Wafer Map Defect Classification with Depthwise Separable Convolutions

Tsung-Han Tsai and Yu-Chen Lee
National Central University, Taoyuan City, Taiwan (R.O.C.)
han@ee.ncu.edu.tw, victorlee@dsp.ee.ncu.edu.tw

***Abstract*--In the IC design process, the test process is the main factor of production cost. Existing tests rely on additional analysis of testing result data by the engineer to determine the status of the process. Thus it could take an additional amount of time and cannot make adjustments of the process immediately. Wafer map defect recognition is an import part of semiconductor. There is lots of information in wafer maps which can quickly help engineers to identify what failure type it is. The location of the error point is graphical represented and the relationship of these points contains the feature of this map. In this paper, we proposed a classifier with reduced-weight architecture based on depthwise separable convolutions. The entire work is verified by using the real-world wafer map dataset (WM-811K). The accuracy is 96.63% in test set.**

I. INTRODUCTION

The semiconductor industry has grown rapidly in recent years. Although technology is constantly improving, the wafer map defects are unavoidable even in precise process. Wafer map is a collection of visual data about physical parameters that experienced engineers can judge when failure occurred. However, it wastes a lot of time and human. The complexity of chip design is also increasing [1], and the mount of analysis has become large and the intelligent assistant judgment will be essential.

In most case, utilizing the wafer map to classify the error type is used. There are number of studies about the wafer map classification [2]. The classification is mainly with two steps: 1) map recognition, 2) feature extraction. The map recognition is through probability distribution function to model each defect map and chose the most matching one. After the map recognition step, the feature extraction will get the important information in the map. Once the features are acquired, the classifier can be trained. There are many classifier such as SVM (support vector machines), Random forest, neural networks, etc.

Nowadays, deep convolutional neural networks (CNN) is a state-of-the-art method. It has shown excellent performance in various areas such as audio processing and image processing tasks, including image classification [3], [4]. In wafer defect area, Nakazawa *et al.* [5] present a method for wafer map defect pattern classification and image retrieval using convolutional neural networks (CNNs). Kim *et al.* [6] propose a neural network-based bin coloring method called Bin2Vec to make similar bin codes are represented by similar colors. Saqlain *et al.* [7] propose a voting ensemble classifier with multi-types features to identify wafer map defect patterns in semiconductor manufacturing.

Wu *et al.* [8] build a dataset with real-world wafer (WM-811K). They propose a conventional approaches (SVM) to process wafer map failure pattern recognition (WMFPR). The structure of the dataset is shown as Fig. 1.

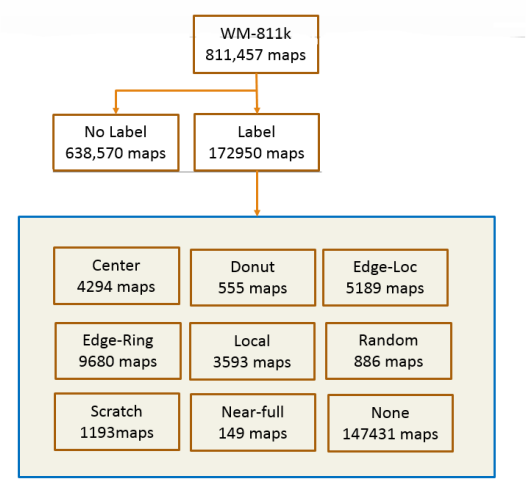

Fig. 1 The structure of the dataset

However, CNN often require a lot of computation because the number of multiplication and addition. There are two ways to face this problem 1). Reducing the weight of architecture, such as MobileNet [9], X-NOR net [10]. 2). Using CUDA acceleration by GPU. The lack of these two cases may lead to slow process and cannot be real-time in some low-computing embedded platforms. We use the first way to optimize our work because it directly modify the main architecture without extra cost.

This paper is organized as follows. In Section II, we describe the data, depthwise separable convolutions and the model architecture we used. The preprocessing of the data is also mentioned. In Section III, we present the low-weight CNN training and show the CNN training parameters. In Section IV, there is a result of comparison with standard convolution and other methods. The conclusion is given in Section V.

