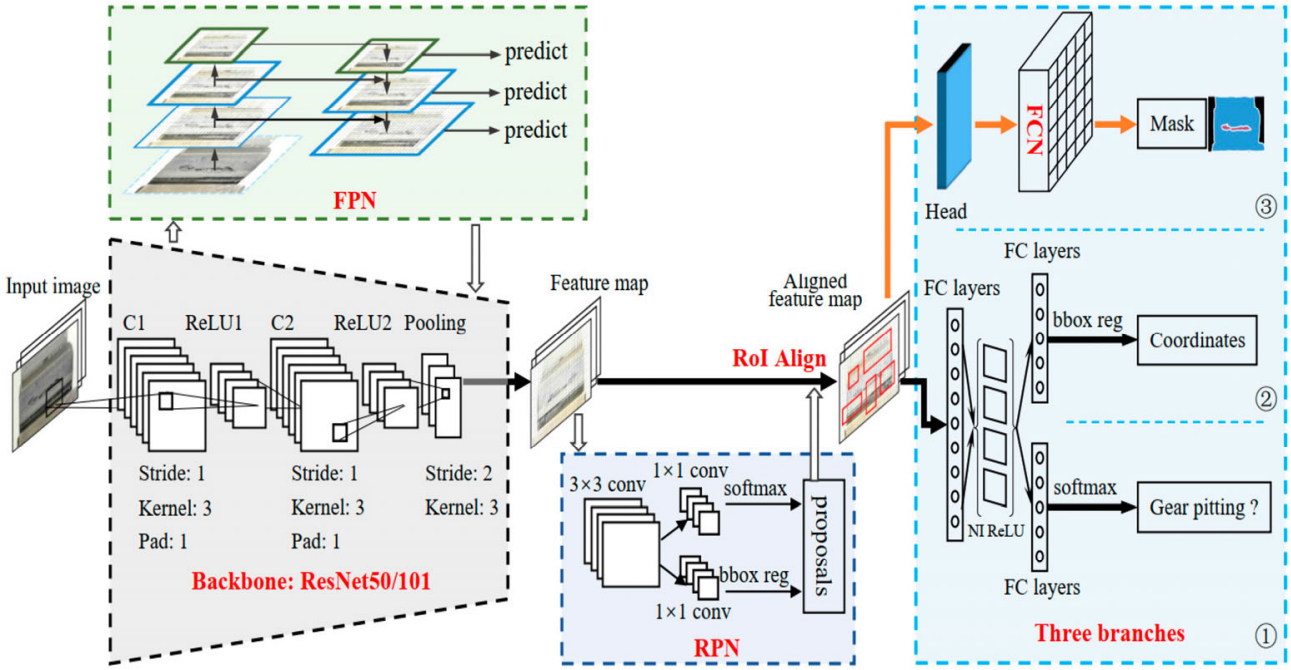


Figure 5. Structure of a Convolutional Neural Network Based on Depth Mask Region (Mask R-CNN) [87].

Xi et al. [88] introduced a deep Mask R-CNN-based method for gear pitting visual measurement. Using CNNs for automatic feature extraction, this method achieves high-precision detection of gear pitting. Extensive experiments across various scenarios demonstrate an average PSP of 88.2%. Some specific performance metrics of these algorithm models are listed in Table 1. The meanings and calculation formulas of these indicators can be found in Appendix A.

Table 1. Relevant data and performance metrics for gear defect detection.

Ref.	Year	Model	Original Model	Dataset	Experimental Results	Advantages	Disadvantages
[55]	2022	SR-ResNetYOLO	ResNet21	610 images for training and validation, and 100 images for testing	Recall 97.07%, mAP 96.66%	Improved robustness and accuracy of defect detection	Long training time and complex model structure
[57]	2024	Improve the ResNet101 network	ResNet101	10,080 images in the dataset for training and 4320 images in the dataset for testing	Recall 96.54%, Precision 97.33%, Accuracy 96.78%	Excellent feature extraction capability and adaptability	Weak ability to detect small defects
[59]	2023	MSSA U-Net	U-Net	A training set of 2000 images and a test set of 200 images	Recall 89.82%, Precision 87.97%	Excellent performance in multi-scale feature extraction	Slow training and reasoning due to high model complexity
[64]	2024	YOLOv5-CDG	YOLOv5	1206 images for Training, 206 images for Validation, and 397 images for Testing	Accuracy 99.46%, Average Precision 97%	Strong real-time detection capability, suitable for industrial applications	Dependent on background conditions and requires pre-optimization processing
[77]	2023	Improved YOLOv5	YOLOv5	1200 images of gears with defects	Recall 76.7%, mAP 86.3%, Precision 91.6%	Enhanced detection rate for small defects and good resistance to noise interference	High model complexity and long training time
[82]	2021	Faster R-CNN	Faster R-CNN	1405 images in the training set and 306 images in the test set	Recall 72%, Precision 95%	Wide applicability, capable of handling various types of gear defects	Requires a large amount of labeled data and is difficult to train
[87]	2020	Deep Mask R-CNN	Mask R-CNN	1050 images in the training set and 450 images in the test set	Recall 87.9%	Strong scenario adaptability	Long model training time

To sum up, The ResNet series has significant advantages in deep feature extraction and is suitable for detecting complex defects (such as the coexistence of multiple types of defects), while U-Net excels in defect segmentation and is applicable to scenarios that require quantifying defect sizes (e.g., pitting corrosion depth measurement); the YOLO series, by contrast, performs optimally in real-time detection scenarios (such as high-speed detection on production lines), and Faster R-CNN/Mask R-CNN offer high accuracy but low speed, a characteristic that makes them suitable for offline high-precision detection (e.g., factory quality inspection of gears).

4. Intelligent Algorithms for Addressing Challenges in Vision Algorithms

In the field of gear defect detection, although the technology has evolved from early image processing to the deep learning stage, both traditional vision algorithms and traditional deep learning methods still struggle to meet the high-precision detection requirements in complex industrial scenarios [89–91]. Traditional vision algorithms rely on manually designed features and have weak resistance to light and background interference, while traditional deep learning methods suffer from problems such as high computational resource

consumption, dependence on large-scale labeled data, and poor generalization in few-shot scenarios [92,93]. Currently, the core breakthrough direction of researchers focuses on exploring new intelligent technical paths: reducing reliance on hardware computing power through lightweight network architecture design (e.g., model pruning, parameter quantization); addressing the pain point of scarce labeled data by combining few-shot learning; and simultaneously enhancing the model's anti-interference ability with multi-modal fusion (e.g., integration of visual and vibration signals) and attention mechanisms. These efforts form a targeted solution that effectively breaks through existing technical bottlenecks and promotes the upgrading of gear defect detection technology toward higher precision and better adaptability.

4.1. Research on Gear Defect Detection Based on Cross-Modal Feature Alignment

In complex industrial settings, traditional unimodal detection methods, such as those relying solely on visual images or vibration signals, show clear limitations [94]. Frequent lighting changes can severely affect the quality of visual images, making it hard for vision-based methods to accurately extract defect features. Meanwhile, noise interference can significantly impact vibration signals, reducing the accuracy of signal analysis [95,96]. Additionally, the diversity of gear defects increases detection difficulty, as single-modal data often fails to fully reflect defect characteristics (e.g., visual images cannot capture internal stress changes caused by defects, and vibration signals cannot show the specific shape of surface defects). To address these challenges, cross-modal feature alignment technology has become a research focus. By integrating multisource data (e.g., visual images, vibration signals, and thermographic data) and leveraging the strong global modeling ability of Transformer-based multimodal networks, this technology significantly enhances the robustness and accuracy of gear defect detection [97,98].

