machines

MDPI

Review

Advancements in Smart Nondestructive Evaluation of Industrial Machines: A Comprehensive Review of Computer Vision and AI Techniques for Infrastructure Maintenance

Samira Mohammadi [1,2,3], Sasan Sattarpanah Karganroudi [1,2,*] and Vahid Rahmanian [1]

[1] Centre National Intégré du Manufacturier Intelligent, Université du Québec à Trois-Rivières, 575 Boul de l'Université, Drummondville, QC J2C 0R5, Canada; samira.mohammadi2@uqtr.ca (S.M.); vahid.rahmanian@uqtr.ca (V.R.)
[2] Équipe de Recherche en Intégration CAO-Calcul, Department of Mechanical Engineering, Université du Québec à Trois-Rivières, 575 Boul de l'Université, Drummondville, QC J2C 0R5, Canada
[3] Departement of Mathemathics, Informmathematicsattics and Engineering, Université du Québec à Rimouski, 1595 Bd Alphonse-Desjardins, Lévis, QC G6V 0A6, Canada
* Correspondence: sattarpa@uqtr.ca

Abstract: Infrastructure maintenance is critical to ensuring public safety and the longevity of essential structures. Nondestructive Evaluation (NDE) techniques allow for infrastructure inspection without causing damage. Computer vision has emerged as a powerful tool in this domain, providing automated, efficient, and accurate solutions for defect detection, structural monitoring, and real-time analysis. This review explores the current state of computer vision in NDE, discussing key techniques, applications across various infrastructure types, and the integration of deep learning models such as convolutional neural networks (CNNs), vision transformers (ViTs), and hybrid models. The review also highlights challenges, including data availability and scalability. It proposes future research directions, including real-time monitoring and the integration of Artificial Intelligence (AI) with Internet of Things (IoT) devices for comprehensive inspections.

Keywords: nondestructive evaluation; structural health monitoring; defect detection; deep learning models; predictive maintenance

check for updates

Academic Editor: Hui Ma

Received: 21 November 2024
Revised: 17 December 2024
Accepted: 24 December 2024
Published: 28 December 2024

Citation: Mohammadi, S.; Sattarpanah Karganroudi, S.; Rahmanian, V. Advancements in Smart Nondestructive Evaluation of Industrial Machines: A Comprehensive Review of Computer Vision and AI Techniques for Infrastructure Maintenance. *Machines* **2025**, *13*, 11. https://doi.org/10.3390/machines13010011

Copyright: © 2024 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

Infrastructure is essential in modern society, supporting transportation, communication, and urban development. From bridges and highways to pipelines and buildings, the integrity and safety of these structures are essential [1]. The regular inspection and maintenance of such infrastructure are vital to prevent accidents, reduce maintenance costs, and extend the lifespan of structures [2]. Traditionally, these inspections have relied on manual methods such as visual inspections and basic testing techniques. However, these approaches are time-consuming, labor-intensive, and prone to human error [3]. Nondestructive Evaluation (NDE) refers to various techniques used to assess the condition of infrastructure without damaging structures [4]. These methods are crucial for evaluating critical infrastructure such as bridges, buildings, pipelines, and railways, as they allow for the early detection of defects and deterioration, thus preventing failures [5]. Traditional NDE methods include ultrasonic testing, radiography, thermography, and acoustic emission testing, which have been widely used for decades [6]. Each technique provides valuable insights: ultrasonic testing utilizes high-frequency sound waves to detect internal

defects [7], while radiography uses X-rays or gamma rays to identify flaws within a material [8]. Thermography detects temperature variations caused by defects, such as cracks or corrosion, by analyzing heat distribution [9].

However, while these methods are effective, they often require expert operators and can be time-consuming, particularly for large-scale infrastructure projects. As a result, there is an increasing demand for more efficient, scalable, and accurate inspection techniques that can process large datasets and provide real-time analysis. Although many infrastructure defects, such as corrosion or cracks, may progress slowly, real-time analysis remains advantageous for ensuring operational efficiency and proactive maintenance. By enabling on-the-fly detection and decision-making, real-time systems can promptly identify anomalies that, if left unaddressed, could lead to costly failures or safety risks under dynamic loading or environmental changes. Furthermore, real-time integration supports automated workflows, reduces inspection delays, and enhances scalability for large-scale infrastructure projects [10].

This need for modernization has driven the development of computer vision technologies, which are transforming NDE by automating the process of detecting surface and internal defects such as cracks, corrosion, and structural deformations using advanced imaging systems [11]. Computer vision applies image processing, pattern recognition, and machine learning to automate defect detection, significantly reducing the reliance on human inspection [12]. Modern deep learning algorithms enable automated systems to learn from vast datasets of infrastructure images, identifying even minor defects with high precision [13]. Recent studies highlight the integration of computer vision in various NDE applications, such as crack detection in bridges and corrosion detection in pipelines [14]. Moreover, 3D reconstruction techniques like photogrammetry and LiDAR, as shown in Figure 1, have been instrumental in providing detailed structural models that allow for more accurate geometric analysis and defect detection [15].

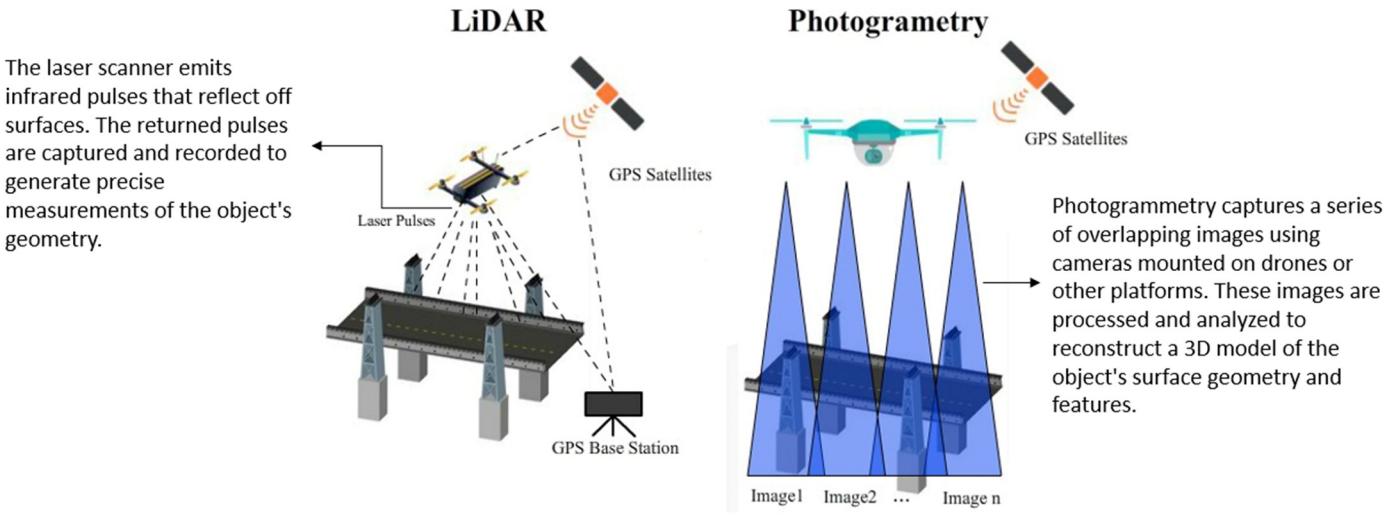


Figure 1. LiDAR uses laser pulses and GPS coordination to generate highly accurate, detailed 3D models of terrain or structures by measuring the time it takes for the laser pulses to return to the sensor. Photogrammetry involves capturing multiple images from different angles using GPS-coordinated drones and then processing these images to create a 3D model by analyzing photo overlaps.

Beyond identifying visible damage, there is a critical emphasis on detecting subtle indicators of potential failure before defects become apparent to the naked eye. This involves integrating computer vision with complementary sensing technologies like high-resolution thermal imaging, hyperspectral sensors, and vibration or strain measurement devices. By continuously comparing new sensor data to established "healthy" baselines,

these combined approaches can identify micro-fractures, stress distributions, temperature anomalies, and chemical changes in materials long before corrosion or cracks form. Such proactive monitoring enables infrastructure managers to take preventive measures early, mitigating risks and reducing long-term maintenance costs [10].

Computer vision-based techniques are transforming Nondestructive Evaluation (NDE), providing advanced defect detection and monitoring capabilities. Complementing these advancements, in situ sensor technologies—such as fiber optic, acoustic, and ultrasonic sensors—play a critical role in real-time, localized structural health monitoring. When integrated with computer vision and edge processing, these technologies create sophisticated hybrid systems capable of detecting a broader spectrum of defects, including subsurface anomalies and dynamic changes. While the primary focus is on vision-based methods, the potential of these hybrid systems is acknowledged as a promising frontier for future innovation. In situ sensors have emerged as indispensable tools in structural health monitoring, providing critical real-time data for early defect detection. Technologies such as fiber optic sensors, which measure strain and temperature variations, and acoustic sensors, capable of detecting minute vibrations and emissions, offer unparalleled insights into material behavior [16–18]. For example, fiber optic sensors embedded in bridge components can continuously monitor strain distributions, alerting operators to potential structural issues [19].

Integrating these sensors with computer vision techniques enhances the effectiveness of NDE by combining localized, real-time monitoring with automated defect detection. Paired with strain sensors, thermal imaging systems can analyze deformation under variable environmental conditions. Similarly, acoustic emission data can complement vision-based crack detection to identify and classify subsurface defects [20]. Furthermore, advancements in edge computing technologies enable real-time sensor data processing, reducing latency and improving decision-making during inspections [21,22]. The practical applications of these hybrid systems include monitoring pipelines for corrosion and cracks using ultrasonic sensors in conjunction with computer vision models or assessing bridges through strain sensors integrated with drone-based imaging systems. These combined approaches have shown significant potential for enhancing infrastructure inspections' accuracy, scalability, and cost efficiency, addressing limitations inherent to standalone methods [23].

The combination of computer vision and machine learning has proven advantageous for processing large-scale datasets typically gathered in modern infrastructure inspections. For instance, aerial drones with high-resolution cameras can collect thousands of images of large infrastructures such as bridges and power plants. These images are then processed by computer vision algorithms, which autonomously detect defects [24,25]. Machine learning models may be employed to improve the accuracy of defect classification based on the patterns found in the images. This approach enables more frequent inspections at lower costs, enhancing critical infrastructure's overall safety and reliability [26]. As infrastructure ages and urban environments expand, there is an ongoing need to scale inspection techniques. Automation through computer vision accelerates the inspection process and enhances accuracy, making it possible to identify defects that traditional methods could miss. This transition from manual to automated inspection has been crucial in meeting the increasing demand for rapid and comprehensive infrastructure monitoring [27]. Computer vision models have transformed NDE by enhancing the ability to detect, classify, and quantify defects. Convolutional neural networks (CNNs) have been used for infrastructure inspections, excelling in surface-level defect detection [28]. More recently, vision transformers (ViTs) have shown potential for analyzing large-scale datasets and processing complex structures with increased accuracy [29]. However, each

model has limitations: CNNs are typically more effective for local feature extraction, while transformers excel in capturing long-range dependencies [30].

