POLITECNICO DI TORINO
Repository ISTITUZIONALE

Unsupervised Defect Detection in Automotive Quality Inspection with Convolutional Autoencoder

Original
Unsupervised Defect Detection in Automotive Quality Inspection with Convolutional Autoencoder / Casella, A., Randazzo, V., Pasero, E.. - ELETTRONICO. - (2025), pp. 1-8. (International Joint Conference on Neural Networks (IJCNN) Roma (Ita) 30 June - 5 July 2025) [10.1109/ijcnn64981.2025.11228358].

Availability:
This version is available at: 11583/3005182 since: 2025-12-18T11:02:16Z

Publisher:
IEEE

Published
DOI:10.1109/ijcnn64981.2025.11228358

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

19 November 2026

Unsupervised Defect Detection in Automotive Quality Inspection with Convolutional Autoencoder

Alessandro Casella
DET
Politecnico di Torino
Turin, Italy
alessandro.casella@polito.it

Vincenzo Randazzo
DET
Politecnico di Torino
Turin, Italy
vincenzo.randazzo@polito.it

Eros Pasero
DET
Politecnico di Torino
Turin, Italy
eros.pasero@polito.it

***Abstract*—This paper introduces an unsupervised defect detection system using a Convolutional Autoencoder (CAE) for brake caliper quality control in automotive manufacturing. A CAE is trained to reconstruct defect-free images, leveraging a Structural Similarity Index Measure (SSIM) loss to isolate anomalies between original and reconstructed images. Candidate defect regions are refined using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering to distinguish true defects (e.g., deformities, scratches) from noise.**

Experiments conducted on a custom imaging workstation demonstrated strong performance in diverse regions of interest (ROIs) of the brake calipers. The system achieved F1-scores of 0.92, 0.95, and 0.76 on the logo, flat (2D), and non-flat (3D) ROIs, respectively. Data augmentation improved generalization, and clustering reduced false positives (FPs).

By offering an alternative to traditional supervised methods, this CAE-based approach enables reliable defect detection, reducing manual dependence. Its flexible design supports integration into automotive production, leading to improved real-time monitoring and cost savings.

***Index Terms*—Convolutional Autoencoder, Unsupervised Anomaly Detection, Industrial Quality Control, Automotive Manufacturing, Machine Vision**

I. Introduction and Related Work

Automated defect detection plays a pivotal role in ensuring quality control within the industrial sector, especially in automotive manufacturing. Traditional inspection methods, often manual, suffer from being time-consuming, costly, and susceptible to human error [1], [2]. Machine Vision (MV) systems offer an attractive alternative by automating inspection processes, with applications ranging from surface checks to assembly verification [3]. Among automotive components, brake calipers hold particular significance, as their quality directly impacts vehicle safety and aesthetic appeal, the latter especially in luxury vehicles. Brake calipers are critical components of a vehicle's braking system, and their quality must meet high standards to ensure safety and reliability [4].

Recent progress in Deep Learning (DL) has enabled innovative approaches to defect detection. Convolutional Autoencoders (CAEs), a form of unsupervised learning, are able to learn complex features and identify anomalies in images without requiring labeled data [5]. Feature autoencoders have been applied to anomaly detection in industrial machines, enabling anomaly detection systems to identify operational irregularities and improve quality control processes [6]. This capability is invaluable in industrial contexts where defective parts are rare or difficult to classify [7]. Bergmann et al. [8] demonstrated the effectiveness of CAEs by using the Structural Similarity Index Measure (SSIM) loss function to improve image reconstruction quality and highlight defect areas. Wang et al. [9] originally introduced SSIM, a method later applied by Carrera et al. [10] to defect detection in high-resolution microscopy images.

Building on these advancements, numerous studies have refined CAE architectures to improve performance across diverse industrial settings. Techniques such as multi-scale denoising have been introduced to address variations in lighting conditions [11], while generative adversarial networks (GANs) with dual attention mechanisms enhance defect localization by focusing on salient regions [12]. GANs have been successfully employed for small surface defect detection, leveraging exaggerated local variations to enhance model sensitivity to subtle anomalies [13]. Clustering-based methods, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN), have also been used to minimize FPs caused by variations in texture or illumination [14].

MV applications in automotive manufacturing have historically relied on barcode scanners, 3D sensors, and traditional image processing, such as edge detection for weld inspection, circle-based methods for joint quality evaluation, and deflectometry techniques for paint surface analysis [15]. AI-driven quality control systems have been widely adopted in the automotive industry to detect and classify manufacturing defects, enhancing efficiency and reducing reliance on manual inspection. Mazzetto et al. [16] explored the use of DL models for visual inspection in automotive assembly lines, showing that hybrid approaches combining supervised and unsupervised learning achieve higher defect detection rates than purely rule-based systems. Additionally, research has highlighted the role and importance of MV systems in industrial applications, emphasizing their capability to perform automated visual inspections, process control, and parts identification, thereby improving product quality and production reliability [17].

