Feature extraction and classification of semiconductors

------Advance topics about deep learning

Ren Kai TAN, QIAN Chao

Introduction

Manufacturing processes of semiconductors are complex processes which are completed in various series of stages. Through the stages, integrated circuits with desired functions are typically manufactured in large quantity. Since semiconductors have high demand in various fields which could be very distinctive from each other, it is important to ensure that functionality matches desired function during production stages. In order to check the performance of the semiconductors, yield analysis will be performed during the stages of manufacture, specifically after the wafer fabrication stage as it is found that the primary loss of yield usually happens during the stage. Yield analysis measure the yield loss of the fabricated devices, which is the proportion of devices on the wafer which are tested to be able to achieve the required function properly. Therefore, it is very important for the yield test to be performed and having the yield as high as possible is desirable to ensure the performance of the integrated device.

However, it is found that most semiconductors manufacturing companies are having difficulty in performing the yield analysis. This is due to the fact that yield analysis is usually time consuming to be performed. In order to obtain a good yield test, process engineers would need to compile wafer test data from different sources and perform heuristic analysis which are both laborious and time-intensive. Therefore, if machine learning techniques could be applied to classify well fabricated wafer from the bad ones, the total manufacturing time for semiconductors could be greatly reduced. This method will be explored in this project along with the performance and effectiveness of the method.

Method

In this project, a few machine learning methods are applied for the classification purpose. The image data and labels of the semiconductors are provided by Nextperia, one of the biggest worldwide semiconductors company.

First, the images of the semiconductors are fed to scattering network and a pretrained VGG19 network to perform the extraction of features. Then, the extracted features from both networks are passed over to SVM and LDA along with the corresponding labels for classification training. We also try a simple CNN to do the classification problem directly. The performance of these methods are then studied and analyzed.

Feature extraction

Scattering Net

A wavelet scattering network computes a translation invariant image representation, which is stable to deformations and preserves high frequency information for classification.

The training images are randomly permutated. The good ones are labeled as 0 and the bad ones are labeled as 1. A Scattering Net is performed to extract the features of the each image.

Image 1 [1567.1 34.269 31.559 ...]
Image 2 [1603.8 43.305 39.996 ...]
Image 3 [1544.8 34.196 31.321 ...]
Image 4 [1420 29.26 26.096 ...]
Image 5 [1544.2 29.604 28.725 ...]
Image 6 [1542.8 36.247 36.637 ...]
...

VGG19

VGGNet is a deep convolutional network developed and trained by Oxford's renowned Visual Geometry Group (VGG), which performs very well in image recognition problems

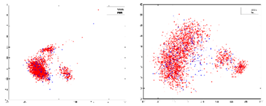
Here a pre-trained VGG19 network is adopted for feature extraction task. For each image, 25508 features are extracted. Due to the limitation of this pre-trained VGG19 network, the size of extracted features are very large and quite a lot of features are zero.

Image 1 [0.000 0.000 0.000 ...]
Image 2 [0.000 7.967 0.000 ...]
Image 3 [0.000 0.000 0.000 ...]
Image 4 [0.000 0.000 0.000 ...]
Image 5 [0.000 0.000 0.000 ...]
Image 6 [0.000 0.000 0.000 ...]
....

visualization

PCA

Principal Component Analysis is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still preserves most of the information in the large set.



(a) Scattering Net (b) VGG19
Visualization of extracted features using PCA

Here PCA is adopted to reduce the dimension of features extracted from scattering Net and VGG19 to two . As we can see, the reduced features are not able to be used to visually distinguish good and bad semiconductors.

Training and Classification

SVM

support vector machine is a popular supervised learning model widely used in classification.

LDA

linear discriminant analysis is a method use din statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separated two or more classes of objects or events.

The images and their labels are divided into training and testing data and then the training data are fed to the supervised learning algorithms, SVM and LDA to train the model. And testing data are used to verify their generalization ability. As we know, the data set is very unbalanced, where the good ones account for 90% while the bad ones are only 10%. In this case, these machining learning methods can get biased results. To solve the problem, we expand the training data by repeating the bad ones by 9 times. And then the training data are randomly permutated again.

