Rethinking Automated Visual Quality Control: A Systematic Benchmark of State-of-the-Art Anomaly Detection Methods for Assembled PCB Inspection

Dário Alef Barros Lima[1], Karen Letícia Santana da Silva[1], Willian do Nascimento Severiano[1], Antonio Fontenele[1], Wington L. Vital[1], Hendrio Bragança[1], Ítalo P. Caliari[1]

[1]Núcleo de Capacitação em Inteligência Artificial (NCIA), Fundação Paulo Feitoza - FPFtech, Av. Danilo de Matos Areosa, 1170, Distrito Industrial, Manaus, Amazonas, 69075-351, Brazil.

{dario.220714, karen.252324, willian.252319, antonio.fontenele, wington.vital, hendrio.braganca, italo.caliari}@fpf.br

Abstract. *This work benchmarks four state-of-the-art unsupervised anomaly detection methods CFA, DFM, FRE, and PatchCore for automated PCB inspection under a strict one-class learning protocol. All models were trained exclusively on nominal samples and evaluated using default configurations to reflect realistic industrial deployment. Experiments on the MVTec AD Transistor dataset and a high-density Raspberry Pi PCB dataset assessed image- and pixel-level performance using AUROC, F1-Score, and AUPRO. PatchCore achieved the strongest image-level discrimination, CFA provided the most balanced detection–localization trade-off, FRE showed sensitivity to localized structural defects, and DFM performed well globally but with weaker boundary precision. Results demonstrate the practical viability of unsupervised deep anomaly detection for modern AOI systems in dynamic manufacturing environments.*

1. Introduction

Driven by the widespread adoption of advanced electronic technologies, including smartphones, autonomous and connected vehicles, industrial automation systems, and IoT-enabled smart devices, the global Printed Circuit Boards (PCB) market surpassed USD 75 billion in 2023 and is projected to maintain sustained growth in the coming years [SNS Insider Pvt Ltd 2025]. In this context, progressive miniaturization enabled by Surface-Mount Device (SMD) technology, coupled with increasing integration density, has rendered quality assurance a critical component of electronics manufacturing. In this scenario, even sub-millimetric soldering defects or minor component misplacements may compromise electrical connectivity, signal integrity, thermal behavior, or mechanical reliability, potentially leading to system-level failures and significant economic losses.

Quality control in electronic assembly lines has traditionally relied on Automated Optical Inspection (AOI) systems. While AOI substantially improves inspection speed, repeatability, and consistency relative to manual visual inspection, most industrial deployments remain predominantly dependent on handcrafted feature engineering, deterministic thresholding strategies, or supervised classifiers trained on predefined defect categories. Consequently, system performance is often tightly coupled to prior knowledge of defect taxonomies and to the availability of representative labeled samples.

These characteristics introduce structural limitations in highly dynamic Industry 4.0 manufacturing environments [Wang et al. 2018]. The need to explicitly model each potential defect type requires extensive parameter tuning, frequent recalibration, and continuous dataset curation. In high-mix, low-volume production scenarios, which are common in modern electronics manufacturing, such requirements reduce scalability and hinder the reliable detection of previously unseen anomalies, referred to as zero-day defects [Pang et al. 2021].

In industrial practice, defective samples are typically scarce and highly imbalanced relative to nominal production data. Under these conditions, unsupervised and semi-supervised learning architectures, particularly within the One-Class Classification framework, have emerged as promising alternatives [Ruff et al. 2021]. Unlike supervised approaches, these models are trained exclusively on nominal, defect-free samples to learn a representation of the underlying data distribution. During inference, statistically significant deviations from this learned distribution are interpreted as anomalies [Bergmann et al. 2019, Li et al. 2021].

Within this context, the present study performs a systematic comparative evaluation of five state-of-the-art unsupervised anomaly detection algorithms applied to assembled PCB inspection. The primary objective is to assess the technical feasibility of replacing conventional AOI decision logic with data-driven unsupervised models that operate without requiring annotated defect samples or prior exposure to damaged components.

The main contributions of this work are threefold: (i) A fair and reproducible benchmark of state-of-the-art unsupervised anomaly detection methods applied to assembled PCB inspection under strictly nominal-only training conditions; (ii) A cross-scenario evaluation covering both controlled single-component defects and high-density multi-component PCB assemblies; (iii) An industrially grounded analysis of image-level versus region-level performance trade-offs, providing deployment-oriented guidance for AOI system modernization.

