PCBA Standard-to-Real Challenge: Cross-domain VQA for Real-world Manufacturing Inspection

Abstract

Surface Mount Technology (SMT) is a critical stage in Printed Circuit Board Assembly (PCBA) manufacturing, where Automated Optical Inspection (AOI) systems are typically employed to make correct decisions that depend on both component-level semantics and process rules (e.g., package constraints, polarity, and handling), and on low-level texture cues. While industrial anomaly detection benchmarks have driven progress in anomaly localization, and recent visual-text resources and MLLM benchmarks begin to probe multimodal reasoning in inspection settings, there remains a gap in standardized evaluation for *actionable* defect under the significant cross-domain shift from manufacturing standards to real-world PCBA environments. We propose the **PCBA Standard-to-Real Challenge** (Link), a new grand challenge that evaluates multimodal models on cross-domain visual question answering grounded in real-world inspection imagery. The challenge covers: (i) perception-centric recognition and detection (component identity, mount-side, defect presence, defect type), (ii) quantitative reasoning (component/pin counting), and (iii) defect cause & handling decisions. Our goal is to accelerate research toward industrially deployable multimodal systems that can generalize from design principals and manufacturing standards, across product variations and inspection conditions, while producing interpretable, actionable answers aligned with manufacturing practice. The challenge provides public development data, an official evaluation toolkit with robust answer normalization, and a hidden test protocol to encourage generalizable and reproducible comparisons.

Keywords

Industrial inspection, domain generalization, visual question answering, multimodal evaluation, manufacturing, PCBA

1 Introduction

Industrial visual inspection is increasingly expected to support decisions rather than only flag anomalies. On real production lines, engineers must localize the affected region, identify the defect mode, understand why it occurs, and determine an appropriate handling action that satisfies process constraints and quality standards. This requirement is particularly pronounced for surface-mount technology printed circuit board assembly, where components are densely packed, defect signatures can be subtle, and acceptability depends on manufacturing knowledge such as package conventions, polarity rules, and rework policies.

While general industrial anomaly detection (AD) benchmarks like MVTec AD [1] and VisA [9] have driven progress in localization, existing PCBA-specific datasets such as DeepPCB [7] and others [4–6] remain limited to simplified settings. These datasets typically feature single-component products on clean backgrounds with rudimentary bounding-box or segmentation annotations. They treat inspection as a pure geometric detection task, completely stripping away the Visual Question Answering (VQA) dimension and the domain knowledge required for professional-grade inspection.

Furthermore, recent multimodal benchmarks like MMAD [3] and MANTA [2] begin to probe VLM reasoning, yet they lack the *Standard-to-Real* cross-domain perspective. Factory decision-making routinely relies on two complementary knowledge sources that are rarely modeled jointly. The first is *standards-based* knowledge, where canonical illustrations and normative text specify what is acceptable, how defects are defined, and how they should be handled. The second is *real-world* data, where images captured on production lines exhibit high diversity in boards, imaging setups, and operating conditions, while explicit textual descriptions of root causes and handling are often absent. As a result, a model that performs well on one source can fail under deployment due to domain shift in visual appearance and a mismatch in available textual supervision.

We propose the *PCBA Standard-to-Real Challenge* for industrial domain generalization in SMT-PCBA inspection. The benchmark is organized into two challenge suites that reflect distinct generalization objectives. One suite emphasizes transferring standards-based anomaly knowledge to real production imagery, focusing on factual anomaly recognition and actionable cause and handling questions. The other suite emphasizes robustness within real production imagery, where visual diversity is high and spatial reasoning is required for tasks such as counting and relational queries. To discourage answer memorization and promote robust reasoning, the hidden evaluation also includes a small probe set drawn from general-domain VQA benchmarks and hallucination diagnostics.

2 Related Work and SOTA Context

Progress in industrial anomaly understanding has transitioned from unsupervised localization to multimodal interpretation. Early benchmarks like MVTec AD [1] established the foundation for anomaly detection. Subsequent works like VisA [9] and Real-IAD [8] introduced more complex structures and multi-view captures. But they are increasingly saturated for some settings and do not directly measure cross-domain reasoning about defects.

PCBA datasets like DeepPCB [7], PCB-SAID [5], and FPCB-related benchmarks [4] simplify the problem into a standard object detection or template matching task. These resources often focus on isolated components or simplified board layouts, providing only spatial labels (bounding boxes). They overlook the **semantic complexity** of PCBA manufacturing—such as why a particular solder joint is "insufficient" or how a "flipped" component necessitates specific rework protocols.