Hybrid models, which combine the strengths of CNNs and transformers, have emerged as a promising approach to address these limitations. By leveraging the hierarchical feature extraction capabilities of CNNs and the global context understanding of transformers, hybrid models improve defect detection in complex infrastructure environments. For instance, integrating CNN backbones with transformer layers allows for better handling of high-resolution images, which is critical for detailed infrastructure inspections [31]. Swin Transformers, a variant of vision transformers, have recently gained attention for their ability to capture both local and global features through a shifted windowing mechanism. This hierarchical architecture makes Swin Transformers highly effective for high-resolution image processing, as they can perform both fine-grained defect detection and large-scale structural analysis. Their application in NDE is particularly promising for complex infrastructures like bridges and tunnels, where detailed analysis and scalability are both essential [32]. Figure 2 shows different stages of defect detection using deep learning models, including data collection, preprocessing, model training, and deployment, along with examples of detected defects such as cracks, missing fasteners, internal pipeline cracks, defective welds, surface defects, and corrosion.

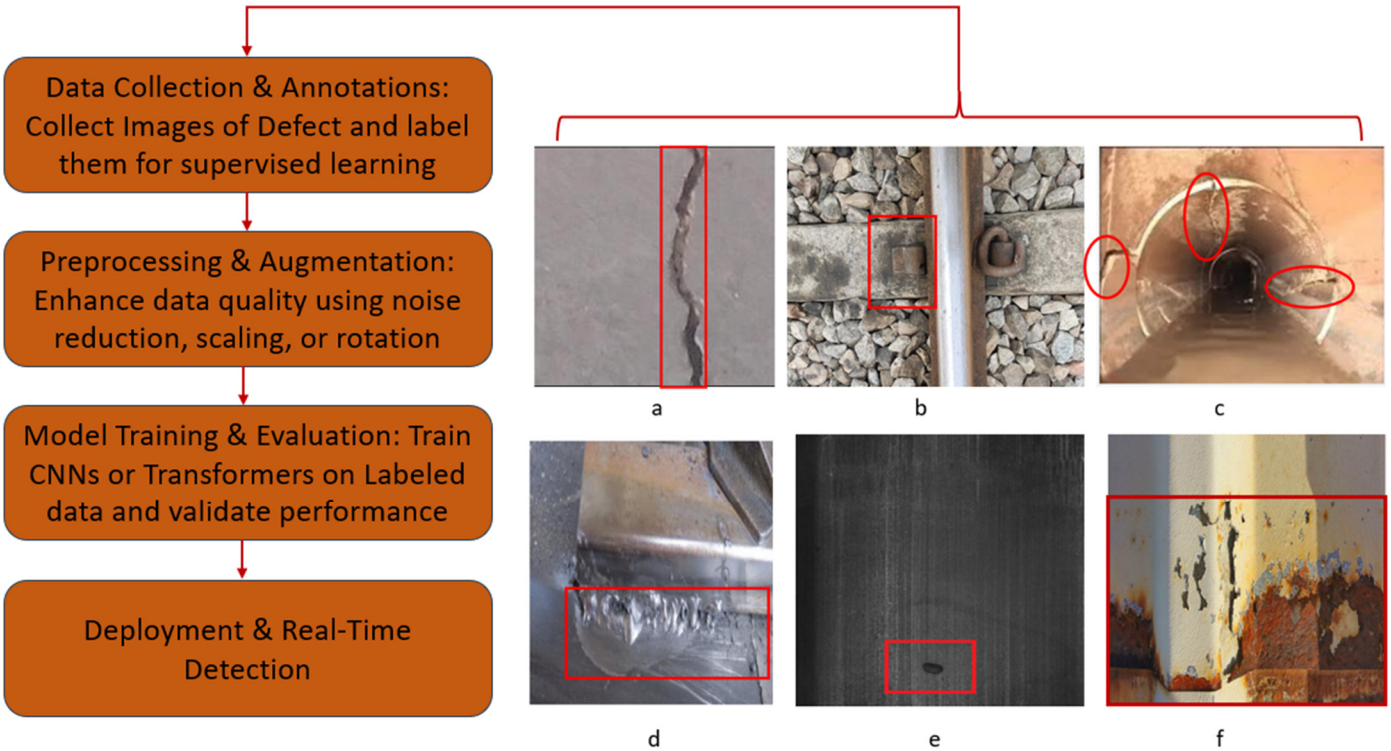


Figure 2. Overview of defect detection using deep learning models. (**a**) Crack on a concrete surface, (**b**) missing fastener on a railway track, (**c**) internal cracks in a sewer pipeline, (**d**) defective weld joint, (**e**) surface defect on a metallic component, and (**f**) corrosion and peeling on a painted metal surface.

Structural health monitoring (SHM) has emerged as a critical field for maintaining and ensuring the safety of civil infrastructure such as bridges, dams, tunnels, and other concrete structures. Traditional methods for detecting defects, including visual inspections and manual data analysis, are increasingly seen as insufficient due to their time-consuming nature, subjectivity, and inability to detect subsurface or minute defects [33]. Advances in deep learning, mainly through convolutional neural networks (CNNs), transformers, and hybrid models combining both approaches, have revolutionized the defect detection

process by providing automated, scalable, and more accurate solutions. This review explores the application of CNNs, transformers, and hybrid models in SHM, comparing their performance, strengths, limitations, and potential for future research. This review explores the role of computer vision, including hybrid models and transformers, in NDE infrastructure. It discusses key techniques, applications across different infrastructure types, and the integration of deep learning architectures like CNNs, vision transformers, and Swin Transformers. Additionally, the paper examines challenges, such as data availability and scalability, and proposes future directions for improving computer vision-based NDE systems.

Several prior reviews have established a solid foundation in the field of Nondestructive Evaluation (NDE). Early comprehensive works, such as those by Shull [34] and Balageas et al. [35], examined classical NDE techniques—ultrasonic testing, radiographic methods, thermography—and their integration into broader structural health monitoring frameworks. More recently, reviews have addressed the rise of AI and computer vision within NDE. For instance, Ye et al. [36] and Bao et al. [37] discussed machine vision-based approaches and deep learning applications for structural assessments, focusing mainly on infrastructures like bridges. Harley et al. [38] highlighted emerging machine and deep learning architectures for NDE in quality inspection and nondestructive testing contexts, while Masino et al. and Valeske et al. [39] explored AI-driven methods within the paradigm of Industry 4.0. Malekloo et al. [40] also provided a broad, data-driven perspective on NDT and SHM, indicating how AI techniques are progressively reshaping traditional inspection workflows.

While these recent surveys acknowledge the growing importance of machine learning and deep learning in NDE, most primarily focus on either conventional CNN-based frameworks, specific infrastructure types, or general machine vision approaches. They often do not delve deeply into the latest advancements in model architectures—such as vision transformers (ViTs) and hybrid CNN–transformer models—nor do they thoroughly examine the integration of these methods with Internet of Things (IoT) infrastructures, edge computing solutions, and large-scale sensor data fusion techniques. Moreover, gaps remain in discussing real-time analytics and scalability issues and adapting these new technologies to handle complex, dynamic environments.

This review offers a more comprehensive and future-oriented perspective on NDE by synthesizing the most up-to-date literature. We emphasize not only cutting-edge deep learning frameworks (including ViTs and hybrid models) but also the synergistic interplay between advanced computer vision algorithms, diverse sensing modalities, and real-time IoT-driven monitoring systems. Our review highlights current challenges—such as data scarcity, computational overhead, and the need for standardization—and maps out opportunities for accelerating innovation. This holistic approach ultimately provides researchers, practitioners, and decision-makers with clearer guidance on deploying and further developing next-generation NDE more accurate, efficient, and scalable solutions, thereby enhancing critical infrastructures' safety, reliability, and operational efficiency.

Unlike other recent reviews that focus predominantly on either the theoretical underpinnings of advanced NDE techniques or the performance characteristics of isolated AI models, the primary objective of this review is to holistically examine the synergy between computer vision, deep learning methodologies (including CNNs, ViTs, and hybrid models), and NDE applications across a broad range of infrastructure types. This review offers a more integrative and application-oriented perspective by critically comparing diverse model architectures, exploring their practical deployment challenges, and highlighting underrepresented areas—such as early, proactive damage detection through sensor-fusion and edge computing. It aims not only to elucidate the current state-of-the-art in AI-driven NDE

but also to identify the gaps, scalability issues, and data constraints that must be addressed to accelerate the real-world adoption and effectiveness of these emerging technologies.

2. Review Methodology

This review synthesizes recent advances in applying deep learning models for Nondestructive Evaluation (NDE) across essential infrastructure sectors, including civil structures, transportation systems, and industrial facilities. As infrastructure has aged and urban areas have expanded, the demand for accurate, efficient, and scalable inspection methods has grown. Thus, this review targets studies that leverage deep learning models—primarily convolutional neural networks (CNNs), vision transformers (ViTs), and hybrid approaches—to enhance defect detection, structural monitoring, and predictive maintenance in various infrastructure contexts.

We limited our scope to studies published within the past five years to ensure a focused and up-to-date review. These recent publications reflect advancements in computer vision techniques and their practical applications to NDE, with a particular focus on sectors where infrastructure stability and safety are critical. The studies included were selected based on their effectiveness in applying deep learning to detect and classify structural defects such as cracks, corrosion, material degradation, and misalignments. This scope allows for a concentrated assessment of the potential and limitations of each deep learning model within real-world infrastructure inspection scenarios.

The literature search was conducted through a systematic review of major academic databases, including IEEE Xplore, ScienceDirect, and Google Scholar, employing keywords such as “Non-destructive Evaluation”, “Deep Learning Models”, “Computer Vision in Infrastructure”, “Defect Detection”, “Convolutional Neural Networks”, and “Vision Transformers”. Each selected study met specific inclusion criteria to ensure relevance and rigor. First, only studies applying deep learning methods directly to NDE tasks were considered. Articles were required to provide quantitative performance metrics—such as accuracy, precision, recall, mean Intersection over Union (mIoU), and mean average precision (mAP)—that enabled an objective assessment and comparison of model effectiveness. Additionally, preference was given to studies that presented comparative analyses, highlighting the strengths and weaknesses of various deep learning models for specific infrastructure types. This focus on empirical evidence and comparative results allowed us to comprehensively evaluate each model’s practical contributions and limitations.

To organize the findings and draw meaningful comparisons, studies were categorized according to the type of deep learning model applied: CNN-based, transformer-based, or hybrid models. CNN models were examined for their ability to detect localized defects, which is especially useful in identifying surface-level issues such as cracks and corrosion. Transformer models, such as ViTs and Swin Transformers, were reviewed for their strength in capturing complex patterns and global features within large-scale datasets, making them suitable for infrastructure elements where large-scale monitoring is necessary. Hybrid models, which integrate the CNN and transformer architectures, were evaluated for their combined benefits of local feature extraction and global context recognition, particularly in complex or large infrastructure systems where both aspects are critical. Each study was analyzed based on several key performance metrics. Accuracy, precision, and recall were extracted to evaluate the models’ general effectiveness in detecting and classifying defects. For models that incorporated segmentation tasks, metrics like IoU and mAP were emphasized, given their importance in assessing segmentation quality and localization precision. Additionally, given the increasing demand for real-time defect detection in large-scale infrastructure inspections, studies detailing computational efficiency were prioritized, as these metrics provide insight into each model’s suitability for field deployment.

