A Deep Learning-based Multi-model Method for Etching Defect Image Classification

Shih-Kuei Chou[1], Yuan-Hsin Lin[1] ,Wen-Hsing Liao[1],
Yu-Min Hsu[1], Chi-Hsiang Kuo[1], Cheng-Kuo Lin[1]

[1]WIN Semiconductors Corp., No. 358, Huaya 2nd Rd., Hwaya Technology Park, Guishan District, Taoyuan City 333411, Taiwan
E-mail: skchou@winfoundry.com and ymhsu@winfoundry.com

KEYWORDS: CNN (Convolutional Neural Network), Deep-learning, Etching defect, Multi-model

Abstract

Many types of defects can occur during the etching process. Some defects are obvious, while others are not. In addition, the characteristics of some defect types are similar to each other. Traditionally, the classification of defect types fully relies on a well-trained engineer. To overcome the disadvantage of manual classification, we introduced a multi-model method to achieve a high-speed classification computer system.

We propose a multi-model classification method, which combines two models to enhance the classification property. The model consists of two InceptionV3 models. To improve the accuracy of partial classification, the second-stage is determined by some critical data from the first-stage model. Finally, we have significant accuracy improvement on metal defect data through the two-stage model.

INTRODUCTION

Etching quality has always been one of the most important items in quality control in the semiconductor industry. The quality of the product affects not only the price of the product, but also the purchase intention of customers. Traditionally, the quality monitoring method in the etching process uses an automatic inspection machine to detect defects, and then the engineer manually classifies the defective photos. However, inconsistent judgment standards of different engineers usually lead to inconsistent results of defect classification. In addition, the speed of personnel judgment is slow. For example, when a factory is working at full capacity, it must use a lot of human resources for defect detection, which will cost the company a lot of money, but it has not yet obtained fast and accurate judgments.

According to the special classification requirements of the factory in defect analysis, some of the defect classification results will have special judgment rules. For example, Defect type A has a higher priority than other any defect type. Therefore, "A" classification result should always be classified, when defect A and Defect B or other defect appear in the etching image at the same time. There is a high correlation by such special rules. Consequently, the accuracy of some classes can be difficult to improve when only one model is used.

Therefore, we propose a compound model to improve the accuracy which may be influenced by special classification rules. In the process, all the etching images are classified by the first-stage model, and some of the classification classes are chosen by criteria. We used these classes to classify the images in the second-stage model again. After the second model classification, we took the classification result of the second-stage model as the final answer. If the image was classified into the residual classes from first-stage model, we do not change the classification result. Our compound model structure can greatly improve the accuracy of specific classification requirements to comply with special classification rules. As demonstrated in Figure 1, the proposed architecture is constructed by two independent CNN (Convolutional Neural Network) models architectures.

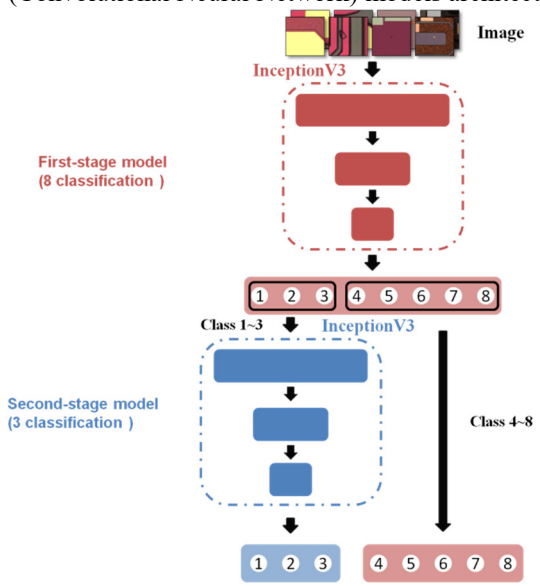

Fig. 1. Flowchart of the proposed deep-learning based multi-model method.