2. Methodology

This section details the datasets, algorithms, and evaluation metrics used to assess unsupervised visual anomaly detection for PCB-level inspection. Experiments use two complementary datasets to capture both low- and high-complexity inspection scenarios: the **Transistor** category from the MVTec AD dataset [Bergmann et al. 2019] and a custom dataset for **Raspberry Pi** boards.

2.1. Datasets

The **Transistor** dataset corresponds to the official MVTec AD Transistor category and includes 213 nominal (good) images and four defect classes, namely *bent lead*, *cut lead*, *damaged case*, and *misplaced*, each containing 10 anomalous images with corresponding ground-truth segmentation masks, totaling 40 defective samples. Each image depicts a single transistor mounted on a through-hole PCB, captured under controlled acquisition conditions. The dataset therefore represents a low-complexity inspection scenario centered on a single through-hole component, where defects primarily affect the leads or package integrity. The availability of precise pixel-level annotations enables accurate evaluation of localization performance in a controlled, single-object setting.

The **Raspberry Pi** dataset consists of 99 nominal images and 20 anomalous images, each accompanied by a pixel-level segmentation mask. It comprises high-resolution images of a densely populated Raspberry Pi PCB, representative of complex surface-mount device (SMD) assemblies. Due to the limited availability of naturally occurring defective samples in industrial environments, anomalies were synthetically generated by digitally removing and repositioning components to simulate misplacements, short circuits, printing defects, and component failures; with masks created simultaneously to ensure exact spatial correspondence. While the Transistor dataset evaluates real component-level defects, this dataset represents a high-density PCB inspection scenario characterized by increased structural complexity and controlled defect induction.

2.2. Algorithms

This section presents a set of unsupervised visual anomaly detection algorithms evaluated in this work. Training for every method uses only defect-free (nominal) samples so the learned model captures the distribution of healthy appearances; during inference, deviations from this distribution are interpreted as anomalies and converted into pixel- or region-level anomaly scores to produce spatial heatmaps.

2.2.1. Coupled-Hypersphere-based Feature Adaptation (CFA)

Coupled-Hypersphere-based Feature Adaptation (CFA) is an unsupervised anomaly detection algorithm that integrates a pretrained convolutional neural network (CNN) backbone with a hypersphere-based feature adaptation module. The CNN first extracts discriminative feature maps from input images, which are subsequently projected into a domain-specific latent space through a trainable adaptation network. Within this space, feature embeddings are organized around multiple coupled hypersphere centers that model the distribution of nominal patterns. As training is conducted exclusively on defect-free samples, the model learns compact representations constrained within the radii of these hyperspheres. During inference, anomalous regions produce embeddings that deviate from the learned centers, resulting in increased latent-space distances. These distances are interpreted as pixel- or region-level anomaly scores, enabling the generation of spatial heatmaps for defect localization [Lee et al. 2022].

2.2.2. Deep Feature Modeling (DFM)

Deep Feature Modeling (DFM) combines a pretrained CNN backbone with statistical modeling based on Principal Component Analysis (PCA) and multivariate Gaussian density estimation. The CNN extracts deep feature representations capturing structural and semantic information from the input images. These features are projected into a lower-dimensional subspace using PCA to reduce redundancy and retain the principal variance directions. In the reduced space, nominal feature distributions are modeled as a multivariate Gaussian whose parameters are estimated solely from defect-free samples. During inference, anomalous regions exhibit low likelihood under the learned Gaussian model or significant deviations within the PCA subspace, often reflected as elevated reconstruction errors. These deviations are used to compute spatial anomaly scores and generate pixel-level heatmaps for defect localization [Ahuja et al. 2019].

2.2.3. Feature Reconstruction Error (FRE)

Feature Reconstruction Error (FRE) combines a pretrained CNN backbone with a tied-weight autoencoder operating in feature space. The CNN extracts deep feature maps encoding structural characteristics of the input images, which are then projected into a lower-dimensional latent space through the encoder and reconstructed via a decoder whose weights are defined as the transpose of the encoder weights. Training is performed exclusively on nominal samples, enabling the model to learn a reconstruction function that preserves normal feature representations with high fidelity. During inference, anomalous regions yield elevated reconstruction errors due to their deviation from the learned feature distribution. The discrepancy between original and reconstructed feature maps is quantified to produce spatial anomaly scores, which are used to generate heatmaps for defect localization [Batzner et al. 2022].