Despite these advancements, challenges remain in deploying CAE-based systems in real-world manufacturing environments. Variations in lighting conditions can significantly

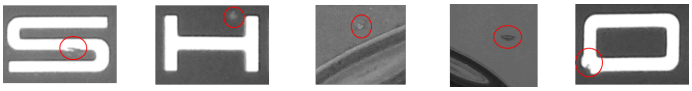


Fig. 1. Example of some analyzed real defects on brake caliper's surface.

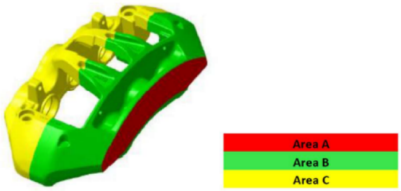


Fig. 2. Brake caliper structure representation annotated with the three sections: logo (A), flat 2D and non-flat 3D (B), rear (C).



Fig. 3. Image acquisition setup with LED lighting and overhead camera.

affect defect reconstruction accuracy, necessitating adaptive illumination techniques [16]. Highly reflective or complex surfaces can introduce unintended artifacts, leading to FPs, which remain a challenge in unsupervised defect detection models, particularly when training datasets lack sufficient diversity in anomaly-free samples [18]. Furthermore, the computational demands of DL models can pose scalability issues for real-time applications. Addressing these challenges requires innovative solutions that combine the strengths of CAEs with robust pre- and post-processing techniques.

Multistage quality control systems incorporating machine learning have demonstrated significant potential in improving defect detection rates within automotive manufacturing processes [19]. Similar approaches have also been explored in other industrial domains, where unsupervised neural models combined with dimensionality reduction have shown effectiveness in detecting faults in power electronics and rotating machinery [20]–[22]. Jin et al. applied DBSCAN to semiconductor defect detection, successfully filtering out non-defect variations by grouping candidate regions based on similarity metrics [23].

This paper proposes an improved CAE-based system for unsupervised defect detection in automotive manufacturing, specifically for the brake caliper quality control. Some examples of defects taken into account (bubbles, scratches, stains...) are shown in Fig. 1. The key contributions are:

- Optimized CAE architecture: Bergmann et al. approach [8] was refined by optimizing the use of SSIM-based loss through threshold tuning, improved defect selection with DBSCAN clustering, and additional regularization techniques to enhance model robustness.
- Comprehensive evaluation: The model was tested on a real-world automotive dataset, analyzing performance under different lighting and surface conditions.
- Improved training strategy: The impact of data augmentation was explored to enhance the performance of the CAE, improving its ability to generalize across different defect types.

The paper is structured as follows: Section II describes the proposed method, including network architecture and dataset details; Section III presents the most significant results derived from this approach; finally, Section IV provides concluding remarks on the contributions and implications for automotive quality control, and outlines possible future developments.

II. Methods

To provide a reliable system, its validity had to be demonstrated in a scenario that faithfully simulated the real production cycle of quality control. The analysis focused on the three specific regions of the brake caliper shown in Fig. 2. It started with the most critical area of interest, the logo region (Area A), and then extended to two distinct regions within Area B: the flat (2D) and the non-flat (3D) areas. The rear part of the brake caliper, that is the part behind the logo (Area C), was not taken into consideration. The flat regions are areas of the brake caliper that lie entirely on a single plane, meaning they do not contain transitions between different surface levels. These areas generally exhibited uniform lighting conditions, making them less challenging for defect detection. Conversely, non-flat regions included sections that span multiple planes or curved surfaces, where shadows, reflections, and variations in illumination introduce additional complexity.

This distinction was necessary because acquiring and training a model on the entire brake caliper was not feasible, since no completely defect-free brake calipers were available. Specific Regions of Interest (ROIs) had to be selected for training, ensuring that the model could learn a reliable baseline representation of defect-free surfaces.

A. Dataset Acquisition and Image Preprocessing

The dataset was acquired in a controlled industrial environment using a dedicated imaging workstation designed to capture high-resolution images of brake calipers under variable lighting conditions. The goal was to develop a robust dataset that allows the CAE to generalize well to real-world variations in defect appearance.

The images were collected using a *RealSense* camera [24] positioned inside a controlled workbench with adjustable LED lighting (see Fig. 3). The workbench was enclosed on three sides, with an open front for inserting the brake calipers.

TABLE I
AUGMENTED DATASET DISTRIBUTION FOR THE LOGO REGION

Set	Original Images	Augmentations	Total Images
Training	60	880	940
Validation	15	220	235
Test	116	-	116

TABLE II
AUGMENTED DATASET DISTRIBUTION FOR A FLAT (2D) ROI

Set	Original Images	Augmentations	Total Images
Training	48	752	800
Validation	12	188	200
Test	130	-	130

For this reason, ambient light could still influence image acquisition. To address this, a dynamic lighting cycle with adjustable LED illumination was set inside the bench.

