From Method-Centric Optimisation to Failure-Oriented Robustness: A Framework for Industrial Inspection Systems

Ying Zhou[1], Mei Choo Ang[2], Ummul Hanan Mohamad[3],
Ah-Lian Kor[4], Kok Weng Ng[5], Saraswathy Shamini Gunasekaran[6], Halimah Badioze Zaman[7]
Institute of Visual Informatics (IVI), Universiti Kebangsaan Malaysia (UKM),
43600 UKM Bangi, Selangor, Malaysia[1,2,3]
School of Built Environment, Engineering and Computing,
Leeds Beckett University, Leeds LS1 3HE, United Kingdom[4]
Faculty of Science and Engineering, Department of Mechanical, Materials and Manufacturing Engineering,
University of Nottingham Malaysia, Semenyih, Selangor, Malaysia[5]
Institute of Informatics and Computing in Energy (IICE),
Universiti Tenaga Nasional (UNITEN), Kajang, Selangor, Malaysia[6,7]
College of Computing and Informatics (CCI),
Universiti Tenaga Nasional (UNITEN), Kajang, Selangor, Malaysia[6]
School of Computer Science, Faculty of Innovation and Technology, Taylor's University, Malaysia[6]

***Abstract*—Robust inspection of aeroengine turbine blades remains a critical challenge in safety-critical industrial environments, where limited data availability, class imbalance, optimisation bias, and imaging degradation can reduce the reliability of learning-based inspection systems. Existing studies mainly improve robustness through isolated algorithmic optimisation, while providing limited understanding of why inspection failures occur under different industrial conditions. This study proposes a Failure-Oriented Robustness Framework that categorises robustness degradation into three dominant failure sources: distribution failure, decision failure, and representation failure. The framework establishes a systematic relationship between failure diagnosis, intervention selection, and robustness interpretation. The applicability of the proposed framework is demonstrated through representative turbine blade inspection scenarios involving two-dimensional surface defect detection and three-dimensional CT image enhancement. Data-level experiments investigate augmentation behaviour under different dataset characteristics, model-level experiments analyse optimisation objectives and architectural factors, and imaging-level experiments evaluate CT enhancement under degradation conditions. Experimental results demonstrate that robustness improvement depends on the alignment between intervention strategies and dominant failure sources rather than algorithmic complexity alone. Under severe distribution failure, targeted augmentation improved mAP@0.5 from 0.290 to 0.599, while objective alignment increased detection performance from 0.588 to 0.710 without increasing model complexity. For CT representation recovery, the proposed enhancement approach achieved a 3.21 dB PSNR improvement compared with U-Net. The findings demonstrate that robustness is a failure-dependent property rather than an intrinsic characteristic of individual methods. The proposed framework provides an interpretable analytical perspective for diagnosing robustness degradation and guiding failure-oriented intervention design in safety-critical industrial inspection applications.**

Keywords*—Industrial inspection; robustness; Failure-Oriented Framework; deep learning; computer vision; turbine blade inspection; artificial intelligence

I. Introduction

Industrial inspection systems play a critical role in ensuring product quality, operational reliability, and safety across manufacturing, aerospace, energy, and infrastructure sectors [1]. Recent advances in computer vision and artificial intelligence have significantly improved the capability of automated inspection systems, enabling accurate detection of surface defects, structural anomalies, and internal material degradation [2]. As a result, deep learning-based inspection methods have become increasingly prevalent in safety-critical industrial applications [2].

Despite these advances, robustness remains a persistent challenge in industrial inspection [3]. Unlike controlled benchmark environments, real-world inspection systems frequently operate under adverse conditions, including limited training data, severe class imbalance, domain variability, optimisation instability, and imaging degradation [4], [5]. Under such conditions, inspection performance may deteriorate substantially, reducing system reliability and limiting industrial deployment [3].

Consequently, robustness has become a central research objective in industrial inspection [3]. A robust inspection system should not only achieve high performance under ideal conditions but also maintain reliable behaviour when confronted with uncertainty, data scarcity, and environmental variability.

Existing research has sought to improve robustness through various methodological approaches [1], [2]. Data-level studies primarily focus on augmentation, synthetic data generation, and domain adaptation to address data scarcity and imbalance [5], [4], [30]. Model-level studies investigate optimisation objectives, loss functions, architectural refinements, and attention mechanisms to improve learning behaviour and training stability [2], [7], [6]. Imaging-level studies employ enhancement, restoration, and reconstruction techniques to mitigate the effects of degraded image acquisition [1].

Although these approaches have achieved notable performance improvements, they are typically investigated as independent technical solutions. Consequently, robustness is often interpreted as a by-product of algorithmic improvement rather than as a phenomenon that can be systematically analysed and explained.

A fundamental limitation of existing research is the lack of a unified perspective on why certain interventions improve robustness under some conditions but fail under others [3], [6]. Methods that demonstrate strong performance on one dataset frequently exhibit inconsistent behaviour when applied to different inspection scenarios. Such observations suggest that robustness behaviour may be governed by underlying failure mechanisms rather than by algorithmic superiority alone.

However, current industrial inspection literature provides limited guidance on identifying these failure mechanisms or on selecting interventions based on the source of robustness degradation [1], [2]. As a result, intervention selection often relies on empirical trial and error rather than systematic failure diagnosis.

To address this challenge, this study proposes a Failure-Oriented Robustness Framework for industrial inspection. Rather than analysing robustness from the perspective of individual algorithms, the proposed framework organises robustness challenges according to three dominant sources of failure: distribution failure, decision failure, and representation failure. The framework further establishes explicit relationships between failure diagnosis, intervention selection, and robustness interpretation, providing a systematic mechanism for understanding robustness behaviour across different inspection conditions.

The main contributions of this work are summarised as follows:

- A Failure-Oriented Robustness Framework is proposed to organise industrial inspection robustness challenges according to dominant failure sources rather than methodological categories.
- A failure-to-intervention mapping mechanism is introduced to connect robustness degradation with appropriate intervention strategies across data-level, model-level, and imaging-level inspection stages.
- The framework is validated through representative turbine blade inspection scenarios covering distribution failure, decision failure, and representation failure.
- A failure-oriented interpretation of robustness is established, demonstrating that robustness emerges from alignment between intervention strategies and dominant failure sources rather than increasing algorithmic complexity alone.

II. Reframing Robustness in Industrial Inspection

A. Existing Paradigms of Industrial Inspection

Industrial inspection systems have evolved from procedure-driven inspection towards intelligent and reliability-aware paradigms, driven by continuous advances in sensing technologies, machine vision, and artificial intelligence, as well as increasing industrial reliability requirements [1], [13], [15]. Rather than replacing previous approaches, each new paradigm extends inspection capability by introducing new perspectives on robustness and reliability. As a result, several dominant paradigms have emerged throughout the evolution of industrial inspection.

The earliest paradigm is the procedural NDT-centred framework, in which reliability is achieved through standardised inspection procedures, operator qualification, and Probability of Detection (PoD) validation [8], [9]. Inspection performance primarily depends on compliance with established inspection protocols and human expertise. Within this paradigm, robustness is interpreted as operational repeatability, procedural consistency, and regulatory compliance, with limited capability to model computational failure or data-driven uncertainty.

Subsequently, inspection research shifted towards an accuracy-centric, learning-based framework with the rapid adoption of machine vision and deep learning technologies [12], [2], [10]. Research within this paradigm primarily focuses on feature representation, model optimisation, and quantitative performance metrics such as mAP, precision, and recall [14], [11]. Robustness is therefore typically regarded as an implicit consequence of improved detection accuracy, assuming that better-performing models naturally exhibit greater robustness. Consequently, robustness is often evaluated through performance improvement rather than through explicit analysis of the mechanisms responsible for robustness degradation.

More recently, a reliability-oriented inspection paradigm has emerged to address the increasing demand for trustworthy artificial intelligence in safety-critical applications [17], [19], [16]. Recent studies have incorporated uncertainty quantification, confidence estimation, explainability, verification procedures, and risk-aware decision-making mechanisms to improve inspection reliability under uncertain operating conditions [3]. Although these approaches elevate reliability as a primary objective, they generally remain modality-specific and concentrate on uncertainty management rather than on identifying the underlying mechanisms responsible for robustness degradation.

