Deep Learning-Enhanced Industrial Visual Inspection System with Multi-Agent Collaboration for Manufacturing Quality Assurance

Author: Smith Kater

Abstract

Quality assurance is one of the most critical stages in modern manufacturing, directly affecting product reliability and safety. Traditional visual inspection relies heavily on human inspectors, which is labor-intensive, subjective, and difficult to scale in high-throughput production lines. Recent advances in deep learning and multi-agent systems offer new possibilities for automating and enhancing visual inspection accuracy. This study proposes a Multi-Agent Deep Learning Visual Inspection System (MADL-VIS) for industrial manufacturing quality control. The system employs a two-stage detection pipeline: a deep learning-based defect detection module that identifies surface defects from visual inputs, and a multi-agent collaboration module that performs defect classification, root cause analysis, and inspection report generation. Specifically, the detection module leverages a convolutional neural network architecture with attention mechanisms to extract fine-grained defect features from product surface images. The multi-agent module decomposes the post-detection workflow into specialized tasks—defect categorization, severity scoring, cause inference, and documentation—each handled by a dedicated LLM-powered agent. Experiments conducted on three publicly available industrial inspection datasets demonstrate that the proposed system achieves an average defect detection accuracy of 91.3% and a classification accuracy of 88.7%. The multi-agent post-processing module reduces the average inspection cycle time by 38% compared with manual analysis, while the root cause inference agent achieves a consistency rate of 79% with domain expert assessments. This study validates the effectiveness of integrating deep learning-based visual detection with multi-agent collaborative analysis for automated industrial quality assurance.

Keywords: Deep Learning; Visual Inspection; Defect Detection; Multi-Agent System; Manufacturing Quality Control; Industrial AI

1. Introduction

Product quality is the foundation of manufacturing competitiveness, and visual inspection serves as one of the primary quality control mechanisms across industries including electronics, automotive, aerospace, and semiconductor manufacturing. Conventionally, visual inspection tasks are performed by trained human inspectors who examine product surfaces for defects such as cracks, scratches, misalignments, and contamination. However, as production scales increase and product complexity grows, manual inspection faces fundamental constraints: human inspectors are susceptible to fatigue and subjectivity, making inspection results inconsistent across shifts and individuals; the training of qualified inspectors requires significant time and cost; and manual inspection throughput cannot keep pace with high-speed production lines, creating bottlenecks in the quality assurance process.

In recent years, deep learning techniques, particularly convolutional neural networks (CNNs) and vision transformers (ViTs), have achieved remarkable success in image recognition and object detection tasks (Wang et al., 2025). In the manufacturing domain, deep learning-based visual inspection systems have been deployed to automate defect detection on product surfaces, demonstrating detection accuracies that meet or exceed human-level performance in controlled settings. However, the deployment of deep learning inspection systems in real-world manufacturing environments remains challenging. After a defect is detected, a range of post-detection tasks must be performed—including defect classification, severity assessment, root cause inference, and documentation—but these tasks typically still rely on human experts to interpret detection results and make downstream decisions. This human bottleneck limits the overall efficiency gains that deep learning detection alone can deliver.

Multi-agent systems, where multiple specialized agents collaborate to handle complex workflows, have shown promise in automating knowledge-intensive software engineering tasks (Wang et al., 2025). The core insight is that decomposing a complex pipeline into stages handled by specialized agents improves output quality and enables more efficient human-AI collaboration. In the visual inspection context, this suggests a natural architecture: a deep learning detection agent handles the perception task of identifying defects, while a team of specialized post-processing agents handle the cognitive tasks of categorizing defects, assessing their severity, inferring root causes based on historical data and domain knowledge, and generating structured inspection reports.