978-1-7281-5186-1/20/$31.00 ©2020 IEEE

II. METHOD

A. Dataset

The dataset is a real word wafer map named WM811K which is built by Wu *et al* [8]. There are 811,457 original wafer maps collected from 46,293 LOTs in the set. The failure type classes include None, Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-full, shown as Fig.2 It is the open source dataset available at website [11].

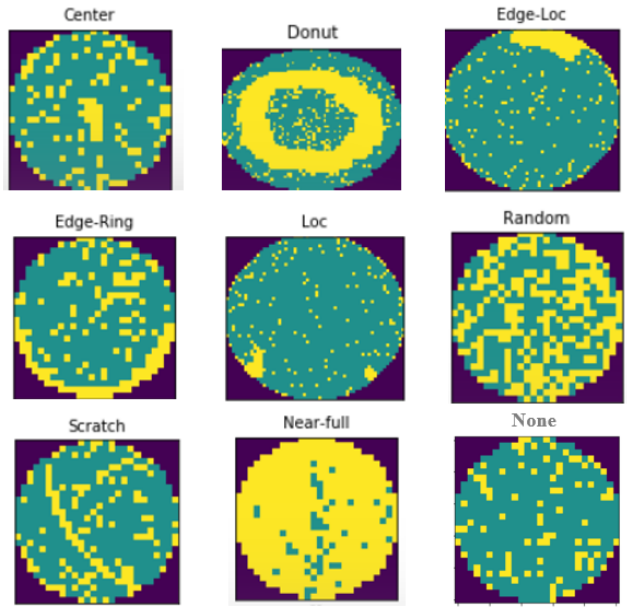

Fig.2. Visualization of the 9 classes

B. Preprocess

Before the neural network training, the input data must be resized to the same size to feed the network. Different size maps are resized to (50, 50, 3) as input shape. Wafer data's each pixels have a state variable that express 0: not wafer, 1: normal, 2: faulty. Extend extra dimension with one-hot-encoded categorical data as channel [12].

C. Depthwise separable convolutions

Depthwise Separable Convolutions are an important block in many efficient neural network architectures and we also use them in our work. The main idea is to factorize standard convolution to depthwise convolution and pointwise convolution. The first step is a light-weight filter by applying a single convolutional filter per input channel. The second step is a 1×1 convolution and the filter number will be the same as basic convolution filter number. Standard convolution takes an $h \times w \times d$ input shape, and applies convolutional kernel $k \times k \times d \times n$ to produce an $h \times w \times n$ output shape L. Standard convolutional layers have the computational cost of $h \times w \times d \times n \times k \times k$, which k is the kernel size, n is the kernel number. Depthwise separable convolutions are only cost $h \times w \times d \times (k^2+n)$.

D. Model Architecture

Our model Architecture is based on depthwise separable convolutions. Because our input data is 50x50 and the output classes is 9, we adjust the layer based on MobileNet [9]. In depthwise layer, each stride 2 layer will down sample the map size, so we reduce [conv dw s2, conv s1, conv dw s2, conv s1] before the average pooling layer, shown as

conv / s2	3x3x3x32	50x50x3
conv dw / s1	3x3x1x32 dw	25x25x32
conv / s1	1x1x32x64	25x25x32
conv dw / s2	3x3x1x64 dw	25x25x64
conv / s1	1x1x64x128	13x13x64
conv dw / s1	3x3x1x128 dw	13x13x128
conv / s1	1x1x128x128	13x13x128
conv dw / s2	3x3x1x128 dw	13x13x128
conv / s1	1x1x128x256	7x7x128
conv dw / s1	3x3x1x256 dw	7x7x256
conv / s1	1x1x256x256	7x7x256
conv dw / s2	3x3x1x256 dw	7x7x256
conv / s1	1x1x256x512	4x4x256
5x		
conv dw / s1	3x3x1x512 dw	4x4x512
conv / s1	1x1x512x512	4x4x512
Avg Pool/ s1	Pool 4x4	4x4x512
FC / s1	512x9	1x1x512
Softmax / s1	Classifier	1x1x9

Fig. 3. Model Architecture

III. TRAINING

The main method sued in training work is by the CNN neural network. Our method uses 12 convolution layers, where kernel size is 3x3, and kernel filters is 32, 64, 128, 128, 256, 256, 5x512. Each layers is adding batch normalization, and then connect the 4x4 average pooling layers. Finally received the soft-max function for output, shown as Fig. 4. The epoch is set to 3000, batch size = 32, learning rate =10^ (-4). A total of 172950 maps are used, where 57654 maps for training, 500 maps for validation, and 114796 maps for testing. The final result is 96.63%. Compared with other method, our accuracy is higher than others.