2.2.4. PatchCore

PatchCore is an unsupervised anomaly detection method that combines a pretrained CNN backbone with a memory bank–based nearest-neighbor retrieval mechanism. The CNN extracts localized feature embeddings organized as spatial patches, encoding fine-grained structural patterns of the input image. During training, patch-level features from defect-free samples are stored in a memory bank after a coreset selection procedure that reduces redundancy while preserving coverage of the nominal feature space. Because training relies exclusively on normal data, the memory bank implicitly represents the distribution of healthy patterns. At inference time, each patch from a test image is compared to stored embeddings via nearest-neighbor search in feature space. Patches lacking close correspondences produce higher feature-space distances and consequently higher anomaly scores. These distances are spatially aggregated to generate interpretable heatmaps highlighting regions with increased likelihood of defects [Roth et al. 2022].

2.3. Evaluation Metrics

To evaluate anomaly detection performance, we consider both **image-level** and **pixel-level** metrics. Image-level metrics assess whether an entire PCB image is correctly classified as normal or defective. In contrast, pixel-level metrics evaluate how accurately the model localizes defective regions within an image. Let $y_i \in \{0, 1\}$ denote the ground-truth label of image i, where 1 represents a defective image and 0 a nominal one. Let $s_i \in \mathbb{R}$ denote the anomaly score predicted for that image. At the pixel level, let y_{ij} and s_{ij} represent the ground-truth label and anomaly score for pixel j in image i.

2.3.1. Image AUROC

The Area Under the Receiver Operating Characteristic Curve (AUROC) measures the probability that a randomly selected defective sample receives a higher anomaly score than a randomly selected nominal sample [Fawcett 2006]. Formally,

$$\text{AUROC} = \mathbb{P}(s^+ > s^-),$$

where s^+ and s^- denote anomaly scores from positive (defective) and negative (normal) samples, respectively.

Equivalently, AUROC corresponds to the area under the ROC curve defined by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) for varying thresholds τ:

$$\text{TPR}(\tau) = \frac{TP(\tau)}{TP(\tau) + FN(\tau)}, \qquad \text{FPR}(\tau) = \frac{FP(\tau)}{FP(\tau) + TN(\tau)}.$$

2.3.2. Image F1-Score

Given a fixed decision threshold τ, predictions are defined as $\hat{y}_i = \mathbb{I}(s_i \geq \tau)$, where $\mathbb{I}(\cdot)$ is the indicator function. Precision and recall are defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN}.$$

The F1-Score is the harmonic mean of precision and recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

This metric balances false positives and false negatives under a chosen operating threshold [Sasaki et al. 2007].

2.3.3. Pixel AUROC

Pixel AUROC extends the ROC analysis to the pixel level using anomaly scores s_{ij} and ground-truth labels y_{ij}. It measures:

$$\text{Pixel AUROC} = \mathbb{P}(s_{ij}^+ > s_{ij}^-),$$

where s_{ij}^+ and s_{ij}^- denote scores from anomalous and normal pixels, respectively. As with image-level AUROC, it corresponds to the area under the TPR–FPR curve computed over all pixels [Fawcett 2006].

2.3.4. Pixel F1-Score

At a fixed threshold τ, pixel-level predictions are given by:

$$\hat{y}_{ij} = \mathbb{I}(s_{ij} \geq \tau).$$

Pixel precision and recall are computed analogously:

$$\text{Precision}_{\text{pixel}} = \frac{TP_{\text{pixel}}}{TP_{\text{pixel}} + FP_{\text{pixel}}}, \qquad \text{Recall}_{\text{pixel}} = \frac{TP_{\text{pixel}}}{TP_{\text{pixel}} + FN_{\text{pixel}}}.$$

The Pixel F1-Score [Sasaki et al. 2007] is:

$$\text{F1-Score}_{\text{pixel}} = 2 \cdot \frac{\text{Precision}_{\text{pixel}} \cdot \text{Recall}_{\text{pixel}}}{\text{Precision}_{\text{pixel}} + \text{Recall}_{\text{pixel}}}.$$