Each brake caliper was placed on the acquisition bench for approximately 15 seconds, resulting in around 15 images per brake caliper under varying lighting conditions. A total of 13 brake calipers were included in the dataset; the training dataset was composed exclusively of images from brake calipers that appeared defect-free, while the test dataset contained images from calipers with visible defects, both conditions related to the specific ROI. To ensure consistency in input data and optimize the performance of the CAE, each image was preprocessed using the following standardized pipeline:

1) grayscale conversion: eliminating color variations to focus on structural details;
2) blurring: applying a slight Gaussian blur to reduce high-frequency noise [25];
3) ROI extraction: cropping the image to focus on specific regions (logo, flat, or non-flat);
4) padding: ensuring a consistent input size to the CAE by applying zero-padding when necessary [26].

Preprocessing was applied uniformly to all the acquired images, including those used for testing. The defect-free images were then split into 80% for training and 20% for validation, ensuring a balanced representation of normal samples during model development. A separate set of images containing known defects was acquired independently and reserved for evaluating the model's generalization capabilities.

The final dataset distribution, including augmented samples, is detailed in Tables I, II, and III, which refer to the three studied ROIs: the logo, a flat (2D), and a non-flat (3D), respectively.

B. Data Augmentation

Given the limited availability of defect-free brake calipers, data augmentation was applied to increase dataset diversity and improve model generalization. The goal was to introduce controlled variations that reflect real-world acquisition inconsistencies, such as changes in positioning and lighting.

The applied transformations, summarized in Table IV, are equally divided between geometric adjustments (translation, rotation, rototranslation) and color space modifications (contrast and brightness). Geometric parameters, including shifts of up to 40 pixels horizontally, 20 pixels vertically, and rotations up to 36°, were chosen to reflect typical misalignment caused by manual part placement. Following extensive testing, these parameters were selected as they consistently offered the best balance between model robustness and realistic variation in input data. Similarly, contrast and brightness were varied within a conservative range, designed to reflect typical lighting fluctuations observed in the acquisition setup, without introducing unrealistic distortions.

While the autoencoder was inherently robust to small variations, excessive brightness changes or mispositioning could either obscure real defects or create false differences between the original and reconstructed images, leading to incorrect defect detection. This made augmentation essential to improve the model's reliability as much as possible.

TABLE III
AUGMENTED DATASET DISTRIBUTION FOR A NON-FLAT (3D) ROI

Set	Original Images	Augmentations	Total Images
Training	93	1457	1550
Validation	23	360	383
Test	71	-	71

TABLE IV
DATA AUGMENTATION STRATEGY

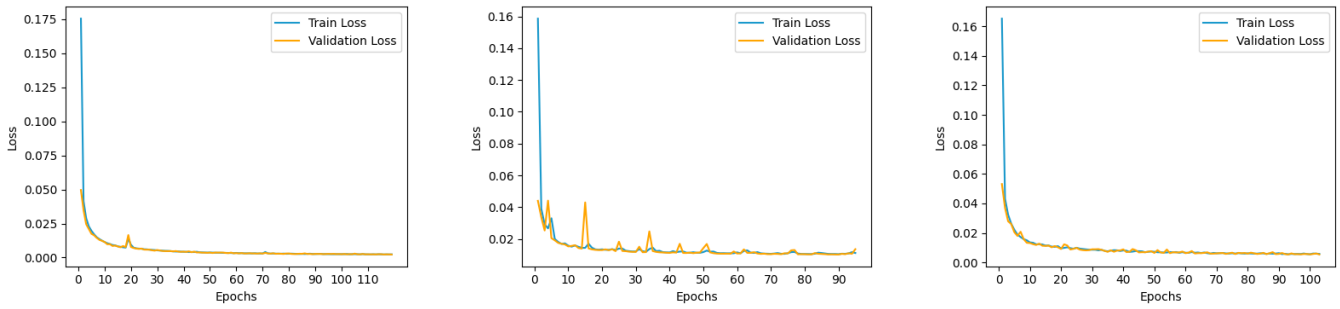
Transformation	Details	Weight
Translation	x = $[-20, 20]$, $y = [-10, 10]$	19%
Rotation	α = $[-18^\circ, 18^\circ]$	12%
Rototranslation	rotation + translation	13%
Contrast & Brightness	$c = [-0.8, 1.2]$, $b = [-0.05, 0.08]$	50%

C. Neural Network Architecture

The CAE architecture employed in this study (see Table V) is inspired by the work of Bergmann et al. [8], who proposed their approach for defect detection on nanofibrous materials and woven fabric textures. However, several targeted modifi-

TABLE V
AUTOENCODER ARCHITECTURE

Layer	Configuration
Input	pixel width × pixel height × 1
Encoder	
Conv2D (1)	32 filters, stride=2, padding=Same, LeakyReLU(0.25)
Conv2D (2)	32 filters, stride=2, padding=Same, LeakyReLU(0.25)
Conv2D (3)	32 filters, stride=1, padding=Same, LeakyReLU(0.25)
Conv2D (4)	64 filters, stride=2, padding=Same, LeakyReLU(0.25)
Conv2D (5)	64 filters, stride=1, padding=Same, LeakyReLU(0.25)
Conv2D (6)	128 filters, stride=2, padding=Same, LeakyReLU(0.25)
Conv2D (7)	64 filters, stride=1, padding=Same, LeakyReLU(0.25)
Conv2D (8)	32 filters, stride=1, padding=Same, LeakyReLU(0.25)
Bottleneck Layer	
Conv2D (9)	10 filters, stride=1, padding=Valid, Linear