From the table below, we can that see that after expanding the training data, the LDA method has higher ability to distinguish the bad semiconductors. But it has almost influence on the SVM method.

We also try a simple convolutional neural network and then fully connected layers to do the classification problem. The training accuracy is 63.54%. The testing accuracy is 77.92%. And for the good ones, the testing accuracy is 82.93% and for the bad ones, it is 70.03%. So the performance is similar with the LDA method when expanding the training data.

Feature extraction method	Scattering net							
Accuracy	SVM				LDA			
Training data set	original		expanded		Original		expanded	
Training accuracy	65.10%		63.13%		93.64%		80.18%	
Testing accuracy	62.80%		75.54%		93.13%		82.32%	
	good	Bad	good	bad	good	bad	good	bad
	64.15%	49.89%	78.54%	46.95%	99.11%	37.46%	83.38%	72.21%
Feature extraction method	**VGG19**							
Accuracy	SVM				LDA			
Training data set	original		expanded		Original		expanded	
Training accuracy	100.00%		100.00%		100.00%		100.00%	
Testing accuracy	90.49%		90.04%		74.37%		62.08%	
	Good	Bad	good	bad	Good	Bad	good	Bad
	94.49%	49.90%	94.49%	49.90%	77.61%	43.51%	63.72%	47.29%

Classification of the disputed paintings

After training SVM and LDA using features extracted from Scattering Net and VGG19, then the well training SVM and LDA models are utilized to do prediction over the given unlabeled images. The number of good ones and the number of bad ones are shown below. Different method can get different results. The predicted label for each image can be read from the attached file.

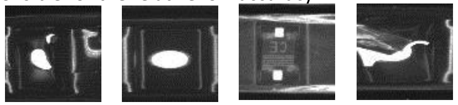
Features extraction methods	Scattering Net				VGG19			
Prediction method	SVM		LDA		SVM		LDA	
Training data set	Original	expanded	original	expanded	original	expanded	original	expanded
Number of good ones	1962	2322	1440	2291	2701	2701	2246	1790
Number of bad ones	1038	688	1560	709	299	299	754	1210

Analysis

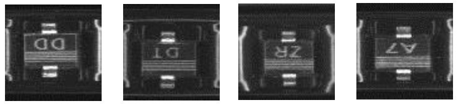
In our case, the size of each image is 224*224=50176. But using Scattering Net and VGG19 can greatly reduce the dimensionality to 391and 25088 according. And then SVM and LDA would be used to do the classification problem from extracted features. They greatly reduces the computation cost against the curse of dimensionality.

Problem

From our understanding, the well organized images are regarded as good while those with some blurs are regarded as bad. However, for some training data, it is even hard/unable to separate the good ones and bad ones visually. It may be responsible for the relative low accuracy.



(c) some weird structures but labeled as good



(d) some clear structures but labeled as bad

Conclusion

Due to the limitation of time and computational resources, we only try a very simple CNN and then adopt some feature extraction methods to do this classification problem. The proposed convolutional neural network method doesn't have advantageous performance compared with the feature extraction method while it is very time-consuming. Scattering Net and VGG19 are very good tools to extract features from high dimension data against the curse of dimensionality. But in this problem, the accuracy is not vey high. In the future, we can try more complex neural network structures. In this way, we may achieve higher prediction accuracy.

Acknowledgment

Professor YAO, Yuan: Course instructor

TAN Ren Kai: feature extraction using VGG19 and classification using CNN

QIAN Chao: feature extraction using Scattering Net and implementation of classification using LDA and SVM

Reference

1. Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1872-1886.

2. Leonarduzzi, R., Liu, H., & Wang, Y. (2018). Scattering transform and sparse linear classifiers for art authentication. Signal Processing, 150, 11-19.

3. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.