In selecting the models for analysis, this review adopted specific inclusion criteria focused on both the practicality of application scenarios and demonstrable performance metrics. Studies were only considered if they directly applied deep learning models—such as CNNs, vision transformers, or hybrid architectures—to real-world or realistically simulated NDE tasks, ensuring their relevance to infrastructure maintenance and defect detection. Additionally, each included model had to report quantitative performance indicators (e.g., accuracy, precision, recall, mIoU, mAP) to facilitate the objective comparison and assessment of their strengths and limitations. Preference was given to models tested in diverse conditions or on large, well-annotated datasets, as these factors better represent the challenges encountered in practical applications. By adhering to these selection criteria, this review offers an evidence-based perspective highlighting the most impactful and developed approaches within the rapidly evolving field of computer vision-driven NDE.

Preliminary keyword-based searches across multiple databases initially yielded a broad set of candidate publications. After removing duplicates and performing abstract-level screening focused on direct NDE applications and the presence of quantitative performance metrics, a more refined set of several dozen studies met our inclusion criteria for full-text review. Studies were excluded if they lacked clear performance indicators, did not apply deep learning directly to NDE tasks, or provided only theoretical insights without practical validation. Among the reviewed literature, bridge and pipeline inspections emerged as the most frequently studied practical applications, reflecting their critical roles in transportation infrastructure and energy distribution networks. Cracks, corrosion, and material degradation were the most commonly addressed defects, given their substantial impact on structural integrity and safety. Convolutional neural network (CNN)-based frameworks—often incorporating widely recognized backbones such as ResNet, VGG, and YOLO—dominated the landscape. Supervised learning approaches prevailed due to the availability of labeled datasets, while unsupervised and semi-supervised methods were less common. These trends underscore the current state of research on AI-driven NDE and identify key areas where future work may broaden the methodological toolkit and expand the range of infrastructure types and defect classes considered.

Table 1 provides a high-level synthesis of what was found across the reviewed literature: which metrics are commonly reported, what types of defects and imaging modalities are used, and which deep learning methods are employed, along with any noted variations in metric definitions.

Table 1. Summary of the range of performance metrics, defect types, imaging modalities, and deep learning approaches encountered across the reviewed NDE literature.

Category	Examples Reported in Analyzed Papers
Common Performance Metrics	Accuracy: standard ratio of correct predictions to total samples. Precision and recall: used to detail defect detection specificity and sensitivity. F1-score: employed occasionally to balance precision and recall. mIoU (mean Intersection over Union): common for segmentation tasks, comparing predicted vs. ground-truth defect areas. mAP (mean average precision): popular for object detection tasks, often following PASCAL VOC or MS-COCO standards. MAE (Mean Absolute Error): used for localization error in some cases. FPS (frames per second): reported in a few studies to indicate real-time viability.
Defect Types	Surface-level defects: cracks, spalling, efflorescence, corrosion spots. Subsurface anomalies: detected via GPR in bridge decks or MFL in pipelines (e.g., internal corrosion, gouges). Material degradation: pits, scratches on metal surfaces, wear on railheads. Manufacturing defects: rare anomalies in industrial quality control scenarios (e.g., PCB defects, irregular surface finish).

Table 1. *Cont.*

Category	Examples Reported in Analyzed Papers
Imaging/Signal Modalities	RGB imagery: most common; used for visible surface defects (bridges, rails, industrial parts). Thermography (Infrared Imaging): occasionally mentioned for detecting temperature anomalies related to defects. 3D LiDAR or photogrammetry: employed in a few studies to reconstruct infrastructure geometry and identify structural deformations. Ground-penetrating radar (GPR): for subsurface defects in concrete decks. Magnetic flux leakage (MFL): common in pipeline inspections for detecting internal defects. Ultrasonic guided waves: used for certain pipelines or material inspections to detect internal flaws.
Deep Learning Techniques	CNN-based architectures: predominantly employed (e.g., ResNet, YOLO, Faster R-CNN) for defect classification and detection tasks. Vision transformers (ViTs): increasingly popular for handling large-scale, complex images and providing improved context and segmentation performance. Hybrid models (CNN + transformer): emerging trend to combine local feature extraction (CNN) with global context modeling (transformer) for improved accuracy and robustness. Transfer learning: often used to overcome dataset scarcity by adapting pretrained models.
Metric Definition Variations	IoU thresholds: some studies use different IoU thresholds for mAP calculations. Class-averaging methods: variation in how classes are averaged for mean metrics (e.g., macro- vs. micro-averaging). Dataset standards: some studies reference PASCAL VOC or MS-COCO protocols for consistency, while others define custom thresholds.

3. Applications of Defect Detection Using Various Deep Learning Models in Different Infrastructure Sectors and Industries

Infrastructure systems, including bridges, railways, and the oil and gas industry, are fundamental to societal functions and require regular maintenance to ensure safety and efficiency. With recent advancements in computer vision and deep learning, specifically using convolutional neural networks (CNNs) and vision transformers (ViTs), automated defect detection has improved significantly across these sectors. By capturing high-resolution images through drones and ground-based systems, these models identify surface-level and subsurface defects like cracks, corrosion, and misalignments, which are often challenging to detect with manual inspections. Furthermore, hybrid models combining CNNs and transformers have emerged, balancing local feature extraction with global context modeling, which is particularly useful in complex and large-scale inspections. This section delves into the application of CNNs, transformers, and hybrid models in defect detection, highlighting their unique advantages and limitations and the impact of these technologies on the infrastructure industry.

3.1. Bridges

Bridges are among the most critical components of infrastructure that require continuous monitoring to ensure their safety and stability. Samples of cracks in concrete are shown in Figures 2a and 3. Computer vision techniques, particularly CNNs and ViTs, have been extensively employed to detect surface-level defects like cracks, corrosion, and spalling, which can compromise structural integrity [41,42]. These models use high-resolution cameras to capture images, which are then processed using deep learning algorithms for accurate defect detection. Combining drone technology with computer vision systems significantly enhances bridge inspection by allowing access to hard-to-reach areas, reducing the need for manual inspections [43]. Recent advancements in 3D reconstruction using photogrammetry and LiDAR, as shown in Figure 1, have enabled detailed analyses of bridge geometries, which are crucial for identifying subtle deformations [44].



Figure 3. Samples of cracks in bridge structures (https://github.com/wanhaifengytu/BRDETR/tree/main/data accessed on 11 October 2024).

CNN-based models have been widely adopted for visual defect detection due to their ability to extract local features from images automatically. For example, Jinsong Zhu et al. (2019) applied CNNs with transfer learning to detect surface defects in bridge structures, achieving an impressive accuracy of 97.8%. Their model proved robust in handling different image conditions and provided a scalable solution for detecting surface issues like cracks and spalling [45]. Similarly, Mahdi Ahmadvand et al. (2021) introduced a 1D CNN for subsurface defect detection using ground-penetrating radar (GPR) signals collected from bridge decks, achieving an accuracy of 84%. This approach outperformed traditional machine learning models but faced challenges in signal variability due to environmental factors like moisture [46]. In another study, Hajar Zoubir et al. (2022) applied deep CNNs combined with transfer learning to classify and localize different defect types, such as cracks and efflorescence, in concrete bridges. The model achieved 97.13% accuracy, demonstrating CNNs' effectiveness in high-precision tasks, though overfitting remained a concern due to the complexity of the network [47]. Other researchers have enhanced CNN models to improve both precision and speed. Shuai Teng et al. (2022) proposed an enhanced version of YOLOv3 for detecting cracks and exposed rebar, achieving an average precision of up to 91% with a detection speed of 23.8 frames per second. This made their model highly practical for real-time defect detection, although it was limited to specific defect types [48]. Similarly, Pravee Kruachottikul et al. (2021) used a modified ResNet-50 architecture for defect detection in reinforced concrete bridges, achieving 90.4% accuracy in detecting defects and 81.1% accuracy in classifying their severity. However, this model was limited by the small dataset, which affected its generalizability [49]. Lastly, Rong Pang et al. (2024) introduced the Multi-Scale Feature Fusion (MFF) model, which integrated EfficientNetV2, significantly improving detection accuracy for small and irregular bridge defects. While the MFF model achieved a mean average precision of 80.1%, it also faced longer testing times, limiting its potential for real-time applications [50].

On the other hand, transformer-based models have gained popularity for their ability to model long-range dependencies and capture global contextual information. Haifeng Wan et al. (2023) introduced the BR-DETR model, a transformer-based architecture for surface damage detection in concrete bridges. This model achieved high precision by incorporating deformable convolutions, with a mean Average Precision (mAP) of 91.9%. Despite its superior performance compared to CNNs, the computational costs and need for larger datasets made the model challenging to implement in resource-constrained environments [51]. Similarly, Elyas Asadi Shamsabadi et al. (2022) applied a vision transformer (ViT) combined with CNNs (TransUNet) for crack detection on asphalt and concrete surfaces. Their hybrid model outperformed CNNs by 61% in terms of mean Intersection over Union (IoU) and demonstrated robustness to noise. However, the high computational complexity of the transformer limited its application for real-time, resource-limited scenarios [52]. Similarly, Hongbing Shang et al. (2023) developed the Defect-Aware

Transformer Network (DAT-Net), which focused on detecting tiny and irregular surface defects such as tool wear and blade deterioration. The model achieved an mIoU score of over 90%, outperforming CNN-based methods in modeling long-range dependencies and improving defect specificity. However, the higher computational requirements of DAT-Net restricted its real-time use [53]. Another comprehensive study by Guidong Yang et al. (2022) highlighted the effectiveness of transformer-based models, especially the Swin Transformer, in defect detection tasks. The model achieved an accuracy of 88% and demonstrated high performance in segmentation tasks. Despite these successes, the authors noted the lack of publicly available datasets for training transformer models, which limited broader research in this area [54].

Hybrid models, which combine the strengths of CNNs and transformers, offer a promising middle ground by leveraging a CNN's ability to extract local features and a transformer's capability to model global dependencies. Wenjun Wang and Chao Su (2022) proposed a hybrid model combining a CNN, transformer, and Multi-Layer Perceptron (MLP) for defect classification in reinforced concrete bridges. This hybrid architecture achieved an accuracy of 85.5%, a recall of 85.0%, and an F1-score of 85.2%, outperforming state-of-the-art CNN-based models. By combining local and global feature extraction, the model proved more robust in detecting defects in real-world environments, although its robustness to noisy conditions was not explored [55]. Similarly, Mohammad Shahin et al. (2022) combined a Visual Transformer with a CNN for detecting concrete cracks. The CNN component achieved an impressive accuracy of 99%, while the transformer enhanced the overall accuracy, albeit at the cost of longer training times and higher computational resource demands [56]. Another hybrid approach was presented by Mingchao Li et al. (2024) in their CrackTrNet model, a lightweight CNN–transformer hybrid designed for mobile devices. CrackTrNet achieved 97.6% segmentation accuracy and had a compact model size of just 34.86 MB, making it suitable for mobile-based inspections of concrete dams. By combining a CNN's local feature extraction with a transformer's global context awareness, the model was highly effective for real-time crack detection in patrol inspections. However, its application was limited to concrete dams, and further optimization may be required to generalize the model to other structures [57].

As illustrated in Table 2, each model type—CNN models, transformer models, and hybrid models—offers distinct advantages and limitations for defect detection in civil infrastructure, with applications ranging from real-time surface crack identification to the robust detection of minor or irregular defects in diverse datasets.

Table 2. Comparison of CNN, transformer, and hybrid models for defect detection in civil infrastructure: advantages, limitations, and applications.