Collectively, these paradigms have substantially advanced the capability of industrial inspection systems. However, they primarily organise robustness around methodologies, algorithms, or evaluation objectives rather than around the sources of robustness degradation themselves. Consequently, existing paradigms provide limited guidance on why particular intervention strategies succeed under certain inspection conditions but fail under others. This limitation motivates the emergence of a failure-oriented perspective, which is introduced in the following subsection.

B. Comparative Perspective on Robustness

Table I illustrates that the interpretation of robustness has progressively evolved from procedural consistency, to performance optimisation, and subsequently to trustworthy and reliable AI systems [17], [18], [19]. Although each paradigm contributes to improving inspection capability, robustness is

TABLE I. COMPARISON OF INDUSTRIAL INSPECTION PARADIGMS FROM A ROBUSTNESS PERSPECTIVE.

Paradigm	Primary Objective	Robustness Interpretation	Intervention Logic	Primary Limitation
Procedural NDT	Compliance and qualification	Operational repeatability	Procedural standardisation	Limited computational failure modelling
Accuracy-Centric Learning	Performance optimisation	Implicit outcome of accuracy improvement	Architecture and performance optimisation	Underlying failure mechanisms remain unexplained
Reliability-Oriented Inspection	Risk minimisation and uncertainty management	Reliable operation under uncertainty	Verification, explainability and uncertainty modelling	Failure diagnosis remains modality-specific
Failure-Oriented (Proposed)	Robustness-oriented inspection design	Explicit failure-dependent robustness	Failure diagnosis and intervention alignment	Requires accurate failure characterisation

consistently organised around methodologies, algorithms, or evaluation objectives rather than around the mechanisms responsible for robustness degradation.

Consequently, robustness is commonly interpreted as an operational outcome, an implicit consequence of improved detection accuracy, or a reliability attribute associated with uncertainty management and explainability [20], [21]. Such interpretations provide valuable guidance for evaluating system performance but offer limited explanatory capability regarding why identical intervention strategies frequently exhibit different behaviour across heterogeneous inspection scenarios.

This limitation becomes increasingly significant as industrial inspection systems incorporate heterogeneous datasets, increasingly complex optimisation strategies, and multi-modal sensing pipelines. These observations motivate a transition from methodology-oriented optimisation towards failure-oriented robustness analysis, which forms the conceptual foundation of the proposed framework.

C. Why Existing Paradigms are Insufficient?

The rapid adoption of AI-based inspection systems has revealed a growing discrepancy between performance optimisation and understanding of robustness. Although recent advances have substantially improved inspection accuracy, numerous studies have shown that methods that perform strongly under one set of conditions often exhibit inconsistent behaviour when applied to different datasets, defect distributions, optimisation settings, or imaging environments [3], [17], [18].

For example, augmentation strategies that successfully mitigate data scarcity in one dataset may provide negligible benefit in another with different distribution characteristics, highlighting the challenges of domain adaptation and domain generalisation [4], [5]. Similarly, architectural refinement does not necessarily improve robustness when optimisation objectives remain misaligned, suggesting that learning dynamics and objective design often have a greater influence on robustness than model complexity alone [6], [7]. In imaging-based inspection systems, image enhancement often contributes more to inspection reliability than modifications to the downstream detector, suggesting that robustness may be constrained more by the quality of the image representation than by the detection model itself.

These observations suggest that robustness behaviour cannot be fully explained through methodological categories alone. Instead, robustness degradation appears to stem from distinct failure mechanisms introduced at different stages of the inspection pipeline. Consequently, improving robustness requires identifying the dominant source of failure before selecting an appropriate intervention strategy, motivating the failure-oriented perspective introduced in the following subsection.

D. Emergence of a Failure-Oriented Perspective

The limitations of existing paradigms motivate a shift from method-centric optimisation towards failure-oriented robustness analysis. Rather than asking which algorithm performs best, a failure-oriented perspective seeks to identify the dominant source of robustness degradation before selecting an appropriate intervention strategy [22].

From this perspective, robustness challenges can be interpreted in terms of three primary sources of failure. Distribution failure originates from data scarcity, class imbalance, and distributional variability. Decision failure arises from misalignment in optimisation during model learning. Representation failure arises from image degradation and information loss prior to downstream analysis.

By explicitly connecting failure diagnosis, intervention selection, and robustness outcomes, a failure-oriented perspective provides a more systematic explanation of robustness behaviour across heterogeneous industrial inspection scenarios.

Fig. 1 illustrates how industrial inspection paradigms have evolved from procedure-oriented inspection and accuracy-centric optimisation towards reliability-aware inspection, culminating in the proposed failure-oriented robustness paradigm. Rather than organising robustness around algorithms or methodologies, the proposed paradigm organises robustness around dominant failure mechanisms, providing a unified perspective for robustness analysis and intervention selection.

E. Positioning of the Proposed Framework

Building upon the paradigm evolution illustrated in Fig. 1, the proposed framework extends existing industrial inspection paradigms by introducing a failure-oriented perspective for robustness analysis. Unlike previous paradigms, which organise robustness around procedures, algorithms, or reliability metrics, the proposed framework organises robustness according to dominant failure mechanisms.

This positioning shifts the research focus from identifying the best-performing method to understanding why robustness degrades and how intervention strategies should be selected according to the underlying source of failure. The proposed framework therefore serves as the conceptual foundation for the failure-oriented robustness architecture presented in the following section.

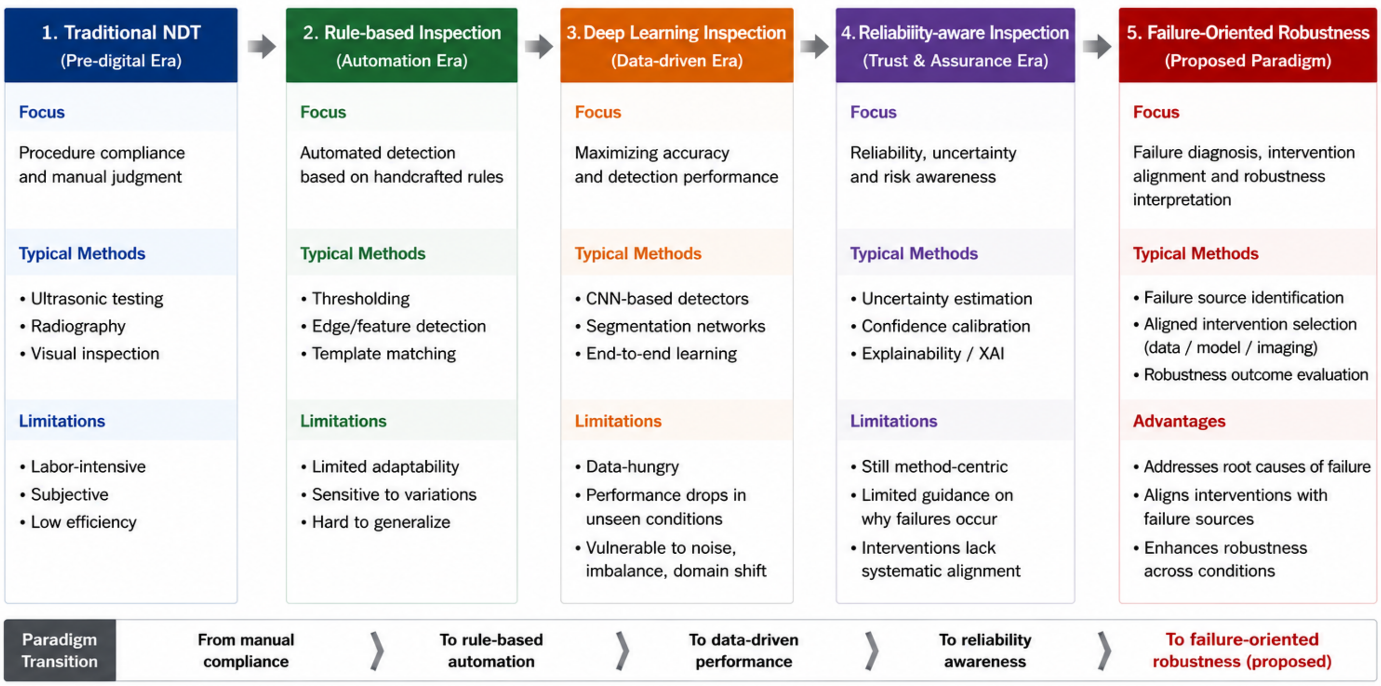

Fig. 1. Evolution of robustness paradigms in industrial inspection.

