Optimizing the Vision-Guard AI models to speed up AI training process

SCAN ME

UIIA

Vu Thanh Tung - s3963172
Dinh Xuan Minh - s3891847
Le Quang Dat - s3927251
Nguyen Linh Chi - s3878609
Le Tran Nhat Nam - s3911298

ACADEMIC SUPERVISOR

Mr. Pham Viet Hung

INDUSTRY SUPERVISOR

Mr. Nguyen Minh Chau
Mr. Pham Cong Vo
Mr. Pham Quang Vinh

Background & Motivation

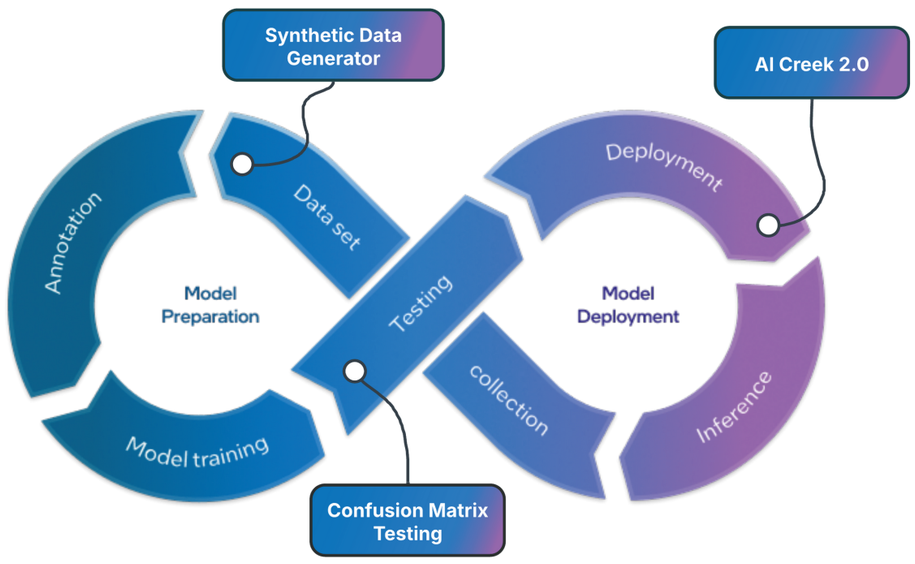
Manual visual inspection in semiconductor manufacturing creates quality risks through inspector fatigue, attention lapses, and subjective assessments. AI-driven inspection eliminates these human factors, providing consistent and objective quality control.

However, AI model development faces bottlenecks in dataset creation and performance validation. Traditional methods require extensive manual image collection, annotation, and time-intensive testing that delays deployment.

Objectives

- Develop an AI-driven defect detection system for Intel's semiconductor production line.
- Improve dataset creation through synthetic generation and active learning to reduce cost and time.
- Train and fine-tune neural networks to achieve ≥95% accuracy while minimizing false positives/negatives.
- Deploy optimized models on edge devices to cut inference costs and ensure real-time performance.
- Build an interactive dashboard to visualize defect rates, confidence scores, and production metrics for decision-making.

Solution Design



Synthetic Data Generator

- Description: Generates additional training images from a small seed dataset.
- Purpose: Automates dataset expansion by replicating original product images and adding realistic defects (e.g., foreign material on pusher, victim scratches). This speeds up dataset preparation and reduces the need for manual photography of every product.

Confusion Matrix Testing

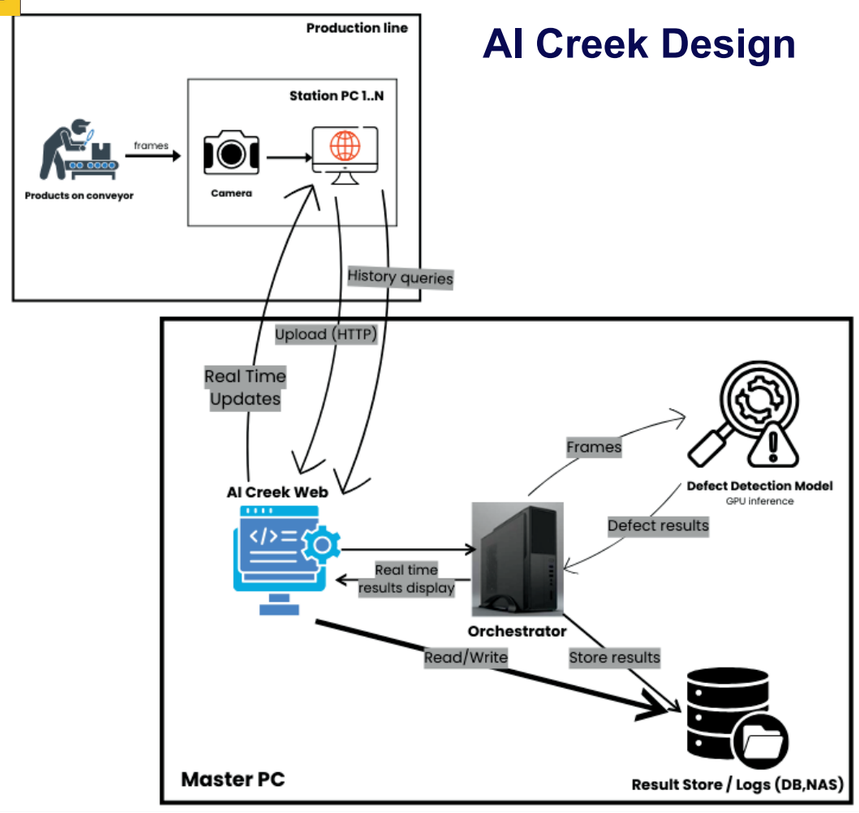
- Description: A validation tool for testing AI models after training.
- Purpose: Instead of manually testing products on Intel Geti one by one, employees can upload a dataset, and the app will automatically display AI predictions against actual product types. Users mark pass or fail, and upon completion, the app exports an Excel file containing images, results, and a confusion matrix score.