To implement cross-modal feature alignment based on Transformer networks, the visual signal is first converted into a sequence of visual feature tokens through patch embedding while the vibration signal is turned into a sequence of vibration feature tokens via time-series segmentation and feature extraction, then the multi-head attention mechanism within the Transformer is employed to calculate the attention weight between the visual feature token sequence and the vibration feature token sequence thereby establishing the global correlation between these two modalities, and finally cross-modal feature vectors are obtained through layer normalization and feed-forward network operations which enables the joint embedding of visual signals. Lu et al. [99] addressed the limitations of traditional deep learning methods in capturing global dependencies across working conditions and generalization ability by proposing a method that integrates Vision Transformers (ViT) [100] with multi-head attention mechanisms. They utilized the global modeling capability of ViT to capture long-range signal dependencies and combined multi-head attention mechanisms to weight and fuse features from multi-source data. Through preprocessing techniques such as signal normalization and time-frequency transformation, time-series data was converted into image-based representations to adapt to ViT input. A lightweight network was designed with domain generalization strategies to reduce computational costs and achieve robust cross-condition diagnosis. Qin et al. [101] tackled the issues of reduced feature extraction capability due to patch embedding (composed of a single convolutional layer with large stride) and high computational costs from complex decoders in Transformer models for semantic segmentation. They proposed a Progressive Downsampling Transformer (PDCDT) framework based on a convolutional decoder. This framework optimizes feature extraction and reduces information loss through progressive downsampling layers and constructs a simple decoder based on convolutional modules for dimension transformation and information interaction. Experiments on gear pitting

showed that PDCDT performed excellently on the ADE20K dataset (achieving 47.9% mIoU) and the Cityscapes dataset (achieving 82.6% mIoU), significantly improving the accuracy of pitting detection. Sun et al. [102] proposed a parallel integrated model combining Swin Transformer and multi-scale convolutions to address potential issues of insufficient global information acquisition, inefficient local feature extraction, and high computational complexity in diagnostics. They designed a cloud-edge collaborative framework. In this model, a spatial interaction block was introduced in the Swin Transformer branch to enhance global information acquisition, while depthwise separable convolutions with different dilation rates were used in the multi-scale convolutional neural network branch to extract local features and reduce computational complexity. A feature interaction module was added between the two branches to improve diagnostic performance. Experimental results showed that this method achieved a diagnostic accuracy of 99.55%.

4.2. Research on Gear Defect Detection Using Lightweight Models

Traditional deep learning models, despite achieving some success in gear defect detection, face significant deployment challenges in resource-constrained industrial equipment (e.g., edge devices with limited CPU/GPU resources and memory) due to their high computational complexity and large parameter sizes [103,104]. To meet the industrial demand for efficient and low-resource-consuming detection models, many scholars have proposed lightweight model architectures for gear defect detection. These architectures incorporate advanced techniques such as depthwise separable convolutions (MobileNet), channel shuffle (ShuffleNet), and dynamic network pruning. Without compromising detection accuracy, they substantially reduce computational costs (reducing parameters by 40–50%), better adapting the models to the resource limitations of industrial equipment.

Table 2 shows the comparison of parameters, FLOPs, and FPS of different lightweight models on the gear defect dataset. It can be seen that compared with the traditional ResNet50 model, the lightweight models such as MobileNetV3, ShuffleNetV2, and YOLOv5s have significantly fewer parameters and FLOPs, and higher FPS, which are more suitable for deployment on edge devices.

Table 2. Comparison of parameters, FLOPs, and FPS of different models [105–107].

Model	Accuracy (%)	Parameters/10^6	GFLOPs/10^9	FPS (Frames/Second)
ResNet50	96.2	25.6	4.1	15
YOLOv5s	96.5	7.2	1.5	45
YOLOv5s + ShuffleNetv2	95.1	3.8	0.3	71
YOLOv5s + MobileNetv3	95.8	4.3	0.4	60

Shen et al. [105] proposed an improved YOLOv5s network model (VSD-YOLOv5s) based on the ShuffleNetV2 backbone. By replacing the C3 backbone with ShuffleNetV2, the size of the model is reduced by 45.5% and the FPS is reduced by 4 fps. The Squeeze–and–Excitation (SE) attention mechanism is introduced to adaptively adjust the weight of each channel feature, enhancing the model's attention to defect features. The Non-Maximum Suppression (NMS) is upgraded to Distance IoU NMS (DIOU-NMS). Compared with traditional NMS, DIOU-NMS not only considers the overlap area between bounding boxes but also considers the distance between the centers of the bounding boxes and the aspect ratio of the bounding boxes. This can effectively solve the problem of missing detection when bounding boxes overlap and improve the detection accuracy by 1.7%. According to the relevant specifications in GB/T 42980 [108] "Performance Evaluation Methods for Machine Vision Systems," the proposed method not only meets the stringent requirements for speed (FPS > 50) and accuracy (mAP > 95%) in the online detection of

surface defects on injection-molded gears in terms of recognition accuracy, precision, and detection speed, but also provides stable and reliable technical support for efficient quality inspection in production lines. Yan et al. [106] introduced the lightweight STMS-YOLOv5 algorithm, which uses ShuffleNetv2 as the backbone to reduce computational load and parameters(parameters reduced by 44.4%, FLOPs reduced by 50.3%). The STMS-YOLOv5 algorithm also introduces a spatial-temporal multi-scale feature fusion module to enhance the fusion of multi-scale defect features and uses a lightweight attention mechanism to improve the model's ability to detect small defects. Experiments on gear surface defect datasets showed that STMS-YOLOv5 maintains high detection accuracy (mAP 97.8%) with faster speed (FPS 16.3) and lower resource requirements (memory usage < 1GB), offering an efficient and lightweight solution for industrial gear defect detection with significant practical value and potential for widespread use in production lines. Wang et al. [107] addressed the low detection accuracy of YOLOv8s for small gear tooth defects by proposing several effective model optimizations. They incorporated the Convolutional Block Attention Module (CBAM) to focus on defect features and ignore irrelevant information, added a dedicated small-object detection layer to retain critical defect information lost during downsampling, and replaced the YOLOv8s backbone with MobileNetV3 (see Figure 6 for the optimized architecture) to reduce parameters (reduced by 40.2%) and computations (reduced by 45.3%). These measures enabled the improved model to excel in small defect detection (detection accuracy of small defects improved by 13.5%), providing a more precise solution for gear tooth surface defect inspection.

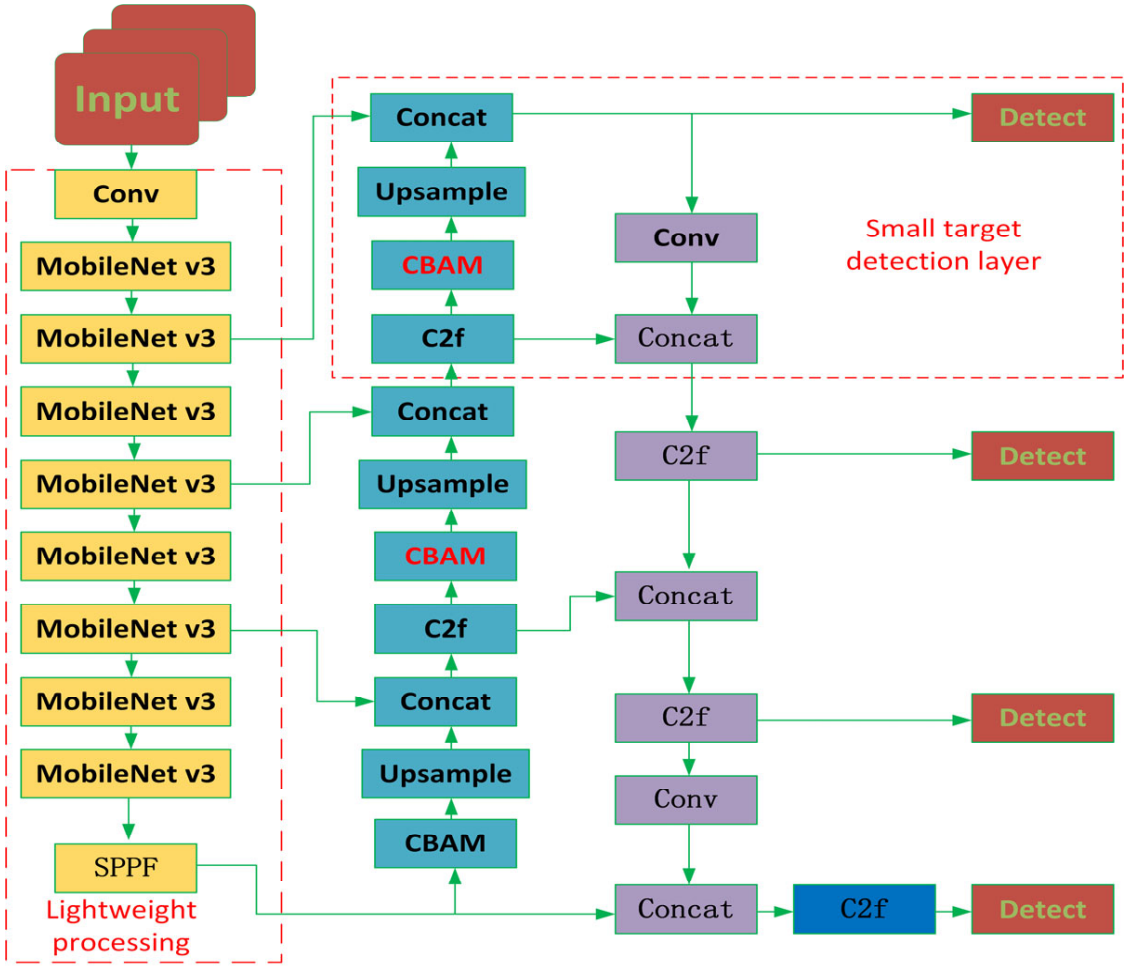


Figure 6. Structure of the optimized YOLOv8s model [107].

