Wafer Defect Pattern Recognition and Analysis Based on Convolutional Neural Network

Naigong Yu, Qiao Xu, and Honglu Wang

***Abstract*—Wafer map defect pattern contains the critical information about semiconductor manufacturing, effective defect analysis technology can improve the yield of products. This paper develops a wafer defect pattern recognition and analysis method based on convolutional neural network (CNN). We build an 8-layer CNN model to inspect wafer map defect and a 13-layer model to classify defect patterns. To analyze the defects, we first extract features based on the classification network, then perform PCA dimensionality reduction and similarity ranking, we infer the root causes of tested samples according to the retrieved wafer maps. In particular, we analyze the impacts of different network layers and feature dimensions on image retrieval performance, it turns out that the appropriate dimensionality reduction can increase the accuracy and speed of wafer map retrieval. The models are trained and tested on the WM-811K wafer map dataset, which are collected from real wafer maps. The experimental results show that the proposed method is able to identify common wafer defect patterns and analyze the root causes accurately and effectively.**

***Index Terms*—Semiconductor defects, pattern recognition, neural networks, information retrieval.**

I. Introduction

THE MANUFACTURING process of semiconductor is quite complicated. As the major raw material for integrated circuit, wafer defects are the primary obstacles affecting product yield. Defective grains on wafer tend to converge into a certain distribution pattern that include critical information, engineers need to identify faults in production process based on the wafer map defect patterns in order to improve yield [1]. However, in the actual manufacturing, wafer map defect pattern recognition and analysis are mainly carried out manually, and the related researches still remain in the theoretical stage.

Traditional methods often fail to achieve a good recognition rate because of complex feature extraction. In addition, most of the methods are limited to pattern classification, which are impossible to analyze the different root causes of defects in the same pattern. However, the root cause analysis of defect is crucial because although the same defect pattern has similar morphological characteristics of defect clusters, different location or size reflect variant causes of defect. The conventional wafer map inspection is to separate the classification and analysis processes, and analyze the defect pattern manually after classification by machine. In this paper, we propose a wafer defect pattern recognition and analysis method using convolutional neural networks (CNN). We train and test the model based on WM-811K wafer dataset, which is the largest open source dataset and comes from the real production process. There is a total of 811,457 samples and only 172,950 samples are manually labeled, including 9 typical wafer map defect patterns (namely *Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-Full* and *None*). Experimental results show that the average detection and classification accuracy of the recognition model achieve 99.98% and 93.25% respectively for test set.

We also develop an online pattern analysis method based on the designed recognition model. We extract the features from the middle layer of classification network and apply the method of PCA to perform dimensionality reduction, retrieve the similar labeled samples from database. Therefore, the root defect causes can be inferred according to the retrieved samples.

Rest of the paper is organized as follows: Section II shows the related work of our study; Section III introduces the approach of wafer defect pattern recognition model and defect causes analysis method; Section IV describes the experimental results and Section V gives the conclusion.

II. Related Work

Early efforts focused on statistical analysis of wafer map defect patterns [2]–[4], Hwang and Kuo proposed a model-based clustering algorithm to analyze the wafer map, they used the principal curve and bivariate normal distribution to model the defect cluster, and compared the logarithmic likelihood probability of two models to determine the shape characteristics [3]. Wang *et al.* used Gaussian EM algorithm to estimate elliptic and linear defects, and spherical shell algorithm to estimate ring defects [4]. The methods based on statistical analysis can only infer the shape of defect clusters. However, different defect patterns may show the same morphological characteristics, for example, *Loc* and *Edge-Loc* have the same shape but different location, statistical analysis method cannot determine the specific defect pattern of a wafer map. With

Manuscript received July 22, 2019; revised August 15, 2019; accepted August 22, 2019. Date of publication August 27, 2019; date of current version October 29, 2019. *(Corresponding author: Naigong Yu.)*

The authors are with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China, also with the Beijing Municipal Commission of Science and Technology, Beijing Key Laboratory of Computing Intelligence and Intelligent System, Beijing 100124, China, and also with the Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China (e-mail: yunaigong@bjut.edu.cn; xq19940704@163.com; 15810214926@163.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSM.2019.2937793

the development of machine learning (ML) technology, the researchers applied ML algorithm to the wafer map defect pattern recognition, which greatly improved the recognition ability. These algorithms are mainly divided into two categories: unsupervised learning and supervised learning.