AI Creek 2.0

- Description: A station-level web application integrated with the Intel Geti SDK.
- Purpose: Allows employees to capture images directly through the web browser, where AI generates predictions. Operators decide pass or fail, while the dashboard displays key production metrics for each product. Supervisors can access logs, including operator names, IDs, and timestamps, to ensure accountability and traceability.

PostgreSQL · intel GETi · intel GETi SDK · OpenCV · Meta SAM 2 · FastAPI

AI Creek Design

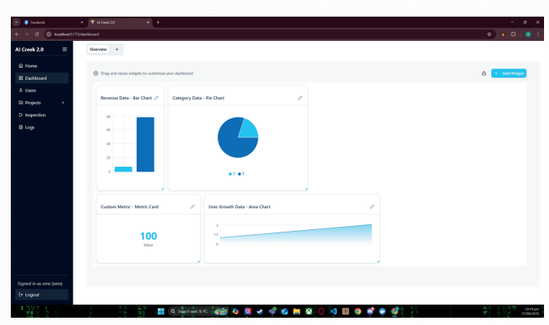


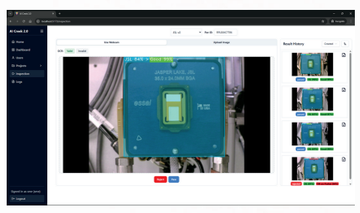
Results

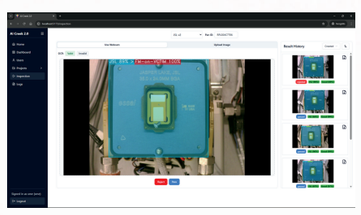
- Developed Dataset Generator Application for dataset creation automation.
- Built AI Creek application for displaying dashboard, showing metrics and graphs from CMMS data.
- Trained and optimized product models, increasing confusion matrix score (Insert specific metrics).

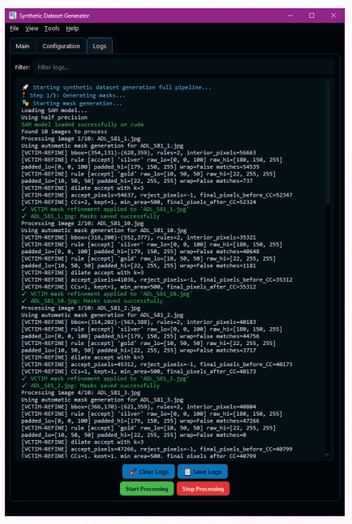
Confusion matrix of JSL products

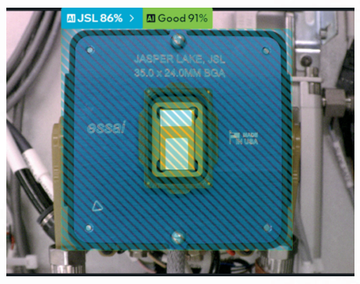
Predicted Reliability	GOOD image	88/100 sample correct (88%)
	BAD image (VCTIM)	46/50 sample correct (92%)
	BAD image (PUSHER)	35/50 sample correct (70%)
	BAD image (Shift/Tear)	26/50 sample correct (52%)
	BAD image (Wrong-Ori)	50/50 sample correct (100%)
	BAD image (Total)	157/200 sample correct (78.5%)

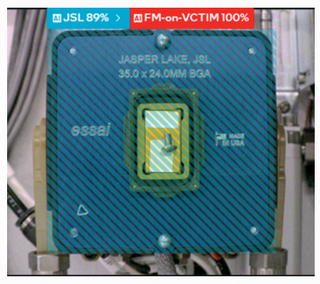












Future works

Future directions:

- Complete fine-tuning for all product models and finalize confusion matrix testing.
- Improve robustness against factory environmental noise (lighting, vibration, dust).
- Deploy the system in Intel's production line for large-scale testing.
- Enhance the dashboard with real-time CMMS data integration.
- Explore hardware upgrades and advanced ensemble methods to further increase reliability.