2.3.5. Pixel AUPRO

The Area Under the Per-Region Overlap curve (AUPRO) evaluates region-wise localization performance. Let $\mathcal{R}_k$ denote the k-th ground-truth defect region and $\hat{\mathcal{R}}_k(\tau)$ the predicted region at threshold τ. The Per-Region Overlap (PRO) is defined as:

$$\text{PRO}_k(\tau) = \frac{|\mathcal{R}_k \cap \hat{\mathcal{R}}_k(\tau)|}{|\mathcal{R}_k|}.$$

The PRO score is averaged across all ground-truth regions and plotted as a function of the false positive rate at pixel level. AUPRO corresponds to the area under this curve:

$$\text{AUPRO} = \int_0^{\alpha} \text{PRO}(\text{FPR})\, d(\text{FPR}),$$

where α is a predefined upper bound on the false positive rate (commonly $\alpha = 0.3$). Unlike purely pixel-wise metrics, AUPRO assigns equal importance to each defect region, preventing large defects from dominating the evaluation. It is particularly sensitive to small-area defects [Bergmann et al. 2019].

2.4. Experimental Setup

Both datasets were standardized to enable direct integration with the *Anomalib v2.2.0* library [Akcay et al. 2022], strictly adopting its official implementation as the execution framework for all evaluated algorithms. All models were instantiated, trained, and evaluated in full compliance with the library's configurations and protocols using *Python v3.12*.

All algorithms were evaluated exclusively under their **default configurations**, without performing dataset-specific hyperparameter tuning, architectural modifications, or domain-adaptive optimization; see the library official documentation for details. This methodological choice was deliberately made to assess the intrinsic generalization capability and true *out-of-the-box* performance of each approach. This evaluation protocol intentionally prioritizes external validity over optimized peak performance. In real industrial settings, anomaly detection systems must operate reliably with minimal hyperparameter tuning, limited expert intervention, and restricted availability of defective samples. Therefore, assessing intrinsic robustness under default configurations provides a more realistic estimate of practical deployability. To enhance model robustness and mitigate overfitting, data augmentation was performed on training data through random photometric transformations using brightness (0.5), contrast (0.25), and saturation (0.25) random variations, while hue was kept unchanged. All experiments were conducted over five independent runs to account for stochastic variability. Performance metrics are reported as the mean ± standard deviation across these repetitions.

The combination of default hyperparameter settings and strictly nominal-only training also establishes a *fair benchmarking* framework across heterogeneous methods. By avoiding asymmetric optimization efforts, this strategy ensures that performance differences can be attributed to the fundamental modeling principles of each algorithm rather than to varying degrees of experimental fine-tuning.

All experiments were conducted on a workstation equipped with an Intel® Core™ Ultra 9 185H processor (22 threads), an NVIDIA RTX 2000 Ada Generation GPU, 64.0 GiB of RAM, and running Ubuntu 24.04.4 LTS (64-bit).

3. Results and Discussion

The results of the metric calculations for the Raspberry Pi and Transistor datasets test sets are presented in Tables 1 and 2, respectively.

Table 1. Comparison of anomaly detection metrics for different algorithms on the Raspberry Pi dataset

Method	Image AUROC	Image F1-Score	Pixel AUROC	Pixel F1-Score	Pixel AUPRO
CFA	0.9375±0.000	0.8411±0.0011	0.9964±0.0016	0.6385±0.0019	0.9156±0.0023
DFM	1.000±0.000	0.8373±0.1328	0.9569±0.0007	0.2195±0.0432	0.5627±0.0232
FRE	0.8750±0.2119	0.4889±0.2434	0.9695±0.0014	0.4560±0.2021	0.5779±0.0074
PatchCore	0.9688±0.0541	0.9333±0.000	0.9984±0.0005	0.6070±0.0075	0.9478±0.0231

Table 2. Comparison of anomaly detection metrics for different algorithms on the Transistor Dataset

