Technische Hochschule Deggendorf

Case study project - Integration and comparison of vision models for smart inspection cell

Bibliography Report

Submitted by:

Sai Prasanth Parnambedu
Yaswitha Pallela
Rakshith Thatikonda
Prarthana Shenoy
Chandrika Tirukkovalluri

Course:

M.Engg. in Applied AI for Digital Production Management

Under the Supervision of:

Prof. Dr.-Ing. Hamidreza Heidari

Date of Submission: 09.11.2025

Contents

List of Tables

Abstract

Modern industrial manufacturing is rapidly evolving under the paradigm of **Industry 4.0**, where **automation, artificial intelligence (AI), and machine vision** are driving the transition toward zero-defect production. Traditional inspection of components such as gears, shafts, and cast parts still relies heavily on human judgment, leading to inconsistent accuracy and limited scalability. To achieve higher efficiency and reliability, industries are increasingly adopting **AI-driven visual inspection cells** capable of detecting and classifying surface defects in real time with minimal human intervention.

This report presents a **bibliographic study** on recent advancements in **deep learning-based surface defect detection** and their integration within robotic automation frameworks. It reviews state-of-the-art **object detection** and **image classification** models—specifically **YOLOv8** and **MobileNetV2**—recognized for their balance between accuracy, speed, and computational efficiency. The analysis evaluates each model's performance in terms of detection precision, inference latency, and adaptability for **edge deployment** in industrial environments.

By synthesizing insights from recent research, the report establishes a foundation for developing a **hybrid AI-robotic inspection framework** that bridges the gap between vision algorithms and real-world manufacturing applications, contributing toward fully **autonomous quality control systems** in next-generation smart factories.

1. Introduction

The industrial shift towards Industry 4.0 emphasizes smart and autonomous systems for quality assurance and defect detection. Traditional manual inspection methods are often slow, inconsistent, and prone to human error. This chapter introduces the motivation behind integrating deep learning models with robotic simulation for real-time inspection, highlighting the need for automated defect detection and sorting in manufacturing environments.

1.1 Objectives

- To review existing literature on visual defect detection using deep learning models.
- To identify research gaps in current defect detection and automation frameworks.
- To justify the use of YOLOv8 and MobileNetV2 models for smart inspection applications.

1.2 Scope of Study

This review covers publications and frameworks from 2019–2025 focusing on surface defect detection using deep learning models, with an emphasis on their adaptability to industrial automation and edge deployment.

2. Bibliography Summary and Literature Alignment

2.1 Overview

This section summarizes key research studies that influenced the selection of **YOLOv8** and **MobileNetV2** for the Smart Inspection Cell project. The reviewed works highlight major developments in deep learning–based defect detection, focusing on lightweight architectures, attention mechanisms, feature fusion, and real-time adaptability in industrial inspection systems.

2.2 Summary of Reviewed Research Works

2.2.1 STMS-YOLOv5: A Lightweight Algorithm for Gear Surface Defect Detection

. **Authors:** Rui Yan et al. (2023), *Sensors (MDPI)*[1]
This paper presents **STMS-YOLOv5**, an enhanced YOLOv5-based model designed for efficient gear surface defect detection under real industrial conditions. The authors replaced YOLOv5's standard backbone with *ShuffleNetv2*, a lightweight feature extractor, and incorporated the **MECA attention mechanism** to strengthen critical feature localization while suppressing redundant information. The resulting model achieved an impressive 98.6% mAP@0.5 and processed images at 130 FPS, outperforming baseline YOLOv5 by 10.8 FPS. Despite these improvements, the model retained its anchor-based structure, limiting scalability for varying defect sizes and shapes.
Key Insight: Although efficient, it remains anchor-based, limiting scalability and flexibility.
Relevance: YOLOv8 overcomes this through its anchor-free, decoupled detection head.