Model Type	Advantages	Limitations	Applications	References
CNN Models	High accuracy, adaptable to different defect types and image sizes, efficient feature extraction, real-time detection in some cases	Limited generalization to unseen defect types, risk of overfitting, struggles with noisy data and small or irregular defects	Surface and subsurface defect detection in bridges, real-time crack detection, defect severity prediction	[45,47–49,58]
Transformer Models	Superior at modeling long-range dependencies, precise for small or irregular defects, robust in handling large and diverse datasets, deformable convolution improves precision	High computational demand, sensitive to dataset-specific training, requires significant computational resources	Crack segmentation, surface damage detection in concrete bridges, defect detection in civil infrastructure	[51–54]

Table 2. *Cont.*

Model Type	Advantages	Limitations	Applications	References
Hybrid Models	Combines strengths of CNNs (local feature extraction) and transformers (global feature modeling), efficient for both small and large defects, mobile-ready models in some cases	Requires large, labeled datasets, computationally heavy, can be complex to implement, limited to certain structures or surfaces	Crack detection and segmentation in concrete structures, mobile-based inspections for infrastructure, real-time defect classification, severity prediction	[55–57]

3.2. Railways

Railway systems are integral to the transportation network, and their safety largely depends on timely defect detection and maintenance. Computer vision-based NDE techniques are crucial in railway inspections, detecting defects such as track misalignments, cracks, and surface wear [59]. Vision-based systems integrated with machine learning algorithms process large volumes of image data from regular track inspections, identifying anomalies in real time and sending alerts for immediate action [60]. Swin Transformers and other hybrid models have been particularly effective in capturing both local and global features, enhancing defect detection accuracy in large-scale railway infrastructures [61]. Automated rail inspection systems reduce downtime and minimize human error, thus improving safety and reducing operational costs [62]. An approach for railway defect detection using drones and deep learning models is shown in Figure 4. A sample of missing fasteners is also shown in Figure 2b.

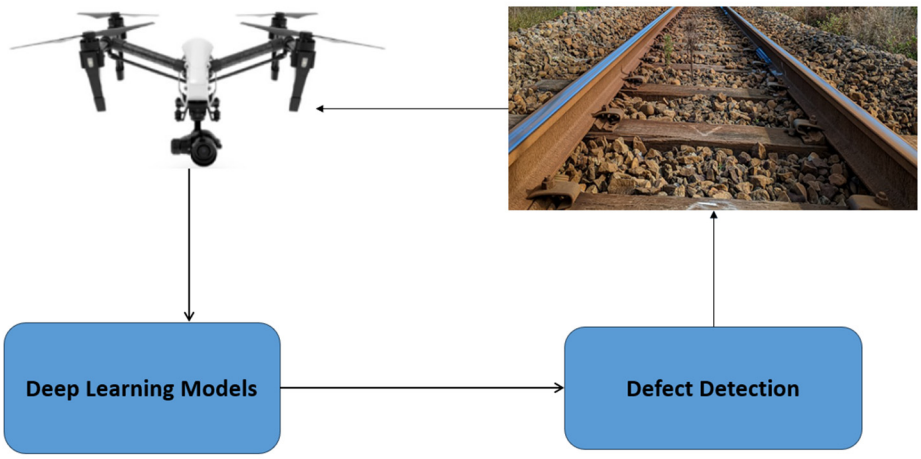


Figure 4. Using drones and deep learning for railway defect detection.

Rail surface defect detection is crucial for maintaining railway safety and efficiency. Recent advancements, particularly in convolutional neural networks (CNNs), have significantly automated this process. Hao Wang et al. (2022) enhanced traditional Mask R-CNN by incorporating a dual fusion feature pyramid network and replacing the standard IOU metric with the complete Intersection over Union (CIOU), achieving a mean average precision (MAP) of 98.7%. This model notably improves defect localization accuracy and adapts to data scarcity through transfer learning, though its computational complexity may hinder real-time applications [63]. Lidan Shang et al. (2021) introduced a two-stage pipeline that first localizes defects using traditional image processing before classification with a CNN. This method effectively balances precision and recall but is limited by its non-real-time capability and specificity to railway images [64]. Rahatara Ferdousi et al. (2024) developed an ensemble framework with CNN models like VGG-19, MobileNetV3,

and ResNet50, achieving a test accuracy of 99% and an F1-score of 0.98. The model quickly stabilizes performance but faces challenges with overfitting and computational demands due to training multiple models [65]. Hui Zhang's MRSDI-CNN, using SSD and YOLOv3, excels in real-time defect detection across multiple types, although it is primarily focused on squat defects and may require retraining for broader applicability [66]. Lastly, Danyang Zheng et al. (2021) proposed a system that combines YOLOv5 with Mask R-CNN and a ResNet-based classifier, enabling high-speed, accurate defect detection and classification within one framework. However, its specialized design and dataset dependency may limit generalization across different railway components [67].

Rail surface defect detection has gained significant attention recently due to its critical role in maintaining railway safety and efficiency. The rise in transformer-based models has presented a new avenue for improving detection accuracy, especially in overcoming some limitations inherent to traditional convolutional neural networks (CNNs). One such advancement is a study by Feng Guo et al. (2024), which introduces RailFormer, a transformer-based semantic segmentation network designed to detect rail surface defects. This model tackles the challenges of CNN-based methods, such as their difficulty preserving small-scale details and capturing hierarchical features from images. RailFormer employs an encoder–decoder architecture with innovative components like efficient self-attention mechanisms and a Criss-Cross attention module, allowing the model to excel at fusing local and global features. Tested on public rail datasets (RSDD Type-I and Type-II) and a customized dataset, RailFormer outperformed competing models like SegFormer and Swin Transformer, particularly regarding the mean Intersection over Union (mIoU) and defect visualization. The strength of RailFormer lies in its ability to capture fine details that CNNs typically struggle with, ensuring better accuracy in detecting small defects. However, its reliance on large, labeled datasets and the inherent computational complexity of transformer architectures present challenges, particularly for real-time applications in resource-constrained environments [68]. Another contribution to the field is the application of a vision transformer (ViT) in the work by Shiyao Lu et al. (2024). This study utilizes the vision transformer to classify rail defects using B-scan images generated from ultrasonic flaw detectors. The ViT model, adapted from its origins in natural language processing, leverages its ability to focus on the global structure of images rather than just localized features, as is the case with CNNs. This approach results in high classification accuracy, reaching up to 98.92%, significantly reducing the manual labor and time traditionally required for rail defect detection. While the basic ViT model shows promise, the study acknowledges that future improvements can be made through dataset expansion and model fine-tuning, which could further enhance its performance [69]. In another study, TrackNet, a transformer-based model, is proposed for detecting defects in ballastless tracks on high-speed railways. TrackNet uses multi-head self-attention to improve global feature extraction, overcoming the limitations of CNNs, which often fail to exploit contextual information fully.

Additionally, TrackNet employs transfer learning to reduce its dependency on large, high-quality datasets by leveraging pretrained weights from publicly available datasets. Experimental results demonstrate TrackNet's superiority over advanced models like Swin Transformer, significantly improving accuracy and F1-score. TrackNet also addresses the "black box" nature of deep learning by providing interpretability through heatmap visualizations, allowing for a better understanding of the model's decision-making process. Despite its strong performance, the model still faces challenges in generalizing to unseen environments and scaling for real-time use, especially in resource-limited settings [70]. Lastly, the Defect-Aware Transformer Network (DAT-Net), proposed by Hongbing Shang et al. (2023), takes an innovative approach to industrial surface defect detection by inte-

grating transformer-based modules instead of traditional convolution layers. This model introduces a defect-aware module (DAM), which enhances the model's ability to perceive the geometric characteristics of defects, addressing a key limitation of CNNs in handling long-range dependencies.

Furthermore, Graph Position Encoding (GPE) provides additional positional information, improving the model's adaptability and detection performance. Tested on datasets involving blade defects and tool wear, DAT-Net achieved impressive results in terms of mIoU. This model's use of transformer-based architectures significantly improves its capacity to handle complex defect shapes. However, the added computational complexity may pose challenges for real-time or large-scale industrial applications [71].

In hybrid models for railway defect detection, combining the CNN and transformer architectures offers a robust solution to the limitations of traditional methods focused on either local or global features. In 2024, Jin He et al. introduced the CNN–Transformer Bridge Mode (CTBM-DAHD) network to detect defects in arcing horns, which are crucial components of the overhead contact system. This network combines CNNs for local feature capture and transformers for the global context, enhancing defect detection under various conditions, including rain and nighttime. The CTBM-DAHD network outperforms existing models, improving recall and precision by 3.5% and 1.5%, respectively, and has been effectively applied on over 500 high-speed trains in China. Despite its effectiveness, adapting this model to different railway environments might require additional adjustments to handle varying environmental and component factors [72]. In 2023, Hui Luo et al. enhanced YOLOv5s for rail surface defect detection, addressing issues like low contrast and insufficient data with techniques like data augmentation and a new CDConv module. They integrated a Swin Transformer to improve scale detection and introduced a global attention mechanism in PANet for better feature integration, achieving 96.9% mAP. Despite its success, challenges remain in its real-time processing efficiency and adaptability to complex environments, highlighting areas for further improvement [73]. In 2024, Chenghai Yu et al. advanced hybrid modeling with their FHB-DETR model, enhancing RT-DETR for railway turnout defect detection. Railway turnouts, crucial for track direction changes, demand high safety standards. Their model integrates CNN feature extraction with a modified transformer that utilizes Hilo attention to lower computational demands and boosts real-time performance. Incorporating BiFPN optimizes feature fusion, improving defect detection in complex scenarios. The model outperformed RT-DETR by 3.5% in terms of mAP50 and cut computational complexity by 6%, promising enhanced efficiency and effectiveness for real-time applications in high-speed rail settings [74]. In 2024, Suli Bai and colleagues developed a vision-based, nondestructive detection network that merges CNN and transformer models to identify rail surface defects. Their approach includes an enhanced Res2Net for capturing detailed local context and a transformer block in the bottleneck layer for improved global context awareness. An attention-based edge enhancement block is also incorporated, focusing on key areas to detect even minor defects accurately. This integration of the CNN's local feature extraction with the global insights from transformers led to impressive results on various datasets, such as NRSD-MN and SEPR, excelling in metrics like mIoU and Pixel Accuracy. Despite its effectiveness in capturing diverse defect shapes, the model's computational demands could limit its application in real-time railway inspections [75]. In their 2024 paper, Shanping Ning and colleagues proposed RailTrack-DaViT, a detection algorithm that combines MobileNetv3 with transformer models to identify rail surface defects with improved precision. Their integration of MobileNetv3's streamlined architecture and transformers' enhanced contextual capabilities led to a detection accuracy of 93.8% mAP at 19.5 FPS, outperforming established models like YOLOv5. By including channel attention through the CA-Bneck and BiFPN-Lite modules, defect features can be

efficiently captured, streamlining detection without increasing system load. Despite its effectiveness, the model faces scalability and real-time application challenges, primarily due to the increased computational demands associated with integrating MobileNetv3 and transformers, potentially limiting its deployment in high-speed rail environments where rapid processing is crucial [76].

As shown in Table 3, CNN-based models are highly accurate and adapt well to scenarios with limited data, enabling the real-time detection of various rail defects. However, their computational demands and potential overfitting issues can limit their broad applicability. Transformer models excel at identifying small defects by leveraging global feature representations, though their high computational requirements and data dependence may pose challenges for real-time applications. Hybrid models combine the strengths of both CNNs and transformers, enhancing precision in complex rail inspection environments. Nevertheless, these hybrid approaches may sacrifice real-time efficiency due to increased computational costs.