Method	Image AUROC	Image F1-Score	Pixel AUROC	Pixel F1-Score	Pixel AUPRO
CFA	0.9779±0.0018	0.8929±0.0181	0.9729±0.0004	0.6339±0.0035	0.9367±0.0005
DFM	0.9750±0.0024	0.7545±0.0851	0.9798±0.0003	0.6175±0.0135	0.7817±0.0015
FRE	0.9594±0.0010	0.7284±0.0543	0.9875±0.0004	0.6869±0.0059	0.9112±0.0023
PatchCore	0.9974±0.0008	0.9588±0.0144	0.9730±0.0001	0.5999±0.0036	0.9417±0.0013

The Raspberry Pi dataset represents a structurally complex inspection environment characterized by a high density of surface-mount components and synthetically induced defects distributed across the PCB. In this context, Pixel AUPRO is the most informative metric once it evaluates region-wise localization while controlling the false positive rate and preventing large defective areas from dominating the evaluation. Under this criterion, PatchCore achieved the highest Pixel AUPRO (0.9478±0.0231), closely followed by CFA (0.9156±0.0023). Both methods also presented excellent pixel-level discrimination, with Pixel AUROC values above 0.996, indicating that deep feature embeddings extracted from WideResNet-50 produce highly separable representations even without dataset-specific tuning.

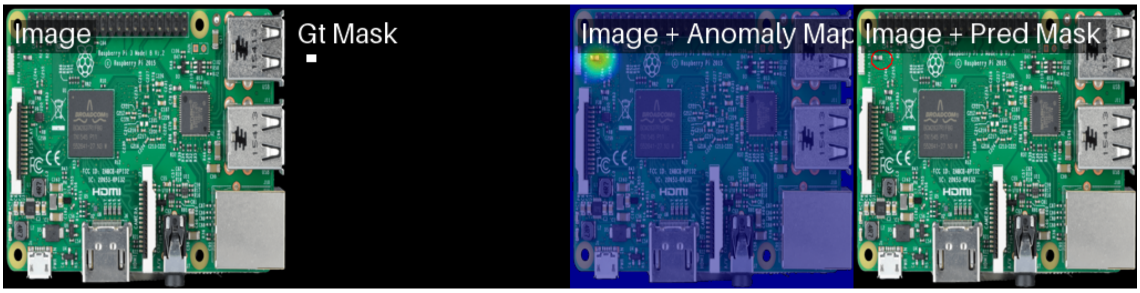

Figure 1. PatchCore Anomaly Analysis of a Faulty Sample from the Raspberry Pi Test Set: misplaced resistor located in the upper-left region of the board.

Although DFM achieved perfect Image AUROC (1.000±0.000), its Pixel AUPRO (0.5627±0.0232) and Pixel F1-Score (0.2195±0.0432) were substantially lower. This discrepancy highlights the distinction between global anomaly scoring and fine-grained spatial localization. The Gaussian modeling strategy employed by DFM appears effective for holistic image-level separation but less capable of precisely delineating defect boundaries in complex multi-component boards. FRE, in turn, showed instability in image-level performance, reflected by higher standard deviations in Image AUROC and F1-Score, which may be attributed to reconstruction sensitivity when confronted with synthetic component removal in cluttered layouts. Overall, the Raspberry Pi results indicate that methods based on local patch comparison (PatchCore) or structured latent constraints (CFA) generalize more robustly to high-density PCB inspection scenarios, whereas purely statistical latent modeling (DFM) may suffer from limited spatial precision when defect morphology diverges from global distribution assumptions.

Figures 1–3 present representative anomalous samples from the Raspberry Pi test set processed using the PatchCore algorithm. Each figure contains four elements: (i) the original input image, (ii) the corresponding ground-truth anomaly mask, (iii) the predicted anomaly heatmap generated by the model, and (iv) the final detection result with the anomalous region highlighted by a red circle. This structured visualization enables direct qualitative comparison between ground-truth defect regions and the model's pixel-level responses, facilitating assessment of detection confidence, localization accuracy, and boundary sharpness. Figure 1 illustrates a misplaced resistor located in the upper-left region of the board. The heatmap shows a concentrated high-response area aligned with the ground-truth mask, indicating strong localization performance even with small area defects. Figure 2 presents a rotated microcontroller in the central-left area. Despite the structural complexity of the component, the model successfully identifies the orientation anomaly, demonstrating robustness to geometric deviations. Finally, Figure 3 shows another misplaced resistor in the lower-central area. The detection remains spatially consistent, reinforcing the method's capability to handle repeated defect types across varying board locations. Collectively, these qualitative examples corroborate the quantitative findings, demonstrating that PatchCore maintains both high image-level discrimination and reliable pixel-level localization in complex PCB inspection settings.