4.3. Research on Gear Defect Detection for Solving Few-Shot Problems

In industrial production, gear defect detection faces the critical bottleneck of few-shot problems. Defective samples (especially rare types such as microcracks and early pitting) are scarce and expensive to label. The number of normal samples far exceeds that of defective samples, which leads traditional deep learning models to be biased towards the majority class and perform poorly in detection. In addition, defect characteristics vary across production lines and gear types, which requires models to quickly adapt to new tasks. To address this issue, research has focused on two complementary technical paths—GAN-based data augmentation (to expand scarce defective samples) and few-shot learning (to enable models to learn efficiently with limited samples)—and this forms a closed-loop solution for few-shot scenarios.

4.3.1. GAN-Based Data Augmentation

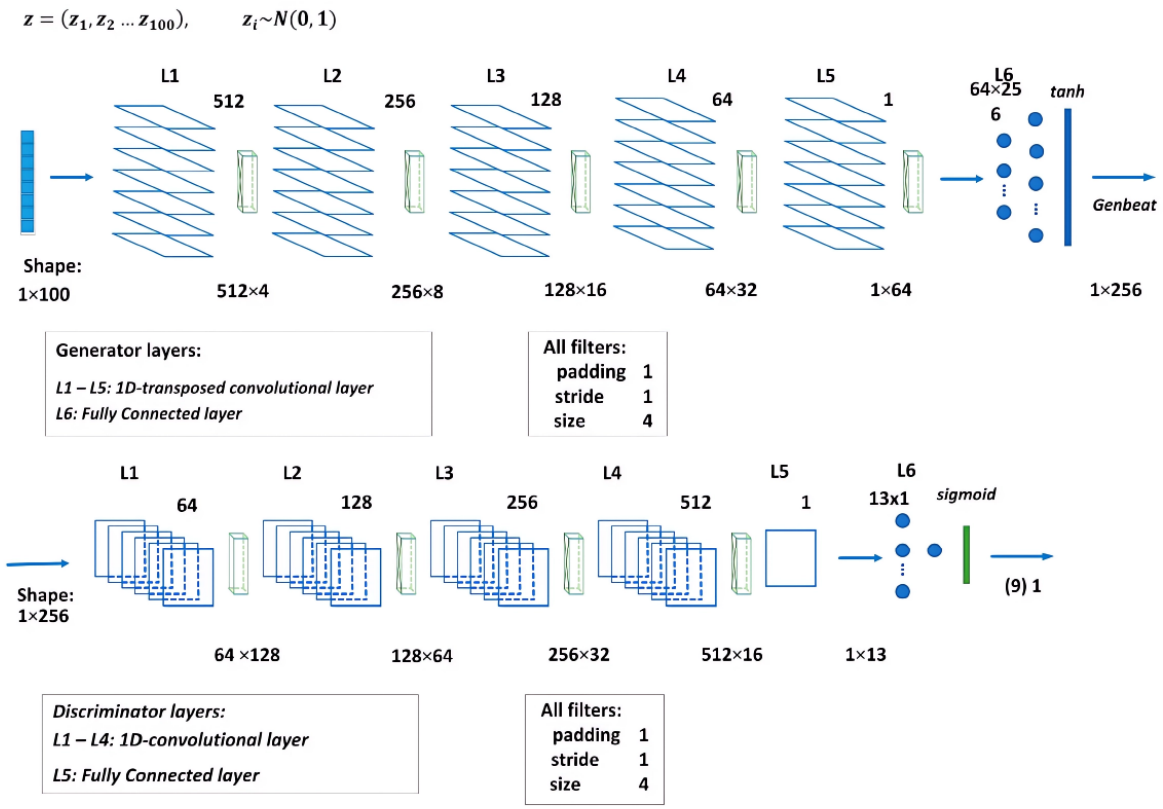
Generative Adversarial Networks (GANs) [109] and their variants fundamentally solve the "sample scarcity" problem by generating high-quality virtual defective samples with authentic physical properties. They first learn the feature distribution of a small number of real defective samples, generate diverse virtual samples, and then combine these virtual samples with real samples to construct an expanded training set, laying the foundation for subsequent few-shot learning. The typical GAN variants widely used in gear defect detection are Deep Convolutional Generative Adversarial Networks (DCGANs) [110] and Cycle-Consistency Generative Adversarial Networks (CycleGANs) [111].

Gear Defect Detection Based on Deep Convolutional Generative Adversarial Network (DCGAN) Architecture

DCGAN [112] is an improved variant of GAN that inherits the core framework of GAN's "adversarial training between generator and discriminator." By introducing convolutional layers (replacing fully connected layers to reduce parameters), Batch Normalization (accelerating training convergence and improving stability), and specific activation functions (using ReLU in the generator and LeakyReLU in the discriminator) [109], it addresses the issues of training instability and low-quality image generation in original GANs, making it better suited for image-related tasks. In gear defect detection, DCGAN is primarily used to tackle the challenges of scarce defect samples and complex defect features.

The architecture of DCGAN is shown in Figure 7. The generator of DCGAN takes a random noise vector as input and generates gear defect images through a series of deconvolution operations. The discriminator takes real gear images (including normal and defective images) and generated images as input and distinguishes them through a series of convolution operations. During the training process, the generator and discriminator play a game against each other, and finally reach a Nash equilibrium, so that the generator can generate high-quality gear defect images.

To address the common need for large amounts of specifically labeled training data when building surrogate models, Zhou et al. [113] utilized Deep Convolutional Generative Adversarial Networks (DCGAN) and incorporated unlabeled data based on semi-supervised learning concepts to enhance performance. By properly designing the architecture of the discriminator and generator in DCGAN, balanced confrontation is achieved, enabling accurate diagnosis even when labeled data is scarce. More crucially, leveraging the rich fault features in unlabeled data, the semi-supervised learning strategy of DCGAN can uncover the physical connections between unknown and known faults, achieving diagnostic generalization for unknown faults outside the training set.



Layer	Generator	Discriminator
1	ConvTr1d(100x512),BN,L-ReLU(0.2)	Conv1d(1,64),L-ReLU(0.2)
2	ConvTr1d(512,256),BN,L-ReLU(0.2)	Conv1d(64,128),L-ReLU(0.2)
3	ConvTr1d(256,128),BN,L-ReLU(0.2)	Conv1d(128,256),L-ReLU(0.2)
4	ConvTr1d(128,64),BN,L-ReLU(0.2)	Conv1d(256,512),L-ReLU(0.2)
5	ConvTr1d(64,1),BN,L-ReLU(0.2)	Conv1d(512,1),FC(13,1),sigmoid
6	FC(64,256),tanh	

Figure 7. DCGAN Network Framework.

Gear Defect Detection Using CycleGAN Architecture

CycleGAN is another important variant of GAN, which is mainly used for unpaired image-to-image translation tasks. In the field of gear defect detection, introducing cyclic consistency loss enables bidirectional domain translation of unpaired images, such as converting normal gear images to defective gear images [114]. The cyclic consistency loss ensures that the image obtained by translating from domain A to domain B and then translating back to domain A is consistent with the original image, which improves the quality and authenticity of the generated images. In practical applications, cross-domain translation can transform normal gear images into images containing specified defects (e.g., wear, broken teeth) without requiring actual defective samples [115]. Defect localization is achieved by generating residual maps through inverse mapping to identify abnormal regions. This approach eliminates the need for strict image alignment, making it suitable for multi-view gear data augmentation captured by industrial cameras. By overcoming the difficulty of obtaining real defective samples, this method enhances data diversity (increasing the number of defect samples by 5–10 times), precisely captures subtle defects (detecting defects with a size of less than 0.1 mm), and processes multi-view and diverse image data, thereby improving the robustness and generalization ability of the detection model [116].