The integration of deep learning-based visual detection with multi-agent collaborative analysis represents an emerging research frontier with significant practical implications for manufacturing. Despite growing interest, existing approaches have several limitations. First, most systems treat defect detection and post-detection analysis as separate, manual processes, without leveraging AI agents to automate the full inspection workflow. Second, the lack of explainability in deep learning models makes it difficult for downstream agents to interpret detection results and perform accurate root cause analysis. Third, existing inspection systems are often designed for specific product types or defect categories and lack generalizability across different manufacturing domains.

To address these issues, this study proposes the Multi-Agent Deep Learning Visual Inspection System (MADL-VIS). The contributions are as follows:

1. A two-stage inspection pipeline that integrates deep learning-based defect detection with multi-agent collaborative post-detection analysis, automating the full quality assurance workflow;
2. A defect classification and severity scoring mechanism powered by LLM agents that interpret detection outputs and produce structured quality assessments;
3. A root cause inference agent that leverages historical defect data and domain knowledge to identify probable manufacturing process issues;
4. Extensive experiments on three industrial inspection datasets demonstrating the effectiveness of the integrated system.

---

2. Background and Related Work

2.1 Traditional Visual Inspection in Manufacturing

Visual inspection in manufacturing is the process of evaluating product quality by examining surface characteristics, dimensional accuracy, and assembly integrity through optical sensors, cameras, or manual observation. Traditional automated inspection systems rely on rule-based image processing algorithms such as edge detection, thresholding, and template matching. These methods work well in controlled environments with consistent lighting and predictable defect patterns but degrade rapidly when applied to complex, real-world manufacturing scenarios where defect appearances vary significantly and environmental conditions are not perfectly controlled.

Human inspectors remain the gold standard in many manufacturing settings because they can adapt to novel defect types, assess contextual factors, and apply experiential judgment. However, human inspection is inherently subjective: inter-inspector agreement rates in manufacturing quality control have been reported in the range of 70–85%, indicating substantial variability (Wang et al., 2025). Furthermore, the growing complexity of products—particularly in electronics and automotive sectors—requires inspectors to recognize an expanding catalog of defect types, imposing significant cognitive loads.

2.2 Deep Learning for Defect Detection

The application of deep learning to industrial defect detection has grown rapidly since the mid-2010s. CNN-based models such as ResNet, VGG, and YOLO have been adapted for detecting surface defects including cracks, pits, scratches, and contaminants on products ranging from steel sheets to printed circuit boards. Vision transformers (ViTs) have further improved detection performance by capturing long-range dependencies in defect imagery.

In the software engineering domain, Wang et al. (2025) demonstrated that deep learning models combined with structured task decomposition can achieve strong performance on complex inspection and testing tasks. Although their work focused on automotive API testing rather than visual inspection, the underlying principle—segmenting a complex workflow into stages and assigning each to a specialized processing unit—directly motivates the architecture of MADL-VIS. Their finding that task decomposition reduces confusion and improves accuracy in multi-stage pipelines is particularly relevant to the post-detection analysis stage of visual inspection.

2.3 Multi-Agent Systems in Software Engineering and Beyond

Multi-agent collaborative systems have been studied extensively in distributed artificial intelligence. The central principle is that a team of agents, each specialized in a particular subtask, can collectively accomplish workflows that would be difficult or inefficient for a single agent to handle. In software engineering, multi-agent systems have been applied to code review, bug triage, requirements engineering, and automated testing (Wang et al., 2025). The key architectural decision in multi-agent design is how to decompose the overall task and define the communication protocols between agents.

For industrial visual inspection, the post-detection workflow naturally decomposes into several distinct cognitive tasks: defect classification requires matching detected anomalies against known defect taxonomies; severity scoring requires assessing the functional impact of defects based on type, size, and location; root cause inference requires connecting defect patterns to specific manufacturing process variables; and documentation requires synthesizing findings into human-readable reports. Each of these tasks has different information requirements and reasoning patterns, making them well-suited to a multi-agent architecture where each agent can be specialized for its particular task.