In the Transistor dataset, which represents a controlled single-component inspection scenario with real physical defects, performance across methods was more homogeneous. All models achieved Image AUROC values above 0.95, confirming that one-class training on nominal samples is sufficient to detect clear structural abnormalities in low-

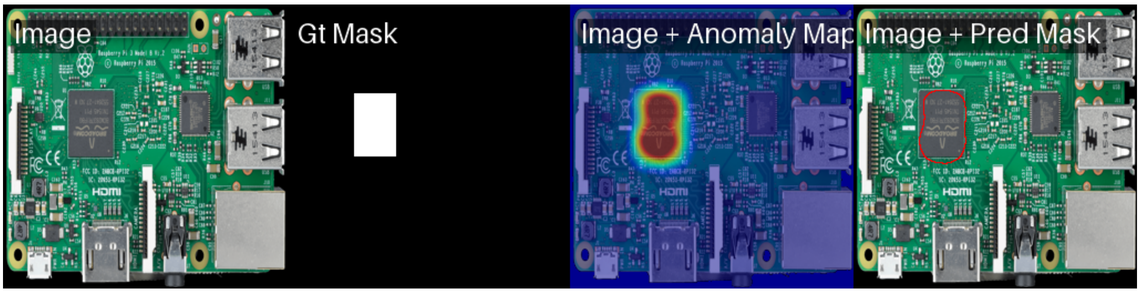

Figure 2. PatchCore Anomaly Analysis of a Faulty Sample from the Raspberry Pi Test Set: rotated microcontroller in the central-left region, representing a geometric orientation defect.

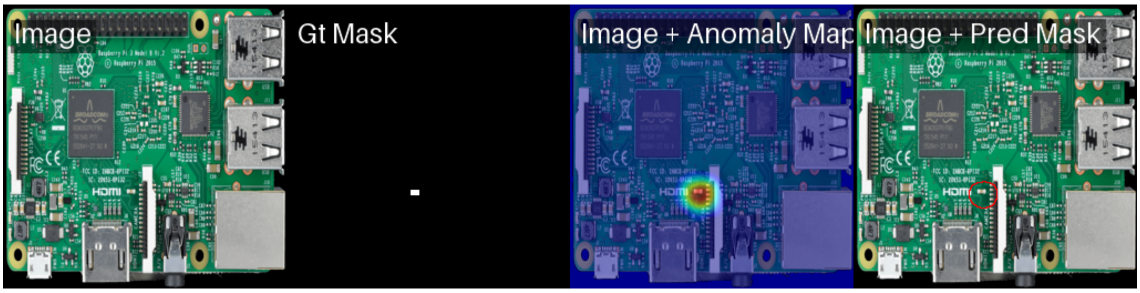

Figure 3. PatchCore Anomaly Analysis of a Faulty Sample from the Raspberry Pi Test Set: misplaced resistor in the lower-central region.

complexity assemblies. PatchCore obtained the best Image AUROC (0.9974±0.0008) and Image F1-Score (0.9588±0.0144), reinforcing its strong global discrimination capability. However, its Pixel F1-Score (0.5999±0.0036) was lower than that of FRE (0.6869±0.0059), despite similar Pixel AUROC values. This suggests that reconstruction-based modeling may better capture subtle lead deformations and localized structural distortions typical of transistor defects. FRE also achieved a competitive Pixel AUPRO (0.9112±0.0023), indicating consistent region-level localization. CFA maintained balanced performance across all metrics, particularly strong Pixel AUPRO (0.9367±0.0005), highlighting the stability of hypersphere-constrained feature adaptation even in simpler inspection tasks. DFM again demonstrated solid pixel discrimination (Pixel AUROC 0.9798±0.0003) but comparatively weaker region consistency (AUPRO 0.7817±0.001), reinforcing the observation that Gaussian-based modeling may lack robustness in boundary delineation.