III. Proposed Failure-Oriented Robustness Framework

A. Motivation for a Failure-Oriented Perspective

Recent advances in industrial inspection have largely focused on developing new detection architectures, optimisation strategies, and augmentation techniques to improve detection accuracy. Although these approaches have achieved notable performance gains, robustness remains difficult to explain and predict under realistic industrial conditions. Methods that perform well on one dataset frequently exhibit inconsistent behaviour when applied to different defect distributions, data scales, or imaging conditions [17], [18].

This inconsistency suggests that inspection performance is often governed by underlying failure mechanisms rather than algorithmic complexity alone [22], [23]. In industrial inspection systems, failures may originate from multiple sources, including data scarcity and class imbalance, optimisation bias during model training, and imaging degradation caused by acquisition artifacts [4], [5], [6], [7], [24], [25]. These sources of failure affect different stages of the inspection pipeline and cannot always be mitigated by a single methodological improvement.

However, existing studies typically analyse augmentation, model optimisation, and image enhancement as independent technical solutions [5], [7], [24]. As a result, robustness is often treated as an implicit outcome of performance improvement rather than an explicit design objective [17], [19]. This motivates the need for a failure-oriented perspective that focuses on identifying dominant sources of failure and selecting interventions based on the nature of the failure.

Based on this motivation, a Failure-Oriented Robustness Framework is proposed to systematically connect failure diagnosis, intervention selection, and robustness interpretation across data-level, model-level, and imaging-level inspection stages.

B. Failure Taxonomy and Definitions

To improve the reproducibility and conceptual clarity of the proposed framework, the three failure categories are explicitly defined according to the stage at which robustness degradation is introduced into the industrial inspection pipeline. Rather than categorising robustness according to algorithmic families, the proposed framework classifies failures according to their origin and the corresponding intervention stage.

1) Distribution failure: It refers to a degradation of robustness arising from deficiencies in the training data before model optimisation. Typical causes include limited sample availability, class imbalance, distribution shift, and insufficient data diversity. Such failures primarily affect the statistical characteristics of the learning data and are generally mitigated through data-oriented interventions such as augmentation, sampling, and domain adaptation.

2) Decision failure: It refers to robustness degradation introduced during model optimisation and inference. Even when sufficient data are available, inappropriate optimisation objectives, loss functions, class weighting strategies [33], or decision boundaries may lead to unstable learning behaviour and biased predictions. Decision failure therefore originates from misalignment in optimisation rather than data insufficiency.

3) Representation failure: It refers to robustness degradation caused by information loss before feature learning begins. This category includes image degradation resulting from acquisition noise, motion blur, illumination variation, CT artifacts, low contrast, and other sensing imperfections. Representation failure limits the quality of information available to downstream learning algorithms and is typically addressed through image enhancement or restoration techniques. Unlike distribution failure and decision failure, representation failure is evaluated at the image representation level rather than directly through downstream detection performance. The objective of a representation-oriented intervention is to recover degraded information and improve the quality of inputs available for subsequent inspection tasks.

Accordingly, the proposed framework distinguishes the three failure categories according to the stage at which robustness degradation is introduced: before model optimisation (distribution failure), during optimisation and inference (decision failure), and before feature learning due to acquisition degradation (representation failure). The taxonomy and formal definitions of the three failure categories are summarised in Table II. This stage-oriented definition establishes the conceptual boundary between failure categories and provides the basis for failure-oriented intervention selection.

In practical industrial inspection systems, multiple failure sources may coexist simultaneously. The proposed framework does not assume that distribution failure, decision failure, and representation failure are mutually exclusive. Instead, these categories provide a structured mechanism for diagnosing the dominant source of robustness degradation under a specific inspection scenario.

The dominant failure refers to the failure source that contributes most significantly to the observed robustness degradation and should therefore receive priority during intervention selection. Other failure sources may still exist as secondary factors and may require additional interventions when their influence becomes significant.

C. Framework Architecture

The proposed Failure-Oriented Robustness Framework is designed to analyse robustness from the perspective of failure generation rather than algorithm selection. Instead of treating augmentation, optimisation, and enhancement as independent technical solutions, the framework organises robustness challenges according to the dominant failure mechanisms responsible for inspection performance degradation [5], [7], [24].

As illustrated in Fig. 2, the framework consists of three hierarchical layers: Distribution Failure, Decision Failure, and Representation Failure. These layers correspond to different stages of the industrial inspection pipeline and collectively describe how robustness degradation emerges and propagates throughout the system.

1) Distribution failure layer: These layers are arranged according to the stage at which robustness degradation is introduced into the inspection pipeline, ranging from data generation and model decision-making to image representation quality. Distribution failure originates from deficiencies in the training data distribution, including data scarcity, class imbalance, and domain variability. Such failures often manifest as poor generalisation capability, unstable rare-class detection, and sensitivity to distribution shifts [4], [5]. Since the underlying issue is data-related, interventions at this layer primarily focus on modifying the data distribution to improve representation diversity and class coverage.

2) Decision failure layer: Decision failure occurs when the learning objective becomes misaligned with the inspection objective [6], [7]. Even when sufficient data are available, optimisation objectives may favour dominant classes, suppress minority categories, or produce unstable localisation behaviour. These failures are associated with the model's decision-making process rather than with the data itself. Consequently, improvement in robustness at this layer focuses on objective design, loss functions, weighting mechanisms, and optimization strategies.

3) Representation failure layer: Representation failure arises before the learning stage and is caused by degradation in the acquired inspection data [24], [25], [26]. Common examples include image artifacts, noise contamination, contrast reduction, and resolution loss. These effects distort the defect representation and limit the amount of useful information available for downstream analysis. Therefore, interventions at this layer primarily involve enhancement, restoration, and super-resolution techniques [24], [26], [36].

Together, these three layers provide a unified view of robustness in industrial inspection systems, based on failure. Unlike conventional method-centric approaches that evaluate individual algorithms in isolation, the proposed framework emphasises identifying dominant failure mechanisms and selecting interventions that directly address the underlying cause of performance degradation. From this perspective, robustness is understood as the outcome of alignment between intervention strategies and failure sources, rather than a direct consequence of increasing algorithmic complexity.

D. Failure Source–Intervention Mapping

While the framework architecture identifies the dominant sources of robustness degradation, practical improvements in robustness require a mechanism for selecting appropriate interventions. Conventional industrial inspection studies typically adopt a method-centric workflow in which augmentation techniques, optimisation strategies, or enhancement algorithms are evaluated independently [5], [7], [24]. Under such a paradigm, intervention selection is often driven by methodological popularity rather than the underlying cause of failure.

The proposed framework adopts a failure-oriented intervention strategy. Instead of asking which method performs best, the framework first identifies the dominant failure source and subsequently selects interventions that directly target the corresponding failure mechanism. This process establishes an explicit relationship between failure diagnosis and intervention selection.

For distribution failures, robustness degradation originates from deficiencies in the training data distribution, including scarcity, imbalance, and domain variability. Consequently, interventions should focus on modifying the data distribution through augmentation, balancing, or domain adaptation techniques.

