CS230

ANET: Automated Optical Inspection Network

Harold Wang
Department of Electrical Engineering
Stanford University
haroldw@stanford.edu

Abstract

Commercial Automated Optical Inspection(AOI) machines require a substantial capital investment, yet suffer from a high fallout rate. I present ANET, an encoder-decoder based anomaly detection network for AOI application. AOI application impose several challenges: imbalanced data sets, unpredictable failure modes, and localized defection area. I started with an encoder-decoder based generative network and explored various architecture to tailor ANET to address these issues and achieved a similar level of performance as the state of the art AOI machine.

1 Introduction

Electronic manufacturers rely on automatic defect detection machines to improve yield, reduce human intervention, and to lower production cost. Automated Optic Inspection (AOI) is commonly used to detect printed circuit board (PCB) assembly defects, such as missing components, faulty assembly, etc. Existing AOI solutions have three shortcomings: prohibiting capital investment, sensitive to small cosmetic variations, and have a high false triggering rate[1]. As a result, there is a need from both large and small contract manufactures for a solution which offers better precision and with a cheaper entrance bar.

Deep learning models can address these issues effectively. Data is abundant in a repetitive high-volume manufacturing environment, suitable for model development. Also, machine learning models do not require sophisticated machinery to carry out inference, which significantly reduces the capital cost of AOI.

I present ANET, an encoder-decoder based model that can perform PCB defection detection. The input to ANET is an image of the device under test(DUT). The model predicts the likelihood that the DUT is defective or no. The typical supervised training methodology is inadequate for defect detection applications for three main reasons. First, high-volume manufacturing projects can achieve a first-pass yield of 99% or higher. As a result, datasets are heavily skewed toward functional units. The imbalance between classes makes the network difficult to train. In addition, only a small subspace of all possible defection types can be known in advance. Hence, it is impossible for the network to learn to identify all defections. Lastly, the defects can be small and localized which is hard for the network to identify.

ANET addresses these challenges by utilizing the following steps. I framed the model as a novelty detection problem to solve the issue of unexplorable negative space. An attention network is added to improve the network efficiency to focus on small and localized defects. A staged training methodology is employed to boost the overall performance by training the encoder, decoder, and detector network independently. ANET achieved 0.87 recall rate across all test cases.

CS230: Deep Learning, Fall 2018, Stanford University, CA.

2 Related work

Researchers, such as Hiroshi[2] and Krummenacher[3], have made significant progress in applying deep learning algorithms to anomaly detection applications. Both researchers trained generative networks with only positive datasets. Networks learned to accurately represent the positive state-space, generating outputs that closely render inputs. At the same time, since networks have not seen any defective data, the generated output will be dramatically different when presented with a novel input. A detector is employed at the final stage to predict the status based on the difference between the generated output and input images.

The approaches above worked well when anomalous images are notably different from the training set. However, PCB defections can be small and localized. Referring to Figure1, a functional device is shown on the left, whereas the right figure shows a defective unit with a solder bridge at the top right corner of the chip. The defective area is small relative to the overall size of the input image. Vanilla novelty detection network will struggle to pick up this difference.

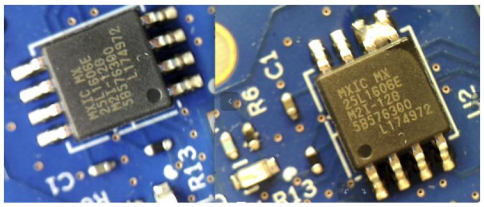

Figure 1: Example images of functional and non-functional devices

3 Data set and Features

A local contract manufacturer(CM) provided a set of 5521 AOI pictures.There are 4 types of devices and each device type has a varying number of functional and non-function images. Table 1 below shows the breakdown of each device category.

Table 1: :Data Set Distribution Breakdown

Device ID	Positive Samples	Negative Samples
0	1582	82
1	492	45
2	1560	6
3	1634	120

Each image sample is hand labeled by CM technicians with a numerical device ID and a boolean test status. The device ID ranges from 0 to 3. It is important to notice that the data set is heavily skewed towards positive samples since it is sampled from a high yield project.

Each image has 480 by 640 pixels stored in jpg format. The RGB images are compressed into grey scale during training, evaluation, and testing.

ANET is trained with positive data only and tested against a balanced test set. The positive images are divided into training, development, and test sets, with a ratio of 0.8, 0.1, and 0.1 respectively. Negative images are used for test and development only.