Compared with DCGAN, the main differences between CycleGAN and DCGAN in architecture and target are as follows: In terms of architecture, DCGAN has only one generator and one discriminator, while CycleGAN has two generators and two discriminators, which are used for bidirectional domain translation, respectively. In terms of target, DCGAN mainly aims to generate realistic images that are similar to the real data distribution, while CycleGAN mainly aims to realize the translation between two different do-mains and ensure the cyclic consistency of the translation process.

Qin et al. [117] proposed an image enhancement method based on Tree CycleGAN, combined with Maximum Diversity Loss, for gear pitting detection. This method uses Generative Adversarial Networks (GANs) to generate high-quality sample data, thereby enhancing data diversity and improving the robustness and detection performance of the model. Experimental results demonstrate that this method performs exceptionally well in gear pitting detection, achieving a precision of 87.98% and a recall of 83.64%. This indicates that the method effectively identifies pitting areas on gear surfaces, thereby improving detection accuracy and robustness. The method excels in gear pitting detection and is suitable for high-precision inspection, providing critical support for gear maintenance and servicing. By generating high-quality sample data, this method significantly enhances the robustness and detection performance of the model.

4.3.2. Few-Shot Learning

Few-shot learning solves the problem of "efficient learning with limited samples" by changing the traditional deep learning paradigm. Its core logic is to first pre-train the model on multiple similar few-shot tasks to master general "defect feature extraction capabilities", and then quickly fine-tune it with a small number of labeled samples in the target gear detection scenario to finally achieve high-precision detection [118–122]. Common few-shot learning methods include meta-learning and transfer learning combined with physical simulation.

Meta-Learning

The core of meta-learning is to construct a "meta-task set"—packaging few-shot data of different gear types and defect categories into multiple independent tasks. The model learns a general strategy of "how to quickly adapt to new tasks" through alternating training on these tasks, enabling it to converge quickly when facing target few-shot tasks [123,124]. Typical meta-learning methods applied to gear defect detection include Graph Neural Networks (GNNs), Matching Networks, and Relation Networks. Yang et al. [125] proposed a GNN-based method for evaluating the severity of gear tooth fracture in small datasets, which is a typical application of meta-learning in few-shot gear defect detection. The specific process is as follows: first, feature conversion—converting gear vibration signals into 2D time-frequency maps through Short-Time Fourier Transform (STFT), and then using a Convolutional Neural Network (CNN) to extract features related to fault severity; second, meta-task training—constructing multiple "few-shot tooth fracture severity evaluation tasks" (each task contains a small number of samples with different fracture severities), and using GNN to learn the "similarity measurement between sample features" to classify fracture levels (mild, moderate, severe); third, target task adaptation—only 10 labeled samples are used for fine-tuning in the target gear dataset, and the model accuracy can reach 94.34%, which is 5–10% higher than that of traditional Siamese Networks and Matching Networks. This method does not rely on large-scale labeled data; instead, it enables the model to master the general feature laws of gear defects (such as vibration signal mutations caused by tooth fracture and edge discontinuities in images) through "task-level pre-training", thus achieving rapid convergence in new few-shot scenarios. However,

meta-learning has high requirements for the "diversity of meta-task sets"; if the gear types and defect modes in pre-training tasks differ significantly from those in target tasks, the adaptation effect of the model will decrease significantly.

Transfer Learning Combined with Physical Simulation

To further improve the robustness of few-shot models, researchers combine physical simulation with transfer learning. This method uses physical simulation software to generate a large amount of "defect data with physical constraints" (such as data on gear crack propagation processes under different loads and speeds). The model is first pre-trained on the simulated dataset, and then the learned "physical knowledge" is transferred to real few-shot scenarios, reducing reliance on real labeled data. Cohen et al. [11] proposed a gear wear monitoring strategy based on the digital twin framework, which is a representative case of this technical path. The specific process is: first, physical simulation pre-training—generating gear vibration and image data with different wear degrees (e.g., 0.1 mm, 0.5 mm wear depth) through dynamic modeling and experimental analysis, and training a wear severity regression model; second, knowledge transfer—using a small number of real healthy samples and single-category fault samples to construct a mapping function between "simulated data and real data", and transferring the parameters of the pre-trained model to real scenarios; third, few-shot fine-tuning—only 50 or fewer real wear samples are used for fine-tuning, and the model can achieve wear detection and severity evaluation effects comparable to fully supervised models. The value of this method lies in that physical simulation data has the characteristics of "unlimited generation" and "controllable physical authenticity", allowing the model to learn the essential physical laws of gear defects (such as the positive correlation between wear amount and load, and the mechanical characteristics of crack propagation) during pre-training, rather than relying on surface image features. Therefore, the model can still maintain high generalization ability in few-shot scenarios. Experiments show that this "physical simulation + few-shot transfer" scheme can improve the detection accuracy of the model for unknown wear modes by 20–30%.

5. Conclusions

Gear defect detection via machine vision has evolved from traditional image processing and conventional deep learning to intelligent detection, significantly boosting efficiency and accuracy. While traditional image processing is limited by manual feature design (accuracy < 90%), conventional deep learning (e.g., ResNet, Faster R-CNN) solves the problem of automatic feature extraction, although it has high computational complexity and poor generalization ability in small-sample scenarios, so it is only applicable to batch detection scenarios with sufficient computing power and abundant samples (such as regular defect screening in automotive gear mass production lines); among intelligent detection methods, YOLOv5 performs best in balancing accuracy and speed. When combined with MobileNetv3, it can increase the Frames Per Second (FPS) by 4, making it suitable for edge devices and high-speed production lines [105]. Additionally, GAN-based data augmentation technologies (e.g., Tree CycleGAN) increase the recognition rate of small defects such as microcracks and early pitting by 15–20% [118], and cross-modal feature alignment technology improves the robustness by 10–12% in scenarios with variable illumination and high noise [99–102]. However, the current technology still needs to overcome bottlenecks such as insufficient modeling accuracy for defects caused by multi-physics field coupling and real-time response delay exceeding 50 ms under extreme working conditions.

Future research can take several directions. Integrating multi-source data into a multimodal sensing framework can enhance detection stability in complex conditions.

Combining self-supervised contrastive learning with weakly supervised generative adversarial networks can reduce reliance on labeled data. Using edge-cloud collaborative computing can enable low-latency real-time detection and dynamic model optimization. Employing interpretable tools can increase technical trustworthiness. Establishing industry-standard gear defect databases can provide benchmarks for algorithm evaluation and iteration. The integration of these technologies can potentially advance gear detection systems towards intelligence, supporting predictive maintenance and lifecycle management in smart manufacturing and driving industrial quality inspection from partial optimization to comprehensive empowerment.

Author Contributions: D.Z. is responsible for writing the review, and S.Z. is responsible for revising and typesetting the review. Y.Z. is responsible for resources. X.X. is responsible for investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant number No. 52275295) and the Henan Provincial Science and Technology Research Project (Grant number No. 242102230034).

Acknowledgments: This work was supported by the Zhengzhou University of Light Industry and the Henan Provincial Department of Science and Technology.

Conflicts of Interest: Author Xiaoguang Xu was employed by the company China Tobacco Henan Industrial Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

R-CNN	Region-Based Convolutional Neural Networks
DCGAN	Deep Convolutional Generative Adversarial Networks
BLCD	Bidimensional Local Characteristic-Scale Decomposition
LSCM	Laser Scanning Confocal Microscopy
SHGA-PSO	Simulated Annealing Hybrid Genetic Algorithm-Particle Swarm Optimization
CycleGAN	Cycle-Consistency Generative Adversarial Networks
FCN	Fully Convolutional Network
GBSU	Gear Surface Defect Detection U-Net (custom abbreviation for gear surface defect detection U-Net)
BiFPN	Bidirectional Feature Pyramid Network
SAT	Self-Adversarial Training
EMA	Exponential Moving Average
CBAM	Convolutional Block Attention Module
ASFF	Adaptively Spatial Feature Fusion
C2f	Cross-Stage Partial Fusion
ASEF-4H	Adaptive Spatial Feature Fusion-4 Head (custom abbreviation for adaptive spatial feature fusion with 4 heads)
MSSA	Multi-Scale Splicing Attention
ViT	Vision Transformer
PDCDT	Progressive Downsampling Convolutional Decoder Transformer
SE	Squeeze-and-Excitation
NMS	Non-Maximum Suppression
DIOU	Distance Intersection over Union
GNN	Graph Neural Network
FSL	Few-Shot Learning
STFT	Short-Time Fourier Transform

Appendix A

Table A1 presents the evaluation indicators for deep learning models involved in this paper, including their definitions, applicability, and calculation methods.