Figures 4–6 present representative anomalous samples from the Transistor test set processed using the PatchCore algorithm. Consistent with the quantitative evaluation, these examples illustrate the model's strong image-level discrimination alongside its moderately precise, yet not perfectly tight, pixel-level localization. Each figure contains four elements: (i) the original input image, (ii) the corresponding ground-truth anomaly mask, (iii) the predicted anomaly heatmap generated by the model, and (iv) the final detection result with the anomalous region highlighted by a red circle. This structured layout enables direct qualitative comparison between annotated defect regions and the model's spatial response, allowing assessment of localization accuracy, activation concentration, and boundary sharpness. Figure 4 shows a central lead bent to the right. The anomaly

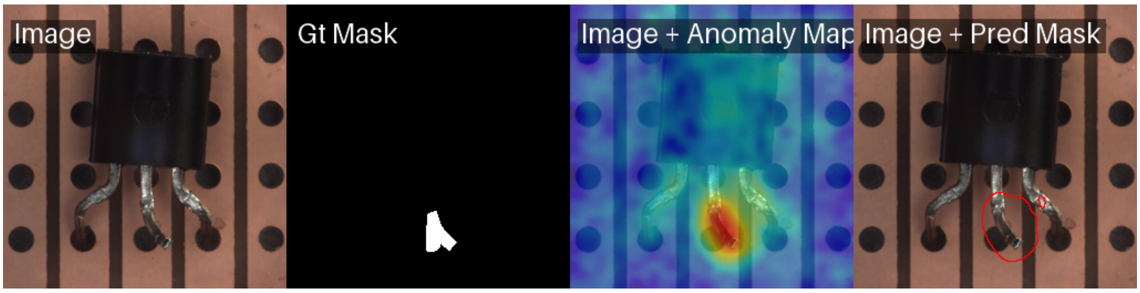

Figure 4. PatchCore Anomaly Analysis of a Faulty Sample from the Transistor Test Set: central lead bent to the right

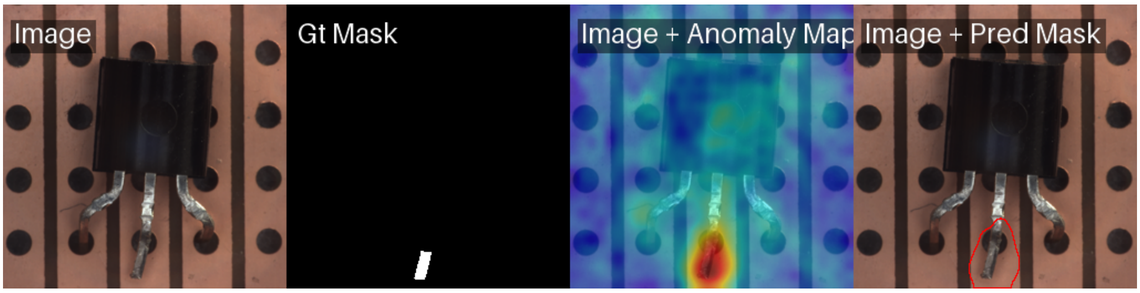

Figure 5. PatchCore Anomaly Analysis of a Faulty Sample from the Transistor Test Set: central lead bent downward

heatmap exhibits a concentrated response along the deformed segment, clearly separating it from the intact neighboring leads. While the activation slightly extends beyond the exact ground-truth boundary, the deformation is unambiguously detected at both image and pixel levels. Figure 5 depicts a central lead bent downward. The model again localizes the geometric distortion effectively, with strongest activation near the curvature region where the deviation from the nominal shape is most pronounced. This demonstrates sensitivity to subtle structural deviations rather than merely gross structural breaks. Figure 6 presents a damaged transistor case. Compared to lead-specific defects, the anomaly response is more spatially distributed, indicating that surface irregularities and casing damage induce broader feature deviations. Nevertheless, the model successfully distinguishes the defective region from the intact housing. Overall, the qualitative results align with the quantitative findings: PatchCore achieves highly reliable global anomaly detection in this controlled single-component scenario, while delivering consistent, though occasionally slightly overextended, pixel-level localization. The examples further confirm that in low-complexity assemblies with real physical defects, structural deviations are sufficiently distinctive to be captured robustly by patch-based feature comparison.

When analyzing both datasets jointly, several consistent patterns emerge. PatchCore consistently achieved the highest image-level discrimination, indicating superior capability in separating nominal and defective samples under default configurations. CFA provided the most balanced trade-off between global detection and region-level localization, particularly in structurally heterogeneous environments. FRE exhibited strong pixel-level sensitivity in simpler defect morphologies, likely because reconstruction error amplifies localized structural deviations. DFM performed well in global anomaly scoring but underperformed in fine-grained localization, especially in complex PCB layouts.