4 Models

4.1 Baseline Model

I constructed a baseline model using a 5-layer encoder-decoder network to establish a reference performance point. Figure2 below shows the overall model architecture.

The encoder consists of five convolution layers. Batch norm and pooling layers are utilized to speed up training and to reduce network parameter size. Two fully connected layers of size 100 connect the encoder output to the decoder input. The decoder has five de-convolution layers that scale the latent space back into a 480x640 gray-scale image.

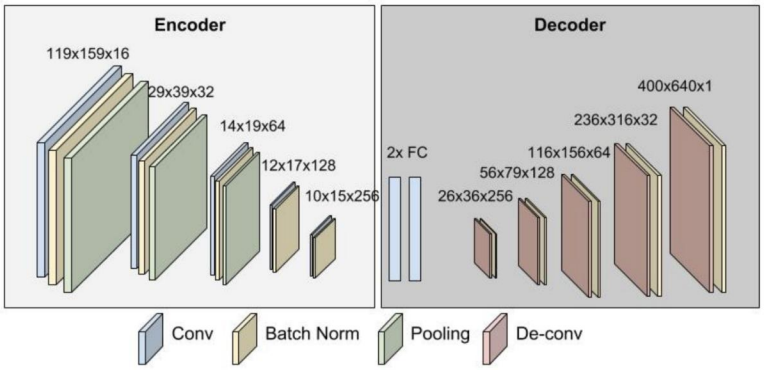

Figure 2: Encoder-decoder Block Diagram

A quick hyperparameter search is performed using the development set to optimize the parameter setting. Table2 shows the final hyperparameters.

Table 2: Hyper Parameter Settings

Hyperparameter	**Value**
Initial Learning Rate	1E-4
Optimizer	Adam
Dropout Rate	0.2
Activation	ReLU
Number of Epoch	100
Batch Size	32

The loss function is defined as the L2 loss of the pixel level difference between the generated image and the input image. With the baseline model, the loss distribution between the positive test images and the negative test images are shown below.

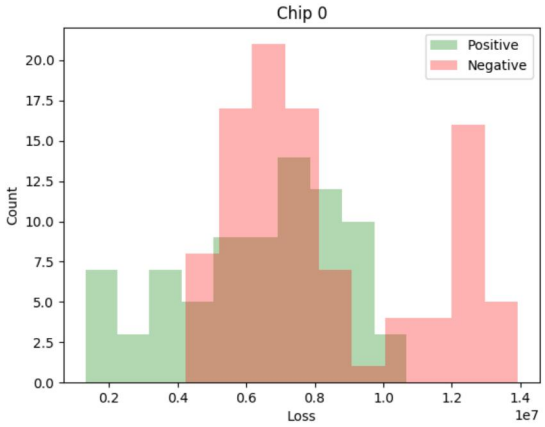

Figure 3: Baseline Model Performance

4.2 Classifier Encoder

It is desirable to have the encoder to compress input image into the latent space and cluster latent representations of the same device close to each other. ANET first trained a classifier network with the labeled device ID. The classifier shares the same convolution layer structure but has a different output fully connected layer. The loss function is defined as the cross-entropy loss between the softmax prediction and the ground truth label. The classifier achieved 97.8% accuracy on the test set.

The pre-trained classifier network (removing the FC layer) acts as an encoder during decoder training with all parameters fixed.

4.3 Attention Network

Since the defective area is small relative to the size of the input, it is desirable for the network to focus on the area of interests that captures the characteristic of the device. ANET uses the class-activation-map[4] output of the classifier as a filter to generate network attention, blackout all uninterested features. The attention matrix is defined by Eq 1 below.

$$f(x,y) = \frac{1}{22500} \sum_{a=x}^{x+150} \sum_{b=y}^{y+150} cam(a,b) \tag{1}$$

A small conv-pool layer is constructed to vectorize the filter implementation: one 150x150 average pool layer is cascaded with a de-conv layer of the same kernel size. The conv layer has a fixed size of with a fixed kernel weight of 1. The pixel value is set to 0 when the average pool output is less than 0.5. The kernel size and filter threshold are determined experimentally. An example class-activation-map (right) and the filtered image(left) is shown in the picture below.