Table A1. Evaluation indexes for deep learning models.

Evaluation Index	Definition and Applicability	Calculation Method
Precision	The proportion of samples correctly predicted as positive by the model to all samples predicted as positive. Suitable for scenarios where false detection is costly, as it helps reduce unnecessary false positives.	$\text{Precision} = \frac{TP}{TP+FP}$
Accuracy	The percentage of total sample that the model correctly predicts. It is suitable for cases where the number of samples without defects and the number of samples with defects are relatively balanced. For a problem with unbalanced categories, it may not effectively reflect the model performance.	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
mAP	The average accuracy of each class in a multi-class or multi-label problem. It is suitable for multi-class defect detection tasks to evaluate the overall model performance.	1. For each class, calculate the area under the Precision-Recall curve; 2. Calculate the arithmetic mean of the APs of all classes to obtain mAP.
Recall	The proportion of samples correctly predicted by the model as positive as a percentage of all samples that are actually positive. Suitable for scenarios with a high cost of omission, as it can detect all defects as far as possible.	$\text{Rec}all = \frac{TP}{TP+FN}$
IoU	The ratio of the intersection area of the predicted area and the real area to the union area, which is used to measure the degree of overlap between the output box of the object detection algorithm and the real box. It is suitable for defect detection tasks with strict boundary requirements.	$IoU = \frac{Area_{Intersection}}{Area_{Union}}$
mIoU	Calculated for each category after the IoU is averaged. It is suitable for comprehensive evaluation of multi-class image segmentation tasks.	1. For each class, calculate the IoU between the predicted region and the real region of that class; 2. Calculate the arithmetic mean of the IoUs of all classes to obtain mIoU.

References

1. Hu, M.Z.; Wang, H.X.; Wei, P.T.; Liu, G.S.; Zhang, L.; He, Z.Q.; Liu, H.J. Multi-objective optimization of a co-rotating twin-screw gear transmission system based on heuristic search. *J. Mech. Sci. Technol.* **2023**, *37*, 5831–5841. [CrossRef]
2. Lu, Z.H.; Reitschuster, S.; Tobie, T.; Stahl, K.; Liu, H.J.; Hu, X.L. Contact fatigue life prediction of PEEK gears based on CTAB-GAN data augmentation. *Eng. Fract. Mech.* **2024**, *312*, 110639. [CrossRef]
3. Chen, T.M.; Zhu, C.C.; Chen, J.X.; Liu, H.J. A review on gear scuffing studies: Theories, experiments and design. *Tribol. Int.* **2024**, *196*, 109741. [CrossRef]
4. Lu, W.D.; Zhang, X.L.; Jiang, X.; Hong, T.Y. Research on point cloud processing of gear 3D measurement based on line laser. *J. Braz. Soc. Mech. Sci. Eng.* **2024**, *46*, 645. [CrossRef]
5. Yao, H.M.; Yu, W.Y.; Wang, X. A feature memory rearrangement network for visual inspection of textured surface defects toward edge intelligent manufacturing. *IEEE Trans. Autom. Sci. Eng.* **2023**, *20*, 2616–2635. [CrossRef]
6. Liu, R.L.; Zhong, D.X.; Lyu, H.Q.; Han, J.Q. A Bevel Gear Quality Inspection System Based on Multi-Camera Vision Technology. *Sensors* **2016**, *16*, 1364. [CrossRef]
7. Xin, Q.Y.; Pei, Y.C.; Luo, M.Y.; Wang, Z.Q.; He, L.; Liu, J.Y.; Wang, B.; Lu, H.Q. A generalized precision measuring mechanism and efficient signal processing algorithm for the eccentricity of rotary parts. *Mech. Syst. Signal Process.* **2023**, *204*, 110791. [CrossRef]
8. Pei, Y.C.; Xie, H.L.; Tan, Q.C. A non-contact high precision measuring method for the radial runout of cylindrical gear tooth profile. *Mech. Syst. Signal Process.* **2020**, *138*, 106543. [CrossRef]

9. Yang, W.Q.; Wang, M.H.; Tang, C.; Zheng, X.; Liu, X.W.; He, K.L. Trustworthy multi-view clustering via alternating generative adversarial representation learning and fusion. *Inf. Fusion* **2024**, *107*, 102323. [CrossRef]
10. Kou, R.K.; Wang, C.P.; Yu, Y.; Peng, Z.M.; Yang, M.B.; Huang, F.Y.; Fu, Q. LW-IRSTNET: Lightweight infrared small target segmentation network and application deployment. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5621313. [CrossRef]
11. Cohen, R.; Bachar, L.; Matania, O.; Bortman, J. Few-shot learning for estimating gear wear severity towards digital twinning. *Eng. Fail. Anal.* **2025**, *170*, 109330. [CrossRef]
12. Zhao, J.; Shi, Y.H.; Tan, F.; Wang, X.F.; Zhang, Y.Q.; Liao, J.; Yang, F.; Guo, Z.H. Research on an intelligent diagnosis method of mechanical faults for small sample data sets. *Sci. Rep.* **2022**, *12*, 21996. [CrossRef]
13. Chen, W.Y.; Tsao, Y.R.; Lai, J.Y.; Hung, C.J.; Liu, Y.C.; Liu, C.Y. Real-Time Instance Segmentation of Metal Screw Defects Based on Deep Learning Approach. *Meas. Sci. Rev.* **2022**, *22*, 107–111. [CrossRef]
14. Liu, B.; Yang, B.; Zhao, Y.L.; Li, J.Q. Low-pass U-Net: A segmentation method to improve strip steel defect detection. *Meas. Sci. Technol.* **2023**, *34*, 035405. [CrossRef]
15. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
16. Zhang, X.H.; Wang, H.; Wang, C.Z.; Liu, M.; Xu, G.W. Time-segment-wise feature fusion transformer for multi-modal fault diagnosis. *Eng. Appl. Artif. Intell.* **2024**, *138*, 109358. [CrossRef]
17. Yuan, H.Y.; Cheng, J.P.; Wu, Y.R.; Zeng, Z.Y. Low-res MobileNet: An efficient lightweight network for low-resolution image classification in resource-constrained scenarios. *Multimed. Tools Appl.* **2022**, *81*, 38513–38530. [CrossRef]
18. Wang, C.C.; Chiu, C.T.; Chang, H.Y. EfficientNet-eLite: Extremely lightweight and efficient cnn models for edge devices by network candidate search. *J. Signal Process. Syst. Signal Image Video Technol.* **2023**, *95*, 657–669. [CrossRef]
19. Achmad, B.; Dhomas, H.F. Lightweight Models for Real-Time Steganalysis: A Comparison of MobileNet, ShuffleNet, and EfficientNet. *J. Resti* **2024**, *8*, 737–747. [CrossRef]
20. Inkawhich, N. A Global Model Approach to Robust Few-Shot SAR Automatic Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 4004305. [CrossRef]
21. Zhao, J.K.; Wang, H.X.; Peng, S.L.; Yao, Y.D. Meta supervised contrastive learning for few-shot open-set modulation classification with signal constellation. *IEEE Commun. Lett.* **2024**, *28*, 837–841. [CrossRef]
22. Farooq, U.; Singh, P.; Kumar, A. A systematic review of quantum image processing: Representation, applications and future perspectives. *Comput. Sci. Rev.* **2025**, *57*, 100763. [CrossRef]
23. Smalley, D.; Lough, S.D.; Holtzman, L.; Xu, K.K.; Holbrook, M.; Rosenberger, M.R.; Hone, J.C.; Barmak, K.; Ishigami, M. Detecting atomic-scale surface defects in STM of TMDs with ensemble deep learning. *MRS Adv.* **2024**, *9*, 890–896. [CrossRef]
24. Guo, L.G.; Wu, S.T. FPGA Implementation of a real-time edge detection system based on an improved canny algorithm. *Appl. Sci.* **2023**, *13*, 870. [CrossRef]
25. Peng-O, T.; Chaikan, P. High performance and energy efficient sobel edge detection. *Microprocess. Microsyst.* **2021**, *87*, 104368. [CrossRef]
26. Chen, Y.; Zhang, L.B.; Zhan, H.W.; Xu, F. Registration algorithm for printed images incorporating feature registration and deformation optimization. *Signal Image Video Process.* **2025**, *19*, 176. [CrossRef]
27. Hosseinian, T.; Saeidi, R.; Motamedi, S.A.; Abdollahifard, M.J.; Mansoori, R. Kaze-sar: Sar image registration using kaze detector and modified surf descriptor for tackling speckle noise. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5207612. [CrossRef]
28. Li, Y.Q.; Zhang, D.H. Toward efficient edge detection: A novel optimization method based on integral image technology and canny edge detection. *Processes* **2025**, *13*, 293. [CrossRef]
29. Kyriakides, G.; Daskalakis, A. A Study of Lighting Models for Automatic Detection of Regions of Interest and Surface Defects in Cast Metal Parts. *Open Access Libr. J.* **2023**, *10*, e10833. [CrossRef]
30. Shao, W.; Shao, Y.Q.; Liu, Q. High-speed and accurate method for the gear surface integrity detection based on visual imaging. In Proceedings of the 2021 International Conference of Optical Imaging and Measurement (ICOIM), Xi'an, China, 27–29 August 2021. [CrossRef]
31. Liu, D.X.; Cheng, J.S.; Wu, Z.T. Bidimensional local characteristic-scale decomposition and its application in gear surface defect detection. *Meas. Sci. Technol.* **2023**, *35*, 025115. [CrossRef]
32. Chang, H.C.; Borghesani, P.; Smith, W.A.; Peng, Z.X. Application of surface replication combined with image analysis to investigate wear evolution on gear teeth. *Wear* **2019**, *430*, 355–368. [CrossRef]
33. Wang, Y.; Wu, Z.H.; Duan, X.Y.; Tong, J.G.; Li, P.; Chen, M.; Lin, Q.L. Design of Gear Defect Detection System Based on Machine Vision. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *108*, 022025. [CrossRef]
34. Wang, Y.; Cheng, Y.J. An Approach to Fault Diagnosis for Gearbox Based on Image Processing. *Shock Vib.* **2016**, *2016*, 5898052. [CrossRef]
35. Saeedi, J.; Dotta, M.; Galli, A.; Nasciuti, A.; Maradia, U.; Boccadoro, M.; Gambardella, L.M.; Giusti, A. Measurement and inspection of electrical discharge machined steel surfaces using deep neural networks. *Mach. Vis. Appl.* **2021**, *32*, 21. [CrossRef]