Due to the difficulty in obtaining wafer samples, unsupervised learning approach has attracted much attention [5]–[10]. Yu and Lu proposed a wafer defect pattern recognition method based on the dynamic integration of manifold learning algorithm and gaussian mixture model. A joint local and nonlocal linear discriminant analysis (JLNDA) was proposed to reduce the dimension of the original defect features, and gaussian mixture model was used to build defect database. Their method improved the average accuracy to 90.5% on WM-811K wafer map dataset [5]. In addition, there are neural network-based clustering [6], [7], k-means [8], hierarchical clustering methods [9], [10] and so on. However, the above methods depend on manual extraction of sample features which may reduce the accuracy.

Supervised learning algorithms [11]–[15] rely on labeled samples, challenge to the acquisition of datasets, but achieve better performance than unsupervised learning. Wu *et al.* used SVM to classify wafer defect patterns, calculated similarity through Euclidean distance and two-dimensional normalized correlation coefficient so as to realize defect root causes analysis [11]. Piao *et al.* extracted the original feature of wafer map based on Radon transform and constructed a decision tree to recognize wafer map defect patterns [12]. Their methods still need complex feature processing, and the models were not good at identifying patterns such as *Loc, Edge-Ring* and *Random* (defect patterns in WM-811K wafer dataset). However, Wu provided an idea to analyze the root causes of defects based on retrieval algorithm. Deep learning has been popular since 2012, and CNN has shown excellent performance in computer vision tasks. Compared with manual feature extraction, the feature expression ability extracted by convolutional neural network is stronger [16]. Nakazawa and Kulkarni constructed an eight-layer CNN model, they obtained a great recognition accuracy after training on simulation wafer dataset [13]. However, the classification ability of the model on the real wafer dataset (such as WM-811K) is weak. Nevertheless, Takeshi Nakazawa's research has verified the potential of deep convolutional neural network in the field of wafer map defect pattern recognition.

III. Methodology

The proposed method can be divided into off-line modeling and online processing. The overall scheme is shown in Fig. 1. During the off-line modeling stage, we build a wafer defect pattern recognition model based on CNN and use WM-811K wafer dataset for training. In the online processing stage, the wafer map is preprocessed and then input to the wafer defect pattern recognition model, which includes two sub-models: wafer map defect pattern detection model and classification model. Compared with the classification model, the other one has a simple structure and a fast computing speed. Due to the low probability of process faults in the actual manufacturing process, this *Detection-Classification* strategy can effectively reduce the pattern recognition time of large-scale wafer maps analysis. Finally, features are extracted from classification model and similar samples are retrieved from the database based on similarity measurement. We sort the retrieved results according to the similarity, and infer the root defect causes of the tested samples according to similar ones.

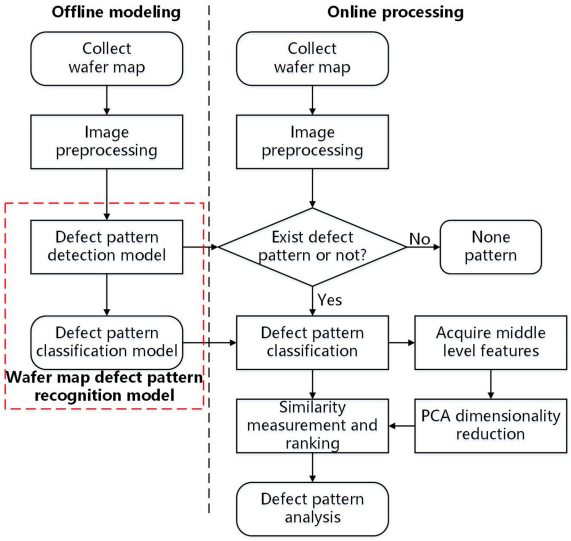

Fig. 1. General scheme of wafer defect pattern recognition and analysis method. Offline modeling stage: construct and train the wafer map defect pattern recognition model. Online processing: detect and classify defect patterns, analyze the root causes through feature extraction and similarity measurement.