For decision failures, performance degradation arises from optimisation bias and objective misalignment during model learning. Therefore, interventions should focus on objective

TABLE II. TAXONOMY AND DEFINITIONS OF THE PROPOSED FAILURE CATEGORIES

Failure Category	Failure Stage	Origin of Failure	Typical Intervention
Distribution Failure	Before model optimisation	Training data deficiencies, including limited samples, class imbalance, and distribution shift	Data augmentation, sampling, and domain adaptation
Decision Failure	During optimisation and inference	Misalignment between optimisation objectives and inspection objectives, including inappropriate loss functions, optimisation strategies, and decision boundaries	Loss function design, optimisation strategy, and class weighting
Representation Failure	Before feature learning	Information degradation caused by image acquisition artifacts, noise, blur, low contrast, and sensing imperfections	Image enhancement, image restoration, and super-resolution

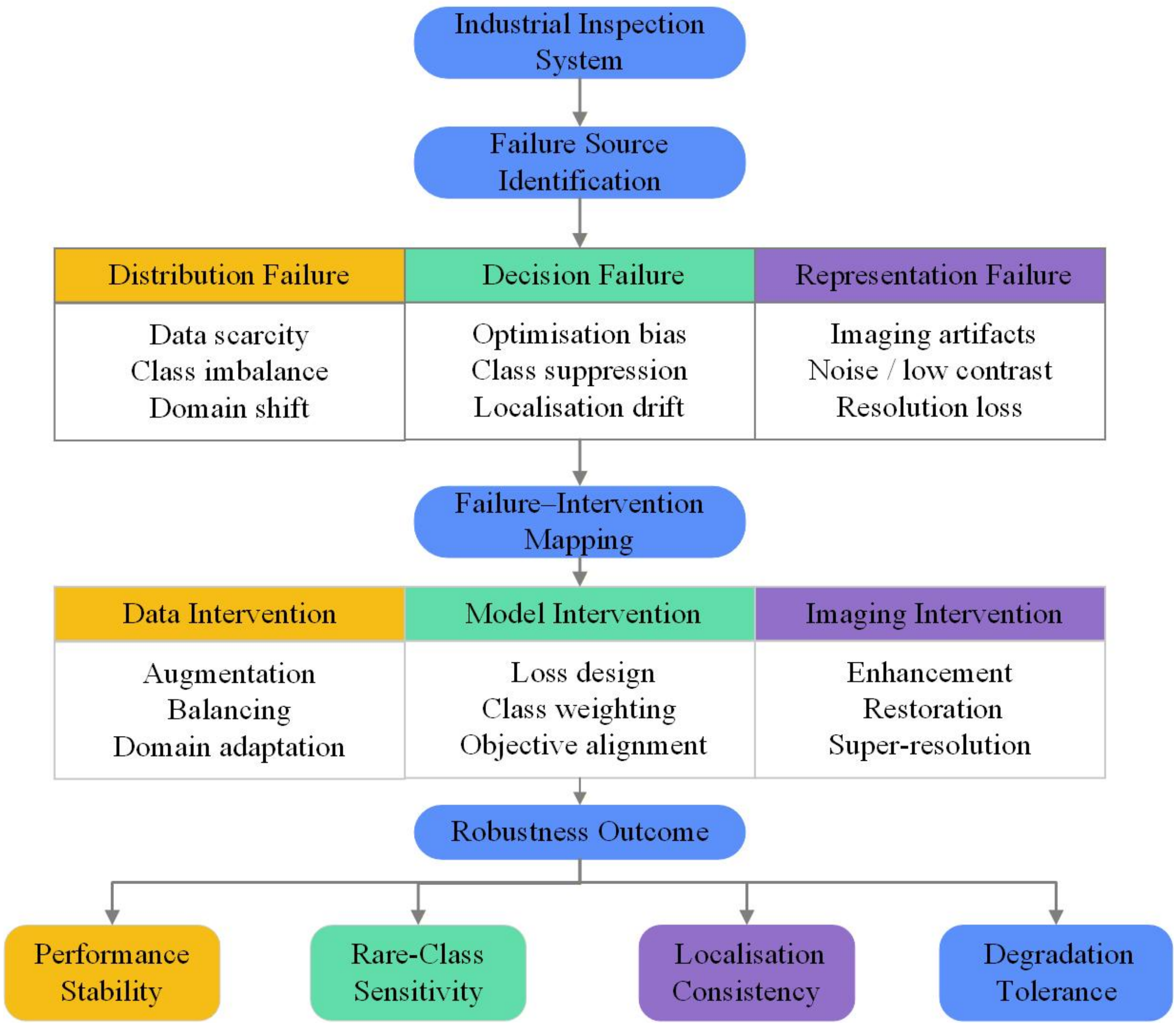

Fig. 2. Proposed Failure-Oriented Robustness Framework for industrial inspection. The framework organises robustness challenges according to three dominant sources of failure and maps them to corresponding intervention strategies and robustness outcomes.

design, loss formulation, class weighting, and optimisation strategies that improve decision consistency and minority-class representation.

For representation failures, degradation occurs before model learning due to corruption of the acquired inspection data. In such cases, robustness improvement should prioritise enhancement, restoration, and super-resolution techniques to recover defect visibility and structural information.

The proposed mapping mechanism does not assume the existence of a universally optimal intervention. Instead, it emphasises alignment between intervention strategy and dominant failure source. Under this perspective, intervention effectiveness is considered failure-dependent rather than method-dependent, providing a systematic basis for robustness-oriented decision making in industrial inspection systems.

E. Robustness Interpretation Mechanism

A central objective of the proposed framework is to reinterpret robustness beyond conventional accuracy-oriented evaluation. Existing industrial inspection studies predominantly assess performance using aggregate metrics such as accuracy, precision, recall, and mean Average Precision (mAP) [17], [18]. While these metrics provide useful indicators of detection performance, they do not necessarily reflect an inspection system's ability to maintain reliable behaviour under adverse operating conditions.

Within the proposed framework, robustness is defined as an inspection system's ability to maintain stable, reliable performance when confronted with dominant failure sources. Consequently, robustness is viewed as a multidimensional property rather than a single performance metric.

To facilitate the interpretation of robustness, four complementary dimensions are considered. Performance Stability evaluates the consistency of inspection performance across varying operating conditions. Rare-Class Sensitivity measures the system's ability to maintain detection effectiveness for minority defect categories. Localisation Consistency reflects the stability and reliability of defect localisation behaviour during model prediction. Degradation Tolerance assesses the inspection system's ability to maintain performance under imaging degradation, noise contamination, or structural corruption.

These dimensions provide an interpretable link between failure diagnosis, intervention effectiveness, and robustness outcomes. Under the proposed framework, robustness is improved not merely by higher accuracy, but by sustained performance under the specific failure conditions that dominate the inspection process. Therefore, robustness is interpreted as the observable outcome of successful alignment between intervention strategies and failure mechanisms.

F. Practical Application Workflow

While Fig. 2 explains what constitutes the framework, Fig. 3 illustrates how the framework is applied during practical inspection. As illustrated in Fig. 3, the proposed framework can be operationalised through a sequential failure-oriented workflow.

To facilitate practical deployment, the proposed framework is implemented through a sequential failure-oriented workflow. Rather than selecting algorithms directly, the workflow begins with diagnosing the dominant source of robustness degradation and subsequently identifying interventions that target the corresponding failure mechanism.

The workflow consists of five stages. First, the inspection scenario is analysed to identify observable performance limitations, such as poor rare-class detection, unstable localisation behaviour, or degraded image quality. Second, these symptoms are traced to their dominant source of failure, which may be categorised as distribution, decision, or representation failure.

Third, intervention strategies are selected according to the identified failure category. Distribution failures are addressed through data-oriented interventions, decision failures through optimisation-oriented interventions, and representation failures through imaging-oriented interventions. Fourth, robustness outcomes are evaluated using the interpretation dimensions defined in the framework, including performance stability, rare-class sensitivity, localisation consistency, and degradation tolerance.

Finally, the intervention results are analysed, and the workflow is iteratively refined if the dominant failure source remains unresolved. This iterative process allows robustness improvement to be guided by failure diagnosis rather than algorithmic trial-and-error.

The proposed workflow transforms the framework from a conceptual robustness model into a structured analytical procedure for industrial inspection applications. By prioritising failure diagnosis before intervention selection, the framework provides a systematic procedure for analysing robustness degradation and guiding the development of robustness-oriented inspection systems.

IV. Framework Validation Through Turbine Blade Inspection

A. Experimental Settings

Since the proposed framework investigates robustness degradation from different failure sources, the experimental configurations were designed according to the characteristics of each validation scenario. The data-level, model-level, and imaging-level experiments correspond to different intervention mechanisms; therefore, their training configurations are not identical. Instead, each experimental setting was controlled to ensure a fair comparison among methods within the corresponding validation scenario.

1) 2D Detection experiments: The 2D industrial inspection experiments were conducted using the YOLOv8 object detection framework implemented with the Ultralytics library. YOLO-based detectors have been widely adopted for real-time object detection because of their balance between detection accuracy and computational efficiency [28]. The experiments were performed using Python 3.9.19, PyTorch 2.4.1 with CUDA 12.4, and an NVIDIA GeForce RTX 4060 Laptop GPU with 8 GB VRAM. The adopted YOLOv8n detector contains 3,006,623 parameters and 8.1 GFLOPs.