Most recently, the field has pivoted toward Vision-Language Models (VLMs) with benchmarks such as MMAD [3] and MANTA [2], which evaluate the descriptive and reasoning capabilities of Large Language Models in industrial contexts. MMAD introduces a comprehensive evaluation suite for multimodal large language models

in industrial anomaly detection, while MANTA provides large-scale multi-view and visual-text anomaly detection for tiny objects. Challenge-style evaluations such as VAND explicitly demonstrate that connecting academia and industry via objective competitions can accelerate progress and reveal deployment-critical bottlenecks.

Despite these advances, two gaps remain under-addressed for industrial inspection: (1) **Actionability:** existing benchmarks rarely evaluate whether a model can provide actionable defect interpretation beyond "anomalous vs normal," such as defect naming, likely causes, and handling actions; (2) **Knowledge grounding:** many inspection decisions require manufacturing semantics (package types, polarity, assembly rules), which are not directly stressed by typical anomaly scoring tasks. Our challenge targets these gaps by framing SMT-PCBA inspection as *cross-domain visual QA*, where success requires require models to demonstrate both fine-grained perception and domain-expert reasoning, moving beyond simple bounding-box detection toward true industrial intelligence.

3 Challenge Overview

We propose a grand challenge: **PCBA Standard-to-Real Challenge**. The benchmark comprises multiple question families that mirror the end-to-end quality assurance process in SMT-PCBA manufacturing, covering visual inspection, defect identification, and standards-aligned decisions for corrective action.

3.1 Dataset Overview

The challenge dataset is bifurcated into two complementary corpora designed to bridge the gap between theoretical manufacturing standards and practical execution. Table 1 summarizes the task taxonomy, while Figure 1 illustrates the instantiation of VQA items.

Standard vs. Real-world Corpora. The *Standard* corpus is grounded in the IPC-A-610 standard, encompassing over 40 distinct defect types to provide a comprehensive knowledge base of theoretical anomalies. In contrast, the *Real-world* corpus consists of high-resolution imagery captured from high-volume production lines. To ensure a focused evaluation for the inaugural challenge, we release five prevalent and actionable defect categories, namely *Missing Component*, *Insufficient Solder*, *Standing*, *Flipped Component*, and *Wrong Polarity*. Unlike existing datasets with controlled setups, our real-world samples are cropped from full motherboard assemblies. This approach introduces significant intra-class variation, such as diverse footprints and markings, alongside environmental complexity including dense layouts, solder mask reflections, and background clutter. These factors directly reflect the challenging "Standard-to-Real" domain shift.

Task Taxonomy. We define four core question families that evaluate both perception-level factuality and high-level actionable reasoning:

(1) **Cause & Handling:** This category utilizes multiple-choice questions to investigate the root cause of defects and determine recommended follow-up actions. Such decisions, including the choice between rework and scrap, are strictly aligned with industrial protocols.
(2) **Factuality:** These perception-centric tasks cover component recognition, mounting-side identification, and the detection of defect presence as well as defect types.

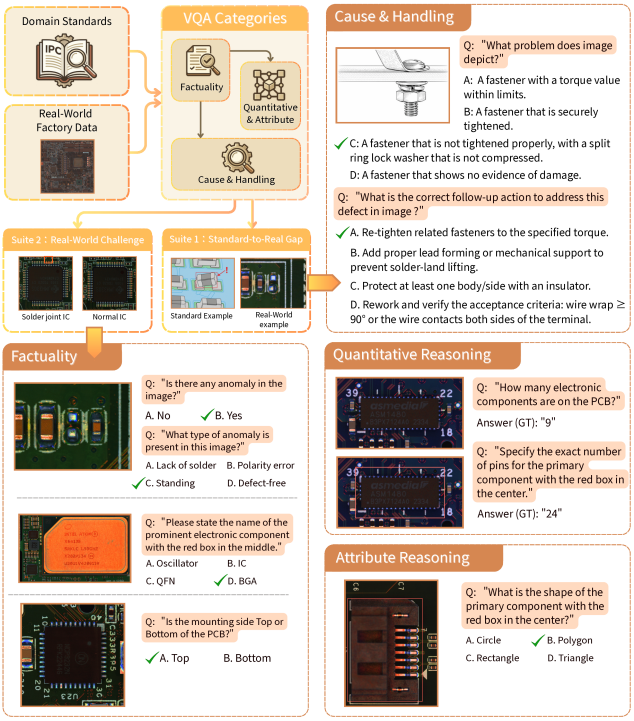

Figure 1: VQA example pipeline and representative questions spanning the two suites and categories.

