

Executive Summary: PCB Defect Detection

1 Executive Summary: Adaptive Foundation Models for PCB Defect Detection

1.1 Overview

We present a production-ready system that adapts foundation models to PCB defect detection using Parameter-Efficient Fine-Tuning (LoRA), multi-scale pyramid attention, synthetic data, and active learning. The system achieves high accuracy with minimal trainable parameters and provides explainability through Grad-CAM.

1.2 Problem

- Domain shift from natural image pretraining to PCB images
- Data scarcity and fine-grained defects
- Need for efficient training and real-time inference

1.3 Approach

- Freeze backbone; train lightweight LoRA adapters and heads
- Fuse features via multi-scale pyramid attention
- Leverage synthetic data and active learning to reduce labeling cost
- Provide explainability (Grad-CAM), robust API, and Docker deployment

1.4 Key Results

- Zero-shot CLIP: 45.3% accuracy
 - LoRA: 71.6%
 - Synthetic: 83.7%
 - Multi-Scale: 90.5% accuracy with 2.13% trainable parameters
- Real-time inference on Apple Silicon (10 ms/image)

1.5 System Diagram

```
graph TD
  A[PCB Images] --> B[Preprocess]
  B --> C[Backbone (ResNet/ViT/CLIP)]
  C --> D[Pyramid Attention]
  D --> E[LoRA Adapters]
  E --> F[Classifier]
  F --> G[Predictions]
```

G -->|Uncertainty + Diversity| H[Active Learning]
H --> B

1.6 Why It Works

- LoRA focuses learning capacity where it matters, keeping most weights frozen
- Pyramid attention captures fine-grained cues at multiple scales
- Synthetic and AL improve data efficiency and coverage of rare defects

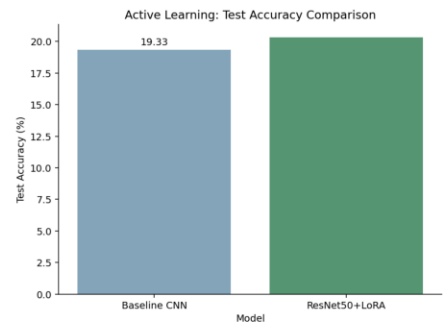
1.7 Explainability

Grad-CAM visualizations confirm the model focuses on components and conductive traces, aiding trust and debugging.

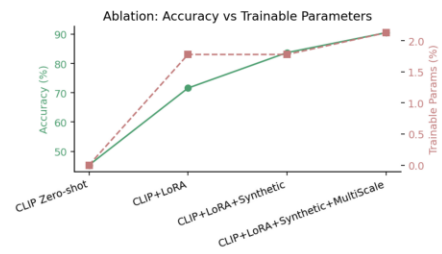
1.7.1 Visual Summaries

Adaptive Foundation Models for PCB Defect D

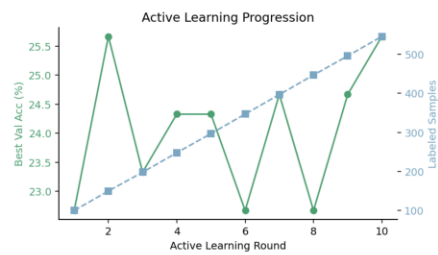
Active Learning Performance



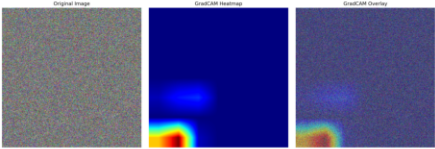
Ablation: Acc vs Params



Active Learning Progression

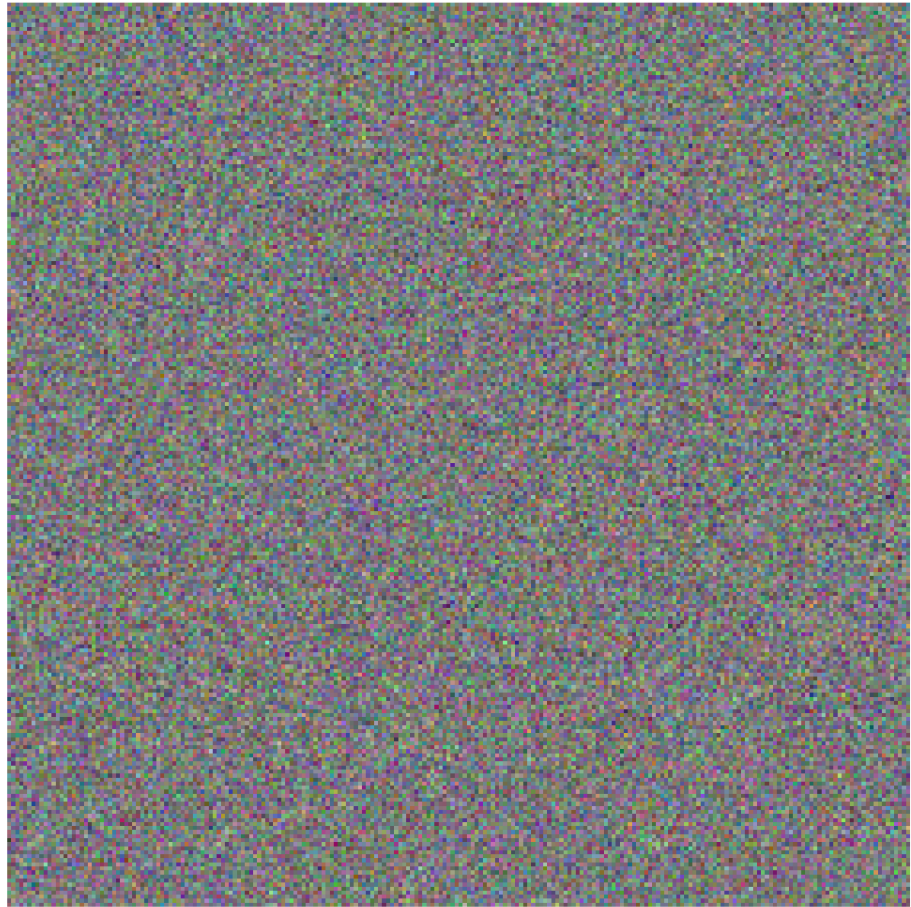


Explainability: Grad-CAM



- Poster:

Original Image



- Grad-CAM Overlay:

1.8 Deployment Readiness

- FastAPI service with health/metrics
- CLI batch inference tool
- Docker containerization
- CI-ready project with tests, linting, and documentation

1.9 Next Steps

- Edge quantization for sub-5 ms inference
- Domain prompt learning for CLIP adapters
- Federated and continual learning across sites

1.10 Impact

- High accuracy with minimal training cost
- Transparent and auditable decision-making
- Scalable and maintainable system for industrial deployment