IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/46AI/2025/4642825/
Vol. 2025 (34), Article ID 4642825

Deep Learning Models for Real-Time Quality Defect Detection in Assembly Lines*

Kamil MUSIAŁ [0000-0002-7615-7165], Artem BALASHOV [0000-0002-2937-1070], Joanna KOCHAŃSKA [0000-0002-4695-1640] and Dagmara ŁAPCZYŃSKA [0000-0002-4745-0768]

Wroclaw University of Science and Technology, Wroclaw, Poland

Correspondence should be addressed to: Kamil MUSIAŁ, kamil.musial@pwr.edu.pl

* Presented at the 46th IBIMA International Conference, 26-27 November 2025, Ronda, Spain

Abstract

The high competitiveness and rapid pace of development of the manufacturing industry requires companies to constantly improve, both in terms of efficiency and quality of production. It is these parameters that directly affect not only the reputation of the company, but also the costs it incurs, ultimately determining the price of the product on the market. Both production efficiency and product quality are affected by the speed and effectiveness of detecting manufacturing defects.

This paper presents a study to evaluate the effectiveness of detecting defective production quality in real time, using Convolutional Neural Networks (CNN), You Only Loon Once v8 (YOLOv8), and Faster R-CNN with ResNet-101 backbone. The study was carried out on the example of data for assembly lines in the automotive industry, supported by advanced techniques of pre-processing, data augmentation and knowledge transfer.

The results of the research made it possible to assess the potential of the analyzed solutions in detecting defects. The highest average precision (mAP_{50}) was achieved using YOLOv8. By using Faster R-CNN with the ResNet-101 backbone, higher precision was achieved only for small defects, but at a lower speed. When using the standard CNN neural network, the average precision was the lowest, while at the same time not providing spatial localization capability. The main barriers to implementation include computational requirements, as well as integration with programmable logic controllers (PLCs).

Keywords: deep learning, object detection, computer vision, Industry 4.0

Introduction

Quality control is one of the main operational functions of manufacturing companies. However, the results of the study (Mazzetto et al., 2020) show that quality controllers detect only 70-90% of defects, with these values being even lower in the case of longer changes. Dynamic technological development means that traditional methods of defect detection turn out to be insufficient in the context of ensuring competitiveness (Guha et al., 2023; Weiher et al., 2023). Increasing the effectiveness of quality control increasingly requires investment in more advanced systems, thus eliminating human errors resulting from the limitations of human senses, fatigue, etc. With the rise

Cite this Article as: Kamil MUSIAŁ, Artem BALASHOV, Joanna KOCHAŃSKA and Dagmara ŁAPCZYŃSKA, Vol. 2025 (34) "Deep Learning Models for Real-Time Quality Defect Detection in Assembly Lines " Communications of International Proceedings, Vol. 2025 (34), Article ID 4642825, https://doi.org/10.5171/2025.4642825

of Industry 4.0, computer vision and deep learning technologies have evolved, opening new opportunities in the context of automating quality control processes (Diers and Pigorsch, 2023; Hussain, 2024).

Despite the enormous opportunities that Industry 4.0 offers in the area of quality control, the implementation of real-time quality control systems on production lines is still quite a challenge (Tewogbade et al., 2024) due to, among others:

- computational constraints requiring inference speed in line with the rate of production,
- changing lighting conditions affecting image quality and visibility of defects,
- variable product orientation,
- lack of data on defects,
- the need for integration with existing production execution systems (MES) and programmable logic controllers (PLCs).

This article compares three leading deep learning architectures for detecting quality defects occurring in production lines. The research was carried out using data from the automotive industry, using the example of assembly processes. Both classification accuracy and implementation factors such as inference speed, technical requirements, precision of defect location, and resistance to changes in image capture conditions were analyzed.

Research methodology

Data Acquisition and Computer Vision

A dataset consisting of 50,000 images with a resolution of 1920x1440 pixels, recorded during 8 months of production of automotive components, was used for the research. The dataset contained images of different product variants, in different lighting conditions. 2,500 manually marked defective components were identified in eight defect categories: surface scratches (18%), dimensional errors (22%), material contamination (15%), assembly misalignment (12%), paint defects (14%), missing components (10%), deformations (6%) and electrical contact problems (3%).