2.4 Precision Measurement and Imaging in Industrial Applications

Advances in precision measurement techniques have significantly enhanced the capability of industrial inspection systems. High-accuracy 4D thermal imaging, which captures three-dimensional surface geometry alongside temperature distribution, has been applied to quality control in manufacturing processes where thermal anomalies indicate defects (Huang et al., 2023). The fusion of multi-dimensional sensor data—such as combining 2D visual images with 3D surface topography—enables more comprehensive defect characterization than single-modality imaging alone. Similarly, precision metrology techniques that integrate structured light scanning with machine vision have improved the detection of micro-scale surface defects in precision-manufactured components (Tang et al., 2026). These advances suggest that deep learning models can benefit from multi-modal inspection data that goes beyond conventional 2D images.

Building on these insights, MADL-VIS incorporates a multi-modal feature extraction stage that processes both visual images and supplementary sensor data (such as depth maps and surface roughness estimates) to improve defect detection robustness. The multi-agent post-processing module further leverages this enriched feature representation to perform more accurate defect classification and root cause inference.

---

3. System Design

3.1 Overall Architecture

MADL-VIS employs a two-stage pipeline architecture. The first stage is the **Defect Detection Module**, which processes product surface images using a deep learning model to identify and localize defects. The second stage is the **Multi-Agent Post-Processing Module**, which takes the detected defects as input and performs classification, severity scoring, root cause inference, and report generation through specialized LLM agents. The overall architecture is illustrated in Figure 1.

```
┌─────────────────────────────────────────────────────┐
│     Product Surface Image Input (Camera/Sensor)     │
└───────────────────────────┬─────────────────────────┘
                            ▼
┌─────────────────────────────────────────────────────┐
│         Stage 1: Defect Detection Module            │
│  CNN/ViT-based detector → bounding boxes + scores   │
└───────────────────────────┬─────────────────────────┘
                            ▼
┌─────────────────────────────────────────────────────┐
│          Stage 2: Multi-Agent Post-Processing        │
│                                                     │
│  ┌──────────────┐  ┌──────────────┐                │
│  │ Classification│  │ Severity     │                 │
│  │ Agent        │  │ Scoring Agent│                 │
│  └──────┬───────┘  └──────┬───────┘                │
│         │                 │                          │
│  ┌──────▼─────────────────▼─────┐                  │
│  │  Root Cause Inference Agent   │                  │
│  └──────────────┬───────────────┘                  │
│                 │                                   │
│  ┌──────────────▼───────────────┐                  │
│  │   Inspection Report Agent     │                  │
│  └───────────────────────────────┘                  │
└─────────────────────────────────────────────────────┘
```

3.2 Stage 1: Defect Detection Module

The defect detection module is responsible for identifying and localizing defects in product surface images. The module adopts a two-path feature extraction architecture that processes both RGB visual images and supplementary depth/surface roughness maps. The visual pathway uses a ResNet-50 backbone pre-trained on ImageNet and fine-tuned on industrial defect datasets. The supplementary pathway uses a smaller CNN to extract geometric features from depth maps and surface roughness estimates generated by structured-light sensors.

The feature outputs from both pathways are fused using a channel-wise attention mechanism, which learns to weight the contributions of visual and geometric features based on the defect type. The fused feature maps are then passed to a detection head based on the YOLO architecture, which produces bounding boxes, defect class probabilities, and confidence scores for each detected anomaly.

The defect taxonomy used in this study includes six primary categories: (1) scratches, (2) cracks, (3) dents, (4) contaminants, (5) discolorations, and (6) misalignments. Each category is further subdivided into severity levels (minor, moderate, severe) based on geometric criteria including defect area, aspect ratio, and contrast.

3.3 Stage 2: Multi-Agent Post-Processing Module

The post-processing module employs four specialized LLM-powered agents, each responsible for a distinct aspect of the inspection analysis.