BACKGROUND

Image classification

Figure 2 shows 8 classes. Class 1 is Normal image w/o any Etching defect issue, and class 2 to 8 is different defect types. Class 2 is Metal scum, which is defined as any foreign matter remaining on the metal layer, such as photoresist, stains or water marks. Class 3 is Passivation scum, which defines that there are foreign matter residues on the protective layer, such as photoresist, stains or water lines.
Foreign matter often remains on the protective layer and the metal layer. When it occurs at the same time, we think it should be classified in class 2. So class 2 is a special class, which has a higher priority than other defect types. The schematic diagram of such defect characteristics is shown in Figure 3 and Figure 4.

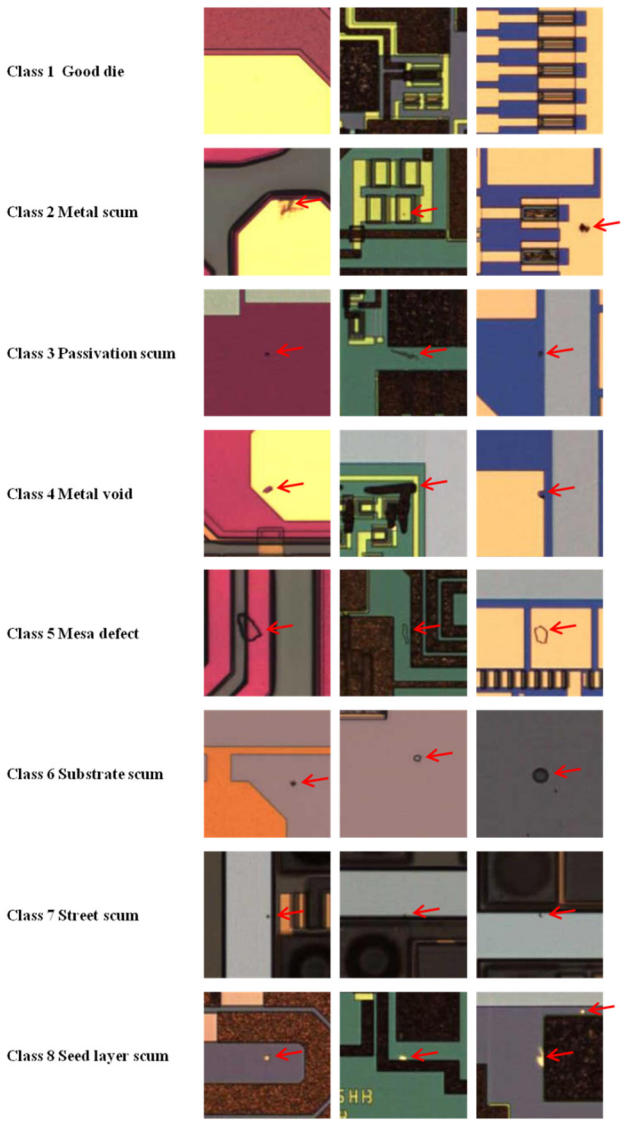

Fig. 2. Examples of etching defect patterns. (Defects are pointed by red arrow)

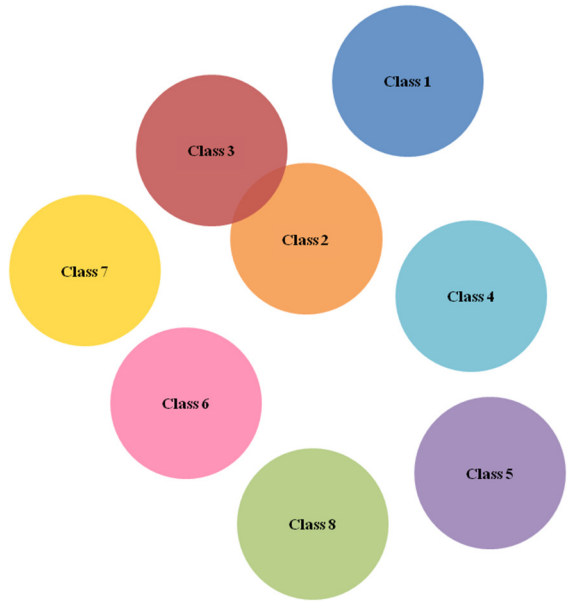

Fig. 3. The relationship between the different classes. Class 2 has a partial intersection with class 3.