2.2.2 Improved YOLOv5 for Metal Shaft Defect Detection Using Transfer Learning

Authors: Bi Li and Quanjie Gao (2023), *Sensors (MDPI)*[2]
This study focuses on enhancing YOLOv5's capability for metal shaft surface defect detection by integrating **CBAM (Convolutional Block Attention Module)** and **BiFPN (Bidirectional Feature Pyramid Network)** for efficient feature aggregation. Transfer learning was applied using a pre-trained YOLOv5 backbone to improve model convergence with limited training data. The enhanced model demonstrated a 93.6% mAP

and stronger small-defect recognition compared to standard YOLOv5, but suffered from a 15% drop in frame rate due to additional feature extraction overhead.

Key Insight: Higher accuracy but increased complexity.

Relevance: YOLOv8 offers real-time adaptability without the computational penalty of CBAM–BiFPN designs.

2.2.3 Surface Defect Detection of Industrial Products — Review and Future Directions

Authors: Y. Ma et al. (2024), *Artificial Intelligence Review (Springer)*[3]

This extensive literature review discusses advances in both one-stage and two-stage detection frameworks (YOLO, SSD, Faster R-CNN, etc.) for industrial defect inspection. The paper identifies three major bottlenecks: (1) the scarcity of labeled industrial datasets, (2) limited adaptability of anchor-based detectors to new defect geometries, and (3) challenges in maintaining real-time inference on embedded devices. The review emphasizes the urgent need for hybrid, modular, and lightweight solutions that balance detection accuracy and computational efficiency.

Key Insight: Future models must achieve a balance between speed and accuracy for Industry 4.0 deployment.

Relevance: YOLOv8 and MobileNetV2 directly target this balance with lightweight, high-speed architectures.

2.2.4 Deep Learning-Based Steel Surface Defect Detection: Review

Authors: Feiyu Chen et al. (2025), *Academic Journal of Science and Technology*[4]

This study analyzes steel surface defect detection methods using CNN and YOLO families. It compares performance across datasets such as NEU-DET and GC10-DET, focusing on accuracy, recall, and robustness under varying lighting conditions. The authors conclude that anchor-based YOLO variants perform well for moderate defects but struggle with extremely small or overlapping anomalies. The paper also highlights a growing trend toward lightweight CNNs and transfer learning for better adaptability.

Key Insight: Lightweight computation and real-time inference are crucial.

Relevance: YOLOv8 provides faster multi-scale detection, while MobileNetV2 ensures embedded deployment efficiency.

2.2.5 S-YOLO: Background-Weakening YOLOv3 for Gear Surface Defects

Authors: Liya Yu et al. (2019), *Journal of Sensors*[5]
This early study introduces **S-YOLO**, an improved YOLOv3 variant that employs a background-weakening algorithm to minimize false detections caused by complex industrial backgrounds. The model achieved higher precision than standard YOLOv3 but inherited the limitations of anchor-based predictions and a coupled detection head, resulting in reduced flexibility for multi-class detection.
Key Insight: Effective, but limited by anchor dependency and coupled detection heads.
Relevance: YOLOv8 eliminates these constraints through anchor-free architecture and decoupled prediction.

2.2.6 MobileNetV2 + Enhanced Feature Pyramid for Fast Object Detection

Author: Hoanh Nguyen (2020), *Journal of Theoretical and Applied Information Technology*[6]
The author explores **MobileNetV2** as a lightweight backbone integrated with a feature pyramid network (FPN) to achieve high-speed object detection on embedded devices. The architecture leverages depthwise separable convolutions and inverted residuals to minimize computation cost. Testing on the Pascal VOC dataset showed competitive accuracy while maintaining high FPS performance.
Key Insight: Confirms MobileNetV2's efficiency and transferability to real-time environments.
Relevance: Supports MobileNetV2's selection for fast, embedded industrial defect classification.

2.2.7 MOLO: Hybrid Model Using MobileNetV2 and YOLOv8 for Edge Devices

Authors: Khushboo Khurana et al. (2025), *Discover Artificial Intelligence (Springer)*[7]
This paper presents a hybrid model named **MOLO**, which fuses MobileNetV2 and YOLOv8 for efficient edge deployment. By applying TensorRT quantization and SIoU loss optimization, the system achieved up to 172 FPS on NVIDIA Jetson Xavier NX with minimal accuracy degradation. The hybrid model reduces overall parameters by 45% compared to standard YOLOv8 while retaining 98% mAP performance.
Key Insight: Demonstrates optimal balance between model size, inference speed, and

detection accuracy.
Relevance: Directly supports the hybrid approach used in this project.