(3) **Quantitative Reasoning:** These items consist of numeric short-answer questions designed to evaluate the precision of component and pin counting.
(4) **Attribute Reasoning:** This family focuses on the recognition of shapes and package-level attributes, serving as a foundational step toward deeper physical property understanding.

Input Modalities. The input configuration is tailored to the nature of the inspection task. *Single-image VQA* serves as the default format for most factuality and attribute items. For logical defects requiring a golden-sample comparison, specifically for categories such as *Missing* or *Wrong Polarity*, we provide a template-based setup consisting of a *reference normal* image alongside the *target* image. Furthermore, *Cause & Handling* items may include up to three multi-angle views to provide sufficient visual evidence for reasoning about damage severity.

Scope and Roadmap. To ensure a well-controlled evaluation for the first year, we withhold fine-grained material texture judgments and complex spatial reasoning tasks. These dimensions are integral parts of our dataset roadmap and will be introduced in subsequent challenge editions after the completion of further governance and validation checks.

3.2 Challenge Suite

We define two challenge suites that differ in corpus composition and target generalization direction. Both suites are built from two corpus sources: *Standard* Derived from standards, with lower visual diversity but containing causal and processing knowledge. And

Table 1: Task taxonomy and per-family evaluation metrics for the public dataset in this challenge. The public dataset contains 16,000 questions in total.

Category	Sub-types	Qty	Type	Metric
Defect Cause & Handling	Defect Cause & Handling	2000	MCQ	Acc
Perception-Level Factuality	Recognition – component type	2000	MCQ	Acc
	Recognition – mount side	2000	MCQ	Acc
	Detection – defect existence	2000	MCQ	Acc, F1
	Recognition – defect type	2000	MCQ	Acc
Quantitative Reasoning	Counting – component quantity	2000	Numeric answer	Acc, MAE
	Counting – pin quantity	2000	Numeric answer	Acc, MAE
Attribute Reasoning	Main component shape	2000	MCQ	Acc

Table 2: Challenge suites are built from two corpus sources and target complementary capabilities.

Corpus Source	Factuality (F)	Spatial Reasoning (SR)	Cause & Handling (C&H)
Standard	Yes / Low diversity	Low amount	Yes
Real-world	Yes / High diversity	Yes / No diversity	No

Real World, from actual factory inspection data, with higher visual diversity.

Suite 1: Standards → Real-world. This suite emphasizes transfer from standards-derived knowledge to factory inspection imagery. The suite is designed to stress-test whether models trained/conditioned on standardized defect knowledge can remain reliable under real-world visual variability. We target primarily F and C&H questions primarily, with a Standard: Real-world composition of **3:1**.

Suite 2: Real-world → Real-world. This suite evaluates generalization across real factory distributions (e.g., new boards, capture pipelines, or production lines), reflecting practical deployment where only limited standard-like supervision is available. It targets F, SR, and C&H, with a Standard:Real-world composition of **1:5**.

3.3 Splits.

Each suite has its own **Train / Test / Hidden** splits with a unified release protocol.

Public data. We will release a public dataset with a Train: Test = 2:1 split. Table 1 summarizes the public dataset scale. Ground-truth annotations/answers are provided for *Train* only. For the public *Test* split, inputs are released, but no ground truth is provided.

Hidden data. We additionally curate a hidden set with roughly 500 questions per question type. We will not release hidden questions and their ground truth. Final evaluation on hidden data is conducted via the official evaluation server submission.

3.4 Metrics

We report task-specific metrics tailored to each question family, and compute an overall score with equal weights. Let N be the number of evaluated items in a question family, y_i the ground-truth answer for item i, and $\hat{y}_i$ the submitted prediction after normalization.

MCQ tasks (recognition; cause & handling). We use Accuracy as the primary metric:

$$\mathrm{Acc} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}[\hat{y}_i = y_i]\,, \tag{1}$$

where $\mathbb{I}[\cdot]$ is the indicator function. We apply answer normalization and synonym mapping before matching whenever a model outputs free-form text rather than an option index.