Manual labeling of the data was carried out according to the established annotation protocol, using labeling software with object marking in bounding boxes, which ensured high location accuracy. Labelling consistency was confirmed by inter-assessment plausibility tests (Cohen kappa coefficient = 0.89). The datasets were divided according to temporal stratification: images from months 1-5 (35,000) were used for the learning stage, from month 6 (8,000) for validation, and from months 7-8 (7,000) were used for timed tests. This approach avoided leakage and provided a realistic assessment of the models' performance on new, previously unseen production batches.

Image pre-processing made it possible to unify the influence of illumination by aligning the histogram and transforming it into the LAB color space, which increased the resistance of the models to the variability of light conditions. The use of adaptive limited-contrast histogram correction (CLAHE) improved the visibility of minor defects while maintaining edge integrity (Mazzetto et al., 2020). Images have been scaled to standard dimensions — 416×416 pixels for YOLOv8 and 800×600 pixels for Faster R-CNN — with proportions thanks to the use of letterboxes.

Data Augmentation and Synthetic Defect Generation

Given that defective components account for less than 3% of production, comprehensive data augmentation was used in the research. Geometric transformations included random rotations (±15°), horizontal and vertical reflections, and cropping while maintaining 80-100% of the original image. Photometric improvements included random brightness adjustments (±20%), contrast changes (0.8–1.2 ratio), and the addition of Gaussian noise. In addition, color jitter is used, introducing controlled, random changes in color, saturation and values to a limited extent.

Deep learning model architectures

The three most common architectures of deep learning models were used for the analysis: Convolutional Neural Networks (CNN), YOLOv8 (You Only Look Once v8) and Faster R-CNN with ResNet-101 backbone network.

The CNN used consisted of five convolutional blocks, each of which included a batch normalization layer and a maximum pooling layer. Such a combination made it possible to gradually isolate the increasingly complex features of the image. The architecture also includes dropout layer (factor 0.5) to reduce the risk of model retraining. Ultimately, the traits obtained were passed on to three fully connected layers (512, 256, and 128 neurons), responsible for binary classification—distinguishing between defective and normal images. The model constructed in this way was a basic solution, providing good performance, although without the possibility of spatial location of defects. This approach is in line with the commonly used CNN architectures used in industrial visual inspection (Marchesiello, 2024).

YOLOv8 is a single-stage object detector that analyzes the image in a single network pass, while predicting the position and category of the elements to be detected. The architecture of the model uses the so-called pyramid of features, thanks to which it can recognize objects at different scales (with down sampling of 8×, 16× and 32×). The YOLOv8 is equipped with a modern, anchorless detection head and an optimized backbone network, which allows for high accuracy while maintaining real-time operation. Recent research (Yi et al., 2024) on defect detection on PCBs confirms the effectiveness of this approach — YOLOv8-based models achieve 97.5–98.7% mAP scores, outperforming earlier versions of YOLO. This study uses a medium-scale variant (YOLOv8-M) that represents a trade-off between accuracy and speed.

The last model analyzed, the Faster R-CNN with the ResNet-101 backbone network, is a two-stage object detector. In the first stage, the Region Proposal Network (RPN) generates areas that potentially contain objects, which are then classified and more accurately located in the second step. The ResNet-101 used consists of 101 residual layers, which allows for the extraction of highly detailed, hierarchical image features. The Faster R-CNN model typically provides higher accuracy compared to single-stage methods, although this involves higher computing power requirements. The network also uses a pyramid of features, so it can effectively analyze images at different scales — from large defects, such as misalignment of elements, to minor surface scratches (Mo and Yan, 2020; Ren et al., 2017).

All models use transfer learning, which uses weights pre-trained on the Common Objects in Context (COCO) dataset. This allowed architectures to effectively learn from a relatively small number of defective images. The hyperparameters were selected based on a grid search performed on the validation set. Optimal values included a learning rate of 0.001, a batch size of 32, and training lasting 100 epochs with an early stop mechanism (15 epochs).