2.2.8 Improved YOLOv5 for Industrial Interference-Resistant Gear Detection

Authors: S. Zhang et al. (2023), *ScienceDirect*[8]
This paper proposes an improved YOLOv5 variant incorporating **CBAMC3 modules** and **BiFPN_concat** to enhance resistance against environmental noise and illumination interference. The model achieved 96.3% detection accuracy but exhibited increased training complexity and slower inference due to multiple feature fusion operations.
Key Insight: Improved robustness but still dependent on anchors and heavy computation.
Relevance: YOLOv8's anchor-free design offers faster, interference-resistant detection.

2.3 Summary of Reviewed Works

Model	**Averaged Accuracy from referred papers (%)**	**FPS**	**Remarks**
YOLOv5[1, 2]	94.8	60	High accuracy, slower inference speed compared to newer versions.
YOLOv8[7, 9]	96.1	85	Anchor-free architecture with decoupled detection head offering faster convergence and real-time adaptability.
MobileNetV2 [10, 11]	90.25	120	Lightweight, low computational cost; ideal for embedded and edge deployments.
Faster R-CNN [12, 13]	92.7	25	Excellent localization precision but computationally heavy and unsuitable for real-time inspection.
EfficientNet-B0 [14, 15]	91.8	70	Scalable CNN backbone providing good balance between accuracy and efficiency.
ResNet50 [16, 17]	89.6	55	Deep residual architecture with strong feature extraction; higher latency for small defect detection.
SSD [18, 19]	88.4	75	Moderate accuracy and speed; efficient for general-purpose detection but less robust for fine defects.

Table 2.1: Reviewed deep learning models commonly used for industrial surface defect detection.

2.4 Synthesis and Observations

The reviewed literature indicates a strong evolution from **YOLOv3** to **YOLOv8**, with increasing focus on balancing accuracy, speed, and hardware efficiency. Key patterns observed include:

- Attention modules (e.g., CBAM, MECA) enhance features but add computational overhead.
- Feature pyramids improve small-defect recognition but increase inference latency.
- Transfer learning supports generalization but slows training convergence.
- Hybrid models (YOLOv8 + MobileNetV2) achieve the best trade-off between accuracy, inference speed, and scalability.

2.5 Alignment with the Proposed Work

Criteria	Research Insight	Proposed Model Justification
Speed vs. Accuracy	YOLOv5 achieves precision but lower speed	YOLOv8 enables real-time, anchor-free detection
Computation Efficiency	Heavy backbones limit deployment	MobileNetV2 ensures low computational load
Industrial Scalability	Anchor-based designs hinder generalization	YOLOv8 scales across varied defect types
Edge Deployability	Limited research for embedded hardware	Both models are optimized for edge inference
Real-time Operation	Attention-heavy models reduce FPS	Proposed setup maintains high FPS and accuracy

Table 2.2: Alignment of reviewed research insights with the proposed model selection.

2.5.1 Identified Research Gaps

- Lack of models optimized for real-time industrial inspection.
- Minimal integration between defect detection and robotic automation.

3. Model Selection Justification

Based on the literature analysis, YOLOv8 and MobileNetV2 were selected as core models for this project. YOLOv8 provides superior real-time detection performance, while MobileNetV2 offers efficiency suitable for embedded hardware. Their combination ensures a balanced trade-off between speed and precision, critical for automated inspection cells.

3.1 Reasons for Selection

- **YOLOv8:** Anchor-free detection, decoupled head, enhanced small-defect recognition, and high inference speed.
- **MobileNetV2:** Lightweight structure, depthwise separable convolutions, optimized for edge hardware like Jetson Nano.
- **Combined Advantage:** Enables hybrid deployment — detection and classification under one smart inspection pipeline.