(a) Logo region (b) Flat ROI (c) Non-flat ROI

Fig. 4. Training and validation loss trends on different regions of interest (ROIs): (a) logo region, (b) flat ROI, and (c) non-flat ROI.

cations were introduced to better adapt the network specifically to the challenges of metallic brake caliper inspection in automotive manufacturing. These architectural differences were introduced based on extensive real-world testing, with the goal of improving robustness under production conditions involving variable illumination and complex 3D surfaces. Specifically, the key modifications are:

- Kernel sizes were standardized to 3×3 across all convolutional layers, rather than the original mix of 4×4 and 3×3 used by Bergmann et al. This simpler design enhances the extraction of subtle features critical for identifying surface anomalies on metallic brake calipers, which differ significantly from defects on woven fabrics or nanofibrous materials.
- Latent space dimension was reduced to 10 filters compared to 100 or 500 filters employed in the original architecture. Extensive tests indicated that this compact latent space effectively improves generalization by avoiding the encoding of irrelevant details, which are particularly problematic due to reflections and lighting variations typical of metallic surfaces.
- The LeakyReLU slope was increased slightly from 0.2 to 0.25 after empirical evaluation. This modification provided improved gradient flow and convergence stability, beneficial when processing uniformly textured metal surfaces with subtle defect features.

Input images were fed into the network without resizing, ensuring the preservation of fine-grained details crucial for defect detection. Downsampling was performed using convolutional layers with a stride of 2 instead of pooling layers, preserving spatial consistency throughout the network. Padding was set to "same" in all convolutional layers, except at the bottleneck where "valid" padding was intentionally used to enforce spatial compression. LeakyReLU activations were consistently used to mitigate vanishing gradients, except in the final decoder layer, where a linear activation function was employed to facilitate unrestricted pixel-value reconstruction.

D. Training Configuration

The model was trained using the Adam optimizer with hyperparameters optimized for stability and convergence, as shown in Table VI. The learning rate value was chosen to balance training speed and stability, while L2 regularization was applied to prevent overfitting. The batch size was set to 4 as a trade-off between memory efficiency and gradient stability.

TABLE VI
BEST MODEL HYPERPARAMETERS

Hyperparameter	**Value**
Optimizer	Adam (lr = 10^{-3}, β_1 = 0.9, β_2 = 0.999, ϵ = 10^{-7})
L2-Regularization	weight_decay = 10^{-5}

$$SSIM(x,y) = l(x,y) \cdot c(x,y) \cdot s(x,y) = \\ = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

To guide the autoencoder toward meaningful reconstructions, training was performed using the SSIM loss, instead of the traditional Mean Squared Error (MSE). SSIM focuses on perceptual image quality by considering luminance, contrast, and structural similarities [9]. The SSIM function between an input image x and its reconstruction y is defined in Eq. 1, where $l(x,y)$, $c(x,y)$, and $s(x,y)$ denote luminance, contrast, and structure, respectively, and are computed as in [8]. μ_x and μ_y are the mean intensities of the two images, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance between the two images. The constants $c_1 = 0.01$ and $c_2 = 0.03$ ensure numerical stability.

The training progress for the three types of ROI analyzed is shown in Figs. 4a, 4b, 4c. The loss curves indicate a rapid initial decrease, followed by stabilization as the model converges. Minor fluctuations in validation loss were observed, particularly in the flat and non-flat ROIs, likely due to surface variations and lighting inconsistencies in the dataset. Overall, the consistency between training and validation loss suggests that the model generalized well without significant overfitting.

E. Clustering Algorithm

To minimize FPs caused by autoencoder reconstruction errors, a clustering-based post-processing step was applied to

the SSIM residual maps. DBSCAN is well-suited for this task as it detects arbitrarily shaped clusters without requiring prior knowledge of the number of defects [14], [23].

The algorithm groups pixels with high anomaly scores while filtering out isolated noise. A point is classified as a *core* point if it has at least $minPts$ neighbors within a radius ϵ; otherwise, it is considered either *border* and it is connected to the core of a specific cluster, or a *noise* point so excluded from defect observations. Unlike traditional thresholding methods, DBSCAN dynamically adapts to local density variations, ensuring robustness among different scenarios.

The values of the two parameters were chosen based on the general defect physiognomies and surface characteristics. $\epsilon = [5, 15]$ parameter range was chosen for flat and non-flat areas, while $\epsilon = 1.5$ value was found to be optimal to detect the expected features on the entire logo region. $minPts = [15, 30]$ parameter was set in this range to balance sensitivity and FP minimization. However, a lower ϵ combined with a higher $minPts$ reduces over-segmentation but may cause subtle defects to be missed, whereas increasing ϵ and decreasing $minPts$ leads to a higher presence of FPs.

