AUTOMATED OPTICAL INSPECTION IN CHIP MANUFACTURING

Florian Kälber, Frank Karstens

SEPTEMBER 2022
AGENDA

The Problem
Proof of concept
Prototype
Live Demo
The Problem
TRADITIONAL WAVEFORM DEFECT DETECTION METHODS

Templates available: isolate defects
Idea: use difference as labels for supervised learning?

Template

Sample

Difference/ label
MOTIVATION/ OBJECTIVE

Quality inspection

Image data

Training on the cloud

Inference on the edge

Anomaly?
Proof of concept
DATASETS

Public Deep PCB Dataset
• 1500 grayscale samples
• Templates available
• 3-12 defects per sample
• Bounding box annotations

Internal NXP Chip defect Dataset
• 1474 color samples
• Templates available
• 1 centered defect per sample
PREPROCESSING – LABEL GENERATION

Pipeline:
• Image registration (≈ align images)
• Defect isolation
• Binarization (chip defect dataset)
increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data

Augmentation transformations:

- Rotation
- Cropping
- Horizontal & vertical flipping
UNET: FULLY CONVOLUTIONAL IMAGE SEGMENTATION MODEL

Source: U-Net: Convolutional Networks for Biomedical Image Segmentation, Ronneberger et al. 2015
PRETRAINING

• Using pretrained weights (ImageNet) improves results
SHUFFELNET & MOBILENET ENCODER

MobileNetv2 Encoder

Unet Decoder
RESULTS/ EVALUATION

Intersection over Union:

- Unet default: 0.79
- MobileNetv2 Encoder 1.0: 0.75
- ShuffleNet Encoder 1.0: 0.73
- ShuffleNet Encoder 0.5: 0.67

Mean Intersection over Union:

- DeepPCB Dataset
- Chip defect dataset
DEFECT DETECTION

- OpenPCB Dataset
- Number of defects in Testset: 19610

<table>
<thead>
<tr>
<th></th>
<th>Shuffle Unet 0.5</th>
<th>MobileUnet v2 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>19040</td>
<td>18919</td>
</tr>
<tr>
<td>FP</td>
<td>1770</td>
<td>718</td>
</tr>
<tr>
<td>FN</td>
<td>570</td>
<td>718</td>
</tr>
<tr>
<td>Precision</td>
<td>0.915</td>
<td>0.964</td>
</tr>
</tbody>
</table>

DBScan Clustering:
## COMPUTATIONAL PERFORMANCE

<table>
<thead>
<tr>
<th>UNet</th>
<th>MobileUNetv2 1.0</th>
<th>ShuffleUNet 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoder</td>
<td>Decoder</td>
<td>Σ</td>
</tr>
<tr>
<td>Parameter</td>
<td>19M</td>
<td>12M</td>
</tr>
<tr>
<td>FLOPs</td>
<td>13G</td>
<td>24G</td>
</tr>
</tbody>
</table>
PROOF OF CONCEPT - CONCLUSION

• Good Image segmentation results, as well as classification:
  - IoU: \( \approx 0.7 \)
  - Precision: \( \geq 0.96 \)

• Model small enough for embedded applications

• Complement traditional defect detection with a more flexible system (react to changes in manufacturing, inbetween steps of an assembly line,..)
Prototype
IDEA: AI ENABLED VISUAL INSPECTION
IDEA: AI ENABLED VISUAL INSPECTION

Drawer with tablet showing Wafer pattern and having Start button in tablet

Dummy components for visualization

Basler’s 8MP MIPI Camera

Basler board based on i.MX 8M Plus

Navigation tablet to go between Demo and Info mode
MODEL RETRAINING – DATA CHALLENGE

Manufacturing related defects vs artificial anomalies

“Photo of Photo on Display”

Intensity 33/255

Intensity 0/255
MODEL RETRAINING – DATA CHALLENGE

Two possibilities:
Utilizes Preprocessing/semisynthetic data to mimic final input
+ no new training data necessary
- Performance suffers

Generate new Training set similar to final input data
+ best performance
- Tedious image registration/train set generation (automation process necessary)

- intensity and contrast noise on train samples
- EMNIST Dataset:
  - Handwritten letters as “Anomalies”
  - 70,000 images 28x28 pixels
MODEL QUANTIZATION/PERFORMANCE

- 32bit, converted from PyTorch to Onnx
- Running on ARM cores
- ARM NN Library
- \( \approx 240\text{ms Inference time} \)

- Fully quantized 8bit TensorFlow Lite model
- Running completely on Neural Processing Unit
- \( \approx 9.5\text{ms Inference time} \)
NXP / Basler Wafer Inspection Demo
@ Embedded world 2022
Hardware Wafer Inspection Demo
Basler Embedded Vision Processing Kit

- Development Kit for Vision Applications
- Suitable also for series production (population options to reduce unit costs)
- Complete interface population:
  - 2x MIPI-CSI 2, 2x USB 3.0, GigE, HDMI, GPIOs, I2C, SPI, LVDS, UART, CAN, USB 2.0, M 2.0, Bluetooth, Wi-Fi
- Equipped with NXP i.MX8M Plus SoC
Hardware Wafer Inspection Demo

NXP i.MX8M Plus Featureset

- Neural Processing Unit (NPU) with up to 2.3 TOPS
- Up to 2 cameras with MIPI CSI-2 interface, 1080p60 stereo vision
- Image Signal Processor (ISP) for 4K vision, HDR, de-warp (e.g. fish-eye lens correction)
- High resolution video compression including video encode (including h.265) and decode
- 3D and 2D graphics acceleration GPU based
- Real-time processing with Cortex-M7 @800MHz
- 4x CortexA53 @1.8GHz
- 14nm FinFet, low power, high performance
- Robust control networks supported by dual CAN FD and dual Gigabit Ethernet with Time Sensitive Networking (TSN)
Hardware Wafer Inspection Demo

Basler dart MIPI CSI-2 camera module

- 8 MPix, 30 fps
- MIPI CSI-2 interface
- Uses NXP i.MX8M Plus Image Signal Processor (ISP)
- GenICam compliant – compatible to Basler camera SDK (pylon) or gstreamer
- Wide angle lens using the i.MX8M Plus Dewarping hardware block