Table 3. Comparison of CNN, transformer, and hybrid models for defect detection in railways: advantages, limitations, and applications.

Model Type	Advantages	Limitations	Applications	References
CNN	High accuracy in defect localization and precision; stabilizes performance quickly; adapts to data scarcity; enables real-time detection across various defect types.	High computational demands, limiting real-time applications; can face overfitting issues; specialized models may require retraining for broader applicability.	Rail surface defect detection, crack detection, fastener inspection, squat defect detection, and track misalignment detection.	[63–67]
Transformer	Advanced small defect detection with global feature focus; enhances accuracy in rail defect detection.	Computationally intensive, often requires large datasets and faces real-time challenges.	Rail defect classification, semantic segmentation of rail surfaces, and detection of defects in ballastless tracks.	[68–71]
Hybrid Models	Combines local (CNN) and global (transformer) feature extraction; enhances precision in complex environments.	Real-time processing efficiency and adaptability can be limited by increased computational load.	Detection of complex defects in overhead contact systems, turnout defect detection, and detailed rail surface inspection	[72–76]

3.3. Industry

In the oil and gas industry, critical infrastructure such as pipelines, storage tanks, and offshore platforms are vulnerable to various types of defects due to harsh environmental conditions and operational stresses [77]. Pipelines, essential for transporting liquids and gases, face several types of mechanical defects that compromise their integrity. Dents, cracks, gouges, and scratches are common, often caused by external interference. When these defects occur close together, they may form combined defects, such as a dent–crack defect, which is particularly hazardous. These compounded defects are more dangerous than individual ones, as they significantly reduce the pipeline's capacity and safety. The effective monitoring and assessment of such defects, including fatigue life assessments, are critical for maintaining safe pipeline operations [78]. Figure 5 shows examples of automated defect detection on production lines, wherein 563 misaligned bottle caps are detected and localized. The red bounding boxes highlight the regions of interest where defects or anomalies are observed.



Figure 5. Missing part in production line [79].

In a 2022 study, Min Zhang et al. introduced a robust deep learning model that enhances pipeline safety detection by integrating convolutional, pooling, and fully connected layers. This autonomous deep learning model processes data from a two-axis magnetic flux leakage detection device, employing a novel two-dimensional data fusion method to improve defect detection accuracy in noisy and harsh operational environments. The study achieves impressive defect identification accuracies, 99.19% during training and 97.38% in testing. However, the real-world application of this model across varied conditions and its computational efficiency requires additional scrutiny [80]. In 2023, Iurii Katser et al. tackled the challenge of defect detection in oil pipelines using magnetic flux leakage (MFL) data. The study introduces advanced convolutional neural network (CNN) architectures and sophisticated data preprocessing techniques to improve anomaly detection efficiency and accuracy. The authors demonstrate potential advancements in pipeline safety diagnostics by validating their approaches with real-world data. While promising, the paper could benefit from broader comparisons to other methods and a more detailed discussion on performance metrics to better understand the models' effectiveness and applicability in various operational environments [81]. In 2023, Hang Xin and Xiaoran Du introduced a novel safety monitoring and defect identification system for oil and gas pipelines utilizing Faster R-CNN technology, aimed at addressing challenges such as difficult defect detection, high false alarm rates, and inaccurate defect positioning in complex environments. Their method employs a Region Proposal Network (RPN) to identify and coordinate image features and anchor points effectively. This is coupled with an ROI Pooling strategy to resolve size inconsistencies between predicted and actual defect boundaries.

Additionally, a classification regression model tailored to the unique characteristics of pipeline defects enhances the accuracy of defect categorization and localization. The system's efficacy was validated through simulations and experiments on the Pascal VOC dataset, demonstrating substantial improvements in defect recognition accuracy [82]. In their 2024 study, Oyedeji et al. developed a multiple-phase convolutional neural network (CNN) to detect and identify corrosion on metallic materials. This model utilizes a dataset of 600 images to execute binary and multiclass classifications, achieving high accuracies of 94.87% for binary detection and up to 96.5% in specific-region identification through patch distribution methods. While the model demonstrates robust performance suitable for industries like aerospace and manufacturing, its complexity and high computational demands may limit its application in smaller settings or real-time operations without substantial hardware support [83]. In 2024, Zhouxiang Fei et al. developed a CNN-based method to enhance crack feature detection in steam cycle components of nuclear power plants. Their research successfully automated the identification of stress corrosion cracks in superheaters with an accuracy of 92.97%. This approach leverages deep learning to improve the efficiency and accuracy of remote visual inspections, a critical component of

nuclear facility maintenance. The study also discusses the challenges related to the black box nature of CNNs and the intensive labeling process required for training data. The findings suggest that this technology could significantly improve anomaly detection in pressure vessel inspections across the nuclear industry [84].

In their 2022 study, Liu et al. developed a transformer-based framework tailored to industrial defect detection in cylinder liners, addressing the limitations of traditional inspection methods. Their approach integrates a Swin Transformer with a block division and mask mechanism to enhance detection accuracy, particularly for small defects. The Swin Transformer captures multi-level features and applies attention mechanisms, while the block division method segments high-resolution images into smaller regions, focusing on fine details. Additionally, a masking mechanism mitigates background noise, enhancing precision. The experimental results demonstrated improved accuracy over established methods, setting a new benchmark for automated, nondestructive detection in cylinder liner inspections [85]. In 2024, Toqa Alaa et al. developed a vision transformer (ViT)-based model for automated defect detection on metal surfaces, achieving 93.5% classification accuracy and effective localization with an MAE of 3.2 pixels and an IOU of 0.72. Using the multi-DET dataset, which includes diverse defect types, their approach leverages a ViT's self-attention for capturing complex patterns, outperforming traditional CNNs. However, limitations remain in detecting irregular defect shapes and real-time processing due to the ViT's complexity, suggesting future improvements through dataset expansion and model optimization [86]. In 2024, Kang An et al. proposed LPViT, a transformer-based PCB classification and defect detection model incorporating mask patch prediction and label smoothing to improve accuracy and robustness. Tested on micro-PCB and DeepPCB datasets, LPViT achieved state-of-the-art results, with 98.8% mean average precision on defect detection tasks. The model outperformed both CNNs and other transformer models, proving effective in PCB manufacturing and recycling. The study highlights limitations in handling larger PCB datasets, suggesting that further data augmentation could enhance the model's future performance [87]. In 2023, Ding et al. introduced a novel vision transformer architecture to enhance industrial defect detection. Their study incorporated Hybrid Window Attention (HWA) and Dynamic Token Normalization (DTN) to effectively balance local and global information processing, achieving high classification accuracies of 96.8% on the NEU dataset and 98.5% on the DAGM dataset. Despite its success, this model may require additional testing across various industrial scenarios to confirm its robustness and generalizability fully. This research marks a significant step forward in applying advanced deep learning techniques to improve efficiency and accuracy in industrial quality control [88].

In 2022, Yao et al. introduced a diagnosis method for oil and gas pipeline defects utilizing transfer deep learning and feature fusion. This method, combining continuous-wavelet and short-time Fourier transform features with a strengthened convolutional neural network (TWSC), aims to enhance detection and diagnosis accuracy under limited and poor-quality data conditions, which are typical in pipeline monitoring. The key advantages of this method include improved diagnostic accuracy and robust performance in noisy conditions, enabled by the effective use of transfer learning to optimize model training with small datasets. However, this approach may face challenges such as complexity in implementation and dependence on high-quality training data, which could limit its practical application across different pipeline systems without further adaptation [89]. In 2022, Zhang et al. introduced a method for estimating defect sizes and cross-sectional profiles in oil and gas pipelines using a Visual Deep Transfer Learning (VDTL) Neural Network. Their approach combines a visual data transformation layer with a transfer-enhanced convolutional neural network and a fully connected layer, achieving high accuracy in defect

predictions. This method transforms 1D magnetic flux leakage (MFL) signals and fuses 2D MFL images, providing a sophisticated and effective solution for improving pipeline maintenance and safety [90]. In 2024, Yuan et al. introduced the Denoising Diffusion Probabilistic Model (DDPM) with a Gated Parallel Convolution Swin Transformer (DGPST) for detecting minute defects in oil pipelines. This innovative approach overcomes challenges like inadequate inspection data and low detection accuracy by generating diverse magnetic flux leakage (MFL) defect samples and employing a gated parallel convolution layer to broaden the model's receptive field.

With an advanced loss function enhancing precision, the DGPST achieves a notable detection accuracy of 95.6% and reduces delay to 7.6 ms, positioning it as a powerful tool for pipeline maintenance, with high computational demands and a potential risk of overfitting due to its complex architecture [34]. In 2024, Chen et al. introduced a cascaded deep learning methodology for pipeline defect detection, integrating pretrained YOLOv5 and vision transformer (ViT) models. Their approach, designed to enhance the accuracy and efficiency of defect identification using magnetic flux leakage data, demonstrated improved performance over single-model applications, achieving the robust classification of various pipeline defects. This method's advantages include its ability to combine detection with classification effectively and its high precision. However, it faces challenges such as increased computational complexity and potential issues with dataset generalizability that could limit its use in diverse real-world settings [91]. In 2024, Cuenca et al. proposed a novel flaw detection approach in carbon steel pipes utilizing ultrasonic guided wave testing (UGWT) combined with an optimized transformer neural network model 111. By selecting the flexural wave mode F(1,1) and tuning various hyperparameters within the transformer architecture, their method effectively captured long-range dependencies and temporal relationships in the signal data—an advantage over traditional CNN or RNN approaches. The optimized transformer model demonstrated a 99% accuracy and F1-score in identifying flaws in diverse pipe lengths and defect locations. This high performance underscores the potential of transformer-based solutions in nondestructive evaluation (NDE) scenarios where robust, scalable detection methods are needed [92]. In their 2024 study, Miriam Alber et al. discuss the application of machine learning for early defect detection in industrial manufacturing through visual quality control systems. They explore integrating state-of-the-art vision transformer models with anomaly detection algorithms to address the challenge of identifying rare defects in unbalanced visual quality control datasets. Their research evaluates various combinations of these models on the MVTecAD and BTAD datasets, aiming to determine the most efficient and effective solutions for reducing costs and human error in repetitive inspection tasks. The paper provides practical insights for selecting models that balance detection quality with inference speed, making it a valuable resource for practitioners in the field [93].

Table 4 highlights that CNN models provide high accuracy and automated defect detection for pipeline safety and aerospace corrosion inspections but can face high computational demands and complexity under varying conditions. Transformers offer state-of-the-art results and precise small-defect detection yet require large datasets and encounter real-time constraints. Hybrid models combine the benefits of CNNs and transformers, achieving high accuracy under noisy conditions and integrating defect classification. However, their complexity, risk of overfitting, and reliance on simulated data remain challenges in industrial applications.

Table 4. Comparison of CNN, transformer, and hybrid models for defect detection in industry: advantages, limitations, and applications.