For the data-level validation, the objective was to investigate whether augmentation strategies could alleviate distribution failure caused by limited samples, class imbalance, and distribution variability. Therefore, the augmentation strategy was considered the primary experimental variable, while the detector training configuration was maintained consistently within each dataset. The models were trained for 100 epochs with an input resolution of 640×640 pixels and a batch size of 1. The default Ultralytics optimization configuration was adopted, and different augmentation strategies with their corresponding parameters were applied according to the experimental design.

For the model-level validation, the objective was to analyse decision failure caused by optimisation misalignment. Unlike

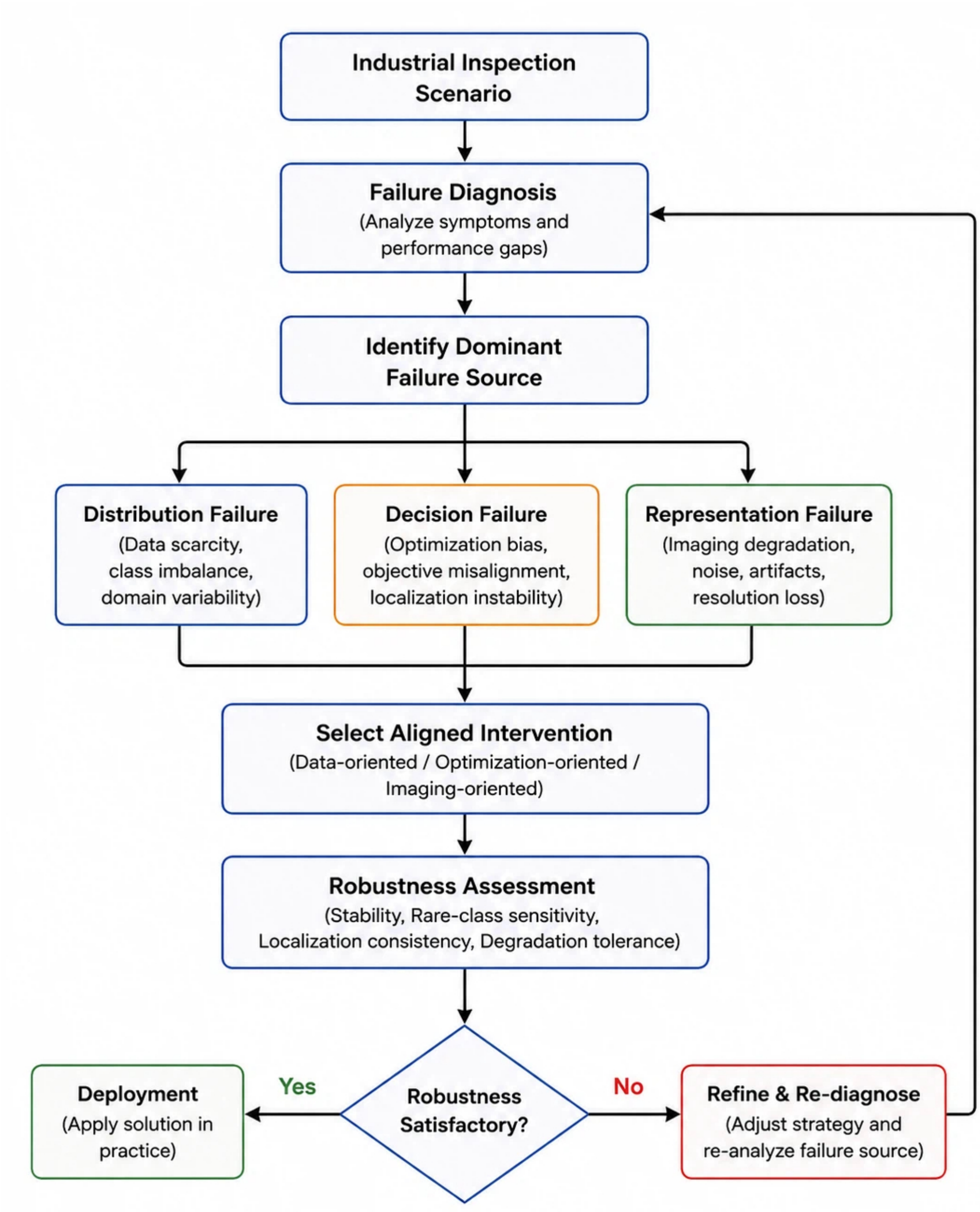

Fig. 3. Failure-oriented robustness workflow for industrial inspection systems.

the data-level experiments, optimisation strategies were treated as the primary experimental variables. Different optimizers, loss functions, class weighting strategies, and architectural modifications were evaluated under controlled training conditions. The models were trained for 300 epochs with an input resolution of 640 $\times$ 640 pixels and a batch size of 16. A fixed random seed of 42 was adopted to improve reproducibility. The optimizer configuration was adjusted according to the specific comparison objectives, including Adam [34] and SGD evaluations. Early stopping was disabled to ensure that all models completed the predefined training schedule.

For all 2D detection experiments, the performance was evaluated using mean average precision at IoU threshold 0.5 (mAP@0.5), mean average precision across IoU thresholds from 0.5 to 0.95 (mAP@0.5:0.95), precision, recall, and class-wise average precision.

2) CT Representation recovery experiments: The CT experiments were designed to validate representation failure caused by image acquisition degradation. Unlike the 2D detection experiments, the objective of this validation scenario was not to directly improve defect detection performance but to recover degraded image representations caused by CT artifacts.

A lightweight GNN-based artifact removal framework was implemented using PyTorch. Several restoration models, including CNN, U-Net [35], and different GNN variants, were compared under identical training conditions. The CT images were converted into normalized grayscale representations with pixel values scaled to the range of [0,1]. The training and validation datasets were loaded according to predefined image-mask pairs.

All restoration models were trained using the Adam optimizer with a learning rate of 1×10^{-3} and the mean squared error (MSE) loss function. The batch size was set to 1, and each model was trained for 50 epochs. The restoration performance was evaluated using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

The CT validation therefore focuses on representation-level robustness, where the objective is to assess whether degraded image information can be recovered before subsequent inspection analysis.

To ensure the consistency and reliability of the reported results, repeated experiments were conducted under identical experimental configurations. For the YOLO-based 2D detection experiments, repeated runs produced identical evaluation results under the fixed training settings. For the CT representation recovery experiments, repeated runs showed minor numerical variations due to the stochastic nature of deep learning optimization; however, the relative performance trends among different restoration models remained consistent. These observations indicate that the reported robustness behaviours are reproducible and primarily associated with the investigated intervention strategies rather than incidental training variations.

B. Validation Scenario Overview

To evaluate the explanatory capability of the proposed framework, validation was conducted using representative industrial inspection scenarios derived from turbine blade inspection. Rather than evaluating the superiority of a specific algorithm, this section investigates whether the proposed framework can consistently explain robustness behaviour across different failure conditions.

Three validation scenarios were considered. Distribution failure was evaluated using two industrial defect datasets exhibiting substantially different levels of data scarcity and class imbalance. We evaluated decision failure through optimisation and architectural studies that analysed how learning objectives influenced detection robustness. Representation failure was evaluated using CT imaging scenarios affected by artifacts and structural degradation.

Collectively, these validation cases provide empirical evidence to examine whether dominant failure sources govern robustness behaviour and whether intervention effectiveness depends on alignment with those sources.

C. Distribution Failure Validation

To validate the distribution-failure component of the proposed framework, two industrial defect datasets with substantially different distributional characteristics were analysed. Dataset 1 (Alibaba Tianchi) represents a small-scale, severely imbalanced inspection scenario, containing 247 images and 388 defect instances, with an imbalance ratio of 7.33. In contrast, Dataset 2 (AeBAD-S) represents a larger-scale, relatively balanced inspection scenario, comprising 1,592 images and 2,557 defect instances.

TABLE III. REPRESENTATIVE AUGMENTATION RESULTS UNDER SEVERE DISTRIBUTION FAILURE.