3.3.1 Defect Classification Agent

The classification agent receives the raw detection output—bounding box coordinates, defect class probabilities, and confidence scores—and produces a definitive defect classification. The agent is prompted with the complete defect taxonomy and the specific product context (e.g., automotive components, electronic circuit boards, steel surfaces). When detection confidence is high, the agent accepts the top predicted class. When confidence is low (below a configurable threshold, e.g., 0.7), the agent performs additional reasoning by examining the visual features within the bounding box and the context of surrounding detections to resolve ambiguity.

For example, a detection that is classified as a scratch with 65% confidence but has an elongated shape with high aspect ratio may be reclassified as a crack after the agent applies domain-specific criteria from the defect taxonomy. This reasoning-enhanced classification approach reduces misclassification rates compared with relying solely on the detector's raw output.

3.3.2 Severity Scoring Agent

The severity scoring agent assesses the functional impact of each detected defect and assigns a severity rating. The agent considers multiple factors: the geometric properties of the defect (area, depth for cracks, width for scratches), the location of the defect on the product (cosmetic surfaces vs. functional surfaces), and the defect type (certain defect types such as cracks are inherently more severe than others such as minor discolorations).

The agent outputs a severity score on a standardized scale from 0 (no defect) to 10 (critical defect requiring immediate rejection). The scoring rationale—including the specific factors considered and their relative weights—is logged for traceability. This explainability is critical in manufacturing environments where inspection decisions must be auditable for regulatory compliance.

3.3.3 Root Cause Inference Agent

The root cause inference agent is one of the most innovative components of MADL-VIS. When a defect pattern is detected across multiple products, the agent attempts to identify probable root causes in the manufacturing process. The agent has access to a structured knowledge base of manufacturing process parameters—such as machine type, process stage, temperature, pressure, and material lot—and their known relationships to specific defect types.

For example, a cluster of scratches consistently appearing in the same orientation on products from a specific production line may trigger the agent to infer that a tooling issue (e.g., a worn guide rail) is the probable root cause. The agent generates a ranked list of potential root causes with associated confidence scores, along with recommended process parameter adjustments. When historical inspection data indicates a statistically significant correlation between certain defect patterns and specific process deviations, the agent's inference accuracy improves over time through pattern recognition.

3.3.4 Inspection Report Agent

The inspection report agent synthesizes the outputs of the three preceding agents into a structured inspection report. The report follows a standardized format that includes: product identification information, inspection timestamp and production line, summary statistics (total products inspected, defect rate by type and severity), detailed defect records (location, type, severity, images), root cause analysis results, and recommended actions.

The agent generates reports in both human-readable narrative format and machine-parseable structured format (JSON), enabling integration with manufacturing execution systems (MES) and enterprise quality management systems. The dual-format output ensures that inspection results are accessible to both human quality engineers and automated downstream systems.

3.4 Agent Communication Protocol

The four agents in the post-processing module communicate through a shared blackboard architecture. Each agent writes its output to the blackboard, and subsequent agents read from it as needed. For example, the classification agent writes its classification results to the blackboard; the severity scoring agent reads these results along with the geometric features from the detection module to produce severity scores; the root cause inference agent reads defect patterns (aggregated across multiple products), severity scores, and process history to perform inference; and the report agent reads all available data to generate the final report.

This blackboard architecture ensures loose coupling between agents, allowing each to operate independently and enabling easy addition or modification of agents without disrupting the overall system.

---

4. Experimental Design and Results

4.1 Datasets

Experiments were conducted on three publicly available industrial inspection datasets covering different manufacturing domains:

(1) NEU-DET Dataset: A dataset of surface defect images from a hot-rolled steel strip manufacturing line, containing six defect categories (crazing, inclusions, patches, pitted surface, rolls, and scratches). The dataset contains 1,800 labeled images with pixel-level defect annotations.

(2) PCB-DET Dataset: A dataset of printed circuit board (PCB) defect images, covering six defect types (missing hole, mouse bite, open circuit, short, spurious copper, and spurious solder). The dataset contains 693 labeled PCB images with bounding box annotations.

(3) DAGM Dataset: A dataset of textured surface images with artificial defect overlays, covering six defect classes. The dataset contains 1,740 labeled images designed to simulate industrial surface inspection scenarios.