3.2 Conclusion

The literature review underscores remarkable progress in AI-based surface defect detection, yet a clear gap remains in effectively integrating real-time vision algorithms with robotic automation for industrial deployment. Conventional models such as Faster R-CNN, SSD, and YOLOv5 demonstrate strong detection accuracy but are limited by anchor-based architectures and heavy computational loads, which restrict their responsiveness and scalability in production environments.

This study identifies **YOLOv8** and **MobileNetV2** as complementary models to overcome these constraints. YOLOv8's anchor-free, decoupled detection head delivers faster inference, improved convergence, and superior recognition of small defects, while MobileNetV2's depthwise separable convolutions offer lightweight, energy-efficient computation suitable for edge deployment. Combined, these frameworks establish a robust foundation for a **hybrid AI–robotic inspection system** that advances **Industry 4.0** objectives through real-time, scalable, and autonomous quality control in next-generation smart manufacturing environments.

Bibliography

[1] R. Yan, H. Liu, Q. Zhao, and Y. Chen, "Stms-yolov5: A lightweight algorithm for gear surface defect detection," *Sensors*, vol. 23, no. 12, p. 5581, 2023.

[2] B. Li and Q. Gao, "Improved yolov5 for metal shaft defect detection using transfer learning," *Sensors*, vol. 23, no. 3, p. 1450, 2023.

[3] Y. Ma, L. Wu, J. Chen, and Y. Zhang, "Surface defect detection of industrial products: Review and future directions," *Artificial Intelligence Review*, 2024.

[4] F. Chen, M. Zhao, and Y. Xu, "Deep learning-based steel surface defect detection: Review," *Academic Journal of Science and Technology*, vol. 14, no. 2, pp. 65–82, 2025.

[5] L. Yu, H. Liu, and L. Zhang, "S-yolo: Background-weakening yolov3 for gear surface defect detection," *Journal of Sensors*, vol. 2019, pp. 1–10, 2019.

[6] H. Nguyen, "Mobilenetv2 + enhanced feature pyramid for fast object detection," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 21, pp. 3341–3352, 2020.

[7] K. Khurana, R. Kumar, and S. Singh, "Molo: Hybrid model using mobilenetv2 and yolov8 for edge devices," *Discover Artificial Intelligence*, vol. 5, no. 1, pp. 1–15, 2025.

[8] S. Zhang, Y. Wang, and H. Li, "Improved yolov5 for industrial interference-resistant gear detection," *Measurement*, vol. 215, p. 112014, 2023.

[9] M. Chen, F. Zhao, Y. Xu, and Q. Li, "Anchor-free yolov8 for small defect detection in industrial surfaces," *IEEE Access*, vol. 12, pp. 125 874–125 885, 2024.

[10] H. Nguyen, "Mobilenetv2 + enhanced feature pyramid for fast object and defect detection," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 21, pp. 3341–3352, 2020. [Online]. Available: http://www.jatit.org/volumes/Vol98No21/1Vol98No21.pdf

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[12] J. Xie, L. Zhou, and T. Wang, "Improved faster r-cnn for surface defect detection in steel plates," *Measurement*, vol. 175, p. 109131, 2021.

[13] Y. Ma, L. Wu, J. Chen, and Y. Zhang, “Comprehensive review of defect detection models for industrial visual inspection,” *Artificial Intelligence Review*, vol. 57, no. 8, pp. 12 563–12 595, 2024.

[14] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114.

[15] W. Zhang, L. Chen, and F. Huang, “Efficientnet-based deep learning model for industrial defect classification,” *IEEE Access*, vol. 10, pp. 98 245–98 256, 2022.

[16] J. Liu, H. Zhang, and W. Sun, “Defect detection of steel surfaces using improved resnet50,” *Measurement*, vol. 183, p. 109886, 2021.

[17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 21–37.

[19] F. Zhao, Y. Wang, and B. Li, “Improved ssd for small defect detection in industrial products,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9876–9885, 2020.