Method	mAP@0.5	Dirty AP	Stain AP
Baseline	0.290	0.000	0.000
MixUp	0.591	0.667	0.672
CycleGAN	0.461	0.667	0.000
C5 Composite	0.599	0.731	0.452

According to the proposed framework, Dataset 1 is primarily characterized by distribution failures arising from data scarcity and class imbalance, whereas Dataset 2 exhibits substantially weaker distribution-failure characteristics. The contrasting properties of the two datasets provide an appropriate validation scenario for evaluating whether intervention effectiveness is determined by the intervention itself or by the dominant failure source.

1) Case A: Small-scale imbalanced dataset: Dataset 1 represents a typical industrial inspection scenario characterised by severe distribution failure. The dataset contains only 247 images and 388 defect instances, with an imbalance ratio of 7.33 between the majority and minority defect categories. Under these conditions, the baseline detector exhibited substantial degradation in robustness, particularly for minority defect classes, in which several categories could not be reliably detected.

As shown in Table III, substantial improvements were observed after applying distribution-oriented interventions. In particular, MixUp [27] and the C5 composite strategy significantly improved minority-class detection performance, transforming previously undetected defect categories into reliably detectable classes.

According to the proposed framework, these improvements are not solely attributable to the augmentation techniques themselves. Instead, the observed robustness gains stemmed from the selected interventions directly addressing the dominant failure sources, namely data scarcity and class imbalance. By increasing semantic diversity and improving minority-class representation [29], [30], the interventions effectively mitigated the distribution failure present in Dataset 1.

This case therefore provides empirical evidence supporting the distribution failure component of the proposed framework. The results suggest that intervention effectiveness depends on alignment with the underlying source of failure rather than on the intervention strategy alone.

2) Case B: Large-scale balanced dataset: In contrast to Dataset 1, Dataset 2 represents a larger-scale industrial inspection scenario with substantially weaker characteristics of distribution failure. The dataset contains 1,592 images and 2,557 defect instances, exhibiting a considerably more balanced class distribution. Consequently, data scarcity and class imbalance are no longer the dominant constraints affecting robustness.

As shown in Table IV, augmentation strategies exhibited markedly different behaviour compared with Dataset 1. While geometric augmentation produced measurable improvements,

TABLE IV. REPRESENTATIVE AUGMENTATION RESULTS UNDER WEAK DISTRIBUTION FAILURE.

Method	mAP_{50}	Ablation AP
Baseline	0.814	0.443
Geometric	0.862	0.684
Mosaic	0.827	0.549
MixUp	0.809	0.443
Diffusion	0.812	0.464

semantic-mixing approaches such as MixUp [27] provided limited benefit and did not improve minority-class performance. Similarly, the overall performance gains achieved through augmentation were substantially smaller than those observed under severe distribution failure conditions.

According to the proposed framework, this behaviour can be explained by the reduced influence of distribution failure in Dataset 2. Since the dataset already provides sufficient sample diversity and a relatively balanced class representation, interventions specifically designed to address data scarcity and imbalance are no longer the most effective robustness strategy, which is consistent with previous observations on dataset-dependent augmentation and domain generalisation [29], [5].

These observations further support the failure-oriented interpretation of robustness. The effectiveness of augmentation was not determined by the intervention itself but by the extent to which it aligned with the dominant failure source in the inspection scenario.

3) Framework interpretation: The contrasting behaviours observed across Dataset 1 and Dataset 2 provide empirical support for the central assumption of the proposed framework. Although similar augmentation strategies were applied in both scenarios, their effectiveness varied substantially depending on the datasets' underlying distributional characteristics.

In Dataset 1, robustness degradation was primarily associated with severe data scarcity and class imbalance. Under these conditions, interventions designed to increase data diversity and improve minority-class representation produced substantial improvements in detection performance, particularly for previously underrepresented defect categories. In contrast, Dataset 2 exhibited substantially weaker characteristics of distribution failure. Consequently, the same augmentation strategies produced reduced benefits and, in some cases, negligible performance improvements.

These observations suggest that augmentation effectiveness is not an intrinsic property of the augmentation strategy itself. Rather, intervention effectiveness depends on the degree of alignment between the intervention and the dominant source of failure. When severe distribution failure is present, distribution-oriented interventions can significantly improve robustness. However, when distribution failure is no longer the primary constraint, the same interventions may provide limited value.

From the perspective of the proposed framework, robustness improvement is therefore failure-dependent rather than method-dependent. The distribution failure validation demonstrates that selecting an effective intervention requires identifying the dominant failure source in advance rather than adopting a universally preferred technique.

TABLE V. REPRESENTATIVE RESULTS FOR OBJECTIVE ALIGNMENT

Configuration	mAP_{50}
BCE + CIoU	0.588
Focal + CIoU	0.676
Focal + CIoU + Weighting	0.710

TABLE VI. REPRESENTATIVE RESULTS OF ARCHITECTURAL REFINEMENT UNDER DECISION FAILURE.

Configuration	mAP_{50}
YOLOv8n	0.588
CBAM	0.303
AIFI	0.342
C2fPSA	0.226

D. Decision Failure Validation

1) Case A: Objective alignment: To validate the decision failure component of the proposed framework, optimisation objectives were examined under identical architectural conditions. In this validation scenario, the detector architecture remained unchanged while different objective formulations, including imbalance-aware loss functions, were applied [31].

As shown in Table V, substantial robustness improvements were achieved through objective redesign without modifying the detector architecture. The results indicate that decision quality was strongly influenced by the optimisation objective, particularly under imbalanced inspection conditions where minority defect categories were vulnerable to suppression. Bounding-box regression objectives were also investigated using IoU-based loss formulations, following previous improvements in localization optimisation [32].

According to the proposed framework, these improvements are associated with mitigating decision failure rather than with changes in model complexity. By improving the alignment between optimisation objectives and inspection objectives, the detector exhibited more stable learning behaviour and improved minority-class representation.

2) Case B: Architectural refinement: A second validation scenario examined whether robustness could be improved through architectural refinement alone. Several attention and feature-enhancement modules were incorporated into the baseline detector architecture.

As shown in Table VI, architectural refinement did not consistently improve robustness. Several architectural modifications resulted in performance degradation despite increasing model complexity. These observations suggest that the robustness limitations were more closely associated with optimisation and decision-making behaviour than with insufficient representational capacity alone.

3) Framework Interpretation: The results of the decision failure validation support the central assumption of the proposed framework. Under identical architectural conditions, objective alignment produced substantial robustness improvements, whereas increased architectural complexity failed to provide consistent benefits.

These observations indicate that decision failure originates primarily from optimisation behaviour rather than model ca-

TABLE VII. Representative Image Quality Results for Representation Failure Recovery.

Method	PSNR	SSIM
Gaussian Filter	21.82	0.7910
CNN	27.83	0.6552
U-Net	25.85	0.9201
GNN v2	28.93	0.9220

pacity. Consequently, interventions targeting objective design, class weighting, and optimisation strategies are often more effective than architectural refinement when decision failure is the dominant source of robustness degradation.

From the perspective of the proposed framework, the improvement in robustness in this scenario emerged from alignment between the optimisation and inspection objectives rather than from increasing architectural complexity. This finding provides empirical support for the framework's decision-failure component and further demonstrates that the dominant source of failure governs intervention effectiveness.

E. Representation Failure Validation

1) Case study CT representation recovery under imaging degradation: To validate the representation-failure component of the proposed framework, industrial computed tomography (CT) scenarios affected by imaging degradation were analysed. Unlike distribution failure and decision failure, which are directly reflected through detection performance, representation failure occurs before model inference and is therefore evaluated through the recovery of degraded image representations.

In industrial CT inspection, image quality may be degraded by ring artifacts, streak artifacts, beam-hardening effects, noise contamination, and reduced contrast [24], [25]. These degradations distort defect-related structural information and reduce the quality of image representations available for subsequent inspection tasks. Consequently, robustness degradation may occur before feature extraction and model inference due to insufficient or corrupted visual information.

According to the proposed framework, such degradation corresponds to representation failure. To address this source of failure, image enhancement and restoration techniques were applied to recover structural information and improve defect visibility [26].

The representative image quality results are summarised in Table VII. The enhancement process improved image quality and restored structural information under degraded CT conditions. These results demonstrate that representation-oriented interventions can alleviate information loss caused by imaging degradation and provide more reliable image representations for subsequent inspection analysis. However, downstream defect detection evaluation after CT enhancement was not conducted in this study, as the objective of this validation case is to assess representation recovery rather than end-to-end detection improvement.