Binary defect existence (yes/no). To account for potential class imbalance, we report both Accuracy and F1-score. Let TP, FP, and FN denote true positives, false positives, and false negatives, respectively. Precision and recall are

$$\mathrm{Prec} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad \mathrm{Rec} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{2}$$

and the F1-score is

$$\mathrm{F1} = \frac{2\,\mathrm{Prec}\,\mathrm{Rec}}{\mathrm{Prec}+\mathrm{Rec}}. \tag{3}$$

Numeric short answers (counting). For component and pin counting, we report Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i|\,. \tag{4}$$

We also report accuracy to reflect near-correct counts. To combine MAE with other higher-is-better metrics in an equal-weight aggregate, we convert MAE into a bounded score using a family-specific normalization range R:

$$\mathrm{MAE}_{\mathrm{score}} = 1 - \min\left(\frac{\mathrm{MAE}}{R},\, 1\right), \tag{5}$$

Finalize R for component/pin counting from the released public set and publish them in the evaluation toolkit.

Cross-domain robustness. To characterize the standard-to-real generalization gap, we report the accuracy difference between

the standard-derived subset and the real-world subset under Suite 1:

$$\Delta_{\text{S2R}} = \text{Acc}_{\text{standard}} - \text{Acc}_{\text{real}}. \tag{6}$$

We report Δ_{S2R} as a diagnostic statistic and do not apply extra weighting in the overall score.

Overall score. We define a higher-is-better score $S_k \in [0, 1]$ for each question family k. For families with a single metric, $S_k = \text{Acc}$.

For defect existence, we use an equal-weight average of Accuracy and F1:

$$S_{\text{exist}} = \frac{\text{Acc} + \text{F1}}{2}. \tag{7}$$

For counting families, we combine accuracy and a normalized MAE score (both higher-is-better):

$$S_{\text{count}} = \frac{\text{Acc} + \text{MAE}_{\text{score}}}{2}. \tag{8}$$

We may additionally report raw MAE on the leaderboard for analysis, but the ranking is determined by the higher-is-better scores above.

The overall score is the unweighted mean across K families:

$$S_{\text{overall}} = \frac{1}{K} \sum_{k=1}^{K} S_k, \tag{9}$$

so that (i) metrics within a family contribute 1:1 when applicable, and (ii) all families contribute equally.

3.5 Evaluation Protocol

We structure the benchmark into two *evaluation suites* (Section 3.2), each representing a different generalization regime, and follow a *unified evaluation protocol* across suites.

Public evaluation. For each suite, we release a public dataset with a Train:Test = 2:1 split. Ground-truth answers are provided for public *Train* only. Public *Test* inputs are released without ground truth; leaderboard feedback is provided through the official evaluation server.

Hidden evaluation. For each suite, we additionally curate a hidden set used exclusively for final ranking. Hidden questions and their ground truth are not released. This design follows standard multi-phase competition practice, where the final phase evaluates submissions on an unseen test set to reduce overfitting.

Final score. We compute suite-level scores on public Test and hidden sets, denoted by $S^{(s)}_{\text{test}}$ and $S^{(s)}_{\text{hidden}}$ for suite $s \in \{1, 2\}$. We then aggregate across suites with equal weight:

$$S_{\text{test}} = \frac{1}{2} \sum_{s=1}^{2} S^{(s)}_{\text{test}}, \qquad S_{\text{hidden}} = \frac{1}{2} \sum_{s=1}^{2} S^{(s)}_{\text{hidden}}.$$

The final ranking score is:

$$S_{\text{final}} = 0.8 \cdot S_{\text{test}} + 0.2 \cdot S_{\text{hidden}}.$$

Submission. Participants will submit a single JSON prediction file following a documented schema, where each entry is keyed by a unique question identifier. A single submission covers the public *Test* items from both suites. We will release an official evaluation script and a containerized evaluation environment to ensure consistent scoring across submissions. To encourage reproducibility, top-ranked teams will be invited to submit a short technical report and code.

Submission platform. We will host the competition on CodaLab Competitions, which supports multi-phase challenge design, multi-score leaderboards, and Docker-based evaluation for reproducibility. We will configure two phases: a public development phase (leaderboard on public Test) and a final phase with hidden evaluation.