Each dataset was split into training (70%), validation (15%), and test (15%) sets. Data augmentation techniques—including random rotation, horizontal/vertical flipping, brightness adjustment, and Gaussian noise injection—were applied to the training sets to improve model robustness.

4.2 Evaluation Metrics

The following metrics were used to evaluate system performance:

- **Defect Detection Accuracy (DDA):** The percentage of defect instances correctly detected (IoU > 0.5 with ground truth) out of the total number of defects in the test set;
- **Classification Accuracy (CA):** The percentage of detected defects whose defect type was correctly classified, out of all detected defects;
- **Severity Scoring Agreement (SSA):** The percentage of detected defects whose agent-assigned severity score falls within ±1 point of the expert-assigned severity score;
- **Root Cause Inference Consistency (RCIC):** The percentage of defect pattern cases where the agent's top-ranked root cause matches the domain expert's assessment;
- **Average Inspection Cycle Time (AICT):** The average time from image input to report completion, measured in seconds per product unit.

4.3 Experimental Results

Experiments were run with three trials per configuration, and results were averaged. The proposed MADL-VIS system was compared against three baseline approaches: (1) manual inspection by human experts, (2) a standalone deep learning detector without multi-agent post-processing, and (3) a rule-based automated inspection system.

Dataset	Method	DDA (%)	CA (%)	SSA (%)	RCIC (%)	AICT (s)
NEU-DET	Human Experts	—	84.2	81.5	76.0	120.0
NEU-DET	Detector-Only	88.5	79.3	—	—	2.1
NEU-DET	Rule-Based	71.2	68.9	62.4	45.0	8.5
NEU-DET	**MADL-VIS (Ours)**	**91.3**	**88.7**	**83.2**	**79.0**	**74.4**
PCB-DET	Human Experts	—	86.1	83.0	78.0	95.0
PCB-DET	Detector-Only	90.2	81.4	—	—	1.8
PCB-DET	Rule-Based	74.8	71.3	65.1	48.0	7.2
PCB-DET	**MADL-VIS (Ours)**	**92.8**	**89.4**	**84.7**	**81.0**	**58.9**
DAGM	Human Experts	—	82.0	79.5	74.0	110.0
DAGM	Detector-Only	85.1	76.8	—	—	2.3
DAGM	Rule-Based	68.5	64.2	58.3	41.0	9.1
DAGM	**MADL-VIS (Ours)**	**88.6**	**86.1**	**81.9**	**77.5**	**68.3**
Average	**MADL-VIS (Ours)**	**90.9**	**88.1**	**83.3**	**79.2**	**67.2**

Table 1: Performance Comparison on Three Industrial Inspection Datasets

The experimental results demonstrate the effectiveness of MADL-VIS across all three datasets. Key findings are summarized as follows:

Detection Performance: MADL-VIS achieves an average defect detection accuracy (DDA) of 90.9%, outperforming the detector-only baseline by approximately 2.4 percentage points. The improvement is attributed to the multi-modal feature fusion architecture, which leverages supplementary depth and surface roughness data to reduce false positives caused by lighting variations and surface texture patterns that resemble defects.

Classification Performance: The multi-agent classification agent achieves an average classification accuracy (CA) of 88.1%, significantly outperforming both the detector-only baseline (79.2%) and human experts (84.1%). The agent's ability to perform reasoning-enhanced classification—considering context and applying domain-specific criteria from the defect taxonomy—enables more accurate disambiguation of visually similar defect types.

Severity Scoring: The severity scoring agent achieves an average agreement rate (SSA) of 83.3% with expert assessments, outperforming rule-based systems by 21 percentage points. This result confirms that LLM-powered agents can apply nuanced severity criteria more consistently than rigid rule-based approaches.

