Assignment 3: Custom Vision–Language Model (VLM) Design for Industrial Quality Inspection

Candidate: Bashwaraj Sonkawade

Domain: Computer Vision, Vision–Language Models, Industrial AI

Introduction

This document proposes a custom offline Vision–Language Model (VLM) for PCB quality inspection.

The system answers natural language questions about PCB defects with structured outputs including

locations and confidence scores under strict latency and reliability constraints.

Problem Statement

Design an offline AI system capable of analyzing PCB images and responding to inspector queries

with defect type, bounding box coordinates, and confidence scores within 2 seconds, while avoiding hallucinations.

Model Selection

A BLIP-2-style modular VLM architecture is selected due to its efficiency, fine-tuning flexibility,

and suitability for domain-specific grounding. Generic VLMs such as LLaVA and Qwen-VL are avoided

due to size, latency, and hallucination risks.

Architecture Design

The system consists of a PCB-specific vision encoder, region-level feature extraction using

bounding boxes, a Q-Former fusion module, and a lightweight language decoder producing structured outputs.

Optimization

Inference speed is optimized using quantization (INT8/INT4), LoRA fine-tuning, pruning, and

ONNX/TensorRT deployment to meet sub-2 second latency constraints.

Hallucination Mitigation

Hallucinations are reduced through structured output enforcement, region-grounded decoding, contrastive grounding loss, negative QA training, and confidence calibration.

Training Strategy

A multi-stage pipeline is used: defect detection pretraining, synthetic QA generation from bounding boxes, VLM fine-tuning with LoRA, and hallucination-aware training.

Validation

The system is validated using mAP@0.5 for localization, counting accuracy, hallucination rate, and end-to-end latency measurements.

Conclusion

This design delivers a reliable, explainable, and fast VLM tailored for industrial PCB inspection, prioritizing engineering robustness over generic intelligence.