A. Image Processing

Even if there is no specific defect pattern in the wafer map, there are a large number of defective grains in it, although they do not form a defect cluster (refer to *None* pattern in Fig. 2 (a)). Noise will affect the recognition performance which is essential to eliminate. Because the wafer map is binary image, therefore we can regard the noise as impulse noise. We applied median filter algorithm to filter out the noise of the wafer map. Median filtering can effectively suppress impulse interference and prevent edge blurring. In order to explore the filtering effect, we tried different size filters. Fig. 2 shows the noise reduction results of the 7×7, 9×9 and 11×11 filters respectively. Compared with the raw wafer maps, the median filter algorithm can filter out most grain noise and effectively retain the morphological characteristics of defect cluster. The 7×7 filter cannot completely filter out the noise of *Center* and *Loc* patterns, the 11×11 filter blurs the edge of Edge-Ring pattern and destroys the morphological features of Scratch pattern, therefore we chose the filter in size of 9×9.

To accelerate the convergence rate during the training stage, we normalized the sample size to 224×224 and standardized pixel values to the [0, 1] interval according to

$$I'(x, y, z) = \frac{I(x, y, z)}{127.5} - 1. \tag{1}$$

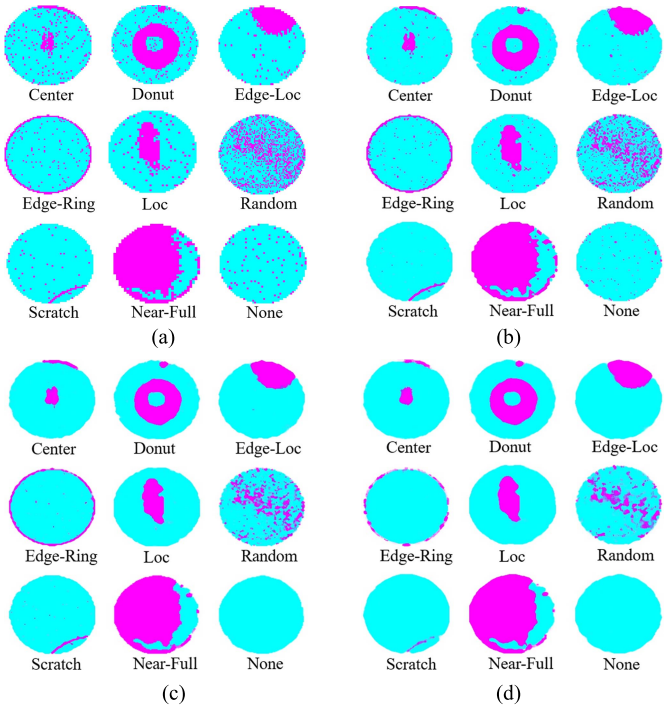

Fig. 2. (a) WM-811K raw wafer dataset with noise. (b) Median filtered wafer maps with a filter size of 7 × 7. (c) Median filtered wafer maps with a filter size of 9 × 9. (d) Median filtered wafer maps with a filter size of 11 × 11.

B. Wafer Defect Pattern Recognition Model

The wafer defect pattern recognition model is composed of two sub-models: detection model and classification model. The detection model is used to judge whether there is defect pattern on wafer map, and then the defect pattern is further determined by the classification model.

1) Wafer Defect Pattern Detection Model: The detection model we constructed consists of three convolutional layers, three pooling layers and two fully connected layers. The schematic diagram of the model structure is shown in Fig. 3. The input of the model is 224×224, the Conv1 to Conv3 are the convolution layers, convolution kernel size is 3×3, the stride value is 1 and use Rectified Linear Unit to activate. After each convolutional layer, a 3×3 maximum pooling layer is used for down-sampling. To avoid overfitting, the Dropout method was introduced to the Fc1 layer during training to randomly inactivate many neurons. The Dropout probability value was set to 0.5 in this method. The activation function of Fc1 layer is Sigmoid, and a Softmax layer is used for calculating class probability.

2) Wafer Defect Pattern Classification Model: We built three convolutional groups, pooling layers and three fully connected layers for the classification model, as shown in Fig. 4, the first two convolution groups are composed of two convolutional layers, and the third one is composed of three convolutional layers. The convolution kernel and the activation function are the same as the detection model. Fc1 and Fc2 layers are activated by Sigmoid function, the Dropout method is introduced to avoid overfitting in these two layers and the probability is set to 0.5.