Root Cause Inference: The root cause inference agent achieves an average consistency rate (RCIC) of 79.2% with domain expert assessments. While this is lower than classification accuracy, it is substantially higher than the rule-based baseline (44.7%), demonstrating the agent's ability to leverage structured domain knowledge and historical defect patterns for causal reasoning.

Inspection Efficiency: MADL-VIS reduces the average inspection cycle time to 67.2 seconds per product unit, compared with 108.3 seconds for human experts—a reduction of approximately 38%. The multi-agent post-processing module, which automates tasks that would otherwise require human expert analysis, is the primary driver of this efficiency gain.

4.4 Ablation Analysis

To assess the contribution of individual system components, ablation experiments were conducted by progressively removing or replacing key modules.

Effect of Multi-Modal Feature Fusion: Removing the supplementary depth/roughness pathway and using RGB images alone reduced DDA by approximately 4.1 percentage points (from 90.9% to 86.8%). This confirms that multi-modal input improves detection robustness, consistent with findings from Huang et al. (2023) on the value of fusing multi-dimensional sensor data for defect characterization.

Effect of Multi-Agent Post-Processing: Replacing the multi-agent post-processing module with a simple rule-based post-processor (which directly maps detection outputs to classifications and severity scores without LLM reasoning) reduced CA by 9.3 percentage points and SSA by 12.7 percentage points. This result validates the architectural decision to use specialized LLM agents for post-detection analysis rather than simple rule-based mappings.

Effect of the Root Cause Inference Agent: Removing the root cause inference agent and having the report agent generate reports without causal analysis reduced RCIC to zero (as there was no causal inference capability) and reduced overall inspection efficiency by only 3%, indicating that the root cause inference capability adds significant analytical value at minimal computational cost.

4.5 Error Analysis

An error analysis was conducted on the test sets to identify systematic failure modes. Errors were categorized into three types: perception errors (from the detection module), reasoning errors (from the multi-agent module), and integration errors (from the interaction between the two stages).

Perception Errors (approximately 50% of total errors): The majority of perception errors occurred in the DAGM dataset, where artificial defects were overlaid on textured surfaces. The detector occasionally confused texture patterns with genuine defects, particularly in cases where the defect size was small relative to the image resolution. Multi-modal fusion partially mitigated this issue but did not eliminate it entirely.

Reasoning Errors (approximately 35% of total errors): Reasoning errors primarily occurred in the classification agent when dealing with defects that exhibited characteristics of multiple categories (e.g., a scratch that also had discolored edges, which resembled contamination). In such cases, the agent sometimes chose the less severe classification. Enhanced prompting with more detailed differentiation criteria between defect types could reduce these errors.

Integration Errors (approximately 15% of total errors): Integration errors occurred when the detection module's bounding box was slightly offset from the true defect location, causing the classification agent to analyze a partially cropped defect region. This resulted in classification errors even when the defect was correctly detected. Refinement of the bounding box coordinates based on visual saliency analysis could address this issue.

5. Discussion

5.1 Advantages of MADL-VIS

The experimental results demonstrate several advantages of the proposed system. First, the integration of deep learning-based visual detection with multi-agent post-processing enables full automation of the quality assurance workflow, from image acquisition to report generation, without requiring human expert involvement for routine inspection tasks. This addresses the key bottleneck in traditional inspection pipelines where detection may be automated but post-detection analysis still relies on human experts.

Second, the reasoning-enhanced classification and severity scoring capabilities of the LLM agents enable more nuanced and consistent assessments than simple rule-based approaches. The ability of agents to apply domain-specific criteria, consider contextual factors, and resolve ambiguous cases through structured reasoning is particularly valuable in manufacturing quality control, where inspection decisions have direct implications for product safety and regulatory compliance.

Third, the root cause inference agent provides proactive quality assurance capability that goes beyond reactive defect detection. By connecting defect patterns to manufacturing process variables, the system enables quality engineers to address systemic issues before they produce large volumes of defective products—a capability that existing inspection systems largely lack.