4 Commitment

We commit to operating the PCBA Standard-to-Real Challenge as a multi-year grand challenge, with the goal of establishing a sustained benchmark that remains useful to the multimedia community and industrial stakeholders over the next several years. In line with the ACMMM Grand Challenge expectations, we will (i) maintain a dedicated website with up-to-date task descriptions, dataset documentation, rules, and evaluation resources, (ii) provide rigorously defined objective evaluation procedures, and (iii) work with the conference organizers to publicize the challenge and broaden participation. We will keep the challenge website, dataset access, and evaluation infrastructure available for at least three years after the 2026 edition.

4.1 Award Plan

To recognize high-quality research and encourage participation:

- **Paper Invitation:** The top 5 performing teams on the final leaderboard will be invited to submit a technical paper (Grand Challenge track) to the ACM Multimedia 2026 proceedings, subject to the conference's peer-review requirements.
- **Industrial Recognition:** Top teams will receive official certificates, acknowledging their contribution to advancing industrial AI.

4.2 Outreach and Promotion

We will leverage our multi-partner network to ensure high visibility within the multimedia and AI communities:

- **Academic and Research Network:** Announcements will be released via the website and GitHub.
- **Industry Collaboration:** The organizing company will promote this challenge through their corporate AI research pipeline and industry partners to attract manufacturing professionals.
- **Social Media and Community:** We will remain active on GitHub and the website. This challenge will also be promoted on the official website and the HuggingFace code repository.

4.3 Timeline and Dataset Availability

We strictly align our organization with the ACMMM 2026 Grand Challenge schedule:

- **Preliminary Data:** Sample images and question-answer pairs are currently available for download via the dataset link on our challenge website.
- **Full Dataset Release:** We commit to releasing the complete PCBA Standard-to-Real dataset in early March 2026.
- **Evaluation Phases:** The public development phase and final hidden-test phase will follow the standard ACMMM timeline to ensure a fair competition.

We commit to maintaining the challenge website, dataset access, and evaluation infrastructure for at least three years.

5 Organizer Team

5.1 Support

This grand challenge is jointly organized by academia, industry, and an international research institute:

- **National Taiwan University (NTU):** Responsible for the overall task design, benchmark protocol, and the development of the evaluation toolkit. Led by **Prof. Wen-Huang Cheng**.
  Wen-Huang Cheng is a University Distinguished Chair Professor in the Department of Computer Science and Information Engineering at National Taiwan University and a Visiting Professor at the Korea Advanced Institute of Science and Technology (KAIST). His current research interests include multimedia, computer vision, and machine learning. He has actively participated in international events and played significant leadership roles in prestigious journals, conferences, and professional organizations. These roles include serving as Editor-in-Chief for IEEE CTSoc News on Consumer Technology, Senior Editor for IEEE Consumer Electronics Magazine (CEM), Associate Editor for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) and IEEE Transactions on Multimedia (TMM), General Chair for ACM MMAsia (2023), IEEE ICME (2022), and ACM ICMR (2021), Technical Program Chair for ACM MM (2025), ACM ICMR (2022), IEEE ICME (2020), IEEE VCIP (2018), Chair for IEEE CASS Multimedia Systems and Applications (MSA) technical committee, and governing board member for IAPR. He has received numerous research and service awards, including the NVIDIA Academic Grant Program Award (2025), the 2024 Best Paper Award of IEEE Consumer Electronics Magazine, the Best Paper Award at the 2021 IEEE ICME and the Outstanding Associate Editor Award of IEEE TMM (2021 and 2020, twice). He is an IEEE Fellow, IET Fellow, and ACM Distinguished Member. [Google Scholar] [Email]

- **ASUS:** Providing real-world SMT-PCBA inspection data, domain expertise, and engineering support for deployment-oriented evaluation. Led by **Tai-I Chen**,
  Tai-I Chen [Main Contact] serves as the Deputy Director of the AI Solution Business Unit (AISolution BU) at ASUS. He leads the strategic development and industrialization of AI technologies for manufacturing. His research interests include Industrial Foundation Models and Multi-Modal Spatial Intelligence. With extensive experience in bridging cutting-edge AI research with production-line requirements, he has overseen the deployment of multiple AOI-related AI systems in high-volume electronics manufacturing environments, focusing on robustness and cross-domain scalability. [Email]