F. Cross-Level Analysis

The three validation scenarios demonstrate that robustness degradation in industrial inspection systems does not originate from a single source. Instead, robustness challenges emerge at different stages of the inspection pipeline and are associated with distinct failure mechanisms.

Distribution failure primarily arises from deficiencies in the training data, including data scarcity, class imbalance, and distributional variability. Decision failure arises from optimisation misalignment during model learning, whereas representation failure is introduced prior to learning through imaging degradation and information loss.

Although these failures occur at different stages of the inspection pipeline, they share a common characteristic: each affects robustness through a distinct mechanism and therefore requires a different intervention strategy. Consequently, robustness cannot be improved through a universally superior algorithm; rather, improvement depends on addressing the dominant constraint present in the inspection scenario.

The three validation cases also reveal an important cross-level observation. Interventions that were highly effective under one failure condition did not necessarily produce comparable improvements under another. Distribution-oriented augmentation substantially improved robustness under severe data imbalance but provided limited benefit when distribution failure was weak. Similarly, optimisation-oriented interventions consistently outperformed architectural refinement only when decision failure dominated, whereas imaging enhancement became the primary contributor to robustness under degraded CT acquisition conditions.

These observations suggest that robustness degradation should be analysed by the stage at which failure is introduced rather than by methodological categories alone. Collectively, these cross-level observations establish a common analytical basis for interpreting robustness across heterogeneous industrial inspection scenarios and naturally motivate the unified framework synthesis presented in the following section.

G. Framework Synthesis

The three validation scenarios collectively provide converging empirical evidence supporting the proposed Failure-Oriented Robustness Framework. Although each validation examined different inspection tasks, datasets, and intervention strategies, all cases exhibited a consistent pattern of robustness that can be interpreted within the proposed failure-oriented perspective.

The relationships identified across the three validation scenarios are summarised in Table VIII.

This synthesis demonstrates that the proposed framework provides a unified explanation of robustness behaviour across heterogeneous industrial inspection systems. Rather than treating augmentation, optimisation, and image enhancement as isolated technical solutions, the framework interprets them as complementary interventions targeting different stages of robustness degradation.

A common robustness principle therefore emerges across all validation scenarios: intervention effectiveness is governed by alignment with the dominant failure source rather than by algorithmic superiority alone. Consequently, robustness should be understood as a failure-dependent property rather than an algorithm-dependent property.

TABLE VIII. Framework Synthesis Across Validation Scenarios

Failure	Validation Observation	Framework Insight
Distribution	Augmentation is effective only under severe imbalance	Distribution alignment
Decision	Objective alignment consistently outperforms architectural refinement	Objective alignment
Representation	Enhancement restores structural visibility under degraded CT	Representation recovery

Taken together, these findings demonstrate that the proposed Failure-Oriented Robustness Framework provides a unified conceptual foundation for robustness-oriented analysis, intervention selection, and future framework development in safety-critical industrial inspection systems.

V. Discussion and Future Perspectives

A. From Method-Centric Optimisation to Failure-Oriented Design

A central implication of the proposed framework is the shift from method-centric optimisation towards failure-oriented robustness design. Existing industrial inspection research has predominantly focused on identifying superior algorithms through architectural innovation, optimisation refinement, and data augmentation. Under this paradigm, robustness is often treated as a secondary outcome of improved performance.

However, the validation results presented in this study suggest that robustness behaviour cannot be fully explained by algorithmic superiority alone. Across all validation scenarios, intervention effectiveness depended strongly on the dominant source of failure. Methods that produced substantial improvements under one failure condition frequently demonstrated limited effectiveness under another.

These observations indicate that robustness should be analysed from the perspective of failure mechanisms rather than individual algorithms. Consequently, future industrial inspection systems may benefit from prioritising failure diagnosis before intervention selection, thereby transforming robustness from an implicit consequence of optimisation into an explicit design objective.

B. Implications for Industrial Inspection Robustness

The proposed framework has several implications for the design and evaluation of industrial inspection systems. Traditionally, robustness improvement has been pursued through the continuous development of new architectures, optimisation strategies, and data augmentation techniques. While such efforts have achieved notable performance gains, the results of this study suggest that robustness cannot be fully understood through algorithmic improvement alone.

Instead, robustness should be interpreted as a consequence of the interaction between failure sources and intervention strategies. Under this perspective, the primary objective of robustness-oriented inspection is not to identify universally superior methods, but to determine which intervention is most appropriate for the dominant failure condition. Consequently, robustness evaluation should incorporate failure diagnosis and intervention alignment in addition to conventional performance metrics.

For industrial practitioners, this perspective provides a practical decision-support mechanism. Rather than relying solely on trial-and-error model selection, inspection systems can be developed through the systematic identification of failure sources and the targeted design of interventions. Such an approach may improve both development efficiency and deployment reliability in safety-critical industrial environments.

C. Generalisability and Limitations

Although the proposed framework was validated using turbine blade inspection scenarios, the underlying failure categories are not specific to this application domain. Data scarcity, optimisation bias, and representation degradation are common challenges across a wide range of industrial inspection tasks, including weld inspection, casting inspection, surface defect detection, and industrial medical imaging. Consequently, the framework may provide a general robustness-oriented perspective for analysing inspection systems beyond turbine blade inspection.

Nevertheless, several limitations should be acknowledged. First, the framework was validated through representative case studies rather than a unified end-to-end implementation. Second, interactions between multiple failure sources were not explicitly investigated. In practical inspection systems, distribution, decision, and representation failures may coexist and influence each other simultaneously. Finally, the framework focuses primarily on robustness interpretation and intervention selection rather than on providing a universal optimisation strategy.

Furthermore, the current framework should be regarded as an analytical and interpretative framework rather than a fully predictive intervention recommendation system. Although the proposed failure-oriented perspective establishes relationships between failure sources and suitable intervention categories, it does not automatically identify failure dominance, rank competing interventions, or quantitatively predict expected robustness improvement before experimentation. Developing such predictive capabilities would require additional data-driven modelling and large-scale cross-domain validation, which represents an important direction for future research.

These limitations provide opportunities for future refinement and broader validation of the framework.

D. Future Research Directions

Several directions can be explored in future research for further developing the proposed framework. First, future studies should investigate cross-level interactions between distribution, decision, and representation failures. Understanding how multiple failure sources jointly influence robustness may lead to more comprehensive intervention strategies.

Second, future research may extend the framework towards predictive robustness management, where failure diagnosis and intervention selection are supported by quantitative models. Such extensions may enable automatic estimation of dominant failure sources, intervention ranking, and expected robustness improvements under different inspection conditions.

Finally, broader validation across different industrial domains is required. Applying the framework to additional inspection scenarios, including manufacturing quality control, infrastructure monitoring, and medical imaging, would help evaluate its generalisability and identify domain-specific robustness challenges.

Through these developments, the proposed framework may evolve from a robustness-interpretation model into a more comprehensive robustness-design methodology for industrial inspection systems.

When robustness degradation is observed, the first question should not be "Which algorithm should be used?" Instead, the dominant failure source should first be identified, followed by failure-aligned intervention selection.

VI. Conclusion

This study addressed a fundamental challenge in industrial inspection: robustness is widely recognised as a critical requirement, yet its behaviour remains difficult to explain using conventional method-centric perspectives. Existing studies have primarily focused on developing new architectures, optimisation strategies, and augmentation techniques, often treating robustness as a secondary consequence of performance improvement rather than an explicit research objective.

To address this limitation, a Failure-Oriented Robustness Framework was proposed. Instead of analysing robustness at the level of individual algorithms, the framework organises robustness challenges into three dominant failure sources: distribution failure, decision failure, and representation failure. By establishing explicit relationships between failure diagnosis, intervention selection, and robustness outcomes, the framework provides a systematic mechanism for understanding and improving robustness in industrial inspection systems.

The framework was validated through representative turbine blade inspection scenarios covering data-level, model-level, and imaging-level robustness challenges. Across all validation cases, a consistent pattern emerged. Distribution-oriented interventions were most effective when robustness degradation originated from data scarcity and class imbalance. Optimisation-oriented interventions produced the greatest benefits when decision failure dominated the learning process. Representation-oriented interventions were most effective when imaging degradation limited the visibility of defects. These observations demonstrate that intervention effectiveness varies according to the dominant source of failure.