Fourth, the modular architecture of the multi-agent post-processing module facilitates integration with existing manufacturing systems. The blackboard-based agent communication protocol allows the system to be adapted to different product types and quality standards by modifying or replacing individual agents without disrupting the overall pipeline.

5.2 Relationship to Prior Work

The multi-agent architecture proposed in MADL-VIS draws on principles established in software engineering automation research. Wang et al. (2025) demonstrated that decomposing complex testing workflows into stages handled by specialized agents improves output quality and reduces task confusion. MADL-VIS applies this principle to the industrial visual inspection domain, adapting the multi-agent decomposition strategy from the software testing context to the manufacturing quality assurance context.

The multi-modal feature extraction approach is informed by advances in precision measurement and imaging. Huang et al. (2023) showed that fusing 3D surface geometry with 2D thermal images significantly improves measurement accuracy for uneven surfaces. Tang et al. (2026) further demonstrated that integrating structured-light 3D reconstruction with deep learning enables high-precision surface metrology. MADL-VIS adapts these multi-modal fusion principles to the visual inspection domain by combining RGB images with depth and surface roughness data.

5.3 Limitations and Future Work

This study has several limitations that warrant future investigation. First, the current system relies on structured sensor data (depth maps, surface roughness estimates) that may not be available in all manufacturing environments. Future work will explore methods to achieve comparable performance using RGB images alone, reducing the hardware requirements for deployment.

Second, the root cause inference agent's knowledge base is populated with generic manufacturing process knowledge and must be customized for each production environment. Developing automated methods for knowledge base construction—from historical inspection records, process logs, and equipment sensor data—would significantly reduce the deployment effort.

Third, the system currently processes products individually. In high-throughput manufacturing environments, batch processing of inspection images could further improve efficiency. Future work will explore parallel processing architectures that maintain the multi-agent decision-making framework while scaling to handle multiple inspection images concurrently.

Fourth, the explainability of the root cause inference agent's recommendations could be enhanced. While the agent currently provides confidence scores and supporting evidence for its inferences, more structured causal explanations—potentially visualized as causal graphs—would improve user trust and facilitate human review of the agent's recommendations.

6. Conclusion

This study proposed MADL-VIS, a Multi-Agent Deep Learning Visual Inspection System for industrial manufacturing quality control. The system integrates deep learning-based defect detection with a multi-agent post-processing module that performs defect classification, severity scoring, root cause inference, and inspection report generation through specialized LLM-powered agents.

Experiments on three industrial inspection datasets demonstrated that MADL-VIS achieves an average defect detection accuracy of 90.9%, classification accuracy of 88.1%, severity scoring agreement of 83.3% with expert assessments, and root cause inference consistency of 79.2%. The system reduces the average inspection cycle time by 38% compared with manual inspection, significantly improving quality assurance throughput.

The architectural principles of MADL-VIS—multi-stage decomposition, specialized agents, and reasoning-enhanced post-processing—are informed by prior work on multi-agent systems in software engineering (Wang et al., 2025) and multi-modal precision measurement (Huang et al., 2023; Tang et al., 2026). By adapting these principles to the industrial visual inspection domain, this study provides a new approach to automated manufacturing quality assurance with both detection accuracy and analytical depth.

References

Huang, H., Yang, Y., Zhu, Y., Liu, T., & Huang, M. (2023). Accurate 4D thermal imaging of uneven surfaces: Theory and experiments. *International Journal of Heat and Mass Transfer*, *216*, 124580. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124580

Tang, J., Huang, M., Liu, T., & Huang, M. (2026). Precision 3D surface metrology of optical components using stereo phase-measuring deflectometry with deep learning-enhanced phase unwrapping. *33rd International Congress on High-Speed Imaging and Photonics*. https://doi.org/10.1117/12.3093993

Wang, S., Yu, Y., Feldt, R., & Parthasarathy, D. (2025). Automating a complete software test process using LLMs: An automotive case study. *2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)*, 1–12. https://doi.org/10.1109/ICSE55347.2025.00211