Assessment metrics and validation strategy

To assess the effectiveness of the developed models, a set of metrics was used to describe both the quality of the classification and the accuracy of defect location.

In the case of binary classification, the following metrics were used: accuracy, precision, completeness, F1 score and the field under the ROC curve (AUROC). They provide a comprehensive assessment of the model's performance, including its susceptibility to false positives and false negatives, which is critical in industrial applications.

For object detection models that further determine the location of defects, average precision metrics (mAP) were used at IoU = 50% and IoU = 75% thresholds. They allow to evaluate the detection effectiveness and location accuracy at the same time. Compute performance was determined by average inference time (ms/image) and throughput (images per second), which is important for real-time deployment on assembly lines processing 30-40 components per minute.

Three additional test kits have been prepared to assess the reliability of the models:

- images with synthetically altered lighting,
- elements with orientations absent from the training set,

- new types of defects that were not previously present in the training data.

This approach allowed to test the ability of the models to generalize knowledge beyond controlled training conditions.

The temporal validation strategy was based on separating the training and test data chronologically. This avoided so-called "data leakage" and provided a reliable assessment of the performance of the models under realistic production conditions.

Results

Compare the performance of models

Table 1 shows the performance results of the analyzed models.

Table 1: modeling performance on a test dataset

Metric	CNN	YOLOv8-M	Faster R-CNN
Accuracy	91.3%	93.8%	92.1%
Precision	0.891	0.943	0.971
Recall	0.847	0.925	0.889
F1 Result	0.868	0.934	0.929
AUROC	0.941	0.969	0.956

Based on the results presented, it can be concluded that all models have achieved high classification performance. The YOLOv8 achieved the highest F1 (0.934) and AUROC (0.969) scores, confirming its good balance between precision and completeness. Faster R-CNN, on the other hand, achieved higher precision (0.971), which means fewer false alarms. This is important in quality control processes, where misdetections can lead to unnecessary downtime or intervention. The CNN base model achieved satisfactory accuracy, but its limitation was the lack of spatial location of defects, which made it difficult to use directly in repair processes.

Object detection performance

Table 2 compares detection performance and inference speed.

Table 2: detection performance and speed of model inference

Model	mAP_{50}	mAP_{75}	Inference (ms)	Bandwidth (fps)
YOLOv8-M	94.7%	89.2%	25	40
Faster R-CNN	93.1%	91.8%	68	14.7
CNN	N/A	N/A	12	83

The YOLOv8 achieved the highest mAP_{50} (94.7%) with an inference time of 25 milliseconds per image, allowing for the processing of approximately 40 components per second. This significantly exceeds the typical assembly line speeds in the automotive industry, which are around 30-40 components per minute. These results are in line with recent research on YOLOv8 applications in a production environment (Liu et al., 2025), which achieved approximately 95% mAP at a detection rate of 30 frames per second.

The Faster R-CNN model, on the other hand, achieved a higher mAP_{75} score (91.8%), which confirms its greater precision in identifying small defects. However, this comes at the expense of computing performance – the average inference time was 68 milliseconds per image, which corresponds to a speed of about 14.7 frames per second.

Performance by defect category

Table 3 presents the F1 points by defect type.

Table 3: F1 points by defect type

Disadvantage category	CNN	YOLOv8	Faster R-CNN
Surface scratches	0.892	0.943	0.959
Dimensional errors	0.831	0.912	0.897
Material contamination	0.847	0.928	0.941
Misalignment of the assembly	0.798	0.889	0.873
Paint disadvantages	0.863	0.934	0.951
Missing components	0.921	0.956	0.948
Deformation	0.789	0.901	0.915
Electrical contact problems	0.756	0.867	0.891

The performance of the models varied depending on the type of defects detected. The highest values of the F1 index (0.921–0.956) were recorded for the category of missing components, which is due to their clear and easily recognizable visual structure. The lowest scores were obtained for electrical contact problems, defects that require analysis at very close range (F1: 0.756–0.891). Of all the models analyzed, YOLOv8 showed the most balanced performance in all categories, confirming its ability to perform stably under a variety of detection conditions.