By integrating DBSCAN with the CAE-based SSIM anomaly maps, the system improves defect discrimination, particularly in highly reflective and non-flat surface areas.

F. Aggregated Detection Method

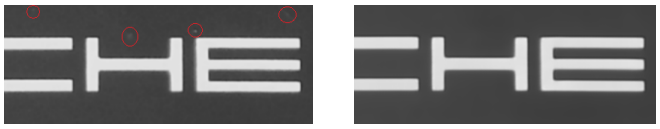
To further enhance defect detection reliability, an aggregated detection method was implemented, leveraging multiple images of the same brake caliper acquired under varying lighting conditions. A cyclic dynamic lighting setup modulated the LED light intensity over five seconds, capturing multiple images per brake caliper.

After having performed the entire previously described defect detection pipeline independently on each image, the results were aggregated by counting clusters recurrence across all the images. A threshold percentage was applied, ensuring that only defects detected in at least 50% of images were considered valid, reducing FPs coming from reflections, dust, or reconstruction errors.

III. Results and Discussion

A. Autoencoder Reconstruction

Fig. 5a illustrates an example of a brake caliper cropped logo containing four defects (circled in red), two clearly visible while the other two very fine. The autoencoder, trained exclusively on defect-free samples, reconstructed an idealized version of the image (see Fig. 5b), where the defects have been removed. By computing the SSIM value between the two images (as explained in Section II), the system highlighted inconsistencies that may correspond to actual defects.



(a) Brake caliper logo cropping (b) Autoencoder reconstruction

Fig. 5. Autoencoder inference process.



Fig. 6. SSIM map with similarity threshold set to 210/255. Four potential defective areas were detected (in red), but two of them are very fine.



Fig. 7. DBSCAN result (similarity threshold = 210/255, min cluster size = 25). Two of the four SSIM-detected areas have been filtered out.

B. SSIM Map and DBSCAN Clusters

This section analyzes the performance of the proposed defect detection pipeline by adjusting the parameters of the SSIM and the DBSCAN clustering algorithm. To illustrate the effect of different configurations, two distinct parameter settings were applied to the same input images (shown in Fig. 5). The figures are grouped in pairs: Figs. 6 and 7 correspond to a more conservative configuration, whereas Figs. 8 and 9 reflect a more sensitive setting. Each SSIM map is followed by its corresponding DBSCAN clustering result.

1) SSIM Threshold: SSIM ranges on a scale from -1 (completely dissimilar) to 1 (perfectly identical). In this study, SSIM values were rescaled to the $[0, 255]$ range to enhance visualization. The SSIM map highlights potential defective pixels with darker values, while the selected threshold determines the maximum similarity allowed for a pixel to be considered defective. Specifically, any pixel with a SSIM below the chosen threshold is marked as a defect candidate and colored in red. By adjusting this threshold, it was pos-



Fig. 8. SSIM map with similarity threshold set to 225/255. All four potential defective areas were detected.



Fig. 9. DBSCAN result (similarity threshold = 225/255, min cluster size = 25). All four SSIM-detected areas have been retained.

TABLE VII
PERFORMANCE METRICS ON THE ENTIRE LOGO.

Brake Caliper (Defects / Images)	Recall	Precision	F1-Score	FPs	FNs
BC1 (1 / 14)	**1.00**	**1.00**	**1.00**	**0.00**	**0.00**
BC2 (2 / 16)	0.97	0.79	0.87	0.50	0.50
BC3 (2 / 14)	0.93	0.93	0.93	0.14	0.14
BC4 (2 / 17)	0.97	**1.00**	0.98	**0.00**	**0.00**
BC5 (2 / 14)	0.90	0.90	0.90	0.21	0.21
BC6 (1 / 12)	**1.00**	0.93	0.96	0.08	**0.00**
BC7 (2 / 13)	0.58	**1.00**	0.73	**0.00**	0.85
BC8 (4 / 16)	0.99	0.97	0.98	0.13	0.06
Average	**0.92**	**0.95**	**0.92**	**0.13**	**0.22**

TABLE VIII
PERFORMANCE METRICS ON A FLAT ROI.

Brake Caliper (Defects / Images)	Recall	Precision	F1-Score	FPs	FNs
BC1 (1 / 15)	**1.00**	**1.00**	**1.00**	**0.00**	**0.00**
BC2 (1 / 16)	0.75	**1.00**	0.86	**0.00**	0.25
BC3 (7 / 14)	0.84	0.89	0.86	0.71	1.14
BC4 (1 / 12)	**1.00**	**1.00**	**1.00**	0.17	**0.00**
BC5 (3 / 13)	0.95	0.97	0.96	0.08	0.15
BC6 (2 / 16)	**1.00**	**1.00**	**1.00**	**0.00**	**0.00**
BC7 (3 / 14)	0.93	**1.00**	0.96	**0.00**	0.07
BC8 (2 / 16)	0.91	**1.00**	0.95	**0.00**	0.19
BC9 (1 / 14)	**1.00**	0.88	0.94	0.14	**0.00**
Average	**0.93**	**0.97**	**0.95**	**0.12**	**0.20**

TABLE IX
PERFORMANCE METRICS ON A NON-FLAT ROI.