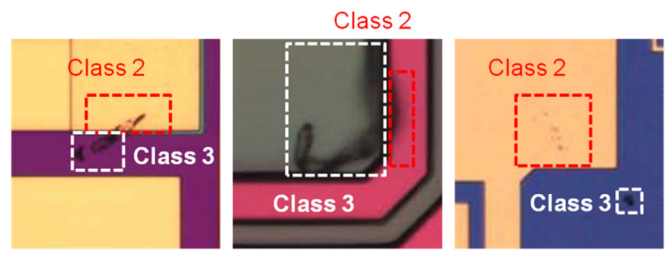

Fig. 4. Examples of class 2 data have a partial intersection with class 3.

EXPERIMENT

Image pre-processing

The image was shot by Rudolph's AVI machine. The shooting rule is comparing the image screen from the camera with the golden wafer image. Through the grayscale and brightness difference comparison between the image screen from the camera and golden wafer image, we can distinguish whether the area of image shot is flawed or not. If the difference between the grayscale and brightness is greater than the setting value, the machine will place the defect in the center and save the image automatically.

AVI camera lens has a magnification of about 10X, and the captured area of each image is about 262×333 um. However, such a wide image range is very difficult for the model to capture and learn for such a small defect. Therefore, we cropped the image and retained the center area of the image. The cropped image is shown in Figure 5.) The cropped image helps the model to learn the defect features easily, because the defect features takes a larger proportion of the entire image. It's like the features selection. We can use the more informative parameters to fit the training model.

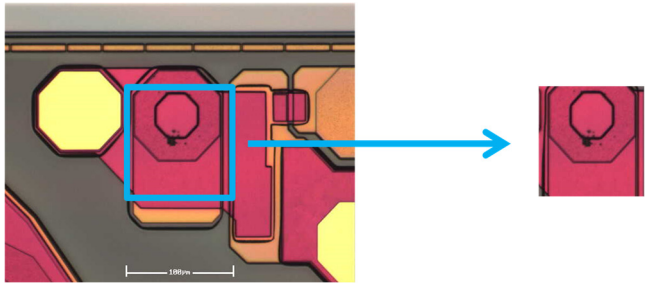

Fig. 5. Crop the image before model training.

Data generation

We consider the precision as well as accuracy the precision of the training model. Because of the limited number of input samples on some classes, CNN training would be over fitting on such more parameters [2]. In order to avoid reducing the precision, one solution is using the training data of the existing training set to generate more data [3]. Applying different techniques such as horizontal flip, rotation, and rotation scaling on the data set, we can create more training samples. The data expansion parameters which were used for all data sets are described in Table 1. The expansion is shown in Figure 6.

Table1. Data augmentation parameters.

Parameter	Value
Rotational Range	-10~10
Zoom Range	0.7~1.3
Vertical Flip	True
Horizontal Flip	True
Range of horizontal translation	±5
Range of vertical translation	±5

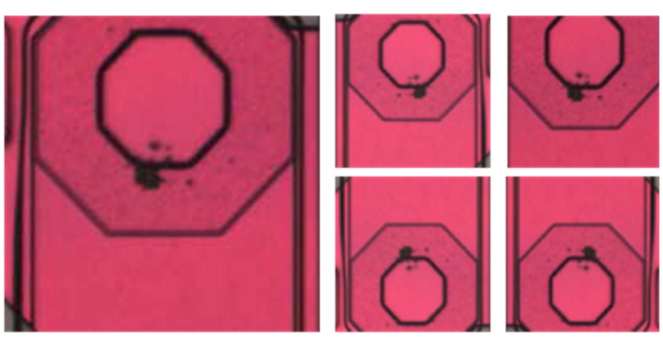

Fig. 6. Images obtained after data generation techniques. The left image is the original image and the right image share the artificially generated image after different data augmentation methods.