Model Type	Advantages	Limitations	Applications	Refs.
CNN	High accuracy and efficiency, effective identification, high classification accuracies, automated crack detection	Needs testing in varied condition, broader comparisons needed, complex environment challenges, high computational demands, intensive training required	Pipeline safety, oil pipeline defect detection, aerospace corrosion detection	[80–86]
Transformer	Enhanced small defect detection, high accuracy and localization, state-of-the-art results	Real-time processing challenges, larger dataset challenges, needs more robustness testing	Cylinder liner inspections, metal surface defect detection, PCB manufacturing and recycling, industrial defect detection	[85–88]
Hybrid	Improved accuracy in noisy conditions, high prediction accuracy, reduced delay and high accuracy, combines detection with classification, high-accuracy and automated process, practical insights for model selection	Implementation complexity, risk of overfitting, relies on simulated data	Pipeline maintenance, oil pipeline defect detection, spiral-welded pipeline detection, industrial manufacturing quality control	[34,89–93]

4. Challenges and Future Directions

The future of Nondestructive Evaluation (NDE) lies in the seamless integration of in situ sensors with advanced computer vision and deep learning techniques. This hybrid approach aims to overcome current limitations by leveraging the complementary strengths of both technologies. While computer vision offers robust scalability and excels at surface defect detection, in situ sensors enable the localized, real-time monitoring of subsurface anomalies. Together, these integrated systems have the potential to yield more comprehensive, reliable, and cost-effective defect detection across a wide range of infrastructure types. However, significant challenges still need to be addressed, including the complexities of data fusion between sensor and vision inputs, the substantial computational demands of the state-of-the-art models, and the absence of standardized frameworks to guide their effective combination. Addressing these challenges requires a multifaceted research agenda. First, the scarcity and quality of datasets must be addressed to support robust machine learning models. Acquiring large, well-annotated datasets is complicated by privacy concerns, restricted access, and the rarity of certain defect types. Additionally, achieving effective generalization across diverse infrastructure categories is critical, as models trained on one dataset may not be effectively transferable without extensive retraining or fine-tuning. Reducing computational overhead is another pressing issue, particularly for models like vision transformers, which are resource-intensive and often impractical in real-world, resource-constrained environments.

As this review acknowledges, the feasibility of applying ViTs in such environments depends on developing more lightweight, optimized model architectures or employing techniques such as model compression, pruning, and quantization. Leveraging edge computing platforms and hardware accelerators can help alleviate this resource burden. At the same time, transfer learning and domain adaptation approaches may streamline fine-tuning for specific tasks and infrastructures without requiring prohibitively extensive computational resources [94–96]. This underscores the need for lightweight, portable systems employing more efficient machine learning algorithms capable of fusing sensor and visual data. Furthermore, enhancing interoperability among sensor types and legacy

infrastructure systems requires standards and methodologies that ensure the seamless integration and adaptation of these technologies, ultimately facilitating more versatile and adaptive monitoring solutions.

As AI-driven NDE systems become more integrated into critical infrastructure monitoring, ensuring data privacy and security is paramount. Large-scale data collection from sensors, imaging devices, and IoT platforms introduces challenges in safeguarding sensitive information against unauthorized access or misuse. Compliance with existing regulations, such as the GDPR or other regional data protection laws, is essential. Developing secure data handling protocols, robust encryption methods, and standardized data-sharing agreements can help maintain public trust. Additionally, interdisciplinary efforts involving engineers, data scientists, policymakers, and legal experts will be necessary to establish ethical frameworks and best practices. By proactively addressing these ethical considerations, we can foster responsible innovation that respects privacy, maintains security, and upholds public confidence in AI-driven NDE solutions.

Scalability poses yet another challenge, as these integrated technologies must function effectively across extensive, heterogeneous infrastructure networks and adapt to diverse environmental and structural conditions. To address these issues, future research should not only focus on developing efficient hybrid models that combine the advantages of convolutional neural networks (CNNs) and vision transformers but also explore real-time processing capabilities and novel approaches for rapid adaptation to changing conditions. Adopting cross-disciplinary strategies that integrate engineering, material science, and data science could lead to more holistic and innovative solutions to the multifaceted challenges of NDE. Encouraging a supportive environment—through funding, policymaking, and educational initiatives—that bridges the gaps between academia, industry, and regulatory bodies will be essential. By directly confronting these technical, organizational, and societal challenges, the NDE community can fully realize the potential of integrated sensor–vision–deep learning systems to ensure the reliability and safety of critical infrastructure worldwide.

Furthermore, several actionable, ongoing initiatives illustrate pathways for future work. Government agencies and industry consortia-funded pilot programs are experimenting with edge-based NDE frameworks, integrating low-power hardware accelerators to perform on-site inference and reduce latency. Standards organizations and research coalitions are developing guidelines and shared platforms for benchmarking NDE datasets and streamlining data exchange, annotation practices, and security protocols. Some infrastructure operators have initiated long-term monitoring campaigns that combine drone-based imaging with embedded fiber optic sensors, providing valuable reference models for cross-disciplinary ventures. These real-world efforts highlight how investing in open-access repositories, establishing community-driven validation benchmarks, and forming public–private partnerships can guide the next generation of NDE innovations and promote best practices for their broad-scale adoption.

Several of the reviewed studies have begun exploring strategies to address data scarcity and enhance model generalization. For example, some employed data augmentation techniques—such as geometric transformations, synthetic defect generation, or generative adversarial networks (GANs)—to artificially expand the variety of training samples [97]. Others leveraged transfer learning, adapting pretrained models from related domains to quickly achieve high accuracy in infrastructure-specific tasks despite limited training data [98]. Domain adaptation methods and meta-learning frameworks have also been reported, enabling models to handle variations in sensor modalities, environmental conditions, and defect morphologies more effectively [99]. Case studies on topics like pipeline inspection and rail defect detection illustrate how combining multispectral imaging with

generative models can simulate rare defect scenarios, while the fine-tuning of pretrained networks with small yet carefully curated datasets has improved performance in bridge crack detection [100]. These approaches collectively indicate that while data scarcity and generalization challenges remain significant obstacles, innovative solutions are emerging to make models more robust and broadly applicable across diverse infrastructure contexts.

5. Conclusions

This review has explored the transformative potential of computer vision and deep learning technologies in the field of Nondestructive Evaluation (NDE). As infrastructure ages and the demands for safety and efficiency grow, integrating these advanced technologies offers significant promise for enhancing the accuracy, speed, and cost-effectiveness of infrastructure inspections. However, as discussed, adopting these technologies is challenging. Data scarcity, model generalization, computational demands, integration complexities, and ethical considerations must be addressed to realize the potential benefits fully. The field of NDE will benefit significantly from collaborative efforts that span across disciplines, involving experts in AI, engineering, and regulatory compliance, to innovate and refine these technologies.

Additionally, advancements in model architectures, particularly through hybrid models that combine various deep learning approaches, are critical for overcoming current accuracy and computational efficiency limitations. It is also imperative that these solutions are developed with a focus on scalability and ethical use to ensure they can be applied effectively across different regions and infrastructures. The methods discussed in this study enhance defect detection accuracy and operational efficiency. While the real-time aspect may not be critical for all defect types, it supports broader applications in infrastructure maintenance where immediate insights can be beneficial.

Ultimately, the continued exploration and adoption of computer vision in NDE signify a crucial step towards more predictive and preventive maintenance strategies that will ensure the safety and longevity of critical infrastructure worldwide. The insights gathered through this review highlight the importance of sustained research and development efforts in this field, encouraging a future where technology and tradition merge to foster safer and smarter infrastructure management.

Author Contributions: Conceptualization, S.S.K. and V.R.; methodology, S.S.K., V.R. and S.M.; software, S.M. and V.R.; validation, S.S.K. and V.R.; formal analysis, S.M. and V.R.; investigation, S.M. and V.R.; writing—original draft preparation, S.M. and V.R.; writing—review and editing, S.S.K. and V.R.; visualization, S.M. and V.R.; supervision, S.S.K.; funding acquisition, S.S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC) grants, grant numbers RGPIN-2023-05578 and DGECR-2023-00336.

Data Availability Statement: All data supporting the results of this project are openly available. No data were generated during the course of this study.

Acknowledgments: The authors would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support and the Centre national intégré du manufacturier intelligent of Université du Québec à Trois-Rivières for their infrastructure support.