Brake Caliper (Defects / Images)	Recall	Precision	F1-Score	FPs	FNs
BC1 (4 / 13)	0.89	**1.00**	0.94	**0.00**	0.46
BC2 (1 / 14)	**1.00**	**1.00**	**1.00**	**0.00**	**0.00**
BC3 (3 / 12)	0.78	0.70	0.74	1.00	0.67
BC4 (1 / 16)	0.19	0.33	0.24	0.38	0.81
BC5 (1 / 16)	0.75	**1.00**	0.86	**0.00**	0.25
Average	**0.72**	**0.81**	**0.76**	**0.28**	**0.44**

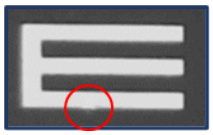


Fig. 10. Example of a FN: fine imperfection.



(a) Logo borders (b) LED reflection

Fig. 11. Examples of FPs detected by the system.

sible to balance detection sensitivity and precision: a lower threshold (e.g., 210/255) reduced noise and minimized FPs but risked missing subtle defects, whereas a higher threshold (e.g., 225/255) increased sensitivity, making even minor anomalies more visible. The threshold was chosen empirically based on experience and extensive testing. Values in the range $[210, 225]$ provided a good trade-off, and slight variations within this range were found to maintain acceptable performance.

2) DBSCAN Parameters: The DBSCAN algorithm was employed to group neighboring defective pixels into coherent regions. The *min cluster size* parameter defines the minimum number of defective pixels required for a group to be considered a valid defect. This step served as a filtering mechanism: lower values (e.g., 25) enable the detection of small defects but increase the likelihood of FPs, while higher values suppresse small, potentially spurious clusters.

3) Experimental Comparison: Fig. 6 presents the SSIM map generated with a similarity threshold of 210/255. This configuration identifies four potentially defective regions, although two of them appear rather subtle. Applying DBSCAN with a minimum cluster size of 25 (see Fig. 7) effectively filtered out the less prominent regions, preserving only the most evident anomalies. Conversely, Fig. 8 shows the SSIM map obtained with a higher threshold (225/255). With this more sensitive configuration, the same four defect regions were detected again, but with greater contrast. The corresponding DBSCAN clustering result (see Fig. 9), using the same parameters, retained all four detected regions, including the smaller ones that were previously discarded.

C. Performance on Different ROIs

The autoencoder's performance was evaluated across three distinct ROIs of the brake caliper: a region that covers the entire logo, a flat ROI, and a non-flat ROI. The effectiveness of defect detection was assessed using Recall, Precision, F1-Score, false positives (FPs), and false negatives (FNs), as summarized in Tables VII, VIII, and IX.

Each row in the tables corresponds to a specific brake caliper (BC1–BC9), where the number of defects and the total number of images used for evaluation are indicated in parentheses (e.g., "2 / 14" means 14 images of the same brake caliper, each containing the same 2 defects). Recall, Precision and F1-Score metrics represent the average values across all images for that brake caliper. Specifically, the number of defects is fixed for each brake caliper and consistent across its images; per-image metrics are computed using this known count and then averaged. Similarly, the FPs and FNs columns indicate the average number of FPs and FNs per image for each brake caliper.

1) Entire Logo: As shown in Table VII, the model achieved an average Recall of 0.92, meaning most defects were correctly identified. The average Precision of 0.95 indicates a low number of FPs, with few normal areas misclassified as defective. This balance resulted in a strong F1-score of 0.92. However, some variability was observed across different brake calipers, particularly in cases with subtle or low-contrast defects (e.g., BC7), which led to missed detections (see Fig. 10). The average number of FPs per image was low (0.13), while the average number of FNs was slightly higher at 0.22, indicating

occasional missed detections. Notably, most FPs in this region were located along the edges of the logo (see Fig. 11a), where minor reconstruction errors introduced by the autoencoder caused misclassifications due to the sharp contrast with the surrounding surface.

2) Flat (2D) ROIs: Detection performance in flat regions was slightly better overall, as shown in Table VIII. These areas are geometrically simpler, with more uniform lighting and less surface variation. The model achieved an average Recall of 0.93 and a Precision of 0.97, resulting in an F1-score of 0.95. The average FPs per image remained low at 0.12, and FNs were slightly reduced to 0.20, confirming the reliability of the system in flat regions. The absence of significant curvature or shadow effects likely contributed to the lower number of both FPs and FNs.

3) Non-Flat (3D) ROIs: These regions posed the greatest challenge for the autoencoder, as shown in Table IX. Complex surface geometries, reflections, and shadows negatively impacted the model's ability to distinguish true defects from natural texture variations. The average Recall dropped significantly to 0.72, indicating more frequent missed detections. Precision also decreased to 0.81, with an increased FPs rate of 0.28 per image. An example of such a FP is shown in Fig. 11b, where there is a specular reflection caused by the LED illumination on the brake caliper surface. Most notably, the FNs rose to 0.44 on average, confirming the system's reduced sensitivity in these areas.