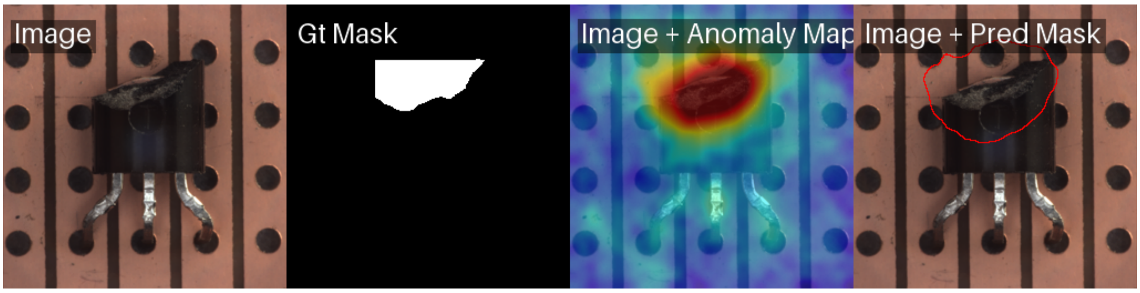

Figure 6. PatchCore Anomaly Analysis of a Faulty Sample from the Transistor Test Set: damaged case

4. Conclusions

This study conducted a systematic benchmark of four state-of-the-art unsupervised anomaly detection methods CFA, DFM, FRE, and PatchCore for assembled PCB inspection under a strict one-class learning protocol and default configurations. Evaluations on the MVTec AD Transistor dataset and a high-density Raspberry Pi PCB dataset demonstrated that modern deep anomaly detection models can achieve robust performance without defect-labeled training data or dataset-specific tuning. PatchCore consistently delivered the strongest image-level discrimination, while CFA provided the most balanced trade-off between global detection and precise region-level localization. FRE showed particular sensitivity to subtle structural deviations in simpler inspection scenarios, whereas DFM, despite strong global scoring, exhibited weaker boundary delineation in complex layouts. Overall, the results indicate that unsupervised deep anomaly detection has reached sufficient maturity to support the replacement or augmentation of rule-based AOI decision logic, especially in dynamic, high-mix manufacturing environments. Future work should focus on evaluating real industrial defect distributions and improving sensitivity to extremely small-scale anomalies.

5. Acknowledgments

This research was conducted with the support of Fundação Amazônica de Amparo à Pesquisa e Desenvolvimento Tecnológico Desembargador Paulo dos Anjos Feitoza (FPFtech) and Foxconn Brasil.

References

Ahuja, N. A., Ndiour, I. J., Kalyanpur, T., and Tickoo, O. (2019). Probabilistic modeling of deep features for out-of-distribution and adversarial detection. *arXiv preprint arXiv:1909.11786*.

Akcay, S., Ameln, D., Vaidya, A., Lakshmanan, B., Ahuja, N., and Genc, U. (2022). Anomalib: A deep learning library for anomaly detection. In *IEEE International Conference on Data Engineering (ICDE)*.

Batzner, K., Heckler, L., König, R., and Steger, C. (2022). Efficientad: Accurate visual anomaly detection at millisecond-level latencies.

Bergmann, P., Fauser, M., Sattlegger, D., and Steger, C. (2019). MVTec AD — A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.

Lee, S., Lee, S., and Song, B. C. (2022). CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access*.

Li, C.-L., Sohn, K., Yoon, J., and Pfister, T. (2021). CutPaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Pang, G., Shen, C., Cao, L., and Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)*, 54(2):1–38.

Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., and Gehler, P. (2022). Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., and Müller, K.-R. (2021). A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5):756–795.

Sasaki, Y. et al. (2007). The truth of the f-measure. *Teach Tutor mater*, 1(5):1–5.

SNS Insider Pvt Ltd (2025). Printed circuit board (pcb) market to worth usd 117.53 billion by 2032. Accessed: 2026-02-12.

Wang, J., Ma, Y., Zhang, L., Gao, R. X., and Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of manufacturing systems*, 48:144–156.