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

References

1. Mukherjee, M.; Abhinay, K.; Rahman, M.; Yangdhen, S.; Sen, S.; Adhikari, B.R.; Nianthi, R.; Sachdev, S.; Shaw, R. Extent and evaluation of critical infrastructure, the status of resilience and its future dimensions in South Asia. *Prog. Disaster Sci.* **2023**, *17*, 100275. [CrossRef]
2. Frangopol, D.M.; Liu, M. Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost. In *Structures and Infrastructure Systems*; Routledge: London, UK, 2019; pp. 96–108. [CrossRef]
3. Mansuri, L.E.; Patel, D. Artificial intelligence-based automatic visual inspection system for built heritage. *Smart Sustain. Built Environ.* **2022**, *11*, 622–646. [CrossRef]
4. Artagan, S.S.; Ciampoli, L.B.; D'amico, F.; Calvi, A.; Tosti, F. Non-destructive Assessment and Health Monitoring of Railway Infrastructures. *Surv. Geophys.* **2019**, *41*, 447–483. [CrossRef]
5. Prakash, G.; Yuan, X.-X.; Hazra, B.; Mizutani, D. Toward a Big Data-Based Approach: A Review on Degradation Models for Prognosis of Critical Infrastructure. *J. Nondestruct. Eval. Diagn. Progn. Eng. Syst.* **2020**, *4*, 021005. [CrossRef]
6. Gupta, M.; Khan, M.A.; Butola, R.; Singari, R.M. Advances in applications of Non-Destructive Testing (NDT): A review. *Adv. Mater. Process. Technol.* **2022**, *8*, 2286–2307. [CrossRef]
7. Xu, X.; Ran, B.; Jiang, N.; Xu, L.; Huan, P.; Zhang, X.; Li, Z. A systematic review of ultrasonic techniques for defects detection in construction and building materials. *Measurement* **2024**, *226*, 114181. [CrossRef]
8. Sathish, K.V.; Manjunatha, H.C.; Vidya, Y.S.; Sridhar, K.N.; Seenappa, L.; Reddy, B.C.; Raj, S.A.; Gupta, P.D. X-rays/gamma rays radiation shielding properties of Barium–Nickel–Iron oxide nanocomposite synthesized via low temperature solution combustion method. *Radiat. Phys. Chem.* **2022**, *194*, 110053. [CrossRef]
9. Bagavathiappan, S.; Lahiri, B.; Saravanan, T.; Philip, J.; Jayakumar, T. Infrared thermography for condition monitoring—A review. *Infrared Phys. Technol.* **2013**, *60*, 35–55. [CrossRef]
10. Duan, L.; Liu, J. Smart composite materials and IoT: Revolutionizing real-time railway health monitoring. *MRS Commun.* **2024**, 1–7. [CrossRef]
11. Jaber, A.; Karganroudi, S.S.; Meiabadi, M.S.; Aminzadeh, A.; Ibrahim, H.; Adda, M.; Taheri, H. On Smart Geometric Non-Destructive Evaluation: Inspection Methods, Overview, and Challenges. *Materials* **2022**, *15*, 7187. [CrossRef] [PubMed]
12. Islam, M.R.; Zamil, M.Z.; Rayed, M.E.; Kabir, M.M.; Mridha, M.F.; Nishimura, S.; Shin, J. Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. *IEEE Access* **2024**. [CrossRef]
13. Baduge, S.K.; Thilakarathna, S.; Perera, J.S.; Arashpour, M.; Sharafi, P.; Teodosio, B.; Shringi, A.; Mendis, P. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **2022**, *141*, 104440. [CrossRef]
14. Ali, A.A.I.M.; Jamaludin, S.; Imran, M.H.; Ayob, A.F.M.; Ahmad, S.Z.A.S.; Akhbar, M.F.A.; Suhrab, M.I.R.; Ramli, M.R. Computer Vision and Image Processing Approaches for Corrosion Detection. *J. Mar. Sci. Eng.* **2023**, *11*, 1954. [CrossRef]
15. Cui, B.; Tao, W.; Zhao, H. High-Precision 3D Reconstruction for Small-to-Medium-Sized Objects Utilizing Line-Structured Light Scanning: A Review. *Remote Sens.* **2021**, *13*, 4457. [CrossRef]
16. Dwivedi, S.K.; Vishwakarma, M.; Soni, P. Advances and Researches on Non Destructive Testing: A Review. *Mater. Today Proc.* **2018**, *5*, 3690–3698. [CrossRef]
17. Alam, F.; Ashfaq Ahmed, M.; Jalal, A.H.; Siddiquee, I.; Adury, R.Z.; Hossain, G.M.; Pala, N. Recent Progress and Challenges of Implantable Biodegradable Biosensors. *Micromachines* **2024**, *15*, 475. [CrossRef] [PubMed]
18. Valavanis, I.; Kosmopoulos, D. Multiclass defect detection and classification in weld radiographic images using geometric and texture features. *Expert Syst. Appl.* **2010**, *37*, 7606–7614. [CrossRef]
19. Bado, M.F.; Casas, J.R. A review of recent distributed optical fiber sensors applications for civil engineering structural health monitoring. *Sensors* **2021**, *21*, 1818. [CrossRef] [PubMed]
20. D'orazio, T.; Leo, M.; Distante, A.; Guaragnella, C.; Pianese, V.; Cavaccini, G. Automatic ultrasonic inspection for internal defect detection in composite materials. *NDT E Int.* **2007**, *41*, 145–154. [CrossRef]
21. Modupe, O.T.; Otitoola, A.A.; Oladapo, O.J.; Abiona, O.O.; Oyeniran, O.C.; Adewusi, A.O.; Komolafe, A.M.; Obijuru, A. Reviewing the transformational impact of edge computing on real-time data processing and analytics. *Comput. Sci. IT Res. J.* **2024**, *5*, 693–702. [CrossRef]
22. Coito, T.; Firme, B.; Martins, M.S.; Vieira, S.M.; Figueiredo, J.; Sousa, J.M. Intelligent sensors for real-Time decision-making. *Automation* **2021**, *2*, 62–82. [CrossRef]
23. Yuan, Q.; Shi, Y.; Li, M. A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges. *Remote Sens.* **2024**, *16*, 2910. [CrossRef]
24. Munawar, H.S.; Ullah, F.; Heravi, A.; Thaheem, M.J.; Maqsoom, A. Inspecting Buildings Using Drones and Computer Vision: A Machine Learning Approach to Detect Cracks and Damages. *Drones* **2021**, *6*, 5. [CrossRef]
25. Gopalakrishnan, K.; Gholami, H.; Vidyadharan, A.; Choudhary, A.; Agrawal, A. Crack damage detection in unmanned aerial vehicle images of civil infrastructure using pre-trained deep learning model. *Int. J. Traffic Transp. Eng.* **2018**, *8*, 1–14. [CrossRef]

26. Ajmi, C.; Zapata, J.; Martínez-Álvarez, J.J.; Doménech, G.; Ruiz, R. Using Deep Learning for Defect Classification on a Small Weld X-ray Image Dataset. *J. Nondestruct. Evaluation* **2020**, *39*, 1–13. [CrossRef]
27. Hütten, N.; Gomes, M.A.; Hölken, F.; Andricevic, K.; Meyes, R.; Meisen, T. Deep Learning for Automated Visual Inspection in Manufacturing and Maintenance: A Survey of Open-Access Papers. *Appl. Syst. Innov.* **2024**, *7*, 11. [CrossRef]
28. Hussain, M. Sustainable Machine Vision for Industry 4.0: A Comprehensive Review of Convolutional Neural Networks and Hardware Accelerators in Computer Vision. *AI* **2024**, *5*, 1324–1356. [CrossRef]
29. Maurício, J.; Domingues, I.; Bernardino, J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Appl. Sci.* **2023**, *13*, 5521. [CrossRef]
30. Ahmed, H.O.A.; Nandi, A.K. Convolutional-Transformer Model with Long-Range Temporal Dependencies for Bearing Fault Diagnosis Using Vibration Signals. *Machines* **2023**, *11*, 746. [CrossRef]
31. Li, S.; Wu, C.; Xiong, N. Hybrid Architecture Based on CNN and Transformer for Strip Steel Surface Defect Classification. *Electronics* **2022**, *11*, 1200. [CrossRef]
32. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
33. Anand, R.; Anand, S. AI-Driven Structural Health Monitoring: Innovations, Challenges, and Future Directions. *Authorea Preprints* 2024. [CrossRef]
34. Lang, X.; Yuan, K.; Cao, J. Small Defect Detection Model of Oil Pipeline Based on Improved Swin Transformer. In Proceedings of the 2024 6th International Conference on Electronic Engineering and Informatics (EEI), Chongqing, China, 28–30 June 2024; pp. 414–417. [CrossRef]
35. Structural Health Monitoring—Google Livres. Available online: https://books.google.ca/books?hl=fr&lr=&id=DXOsGoqtiggC&oi=fnd&pg=PA3&dq=Structural+Health+Monitoring+++Balageas&ots=dKgqePjSar&sig=S8hWGqv7eLLS_cfhXGdT9d0cCZE&redir_esc=y#v=onepage&q=Structural%20Health%20Monitoring%20+%20Balageas&f=false (accessed on 14 December 2024).
36. Ye, X.W.; Dong, C.Z.; Liu, T. A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications. *J. Sens.* **2016**, *2016*, 1–10. [CrossRef]
37. Bao, Y.; Tang, Z.; Li, H.; Zhang, Y. Computer vision and deep learning–based data anomaly detection method for structural health monitoring. *Struct. Health Monit.* **2018**, *18*, 401–421. [CrossRef]
38. Harley, J.B.; Sparkman, D. Machine learning and NDE: Past, present, and future. *AIP Conf. Proc.* **2019**, *2102*, 090001. [CrossRef]
39. Valeske, B.; Osman, A.; Römer, F.; Tschuncky, R. Next generation NDE sensor systems as IIoT elements of industry 4.0. *Res. Nondestruct. Eval.* **2020**, *31*, 340–346. [CrossRef]
40. Malekloo, A.; Ozer, E.; AlHamaydeh, M.; Girolami, M. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Struct. Health Monit.* **2021**, *21*, 1906–1955. [CrossRef]
41. Luo, K.; Kong, X.; Zhang, J.; Hu, J.; Li, J.; Tang, H. Computer vision-based bridge inspection and monitoring: A review. *Sensors* **2023**, *23*, 7863. [CrossRef] [PubMed]
42. Xu, H.; Su, X.; Xu, H.; Li, H. Autonomous Bridge Crack Detection Using Deep Convolutional Neural Networks. In Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019), Chongqing, China, 30–31 May 2019; pp. 274–284. [CrossRef]
43. Perry, B.J.; Guo, Y.; Atadero, R.; van de Lindt, J.W. Streamlined bridge inspection system utilizing unmanned aerial vehicles (UAVs) and machine learning. *Measurement* **2020**, *164*, 108048. [CrossRef]
44. Cabral, R.; Oliveira, R.; Ribeiro, D.; Rakoczy, A.M.; Santos, R.; Azenha, M.; Correia, J. Railway Bridge Geometry Assessment Supported by Cutting-Edge Reality Capture Technologies and 3D As-Designed Models. *Infrastructures* **2023**, *8*, 114. [CrossRef]
45. Zhu, J.; Zhang, C.; Qi, H.; Lu, Z. Vision-based defects detection for bridges using transfer learning and convolutional neural networks. *Struct. Infrastruct. Eng.* **2019**, *16*, 1037–1049. [CrossRef]
46. Ahmadvand, M.; Dorafshan, S.; Azari, H.; Shams, S. 1D-CNNs for autonomous defect detection in bridge decks using ground penetrating radar. In Proceedings of the Health Monitoring of Structural and Biological Systems XV, Online, 22–26 March 2021; Volume 11593, pp. 97–113. [CrossRef]
47. Zoubir, H.; Rguig, M.; El Aroussi, M.; Chehri, A.; Saadane, R.; Jeon, G. Concrete Bridge Defects Identification and Localization Based on Classification Deep Convolutional Neural Networks and Transfer Learning. *Remote Sens.* **2022**, *14*, 4882. [CrossRef]
48. Teng, S.; Liu, Z.; Li, X. Improved YOLOv3-Based Bridge Surface Defect Detection by Combining High- and Low-Resolution Feature Images. *Buildings* **2022**, *12*, 1225. [CrossRef]
49. Kruachottikul, P.; Cooharojananone, N.; Phanomchoeng, G.; Chavarnakul, T.; Kovitanggoon, K.; Trakulwaranont, D. Deep learning-based visual defect-inspection system for reinforced concrete bridge substructure: A case of Thailand's department of highways. *J. Civ. Struct. Health Monit.* **2021**, *11*, 949–965. [CrossRef]
50. Pang, R.; Yang, Y.; Huang, A.; Liu, Y.; Zhang, P.; Tang, G. Multi-Scale Feature Fusion Model for Bridge Appearance Defect Detection. *Big Data Min. Anal.* **2024**, *7*, 1–11. [CrossRef]