C. Defect Root Causes Analysis

We built the database based on WM-811K wafer dataset, it included 8 defect pattern libraries (*Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-Full*). We extracted the features of the samples through the classification model and selected the feature data by Fc2 layer. We randomly took a wafer map sample of *Center* pattern and extracted features of Fc2. After the normalization of the characteristic data, we visualized the features. As shown in Fig. 5, the feature dimension of 1024 has a mass of data redundancy, many elements are zero and the data are mainly distributed at the poles. This will decrease retrieval accuracy and take up excessive computing resources [17]. Literature [18] used PCA to reduce the feature dimension of neural network, and the feature data after dimension reduction showed better performance in image retrieval task, we referenced this method and used PCA [19] algorithm to reduce the high-dimensional features to 256 dimensions, then normalized the features to the interval of [0, 1]. Different from the hash algorithm [20], [21], we mapped the feature vector $\boldsymbol{v}$ into $\boldsymbol{\mu}$, the mapping method is given by:

$$\boldsymbol{\mu}[i] = \begin{cases} \lfloor v[i] \times 10 \rfloor, & if\ v[i] \neq 1 \\ 9, & \text{others} \end{cases} (i \in \{1, 2, 3, \ldots, 256\}), \quad (2)$$

where symbol $\lfloor x \rfloor$ means to round down the value of x. Such a mapping method can retain more feature information than hash algorithm. This is because the hash algorithm maps the vector to 0 or 1 encoding, while our method maps the vector to 0-9 ten values, which will retain more representative information. Finally, we calculated the hamming distance through

$$d(\boldsymbol{\mu}_{sample}, \boldsymbol{\mu}_{database}) = \sum_{j=1}^{256} \boldsymbol{\mu}_{sample}[j] \oplus \boldsymbol{\mu}_{database}[j], \quad (3)$$

where $\boldsymbol{\mu}_{sample}$ is the feature vector from tested sample and $\boldsymbol{\mu}_{database}$ is a feature vector from database. Sorted by distance from smallest to largest, the first three retrieved samples were selected as the reference to infer the defect root causes of tested sample.

IV. Experiments and Results

We trained and tested the model based on the WM-811K wafer dataset, which covers 9 categories of wafer map patterns, including 8 kinds of defect patterns (namely *Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-Full*) and a kind of wafer map with no defect pattern (named *None* pattern), as shown in Fig. 2(a). WM-811K dataset contains 811,457 samples, but the manually labeled samples are only approximately 20%. Table I makes statistics on the number of labeled wafer maps. We can see that the number of *None* pattern is much higher than others. *None* pattern means there is no defective cluster pattern on wafer, compared with the other 8 defect patterns, it is easy to distinguish. Therefore, only 30,000 *None* samples were selected for the experiments. We divided the dataset into training set, validation set and test set according to 60%, 15% and 25%. Fig. 6(a) shows the classification scheme of 9 typical defect patterns, and Fig. 6(b) shows