The central finding of this work is that robustness is not an intrinsic property of a particular method, nor is it achieved simply through increasing algorithmic complexity. Rather, robustness emerges from alignment between intervention strategies and dominant failure sources. This perspective shifts robustness research from method-centric optimisation towards failure-oriented robustness design, providing a more interpretable and practically relevant foundation for industrial inspection systems.

Although the framework was validated using turbine blade inspection scenarios, the proposed perspective is not restricted to a single application domain. Future research may extend the framework to additional industrial and imaging environments, investigate interactions between multiple failure sources, and develop adaptive robustness management strategies. Through such developments, the proposed framework may contribute towards a more systematic and generalisable understanding of robustness in industrial inspection. The proposed framework establishes a conceptual foundation for future robustness-oriented industrial inspection systems, where intervention strategies are selected according to failure mechanisms rather than algorithmic popularity.

Acknowledgment

This work was supported by the Tan Sri Leo Moggie Chair of Energy Informatics Publication Fund, Universiti Tenaga Nasional (UNITEN), under Grant No. 20211EICHAIR (UKM Reference No. ZG-2024-005).

References

[1] A. Ettalibi, A. Elouadi, and A. Mansour, "AI and computer vision-based real-time quality control: A review of industrial applications," *Procedia Computer Science*, vol. 231, pp. 212–220, 2024, doi: 10.1016/j.procs.2023.12.195.

[2] R. Khanam, M. Hussain, R. Hill, and P. Allen, "A comprehensive review of convolutional neural networks for defect detection in industrial applications," *IEEE Access*, vol. 12, pp. 94250–94295, 2024, doi: 10.1109/ACCESS.2024.3425166.

[3] J. Li and G. Li, "Triangular trade-off between robustness, accuracy, and fairness in deep neural networks: A survey," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–40, 2025, doi: 10.1145/3645088.

[4] P. Singhal, R. Walambe, S. Ramanna, and K. Kotecha, "Domain adaptation: Challenges, methods, datasets, and applications," *IEEE Access*, vol. 11, pp. 6973–7020, 2023, doi: 10.1109/ACCESS.2023.3237025.

[5] J. Mai, C. Gao, and J. Bao, "Domain generalization through data augmentation: A survey of methods, applications, and challenges," *Mathematics*, vol. 13, no. 5, Art. no. 824, 2025, doi: 10.3390/math13050824.

[6] N. Suh and G. Cheng, "A survey on statistical theory of deep learning: Approximation, training dynamics, and generative models," *Annual Review of Statistics and Its Application*, vol. 12, no. 1, pp. 177–207, 2025, doi: 10.1146/annurev-statistics-112723-034650.

[7] J. Terven, D. M. Cordova-Esparza, J. A. Romero-González, A. Ramírez-Pedraza, and E. A. Chavez-Urbiola, "A comprehensive survey of loss functions and metrics in deep learning," *Artificial Intelligence Review*, vol. 58, no. 7, Art. no. 195, 2025, doi: 10.1007/s10462-025-11150-6.

[8] M. Gupta, M. A. Khan, R. Butola, and R. M. Singari, "Advances in applications of non-destructive testing (NDT): A review," *Advances in Materials and Processing Technologies*, vol. 8, no. 2, pp. 2286–2307, 2022.

[9] F. Abdollahi-Mamoudan, C. Ibarra-Castanedo, and X. P. Maldague, "Non-destructive testing and evaluation of hybrid and advanced structures: A comprehensive review of methods, applications, and emerging trends," *Sensors*, vol. 25, no. 12, Art. no. 3635, 2025.

[10] R. E. Donatus, "A comprehensive review of artificial intelligence for image- and signal-based nondestructive testing in aerospace structures," *Journal of Computers, Mechanical and Management*, vol. 4, no. 6, pp. 1–14, 2025.

[11] K. Masita, A. Hasan, T. Shongwe, and H. A. Hilal, "Deep learning in defects detection of PV modules: A review," *Solar Energy Advances*, vol. 5, Art. no. 100090, 2025.

[12] Z. Ren, F. Fang, N. Yan, and Y. Wu, "State of the art in defect detection based on machine vision," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 9, no. 2, pp. 661–691, 2022.

[13] S. Kim, T. P. Nguyen, and J. Yoon, “A systematic review of machine vision applications in factory and manufacturing processes: From quality control to predictive diagnostics,” *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 13, no. 2, pp. 745–775, 2026.

[14] Q. Wang, M. Wang, J. Sun, D. Chen, and P. Shi, “Review of surface-defect detection methods for industrial products based on machine vision,” *IEEE Access*, 2025.

[15] A. Jayawardena, J. H. Sul, D. Moratuwage, J. L. Wijayaraja, and L. Piyathilaka, “State of the art in AI-based visual inspection for industrial quality control: Methods, benchmarks, challenges, and autonomous systems,” *Electronics*, vol. 15, no. 12, Art. no. 2727, 2026.

[16] H. Yang, S. Kim, and Y. Choi, “Review of UAV–AI-based automated visual inspection technologies and their applicability to high-risk construction machinery,” *Journal of the Korean Society of Mineral and Energy Resources Engineers*, vol. 62, no. 5, pp. 567–583, 2025.

[17] Y. Li, B. Xie, S. Guo, Y. Yang, and B. Xiao, “A survey of robustness and safety of 2D and 3D deep learning models against adversarial attacks,” *ACM Computing Surveys*, vol. 56, no. 6, pp. 1–37, 2024.

[18] J. Wang, J. Ai, M. Lu, J. Zhao, H. Su, D. Yu, *et al.*, “A survey of neural network robustness assessment in image recognition,” *ACM Computing Surveys*, vol. 58, no. 12, pp. 1–40, 2026.

[19] B. Chander, C. John, L. Warrier, and K. Gopalakrishnan, “Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness,” *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–49, 2025.

[20] N. Kemmerzell, A. Schreiner, H. Khalid, M. Schalk, and L. Bordoli, “Towards a better understanding of evaluating trustworthiness in AI systems,” *ACM Computing Surveys*, vol. 57, no. 9, pp. 1–38, 2025.

[21] K. Nikiforidis, A. Kyrtsoglou, T. Vafeiadis, T. Kotsiopoulos, A. Nizamis, D. Ioannidis, *et al.*, “Enhancing transparency and trust in AI-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of Industry 4.0/5.0,” *ICT Express*, vol. 11, no. 1, pp. 135–148, 2025.

[22] G. Yu, G. Tan, H. Huang, Z. Zhang, P. Chen, R. Natella, *et al.*, “A survey on failure analysis and fault injection in AI systems,” *ACM Transactions on Software Engineering and Methodology*, vol. 35, no. 1, pp. 1–42, 2026.

[23] S. Bi, C. Wang, B. Wu, S. Hu, W. Huang, W. Ni, *et al.*, “A comprehensive survey on applications of AI technologies to failure analysis of industrial systems,” *Engineering Failure Analysis*, vol. 148, Art. no. 107172, 2023.

[24] D. C. Lepcha, B. Goyal, A. Dogra, K. P. Sharma, and D. N. Gupta, “A deep journey into image enhancement: A survey of current and emerging trends,” *Information Fusion*, vol. 93, pp. 36–76, 2023.

[25] Y. Lu, Y. Lin, H. Wu, Y. Luo, X. Zheng, H. Xiong, and L. Wang, “Priors in deep image restoration and enhancement: A survey,” *arXiv preprint arXiv:2206.02070*, 2022.

[26] H. Su, Y. Li, Y. Xu, X. Fu, and S. Liu, “A review of deep-learning-based super-resolution: From methods to applications,” *Pattern Recognition*, vol. 157, Art. no. 110935, 2025.

[27] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond Empirical Risk Minimization,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.

[28] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934, 2020.

[29] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[30] Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong, and S. X. Yu, “Deep Long-Tailed Learning: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 2, pp. 1079–1099, 2024.

[31] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 2999–3007.

[32] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” in *Proc. AAAI Conf. Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[33] Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, “Class-Balanced Loss Based on Effective Number of Samples,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9268–9277.

[34] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.

[35] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[36] C. Ledig *et al.*, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4681–4690.