36. Awtoniuk, M.; Majerek, D.; Myziak, A.; Gajda, C. Industrial Application of Deep Neural Network for Aluminum Casting Defect Detection in Case of Unbalanced Dataset. *Adv. Sci. Technol. Res. J.* **2022**, *16*, 120–128. [CrossRef]
37. Park, J.; Shin, J.; Choi, B. Reduction of False Positives for Runtime Errors in C/C++ Software: A Comparative Study. *Electronics* **2023**, *12*, 3518. [CrossRef]
38. Wang, S.Y.; Duan, P.H. Mesh stiffness calculation of defective gear system under lubrication with automated assessment of surface defects using convolutional neural networks. *Mech. Syst. Signal Process.* **2024**, *216*, 111445. [CrossRef]
39. Shafiq, M.; Gu, Z.Q. Deep Residual Learning for Image Recognition. *Appl. Sci.* **2022**, *12*, 8972. [CrossRef]
40. Jaeger, B.E.; Schmid, S.; Grosse, C.U.; Goegelein, A.; Elischberger, F. Infrared Thermal Imaging-Based Turbine Blade Crack Classification Using Deep Learning. *J. Nondestruct. Eval.* **2022**, *41*, 74. [CrossRef]
41. Olaf, R.; Philipp, F.; Thomas, B. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015. [CrossRef]
42. Liao, D.H.; Cui, Z.H.; Zhang, X.; Li, J.; Li, W.J.; Zhu, Z.X.; Wu, N.X. Surface defect detection and classification of Si3N4 turbine blades based on convolutional neural network and YOLOv5. *Adv. Mech. Eng.* **2022**, *14*, 16878132221081580. [CrossRef]
43. Chen, Z.H.; Juang, J.C. YOLOv4 Object Detection Model for Nondestructive Radiographic Testing in Aviation Maintenance Tasks. *Aiaa J.* **2022**, *60*, 526–531. [CrossRef]
44. Li, X.B.; Wang, W.Q.; Sun, L.H.; Hu, B.; Zhu, L.; Zhang, J.C. Deep learning-based defects detection of certain aero-engine blades and vanes with DDSC-YOLOv5s. *Sci. Rep.* **2022**, *12*, 13067. [CrossRef]
45. Liu, Y.X.; Wu, D.B.; Liang, J.W.; Wang, H. Aeroengine Blade Surface Defect Detection System Based on Improved Faster RCNN. *Int. J. Intell. Syst.* **2023**, *2023*, 1992415. [CrossRef]
46. Zhang, J.; Nong, C.R.; Zhang, H.B.; Zhang, Y.Z. Engine Blade Fault detection Technology based on deep learning. *Aeronaut. Engines* **2022**, *48*, 68–75. Available online: https://link.cnki.net/doi/10.13477/j.cnki.aeroengine.2022.01.011 (accessed on 1 January 2025). (In Chinese)
47. Shang, H.B.; Sun, C.; Liu, J.X.; Chen, X.F.; Yan, R.Q. Deep learning-based borescope image processing for aero-engine blade in-situ damage detection. *Aerosp. Sci. Technol.* **2022**, *123*, 107473. [CrossRef]
48. Shang, H.B.; Sun, C.; Liu, J.X.; Chen, X.F.; Yan, R.Q. Defect-aware transformer network for intelligent visual surface defect detection. *Adv. Eng. Inform.* **2023**, *55*, 101882. [CrossRef]
49. Upadhyay, A.; Li, J.; King, S.; Addepalli, S. A Deep-Learning-Based Approach for Aircraft Engine Defect Detection. *Machines* **2023**, *11*, 192. [CrossRef]
50. Yu, L.Y.; Wang, Z.; Duan, Z.J. Detecting Gear Surface Defects Using Background-Weakening Method and Convolutional Neural Network. *J. Sens.* **2019**, *2019*, 3140980. [CrossRef]
51. Jiang, J.B.; Cao, P.; Lu, Z.C.; Lou, W.M.; Yang, Y.Y. Surface Defect Detection for Mobile Phone Back Glass Based on Symmetric Convolutional Neural Network Deep Learning. *Appl. Sci.* **2020**, *10*, 3621. [CrossRef]
52. Jose, J.; Deepa, O.S.; Saimurugan, M.; Krishnakumar, P.; Praveenkumar, T. Fault Diagnosis of Gearbox using Machine Learning and Deep Learning Techniques. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 3940–3943. [CrossRef]
53. Wu, T.E.; Huang, S.H.; Lai, C.H. Helical Gear Defect Detection System Based on Symmetrized Dot Pattern and Convolutional Neural Network. *IEEE Access* **2024**, *12*, 171328–171333. [CrossRef]
54. He, K.; Zhang, X.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
55. Su, Y.T.; Yan, P.; Yi, R.Z.; Chen, J.; Hu, J.H.; Wen, C. A cascaded combination method for defect detection of metal gear end-face. *J. Manuf. Syst.* **2022**, *63*, 439–453. [CrossRef]
56. Han, M.; Wu, Q.X.; Zeng, X.J. Visual Detection of Tiny Defects in Gears Based on Deep Learning. *Comput. Syst. Appl.* **2020**, *29*, 100–107. Available online: http://www.c-s-a.org.cn/1003-3254/7323.html (accessed on 2 January 2025). (In Chinese)
57. Bao, C.W.; Jiang, W.; Liu, Y.Z.; Xiao, Q.L.; Wu, J. Gear defect detection based on improved ResNet101 network. *Comb. Mach. Tool Autom. Process. Technol.* **2024**, *8*, 145–153. Available online: https://link.cnki.net/doi/10.13462/j.cnki.mmtamt.2024.08.028 (accessed on 3 January 2025). (In Chinese)
58. Zhou, X.; Zhang, Y.C.; Ren, Z.H.; Mi, T.C.; Jiang, Z.Y.; Yu, T.Z.; Zhou, S.H. A Unet-inspired spatial-attention transformer model for segmenting gear tooth surface defects. *Adv. Eng. Inform.* **2024**, *62*, 102933. [CrossRef]
59. Qin, Y.; Xi, D.J.; Chen, W.W.; Wang, Y. Gear Pitting Measurement by Multi-Scale Splicing Attention U-Net. *Chin. J. Mech. Eng.* **2023**, *36*, 50. [CrossRef]
60. Wang, Z.W.; Qin, Y.; Chen, W.W. Vision measurement of gear pitting based on DCGAN and U-Net. *J. Mech. Sci. Technol.* **2021**, *35*, 2771–2779. [CrossRef]
61. Dong, L.; Chen, W.F.; Yang, S.Y.; Yu, H.Y. Automated detection of gear tooth flank surface integrity: A cascade detection approach using machine vision. *Measurement* **2023**, *220*, 113375. [CrossRef]