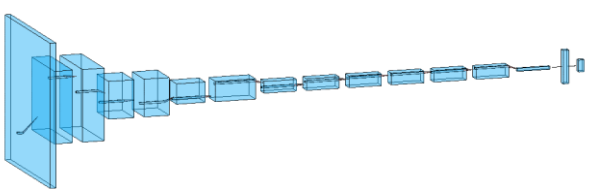

Fig.4. Architecture schematics

IV. RESULT

Table 1. Comparison of accuracy

Method	SVM[8]	SVE[7]	This work
Accuracy	94.63	95.86	96.63

Table 2. Comparison of parameter weight

Architecture	Standard convolution	Depthwise separable
Parameter weight (Million)	14	1.6

V. CONCLUSION

In this paper, we focus on wafer map defect classification with reduced weight method. We use depthwise separable convolutions to train our model, and finally get a classifier to judge the wafer failure type. Compared with two classification method, it shows the low-weight CNN has a better accuracy than SVM and SVE. This method could be used for testing of remaining wafer to reduce the time and the cost required by overall testing process. Of course the method can be improved with this model based on more datasets and more labeling category. In the future work, we will use CNN classifier to recognize wafer map failure in products. It can help to find out the error occurred stage in the first time.

REFERENCE

[1] C. A. Mack, "Fifty years of Moores law," IEEE Trans. Semicond. Manuf., vol. 24, no. 2, pp. 202–207, May 2011.

[2] J. W. Cheng, M. P.-L. Ooi, C. Chan, Y. C. Kuang, S. Demidenko, "Evaluating the performance of different classification algorithms for fabricated semiconductor wafers," In Proceedings of the Fifth IEEE International Symposium on Electronic Design, Test and Application (DELTA'10), Ho ChiMinh City, Vietnam, 13–15 January 2010; pp. 360–366.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Boston, MA, USA, 7–12 June 2015, pp. 1–9.

[4] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, pp. 770–778.

[5] T. Nakazawa and D. V. Kulkarni, "Wafer Map Defect Pattern Classification and Image Retrieval Using Convolutional Neural Network," IEEE Trans. Semicond. Manuf., vol. 31, no. 2, May 2018.

[6] J. Kim, H. Kim, J. Park, K. Mo and P. Kang, "Bin2Vec: A Better Wafer Bin Map Coloring Scheme for Comprehensible Visualization and Effective Bad Wafer Classification," Applied Sciences, vol. 9, issue 3, February 2019.

[7] M. Saqlain , J. Y. Lee and B. Jargalsaikhan, "A Voting Ensemble Classifier for Wafer Map Defect Patterns Identificationin Semiconductor Manufacturing," IEEE Trans. Semicond. Manuf, vol. 32, no. 2, May 2019

[8] M.-J. Wu, J.-S. R. Jang, and J.-L. Chen, "Wafer map failure pattern recognition and similarity ranking for large-scale data sets," IEEE Trans. Semicond. Manuf., vol. 28, no. 1, pp. 1–12, Feb. 2015.

[9] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" in CVPR, 2017

[10] M. Rastegari, V. Ordonez, J. Redmon, A. Farhadi, "XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks" in CVPR, 2016

[11] WM811K Mirlab.org. (2018). MIR Corpora. Accessed: Apr. 7, 2018. [Online].Available: http://mirlab.org/dataSet/public/

[12] Data Science & Business Analytics Lab, School of Industrial Management Engineering, College of Engineering, Korea University[http://dsba.korea.ac.kr/main]