Reliability Rating

Table 4 shows the results of resistance tests, which determine the effectiveness of the models under difficult conditions.

Table 4: resistance tests: effectiveness in harsh environments

Condition	YOLOv8 performance	Faster R-CNN performance
Standard Test Kit	94.7% mAP	93.1% mAP
Lighting variants	92.1% mAP (-2.6%)	89.8% mAP (-3.3%)
New orientations	91.4% mAP (-3.3%)	87.9% mAP (-5.2%)
Untrained defect types	88.7% mAP (-6.0%)	84.3% mAP (-8.8%)

Evaluation of the resistance of the models showed a decrease in their effectiveness under harsher test conditions. Changes in lighting have led to a 2.6% decrease in mAP for the YOLOv8 and a 3.3% decrease in the Faster R-CNN, indicating greater stability for the YOLOv8 in this area. For the new component orientations, YOLOv8 performance decreased by 3.3%, while Faster R-CNN performance decreased by 5.2%, confirming the better generalization ability of the YOLOv8 model. The largest decrease in effectiveness was observed for previously unknown types of defects that were not present in the training data – 6.0% for YOLOv8 and 8.8% for Faster R-CNN. These results are in line with expectations and confirm that changing the distribution of data outside the training range is the most difficult challenge for both models.

Summary

Analysis of the test results confirmed the effectiveness and efficiency of real-time detection of quality defects on assembly lines based on deep learning. The greatest implementation potential was noted for YOLOv8, characterized by high real-time capability (40 frames/s) with high accuracy (94.7% mAP).

On the other hand, the highest precision was demonstrated by Faster R-CNN (mAP_{75} 91.8%), the use of which is recommended only for key products due to the calculation overhead. It is estimated that the implementation of the proposed solutions is profitable primarily for plants with a large production volume (annual return on investment of over 70%).

It's worth noting that the effectiveness of deployment doesn't just depend on accuracy metrics but also on computational requirements, integration with the image acquisition system, lighting stability and organizational issues.

References

- Diers, J and Pigorsch, C (2023) 'A survey of methods for automated quality control based on images,' International Journal of Computer Vision, 131, 2553–2581.
- Guha, B, Moore, S and Huyghe, JM (2023) 'Application and validation of machine vision inspection for efficient in-process monitoring of complex biomechanical device manufacturing,' Journal of Engineering and Applied Science, 70, 72.
- Hussain, M (2024) 'Sustainable machine vision for Industry 4.0: a comprehensive review of convolutional neural networks and hardware accelerators in computer vision,' AI, 5 (3), 1324–1356.
- Marchesiello, M (2024) 'Building a vision inspection CNN for an industrial application,' [Online]. Towards Data Science. Available: https://towardsdatascience.com/building-a-vision-inspection-cnn-for-an-industrial-application-138936d7a34a
- Mazzetto, M, Teixeira, M, Rodrigues, E and Casanova, D (2020) 'Deep learning models for visual inspection on automotive assembling line,' arXiv preprint, 2007.01857.
- Mo, N and Yan, L (2020) 'Improved faster R-CNN based on feature amplification and oversampling data augmentation for oriented vehicle detection in aerial images,' Remote Sensing, 12 (16), 2558.
- Ren, S, He, K, Girshick, R and Sun, J (2017) 'Faster R-CNN: towards real-time object detection with region proposal networks,' IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (6), 1137–1149.
- Tewogbade, S (2024) 'Defect detection in manufacturing: an integrated deep learning approach,' Journal of Computer and Communications, 12, 153–176.
- Weiher, K, Rieck, S, Pankrath, H, Beuss, F, Geist, M, Sender, J and Fluegge, W (2023) 'Automated visual inspection of manufactured parts using deep convolutional neural networks and transfer learning,' Procedia CIRP, 120, 858–863.
- Yi, F, Mohamed, ASA, Noor, MHM, Che Ani, F and Zolkefli, ZE (2024) 'YOLOv8-DEE: a high-precision model for printed circuit board defect detection,' PeerJ Computer Science, 10, e2548.