D. Global ROC Analysis and AUC Evaluation

To evaluate the overall discriminative performance of the proposed system, a global Receiver Operating Characteristic (ROC) curve was computed (see Fig. 12) by concatenating all the pixel-wise anomaly scores and ground truth labels from every image across all brake calipers in the dataset. The resulting curve achieves an Area Under the Curve (AUC) of 0.995. It is worth noting that defective regions occupy only a very small portion of the inspected surface, which can make the ROC curve appear even more favorable due to the relative ease of maintaining a low false positive rate.

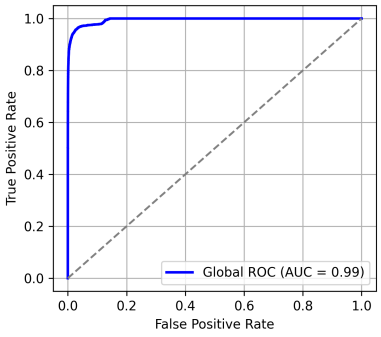


Fig. 12. Global ROC curve computed from pixel-wise anomaly scores across all brake calipers for the entire logo.



Fig. 13. Defect aggregation on 14 acquisitions of a brake caliper (SSIM threshold = 225/255, min cluster size = 15). Colors represent detected defects with respective occurrence rates.

E. Aggregated Detection Improvement

The method discussed in Section II was well-received in the real-world industrial setting, as it balances detection accuracy and production efficiency. A study on the trade-off between result quality and production time showed that analyzing 10 to 20 images per brake caliper effectively captured different brightness levels while keeping inspection time within acceptable limits. Since each image acquisition required about 1 second and processing operated at an average of 40 images per minute, the overall detection process remained efficient and compatible with production times.

A visual representation of this strategy is provided in Fig. 13, which represents the final output of the same starting image displayed in Fig. 5a. Each detected defect is assigned a percentage of occurrence based on the total number of images acquired for the same brake caliper. Defects appearing consistently across different lighting conditions are more likely to be true anomalies, while those detected sporadically are likely artifacts or FPs. The most evident defects were detected in nearly all images, confirming their significance, whereas smaller ones appeared less frequently, indicating their sensitivity to lighting variations. By applying a 50% detection threshold, the system effectively filtered out uncertain detections, ensuring that only consistently detected defects were flagged.

IV. Conclusions and Future Developments

This work proposed an unsupervised defect detection system for brake caliper surface inspection in the automotive sector. Starting from the need to reduce manual inspection times and overcome the limitations of traditional methods, a CAE was designed to reconstruct defect-free samples and detect anomalies via SSIM maps, further refined through DBSCAN clustering.

Extensive experiments conducted on three specific ROIs—logo, flat (2D), and non-flat (3D)—demonstrated the system's ability to adapt across different scenarios with varying geometric complexity and lighting conditions. Quantitatively, the approach yielded promising results: an F1-score of 0.92 on the logo area (see Table VII), 0.95 on flat surfaces (see Table VIII), and 0.76 on more challenging non-flat surfaces (see Table IX). These results highlight the system's robustness, particularly in simple and semi-structured surfaces, and its limitations when dealing with complex geometries and reflections. Nevertheless, systematic comparisons with other DL architectures commonly used for defect detection could help

better assess the added value and limitations of the proposed approach.

Qualitative analysis, shown in Figs. 6–9, demonstrated how the choice of SSIM threshold and DBSCAN parameters affects defect sensitivity and precision. The aggregated detection strategy, based on multiple acquisitions with dynamic lighting, proved effective in filtering out sporadic FPs, ensuring that only consistently detected anomalies were considered valid.

The system's architecture was optimized for the industrial context by tailoring convolutional layers, using SSIM loss instead of MSE, and incorporating an image augmentation strategy that increased the generalization of the model to real-world acquisition variations (see Table IV). Despite these optimizations, the CAE remains relatively simple and could benefit from modern techniques (attention, residual, or skip connections), particularly in complex non-flat ROIs where performance currently lags. The use of DBSCAN allowed dynamic defect clustering, striking a better balance between sensitivity and specificity.

Overall, the system reduces the need for manual inspection, adapts to unseen defects, and shows strong potential for integration into automotive quality control processes. These results suggest that unsupervised CAE models are not only viable but potentially preferable for industrial applications requiring minimal annotation effort.

Some open challenges remain, especially in handling reflections, inconsistent lighting, and complex surface shapes. Future work could focus on improving lighting control during image acquisition for more stable results and on reducing processing times to better match production requirements. Another direction is the combination of image data with additional sensor information to improve defect detection on curved or visually complex areas. Further developments should address the current manual segmentation and model-per-ROI approach, explore cross-validation or other robustness measures for more reliable performance estimates, and provide a clearer description and analysis of augmentation strategies to enhance reproducibility and effectiveness.

REFERENCES

[1] M. R. Islam, M. Z. H. Zamil, M. E. Rayed, M. M. Kabir, M. Mridha, S. Nishimura, and J. Shin, "Deep learning and computer vision techniques for enhanced quality control in manufacturing processes," *IEEE Access*, 2024.