51. Wan, H.; Gao, L.; Yuan, Z.; Qu, H.; Sun, Q.; Cheng, H.; Wang, R. A novel transformer model for surface damage detection and cognition of concrete bridges. *Expert Syst. Appl.* **2023**, *213*, 119019. [CrossRef]
52. Shamsabadi, E.A.; Xu, C.; Rao, A.S.; Nguyen, T.; Ngo, T.; Dias-Da-Costa, D. Vision transformer-based autonomous crack detection on asphalt and concrete surfaces. *Autom. Constr.* **2022**, *140*, 104316. [CrossRef]
53. Shang, H.; Sun, C.; Liu, J.; Chen, X.; Yan, R. Defect-aware transformer network for intelligent visual surface defect detection. *Adv. Eng. Inform.* **2023**, *55*, 101882. [CrossRef]
54. Yang, G.; Liu, K.; Zhang, J.; Zhao, B.; Zhao, Z.; Chen, X.; Chen, B.M. Datasets and processing methods for boosting visual inspection of civil infrastructure: A comprehensive review and algorithm comparison for crack classification, segmentation, and detection. *Constr. Build. Mater.* **2022**, *356*, 129226. [CrossRef]
55. Wang, W.; Su, C. Automatic Classification of Reinforced Concrete Bridge Defects Using the Hybrid Network. *Arab. J. Sci. Eng.* **2022**, *47*, 5187–5197. [CrossRef]
56. Shahin, M.; Chen, F.F.; Maghanaki, M.; Hosseinzadeh, A.; Zand, N.; Koodiani, H.K. Improving the Concrete Crack Detection Process via a Hybrid Visual Transformer Algorithm. *Sensors* **2024**, *24*, 3247. [CrossRef] [PubMed]
57. Li, M.; Yuan, J.; Ren, Q.; Luo, Q.; Fu, J.; Li, Z. CNN-Transformer hybrid network for concrete dam crack patrol inspection. *Autom. Constr.* **2024**, *163*, 105440. [CrossRef]
58. Noshiri, N.; Beck, M.A.; Bidinosti, C.P.; Henry, C.J. A comprehensive review of 3D convolutional neural network-based classification techniques of diseased and defective crops using non-UAV-based hyperspectral images. *Smart Agric. Technol.* **2023**, *5*, 100316. [CrossRef]
59. Helmi, W.; Bridgelall, R.; Askarzadeh, T. Remote Sensing and Machine Learning for Safer Railways: A Review. *Appl. Sci.* **2024**, *14*, 3573. [CrossRef]
60. Min, Y.; Xiao, B.; Dang, J.; Yue, B.; Cheng, T. Real time detection system for rail surface defects based on machine vision. *EURASIP J. Image Video Process.* **2018**, *2018*, 3. [CrossRef]
61. Zhao, J.; Yeung, A.W.-L.; Muhammad, A.; Lai, S.; To, V.; Ng, Y. CBAM-SwinT-BL: Small Rail Surface Defect Detection Method Based on Swin Transformer with Block Level CBAM Enhancement. September 2024. Available online: https://arxiv.org/abs/2409.20113v2 (accessed on 11 October 2024).
62. Vithanage, R.K.W.; Harrison, C.S.; DeSilva, A.K.M. Importance and Applications of Robotic and Autonomous Systems (RAS) in Railway Maintenance Sector: A Review. *Computers* **2019**, *8*, 56. [CrossRef]
63. Wang, H.; Li, M.; Wan, Z. Rail surface defect detection based on improved Mask R-CNN. *Comput. Electr. Eng.* **2022**, *102*, 108269. [CrossRef]
64. Shang, L.; Yang, Q.; Wang, J.; Li, S.; Lei, W. Detection of rail surface defects based on CNN image recognition and classification. In Proceedings of the 2018 20th International Conference on Advanced Communications Technology (ICACT), Chuncheon, Republic of Korea, 11–14 February 2018; pp. 45–51. [CrossRef]
65. Ferdousi, R.; Laamarti, F.; Yang, C.; El Saddik, A. A reusable AI-enabled defect detection system for railway using ensembled CNN. *Appl. Intell.* **2024**, *54*, 9723–9740. [CrossRef]
66. Zhang, H.; Song, Y.; Chen, Y.; Zhong, H.; Liu, L.; Wang, Y.; Akilan, T.; Wu, Q.M.J. MRSDI-CNN: Multi-Model Rail Surface Defect Inspection System Based on Convolutional Neural Networks. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 11162–11177. [CrossRef]
67. Zheng, D.; Li, L.; Zheng, S.; Chai, X.; Zhao, S.; Tong, Q.; Wang, J.; Guo, L. A Defect Detection Method for Rail Surface and Fasteners Based on Deep Convolutional Neural Network. *Comput. Intell. Neurosci.* **2021**, *2021*, 2565500. [CrossRef]
68. Guo, F.; Liu, J.; Qian, Y.; Xie, Q. Rail surface defect detection using a transformer-based network. *J. Ind. Inf. Integr.* **2024**, *38*, 100584. [CrossRef]
69. Lu, S.; Wang, J.; Jing, G.; Qiang, W.; Rad, M.M. Rail Defect Classification with Deep Learning Method. *Acta Polytech. Hung.* **2022**, *19*, 225–241. [CrossRef]
70. Ye, W.; Ren, J.; Li, C.; Liu, W.; Zhang, Z.; Lu, C. Intelligent Detection of Surface Defects in High-Speed Railway Ballastless Track Based on Self-Attention and Transfer Learning. *Struct. Control Health Monit.* **2024**, *2024*, 1–22. [CrossRef]
71. Pereira, G.A.; Hussain, M. A Review of Transformer-Based Models for Computer Vision Tasks: Capturing Global Context and Spatial Relationships. *arXiv* **2024**, arXiv:2408.15178. [CrossRef]
72. He, J.; Duan, R.; Dong, M.; Kao, Y.; Guo, G.; Liu, J. CNN-Transformer Bridge Mode for Detecting Arcing Horn Defects in Railway Sectional Insulator. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–16. [CrossRef]
73. Luo, H.; Cai, L.; Li, C. Rail Surface Defect Detection Based on An Improved YOLOv5s. *Appl. Sci.* **2023**, *13*, 7330. [CrossRef]
74. Yu, C.; Chen, X. Railway rutting defects detection based on improved RT-DETR. *J. Real-Time Image Process.* **2024**, *21*, 1–13. [CrossRef]
75. Bai, S.; Yang, L.; Liu, Y. A vision-based nondestructive detection network for rail surface defects. *Neural Comput. Appl.* **2024**, *36*, 12845–12864. [CrossRef]

76. Rail Track Surface Defect Detection Algorithm Integrating MobileNetv3 and Transformer. Available online: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/13401/1340108/Rail-track-surface-defect-detection-algorithm-integrating-MobileNetv3-and-Transformer/10.1117/12.3038836.full (accessed on 5 November 2024).
77. The Oxford Handbook of Energy Politics—Google Books. Available online: https://books.google.ca/books?hl=fr&lr=&id=dGkLEAAAQBAJ&oi=fnd&pg=PA99&dq=In+the+oil+and+gas+industry,+ensuring+the+safety+and+reliability+of+infrastructure+such+as+pipelines,+storage+tanks,+and+offshore+platforms+is+a+top+priority%5C&ots=L5roE2uZCw&sig=7qMxze1ofuAIc7OTknHQaUCdKHY&redir_esc=y#v=onepage&q&f=false (accessed on 5 November 2024).
78. Alamri, A.H. Localized corrosion and mitigation approach of steel materials used in oil and gas pipelines—An overview. *Eng. Fail. Anal.* **2020**, *116*, 104735. [CrossRef]
79. Bottle Cap Inspection with Computer Vision. Available online: https://blog.roboflow.com/bottle-cap-inspection/ (accessed on 10 October 2024).
80. Zhang, M.; Guo, Y.; Xie, Q.; Zhang, Y.; Wang, D.; Chen, J. Defect identification for oil and gas pipeline safety based on autonomous deep learning network. *Comput. Commun.* **2022**, *195*, 14–26. [CrossRef]
81. Katser, I.; Kozitsin, V.; Mozolin, I. MFL Data Preprocessing and CNN-based Oil Pipeline Defects Detection. September 2023. Available online: https://arxiv.org/abs/2310.00332v1 (accessed on 5 November 2024).
82. Xin, H.; Du, X. Research on Safety Monitoring and Defect Identification Method of Oil and Gas Pipeline Based on Faster R-CNN Technology. In Proceedings of the 2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA), Dalian, China, 27–29 October 2023; pp. 474–479.
83. Abeng, F.; Nyong, B.; Ikpi, M.; Obeten, M. Application of CNN for multiple phase corrosion identification and region detection. *Appl. Soft Comput.* **2024**, *164*, 112008. [CrossRef]
84. Fei, Z.; West, G.M.; Murray, P.; Dobie, G. CNN-based automated approach to crack-feature detection in steam cycle components. *Int. J. Press. Vessel. Pip.* **2023**, *207*, 105112. [CrossRef]
85. Liu, Q.; Huang, X.; Shao, X.; Hao, F. Industrial cylinder liner defect detection using a transformer with a block division and mask mechanism. *Sci. Rep.* **2022**, *12*, 1–14. [CrossRef]
86. Alaa, T.; Kotb, M.; Zakaria, A.; Diab, M.; Gomaa, W. Automated Detection of Defects on Metal Surfaces Using Vision Transformers. In Proceedings of the 21st International Conference on Informatics in Control, Automation and Robotics, Porto, Portugal, 18–20 November 2024; pp. 36–45. Available online: https://arxiv.org/abs/2410.04440v1 (accessed on 23 December 2024).
87. An, K.; Zhang, Y. LPViT: A Transformer Based Model for PCB Image Classification and Defect Detection. *IEEE Access* **2022**, *10*, 42542–42553. [CrossRef]
88. Ding, C.; Tang, D.; Zheng, X.; Wang, Q.; He, Y.; Long, Z. DHT: Dynamic Vision Transformer Using Hybrid Window Attention for Industrial Defect Images Classification. *IEEE Instrum. Meas. Mag.* **2023**, *26*, 19–28. [CrossRef]
89. Yao, J.; Liang, W.; Xiong, J. Novel intelligent diagnosis method of oil and gas pipeline defects with transfer deep learning and feature fusion. *Int. J. Press. Vessel. Pip.* **2022**, *200*, 104781. [CrossRef]
90. Zhang, M.; Guo, Y.; Xie, Q.; Zhang, Y.; Wang, D.; Chen, J. Estimation of Defect Size and Cross-Sectional Profile for the Oil and Gas Pipeline Using Visual Deep Transfer Learning Neural Network. *IEEE Trans. Instrum. Meas.* **2022**, *72*, 1–13. [CrossRef]
91. Chen, P.; Li, R.; Fu, K.; Zhong, Z.; Xie, J.; Wang, J.; Zhu, J. A cascaded deep learning approach for detecting pipeline defects via pretrained YOLOv5 and ViT models based on MFL data. *Mech. Syst. Signal Process.* **2023**, *206*, 110919. [CrossRef]
92. Flores Cuenca, M.F.; Malikov, A.K.; Kim, J.; Cho, Y.; Jeong, K.S. A novel flaw detection approach in carbon steel pipes through ultrasonic guided waves and optimized transformer neural networks. *J. Mech. Sci. Technol.* **2024**, *38*, 3253–3263. [CrossRef]
93. Alber, M.; Hönes, C.; Baier, P. Evaluating Vision Transformer Models for Visual Quality Control in Industrial Manufacturing. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2024; Volume 14950, pp. 116–132. [CrossRef]
94. Dosovitskiy, A. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual Event, Austria, 3–7 May 2021; Available online: https://arxiv.org/abs/2010.11929v2 (accessed on 10 July 2024).
95. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [CrossRef]
96. Han, S.; Mao, H.; Dally, W.J. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016—Conference Track Proceedings, San Juan, Puerto Rico, 2–4 May 2016; Available online: https://arxiv.org/abs/1510.00149v5 (accessed on 15 December 2024).
97. Xu, H.; Li, C.; Rahaman, M.; Yao, Y.; Li, Z.; Zhang, J.; Kulwa, F.; Zhao, X.; Qi, S.; Teng, Y. An enhanced framework of generative adversarial networks (EF-GANs) for environmental microorganism image augmentation with limited rotationinvariant training data. *IEEE Access* **2020**, *8*, 187455–187469. [CrossRef]

98. Han, T.; Liu, C.; Wu, R.; Jiang, D. Deep transfer learning with limited data for machinery fault diagnosis. *Appl. Soft Comput.* **2021**, *103*, 107150. [CrossRef]
99. Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; Loy, C.C. Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4396–4415. [CrossRef] [PubMed]
100. Rezayiye, R.K.; Laurent, K.; Nooralishahi, P.; Ibarra-Castanedo, C.; Maldague, X. Thermal Data Augmentation Approach for the Detection of Corrosion in Pipes Using Deep Learning and Finite Element Modelling. *Eng. Proc.* **2023**, *51*, 20. [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.