62. Flores-Calero, M.; Astudillo, C.A.; Guevara, D.; Maza, J.; Lita, B.S.; Defaz, B.; Ante, J.S.; Zabala-Blanco, D.; Moreno, J.M.A. Traffic Sign Detection and Recognition Using YOLO Object Detection Algorithm: A Systematic Review. *Mathematics* **2024**, *12*, 297. [CrossRef]
63. Hou, W.Q.; Jing, H.C. RC-YOLOv5s: For tile surface defect detection. *Vis. Comput.* **2024**, *40*, 459–470. [CrossRef]
64. Jia, H.K.; Zhou, H.M.; Chen, Z.H.; Gao, R.K.; Lu, Y.; Yu, L.D. Research on Bearing Surface Scratch Detection Based on Improved YOLOV5. *Sensors* **2024**, *24*, 3002. [CrossRef]
65. Zhang, S.W.; Zhong, Z.Y.; Zhu, D.H. Detection Method for Surface Defects of Metal Gears Based on Improved YOLOx Network. *Laser Optoelectron. Prog.* **2023**, *60*, 280–290.
66. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [CrossRef]
67. Wang, C.Y.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023. [CrossRef]
68. Tan, M.X.; Pang, R.M.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [CrossRef]
69. Zhora, G. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**. [CrossRef]
70. Zhu, K.B.; Lyu, H.; Qin, Y.B. Enhanced detection of small and occluded road vehicle targets using improved YOLOv5. *Signal Image Video Process.* **2025**, *19*, 168. [CrossRef]
71. Tong, Y.; Luo, X.F.; Ma, L.Y.; Xie, S.R.; Yang, W.B.; Guo, Y.S. Saliency information and mosaic based data augmentation method for densely occluded object recognition. *Pattern Anal. Appl.* **2024**, *27*, 34. [CrossRef]
72. Wei, C.H.; Bai, L.F.; Chen, X.Y.; Han, J. Cross-modality data augmentation for aerial object detection with representation learning. *Remote Sens.* **2024**, *16*, 4649. [CrossRef]
73. Anitha, S.; Jayanthi, P.; Thangarajan, R. Detection of replica node attack based on exponential moving average model in wireless sensor networks. *Wirel. Pers. Commun.* **2020**, *115*, 1651–1666. [CrossRef]
74. Chang, B.R.; Tsai, H.F.; Chang, F.Y. Applying Advanced Lightweight Architecture DSGSE-Yolov5 to Rapid Chip Contour Detection. *Electronics* **2024**, *13*, 10. [CrossRef]
75. Tu, F.Q.; Qi, Y.Q.; Liu, J.; Wang, S.F. Surface Defect Detection Algorithm of Metal Gears Based on Improved YOLOv8s. *J. Comput. Mod.* **2025**, *353*, 100–106. Available online: https://kns.cnki.net/kcms2/article/abstract?v=uXGtp3S0eCAPlIlubnb-eVNLTq2L-Tm2M-jZLHqnOc1MAmaEIUPkLWf5X3cKv9qn8pK4TFYCMkbTEnRWJQILDhkel7LaSSb47bIRHHAZs7LQ6DyrRTgXVHEntoxCsSG8I_fMquBt6FdG4n0vF7vTaRKmmBab7rNdV_KV2GRfF_Ez1scGQx5uotNNR-twcdofnj2KLtvx6lY=&uniplatform=NZKPT&language=CHS (accessed on 3 January 2025). (In Chinese)
76. Yang, S.A.; Zhou, L.; Wang, C.; Wang, S.H.; Tang, Y. SF-YOLO: An evolutionary deep neural network for gear end surface defect detection. *IEEE Sens. J.* **2024**, *24*, 21762–21775. [CrossRef]
77. Zhang, S.W.; He, M.Q.; Zhong, Z.Y.; Zhu, D.H. An industrial interference-resistant gear defect detection method through improved YOLOv5 network using attention mechanism and feature fusion. *Measurement* **2023**, *221*, 113433. [CrossRef]
78. Qian, C.R.; Su, L.C.; Li, Y.W. Steel surface defect detection based on improved YOLOv8. In Proceedings of the International Conference on Advanced Image Processing Technology (AIPT 2024), Chongqing, China, 31 May–2 June 2024; p. 13257. [CrossRef]
79. Xu, L.Y.; Li, B.Y.; Mi, H.; Lu, X.Z. Improved Faster R-CNN algorithm for defect detection in powertrain assembly line. *Procedia CIRP* **2020**, *93*, 479–484. [CrossRef]
80. Zhou, Z.Z.; Lu, Q.H.; Wang, Z.F.; Huang, H.J. Detection of Micro-Defects on Irregular Reflective Surfaces Based on Improved Faster R-CNN. *Sensors* **2019**, *19*, 5000. [CrossRef] [PubMed]
81. Indasyah, E.; Ibrahim, F.; Syahbana, D.F.; Istiqomah, F. Automated Visual Inspection System of Gear Surface Defects Detection Using Faster RCNN. In Proceedings of the 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), Surabaya, Indonesia, 14–15 November 2023. [CrossRef]
82. Allam, A.; Moussa, M.; Tarry, C.; Veres, M. Detecting Teeth Defects on Automotive Gears Using Deep Learning. *Sensors* **2021**, *21*, 8480. [CrossRef] [PubMed]
83. Miltenovic, A.; Rakonjac, I.; Oarcea, A.; Peric, M.; Rangelov, D. Detection and Monitoring of Pitting Progression on Gear Tooth Flank Using Deep Learning. *Appl. Sci.* **2022**, *12*, 5327. [CrossRef]
84. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014. [CrossRef]
85. Ma, Y.C.; Fu, H.L.; Wu, P.; Chen, X.H.; Wang, D.; Chen, S.; Cao, C.Y. Application Research of Mask R-CNN Model with Adaptive Optimization of Deep Network in Surface Defect Detection of Castings. *Mod. Manuf. Eng.* **2022**, *4*, 112–118. Available online: https://link.cnki.net/doi/10.16731/j.cnki.1671-3133.2022.04.016 (accessed on 5 January 2025). (In Chinese)