[2] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *Journal of manufacturing systems*, vol. 48, pp. 144–156, 2018.

[3] Q. Zhou, R. Chen, B. Huang, C. Liu, J. Yu, and X. Yu, "An automatic surface defect inspection system for automobiles using machine vision methods," *Sensors*, vol. 19, no. 3, p. 644, 2019.

[4] A. Brake, "Brembo— aluminium brake callipers," *Springer*, vol. 119, pp. 32–35, 2017.

[5] A. Saberironaghi, J. Ren, and M. El-Gindy, "Defect detection methods for industrial products using deep learning techniques: A review," *Algorithms*, vol. 16, no. 2, p. 95, 2023.

[6] I. Ahmed, M. Ahmad, A. Chehri, and G. Jeon, "A smart-anomaly-detection system for industrial machines based on feature autoencoder and deep learning," *Micromachines*, vol. 14, no. 1, p. 154, 2023.

[7] J. Zipfel, F. Verworner, M. Fischer, U. Wieland, M. Kraus, and P. Zschech, "Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models," *Computers & Industrial Engineering*, vol. 177, p. 109045, 2023.

[8] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," *arXiv preprint arXiv:1807.02011*, 2018.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[10] D. Carrera, F. Manganini, G. Boracchi, and E. Lanzarone, "Defect detection in sem images of nanofibrous materials," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 551–561, 2016.

[11] G. Zhang, Z. Tang, J. Zhang, and W. Gui, "Convolutional autoencoder-based flaw detection for steel wire ropes," *Sensors*, vol. 20, no. 22, p. 6612, 2020.

[12] X. Li, Y. Zheng, B. Chen, and E. Zheng, "Dual attention-based industrial surface defect detection with consistency loss," *Sensors*, vol. 22, no. 14, p. 5141, 2022.

[13] J. Lian, W. Jia, M. Zareapoor, Y. Zheng, R. Luo, D. K. Jain, and N. Kumar, "Deep-learning-based small surface defect detection via an exaggerated local variation-based generative adversarial network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1343–1351, 2019.

[14] Y. Xie and S. Shekhar, "Significant dbscan towards statistically robust clustering," in *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, pp. 31–40, 2019.

[15] F. K. Konstantinidis, S. G. Mouroutsos, and A. Gasteratos, "The role of machine vision in industry 4.0: an automotive manufacturing perspective," in *2021 IEEE international conference on imaging systems and techniques (IST)*, pp. 1–6, IEEE, 2021.

[16] M. Mazzetto, M. Teixeira, E. O. Rodrigues, and D. Casanova, "Deep learning models for visual inspection on automotive assembling line," *arXiv preprint arXiv:2007.01857*, 2020.

[17] H. Golnabi and A. Asadpour, "Design and application of industrial machine vision systems," *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 6, pp. 630–637, 2007.

[18] J. Qiu, H. Shi, Y. Hu, and Z. Yu, "Unraveling false positives in unsupervised defect detection models: A study on anomaly-free training datasets," *Sensors*, vol. 23, no. 23, p. 9360, 2023.

[19] R. S. Peres, J. Barata, P. Leitao, and G. Garcia, "Multistage quality control using machine learning in the automotive industry," *IEEE Access*, vol. 7, pp. 79908–79916, 2019.

[20] G. Cirrincione, M. Cirrincione, D. Guilbert, A. Mohammadi, and V. Randazzo, "Power switch open-circuit fault detection in an interleaved dc/dc buck converter for electrolyzer applications by using curvilinear component analysis," in *2018 21st International Conference on Electrical Machines and Systems (ICEMS)*, pp. 2221–2225, IEEE, 2018.

[21] R. R. Kumar, V. Randazzo, G. Cirrincione, M. Cirrincione, and E. Pasero, "Analysis of stator faults in induction machines using growing curvilinear component analysis," in *2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, pp. 1–6, IEEE, 2017.

[22] V. Randazzo, G. Cirrincione, G. Ciravegna, and E. Pasero, "Nonstationary topological learning with bridges and convex polytopes: the g-exin neural network," in *2018 international joint conference on neural networks (IJCNN)*, pp. 1–6, IEEE, 2018.

[23] C. H. Jin, H. J. Na, M. Piao, G. Pok, and K. H. Ryu, "A novel dbscan-based defect pattern detection and classification framework for wafer bin map," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 3, pp. 286–292, 2019.

[24] A. Zabatani, V. Surazhsky, E. Sperling, S. B. Moshe, O. Menashe, D. H. Silver, Z. Karni, A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Intel® realsense™ sr300 coded light depth camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 10, pp. 2333–2345, 2019.

[25] E. S. Gedraite and M. Hadad, "Investigation on the effect of a gaussian blur in image filtering and segmentation," in *Proceedings ELMAR-2011*, pp. 393–396, IEEE, 2011.

[26] M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation," *Journal of Big Data*, vol. 6, no. 1, pp. 1–13, 2019.