- **OFFIS – Institute for Information Technology (Oldenburg, Germany):** Providing technical guidance and international outreach support. Led by Dr. Frank Oppenheimer.
  Frank Oppenheimer is a Director in the Manufacturing division at OFFIS - Institute for Information Technology in Oldenburg, Germany. His work focuses on applied research and technology transfer for smart manufacturing, including data-driven production systems, industrial digitalization, and robust deployment of AI methods under practical constraints such as domain shift, heterogeneous hardware, and process variability. He leads and coordinates collaborative projects with industrial partners, translating research prototypes into reproducible evaluation pipelines and deployable software components for manufacturing environments. Within OFFIS, he also contributes to cross-cutting efforts on distributed computing and communication as foundational infrastructure for scalable industrial analytics. His activities emphasize reliable benchmarking, transparent evaluation protocols, and long-term maintainability of challenge resources to support sustained community progress. [Email]

5.2 Other Organizers

The execution and technical rigor of the challenge are supported by a dedicated team of senior researchers and engineering specialists:

- **Yung-Yu Chuang** is a Distinguished Professor in the Department of Computer Science and Information Engineering at National Taiwan University. His research spans computer vision, computational photography, and computer graphics, with an emphasis on bridging imaging, vision, and physically grounded visual computing. He has undertaken extensive editorial responsibilities in top-tier venues, notably serving as an Associate Editor for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) and IEEE Transactions on Visualization and Computer Graphics (TVCG). Furthermore, he has played sustained leadership roles in the international computer vision and machine learning communities, including serving as Senior Area Chair or Area Chair for major conferences such as NeurIPS, ICCV, CVPR, ECCV, ICLR, ICML, and AAAI. His contributions have been recognized by numerous honors, including the CyberLink/Perfect Chair Professorship and several Best Paper awards at premier multimedia and vision conferences. [Google Scholar] [Email]

- **Ruyi Xu [Main Contact]** is a Ph.D. student in Computer Science. She received her M.S. degree in Computer Science and Computer Engineering from National Taiwan University in 2025. Her research focuses on computer vision and deep learning, with an emphasis on industrial visual inspection, domain-specific learning, and generative AI. [Email]

- **Extended Support Team:** The challenge is further supported by **Luyang Lin** from ASUS IoT, alongside the dedicated research team at National Taiwan University, including **Tsai-Yen Chen**, **Yen-Tzu Chiu**, and **Rong-Sheng Lin**. Their tireless efforts in data curation, quality control, and platform testing have been instrumental in bringing this grand challenge to fruition.

Primary contacts. Tai-I Chen and Ruyi Xu serve as the primary points of contact for participants and the conference organizers. Our Shared Contact email: [Email]

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. 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. 9592–9600.

[2] Lei Fan, Dongdong Fan, Zhiguang Hu, Yiwen Ding, Donglin Di, Kai Yi, Maurice Pagnucco, and Yang Song. 2025. MANTA: A Large-Scale Multi-View and Visual-Text Anomaly Detection Dataset for Tiny Objects. 25518–25527.

[3] Xi Jiang, Jian Li, Hanqiu Deng, Yong Liu, Bin-Bin Gao, Yifeng Zhou, Jialin Li, Chengjie Wang, and Feng Zheng. 2025. MMAD: A Comprehensive Benchmark for Multimodal Large Language Models in Industrial Anomaly Detection.

[4] Shengping Lv, Bin Ouyang, Zhihua Deng, Tairan Liang, Shixin Jiang, Kaibin Zhang, Jianyu Chen, and Zhuohui Li. 2024. A dataset for deep learning based detection of printed circuit board surface defect. *Scientific Data* 11, 1 (2024), 811.

[5] Raffaele Mineo, Amelia Sorrenti, Robin Faro, Gabriele Mineo, Francesco Cancelliere, and Alberto Faro. 2025. PCB-SAID: A Low-Cost Camera-Based Dataset for Few-Shot SMD Assembly Inspection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1351–1357.

[6] Minghui Shen, Yujie Liu, Jing Chen, Kangqi Ye, Heyuan Gao, Jie Che, Qingyang Wang, Hao He, Jian Liu, Yan Wang, et al. 2024. Defect detection of printed circuit board assembly based on YOLOv5. *Scientific Reports* 14, 1 (2024), 19287.

[7] Sanli Tang, Fan He, Xiaolin Huang, and Jie Yang. 2019. Online PCB defect detector on a new PCB defect dataset. *arXiv preprint arXiv:1902.06197* (2019).

[8] Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. 2024. Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22883–22892.

[9] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. 2022. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European conference on computer vision*. Springer, 392–408.