86. Zhang, Y.Q.; Chu, J.; Leng, L.; Miao, J. Mask-Refined R-CNN: A network for refining object details in instance segmentation. *Sensors* **2020**, *20*, 1010. [CrossRef]
87. Xi, D.J.; Qin, Y.; Wang, Y.Y. Vision Measurement of Gear Pitting Under Different Scenes by Deep Mask R-CNN. *Sensors* **2020**, *20*, 4298. [CrossRef]
88. Xi, D.J.; Qin, Y.; Luo, J.; Pu, H.Y.; Wang, Z.W. Multipath Fusion Mask R-CNN With Double Attention and Its Application Into Gear Pitting Detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5006011. [CrossRef]
89. Kansal, S.; Tripathi, R.K. Adaptive geometric filtering based on average brightness of the image and discrete cosine transform coefficient adjustment for gray and color image enhancement. *Arab. J. Sci. Eng.* **2020**, *45*, 1655–1668. [CrossRef]
90. Shi, Z.Y.; Fang, Y.M.; Wang, X.Y. Research Progress on Gear Visual Inspection Instruments and Technology. *Adv. Laser Optoelectron.* **2022**, *59*, 1415006. Available online: https://link.cnki.net/urlid/31.1690.tn.20220714.1314.367 (accessed on 5 January 2025). (In Chinese)
91. Shi, Z.Y.; Li, M.C.; Sun, Y.Q.; Yu, B. Development of gear line laser three-dimensional measuring instrument. *Chin. J. Sci. Instrum.* **2024**, *45*, 95–103. Available online: https://link.cnki.net/doi/10.19650/j.cnki.cjsi.J2312129 (accessed on 6 January 2025). (In Chinese)
92. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
93. Yadav, S.P.; Zaidi, S.; Mishra, A.; Yadav, V. Survey on Machine Learning in Speech Emotion Recognition and Vision Systems Using a Recurrent Neural Network (RNN). *Arch. Comput. Methods Eng.* **2021**, *29*, 1753–1770. [CrossRef]
94. Ravimal, D.; Kim, H.; Koh, D.; Hong, J.H.; Lee, S.K. Image-based inspection technique of a machined metal surface for an unmanned lapping process. *Int. J. Precis. Eng. Manuf. -Green Technol.* **2020**, *7*, 547–557. [CrossRef]
95. Feng, K.; Ji, J.C.; Ni, Q.; Beer, M. A review of vibration-based gear wear monitoring and prediction techniques. *Mech. Syst. Signal Process.* **2023**, *182*, 109605. [CrossRef]
96. Mohammed, O.D.; Rantatalo, M.; Aidanpaa, J.; Kumar, U. Vibration signal analysis for gear fault diagnosis with various crack progression scenarios. *Mech. Syst. Signal Process.* **2013**, *41*, 176–195. [CrossRef]
97. Xu, K.; Wang, B.J.; Zhu, Z.; Jia, Z.H.; Fan, C.C. A Contrastive Learning Enhanced Adaptive Multimodal Fusion Network for Hyperspectral and LiDAR Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4700319. [CrossRef]
98. Zheng, S.P.; Liu, J.F.; Jun, Z. MCAFNet: Multiscale cross-modality adaptive fusion network for multispectral object detection. *Digit. Signal Process.* **2025**, *159*, 104996. [CrossRef]
99. Lu, J.Y.; Ji, W.X.; Yu, J.J.; Zhang, C.Y. Data driven deep learning fault diagnosis method based on vision transformer and multi-head attention for different working condition. *Eng. Res. Express* **2025**, *7*, 015205. [CrossRef]
100. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021. [CrossRef]
101. Qin, Y.; Wang, S.; Xi, D. Liang Progressive downsampling transformer with convolution-based decoder and its application in gear pitting measurement. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5008709. [CrossRef]
102. Sun, X.C.; Ding, H.; Li, N.; Dong, X.X.; Sun, J.C.; Zheng, G.Y. Intelligent fault diagnosis method for shearer rocker gear based on swin transformer and multiscale convolution parallel integration. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3519816. [CrossRef]
103. Wang, C.J.; Wang, Y.F. SLGA-YOLO: A Lightweight Castings Surface Defect Detection Method Based on Fusion-Enhanced Attention Mechanism and Self-Architecture. *Sensors* **2024**, *24*, 4088. [CrossRef]
104. Chen, X.; Wu, Y.L.; He, X.Y.; Ming, W.Y. A Comprehensive Review of Deep Learning-Based PCB Defect Detection. *IEEE Access* **2023**, *11*, 139017–139038. [CrossRef]
105. Shen, F.; Zhou, M.; Huang, Z.L.; Li, X.Y.; Zhang, M.Z. Lightweight Detection of Injection Molded Gear Defects Based on VSD-YOLOv5s. *Modul. Mach. Tool Autom. Manuf. Tech.* **2024**, *4*, 145–148. Available online: https://link.cnki.net/doi/10.13462/j.cnki.mmtamt.2024.04.030 (accessed on 7 January 2025). (In Chinese)
106. Yan, R.; Zhang, R.Y.; Bai, J.Q.; Hao, H.J.; Guo, W.J.; Gu, X.Y. STMS-YOLOv5: A Lightweight Algorithm for Gear Surface Defect Detection. *Sensors* **2023**, *23*, 5992. [CrossRef]
107. Wang, H.T. Research on Gear Defect Detection Method Based on Machine Vision. Master's Thesis, Tianjin Polytechnic Normal University, Tianjin, China, 25 September 2024. (In Chinese)
108. *GB/T 42980-2023*; Standardization Administration of the People's Republic of China. Test method for intelligent manufacturing machine vision online inspection system; Standards Press of China: Beijing, China, 2023.
109. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680. [CrossRef]
110. Li, Y.X.; Wang, K.S. An Automatic FIR and DCGAN Model-based Fault Detection Framework for Key Components of Planetary Gearboxes under compartively Stable Conditions. In Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 25–27 October 2019. [CrossRef]

111. Yu, Y.N.; Tang, L.H.; Liu, Z.P.; Xiang, J.W. A Novel Bearing Fault data generation strategy combining physical modeling and CycleGAN Variant for Fault Diagnosis Without Real Samples. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3505717. [CrossRef]
112. Liu, B.Q.; Lv, J.W.; Fan, X.Y.; Luo, J.; Zou, T.Y. Application of an Improved DCGAN for Image Generation. *Mob. Inf. Syst.* **2022**, *2022*, 9005552. [CrossRef]
113. Zhou, K.; Diehl, E.; Tang, J. Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations. *Mech. Syst. Signal Process.* **2023**, *185*, 109772. [CrossRef]
114. Wang, G.; Shi, H.B.; Chen, Y.F.; Wu, B. Unsupervised image-to-image translation via long-short cycle-consistent adversarial networks. *Appl. Intell.* **2023**, *53*, 17243–17259. [CrossRef]
115. Chen, R.C.; Fan, M.Z.; Manongga, W.E.; Sub-r-pa, C. Evaluating image-to-image translation techniques for simulating physical conditions of traffic signs. *J. Adv. Inf. Technol.* **2024**, *15*, 1019–1024. [CrossRef]
116. Shao, G.F.; Huang, M.; Gao, F.Q.; Liu, T.D.; Li, L.D. DuCaGAN: Unified Dual Capsule Generative Adversarial Network for Unsupervised Image-to-Image Translation. *IEEE Access* **2020**, *8*, 154691–154707. [CrossRef]
117. Qin, Y.; Wang, Z.W.; Xi, D.J. Tree CycleGAN with maximum diversity loss for image augmentation and its application into gear pitting detection. *Appl. Soft Comput.* **2022**, *114*, 108130. [CrossRef]
118. Sharma, G.; Kaur, T.; Mangal, S.K.; Kohli, A. Investigating bearing and gear vibrations with a Micro-Electro-Mechanical Systems (MEMS) and machine learning approach. *Results Eng.* **2024**, *24*, 103499. [CrossRef]
119. Niu, P.; Cheng, Q.; Zhang, X.L.; Liu, Z.F.; Zhao, Y.S.; Yang, C.B. Research on High-Precision Measurement Method for Small-Size Gears with Small-Modulus. *Sensors* **2024**, *24*, 5413. [CrossRef]
120. Borghesani, P.; Smith, W.A.; Zhang, X.; Feng, P.; Antoni, J.; Peng, Z. A new statistical model for acoustic emission signals generated from sliding contact in machine elements. *Tribol. Int.* **2018**, *127*, 412–419. [CrossRef]
121. Nico, H.; Peng, Z.X.; Pietro, B. Bridging the Trust Gap: Evaluating Feature Relevance in Neural Network-Based Gear Wear Mechanism Analysis with Explainable AI. *Tribol. Int.* **2023**, *187*, 108670. [CrossRef]
122. Mattera, G.; Polden, J.; Norrish, J. Monitoring the gas metal arc additive manufacturing process using unsupervised machine learning. *Weld. World* **2024**, *68*, 2853–2867. [CrossRef]
123. Ni, X.F.; Ma, Z.J.; Liu, J.W.; Shi, B.; Liu, H.L. Attention Network for Rail Surface Defect Detection via Consistency of Intersection-over-Union(IoU)-Guided Center-Point Estimation. *IEEE Trans. Ind. Inform.* **2022**, *18*, 1694–1705. [CrossRef]
124. Huang, Y.F.; Huang, Z.W.; Jin, T. DEU-Net: A Multi-Scale Fusion Staged Network for Magnetic Tile Defect Detection. *Appl. Sci.* **2024**, *14*, 4724. [CrossRef]
125. Yang, S.; Deng, C.Y.; Chuan, L. Assessment of gear failure severity in wind turbines based on few-shot learning and graph neural networks. *Eng. Res. Express* **2024**, *6*, 045586. [CrossRef]

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