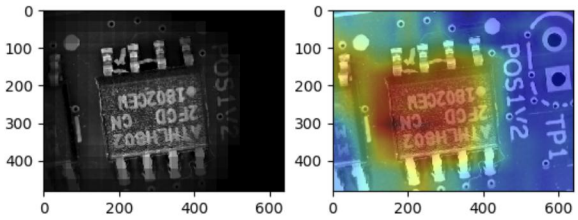

Figure 4: Attention Network Example

4.4 Detector

The detect is implemented using a threshold based binary classifier. The threshold of each device type is determined by optimizing the detector F1 score using the development set.

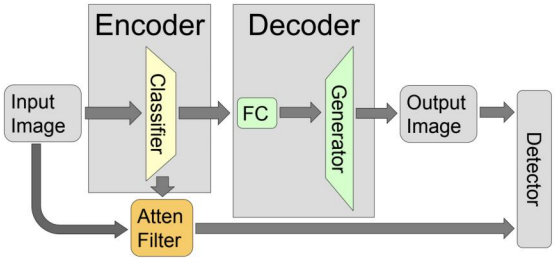

Figure 5: ANET Model

4.5 ANET

The completed ANET model is shown in Figure5.

5 Results and Discussion

ANET achieved a significant performance improvement over the vanilla encoder-decoder baseline model. The pre-trained classifier learned to mask-out abnormal areas of the input image, which altered the distribution of the latent space. The generated output is unrecognizable with the perturbed input. An example is shown in the first row of the Figure 6. The input image is shown on the right, where a solder bridge is observed on the top left pin. The attention network recognized that it is not a feature of interest and masked out the left half of the device. The decoder failed to generate a realistic output with this crippled input image. On the other hand, the generated image based on a functional image captured the key features of the device well, as shown in the second row of Figure 6.

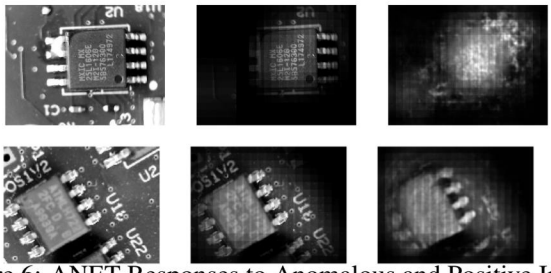
Figure 6: ANET Responses to Anomalous and Positive Inputs

The latent space is critical to the performance of the network. 3 variations of ANET is constructed to study the impact of the size of the fully-connected layers. When the FC layers are too small (ANET-10, with 10 neurons per FC layer), it is insufficient to represent the available positive space. Hence the network failed to distinguish between trained and novel inputs. The network performance improves as the FC layer size increases, and ANET-500 achieved a similar level of performance as existing AOI systems on this test set. However, I believe that a network with sufficiently large latent space is capable of capturing information from both positive and negative input space. The performance improvement with larger fully-connected layers is not warranted to continue. Further investigation is needed to determine the optimal latent space size.

Table 3 shows a comparison between the baseline model, various ANETs, and the existing AOI system.

Table 3: :Performance Comparison

Model	Precision	Recall	F1
Base Line Model	0.85	0.28	0.42
ANET-10	0.7	0.1	0.18
ANET-100	0.95	0.75	0.84
ANET-500	0.93	0.92	0.92
AOI	0.99	0.9	0.92

6 Conclusion/Future Work

In this paper, I presented ANET to demonstrate the feasibility of applying encoder-decoder based anomaly detection models for AOI applications. ANET utilizes a 5-layer encoder-decoder with attention network to focus on small and localized PCB defects. It achieved performance on par to the state of the art AOI solution.

During error analysis, I found that the class-activation-map based attention filter can be erroneous as the network focuses on unrelated background feature of a particular image. A diverse training set of the same device with different backgrounds can improve the stability of the attention filter.

Also, the latent space is unconstrained during training. ANET could explore various loss functions that constraints the training of the fully-connected layers(KL divergence between input and FC layer etc) to improve the stability of the network.

7 Contributions

I, Harold Wang, completed all tasks independently under the guidance of Aarti Bagul. The project source code can be found here on github.

References

[1] Kunte, O. (2018). Automatic Optical Inspection of Printed Circuit Boards. Czech Technical University, https://dspace.cvut.cz/bitstream/handle/10467/74054/F3-DP-2018-Kunte-Ondrej-Automatic_Optical_Inspection_of_Printed_CIrcuit_Boards.pdf

[2] Hirose, N., et al.: Gonet: A semi-supervised deep learning approach for traversability estimation. arXiv preprint arXiv:1803.03254 (2018)

[3] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel Defect Detection With Machine Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 4, pp. 1176–1187, Apr. 2018.

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