Model training settings

According to the input size of the selected InceptionV3 [1] model, we used bicubic interpolation to resize all images to 224x224 pixels. The batch size is set to 32, and the model is trained for 200 epochs. On the Adam optimizer, β1, β2 and learning rate are set to 0.6, 0.8 and 0.001, respectively. Also, we have modified the output of the last layer in all structure to correspond to eight classes, instead of the 1000 classes proposed by ImageNet. In addition, the network weights are initialized by the weights trained on ImageNet. The operating system is Windows with Intel(R) Core(TM) i7-9900KS 4 GHz processor with 32 GB RAM.

The training and testing have implemented the process by the proposed model in this experiment. We used the deep learning toolbox of MATLAB, and run on Nvidia GeForce GTX 2080 Ti GPU with 11GB RAM.

Result

After the two models are trained for 200 epochs, we load them and verify their accuracy. Figure 7 shows the classification results of the first-stage model. It can be seen that the accuracy of the second classification and the third classification are not as good as another classification, which are 65.7% and 83.8% respectively. As previously mentioned, such a special classification rule will affect the model learning. In the experiment results, the accuracy of classification 2 is not being as high as other classifications and it affects the accuracy of other classifications.

		Predicted Class								
	Class	1	2	3	4	5	6	7	8	Acc
True class	1	293581	12443	4348	2282	993	1328	5854	2048	90.9%
	2	97302	482463	59427	13247	58033	5920	1277	16217	65.7%
	3	9579	11170	211773	641	13930	208	451	4988	83.8%
	4	1054	500	47	26195	181	89	5	46	93.2%
	5	227	721	920	192	41576	243	10	96	94.5%
	6	114	126	35	68	76	42136	0	3	99%
	7	222	12	29	5	8	20	19220	5	98.5%
	8	117	63	67	6	18	0	1	14183	98.1%

Fig. 7. Results of the first-stage model classification.

After we used the second-stage model to recognize again the class 1 to 3, the accuracy of these three classes has been greatly improved. The accuracy of class 2 has increased from 65.7% to 72.6%, and the accuracy of class 3 has increased from 83.8% to 89.1%. From the results, it can be clearly proved that the accuracy of a single model is enhanced by the multi-stage model without changing the first-stage model accuracy results of class 4 to 8. Figure 8 shows the accuracy result of the multi-stage model classification.

	Predicted Class								
Class	1	2	3	4	5	6	7	8	Acc
1	292153	14545	3674	2282	993	1328	5854	2048	90.5%
2	47121	532796	59275	13247	58033	5920	1277	16217	72.6%
3	3232	3997	225293	641	13930	208	451	4988	89.1%
4	952	596	53	26195	181	89	5	46	93.2%
5	116	867	885	192	41576	243	10	96	94.5%
6	69	171	35	68	76	42136	0	3	99%
7	213	25	25	5	8	20	19220	5	98.5%
8	92	87	68	6	18	0	1	14183	98.1%

Fig. 8. Results of combining the second-stage model classification.

CONCLUSIONS

We propose a multi-model classification method for etching defect detection. This method of combining two models can improve the accuracy of special classification rules, while the accuracy of other classes will remain unchanged.

The experimental results showed that the proposed multi-model outperformed the single CNN model. The accuracy of Metal scum has increased from 65.7% to 72.6%, and the accuracy of passivation scum has increased from 83.8% to 89.1%. Therefore, the multi-model method based on deep learning can make full use of different model combinations to improve the prediction accuracy of the special classification rules.

REFERENCES

[1] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016

[2] Hua Li, Shasha Zhuang, Deng-ao Li, Jumin Zhao, and Yanyun Ma. 2019. Benign and malignant classification of mammogram images based on deep learning. Biomedical Signal Processing and Control 51 (2019), 347–354.

[3]Sara Hosseinzadeh Kassani and Peyman Hosseinzadeh Kassani. 2019. A comparative study of deep learning architectures on melanoma detection. Tissue and Cell 58 (2019), 76–83.

ACRONYMS

CNN: Convolutional Neural Network
AVI: Auto Visual Inspection