ANET: Automated Optical Inspection Network

haroldw@standford.edu
Harold Wang, Stanford University
Fall 2018 CS230 Final project

Introduction

Electronic manufacturers rely on computer vision based Automated Optical Inspection (AOI) machines to detect circuit failures. However, state of the art AOI systems has a poor recall rate. I introduce ANET, a CNN based encoder-decoder network that can accurately identify defective units based on inspection images.

Data Set

A local contractor manufacture provided a data set consists of 5268 positive images taken from functional units and 253 defective images.

There are 4 types of devices among this data set, and each image is stored as a grayscale image of size 480*640.

All images are hand labeled by technicians with the device ID and functional status.

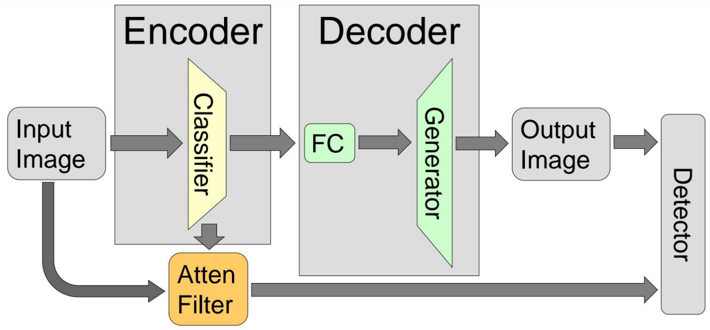
Chip ID	Positive Samples	Negative Samples
0	1572	82
1	492	45
2	1570	6
3	1634	120

Training Methodology

ANET is trained using positive images only for the following reason:
- There is a finite functional state-space
- An infinite space of defective conditions
- The dataset is heavily skewed

ANET training procedure:
1. Train the encoder classifier using labeled device ID
2. Train the decoder network with the pre-trained encoder network, minimizing pixel level loss
3. Train the detector network

Model

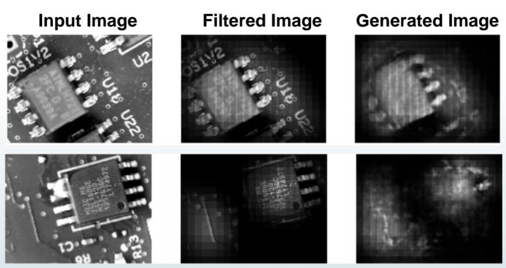


The ANET encoder block consists of a 5-layer CNN classifier. The output of the last convolution layer feeds into the attention filter and the decoder network fully-connected layers.

The attention filter utilizes the classifier output and class-activation-map[1] to force the network to focus only on areas that matter.

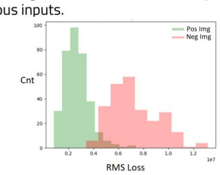
The ANET decoder block consists of two fully-connected layers cascaded with 5-layer de-convolution layers. The generated image feeds into the detector along with the filtered input image.

The detector is a simple threshold-based decision block based on pixel level RMS difference between the generated image and the filtered image.



Results

The loss distribution between positive images and negative images are shown below. ANET reconstructs positive images much more accurately compared to anomalous inputs.



With a proper decision threshold, the precision, recall, and F1 score of each chip type are shown below.

Chip ID	Precision	Recall	F1 Score
0	0.99	0.91	0.95
1	0.95	0.99	0.97
3	0.77	0.92	0.84

Discussion

1. The attention network can mask anomalous area
2. The most accurate generator network is not suitable for this application
3. The size of the FC layer is critical, and the network failed to learn when it is too large or too small
4. Inaccurate attention filter leads false defection trigger

Future Work

1. Investigate into the contribution of the classifier, FC layers, and generator to anomaly detection performance
2. Expand the number of classes of support devices
3. Improve the accuracy and stability of the attentional filter

[1] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva and Antonio Torralba, Learning Deep Features for Discriminative Localization, 2015; arXiv:1512.04150.