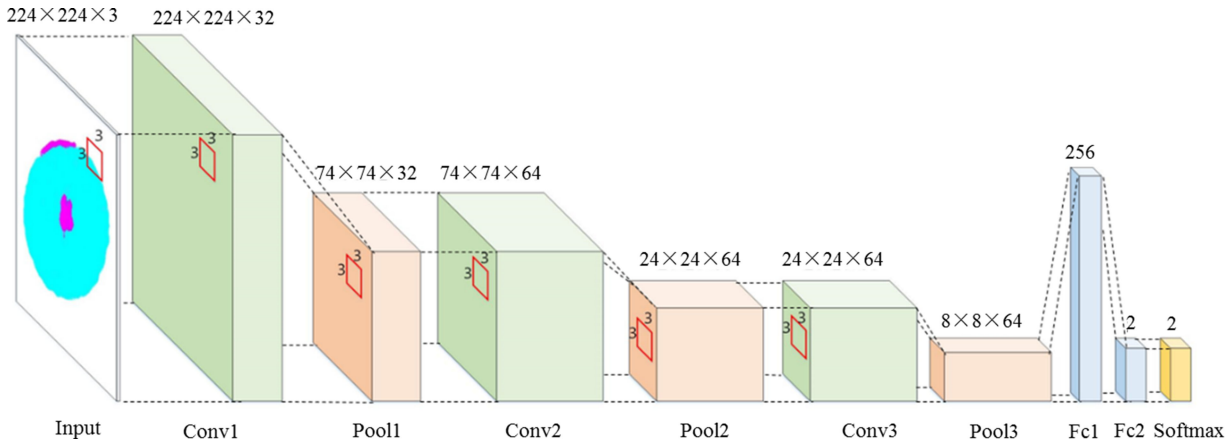

Fig. 3. Wafer defect pattern detection model diagram, the model consists of 3 convolutional layers, 3 pooling layers and 2 fully connected layers.

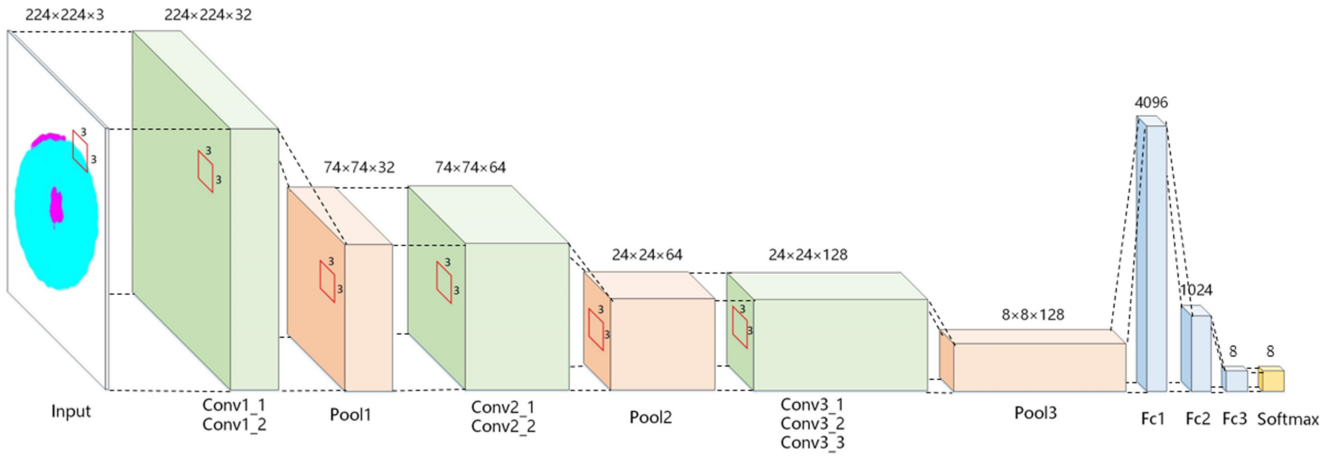

Fig. 4. Wafer defect pattern classification model diagram, the model is composed of 7 convolutional layers, 3 pooling layers and 3 fully connected layers.

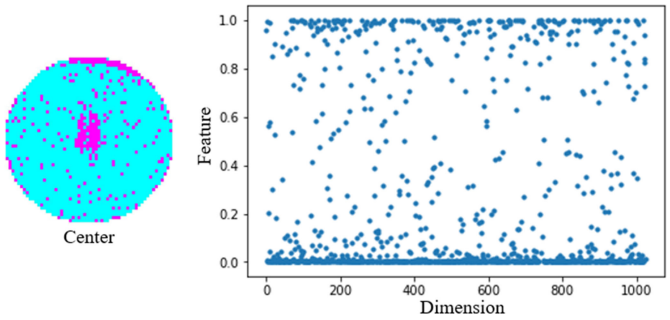

Fig. 5. Visualization results of Fc2 layer features. Most of the data is 0 and 1, the information is redundant.

TABLE I
SAMPLE SIZE OF EACH TYPICAL WAFER MAP PATTERN

Pattern	Sample size
Center	4294
Donut	555
Edge-Loc	5189
Edge-Ring	9680
Loc	3593
Random	866
Scratch	1193
Near-Full	149
None	147431
Total number	172950

the sample division of defective pattern (a set of 8 defective patterns samples) and *None* pattern.

A. Wafer Defect Pattern Detection Accuracy

We first detect whether there is defect pattern on the wafer map. This is a binary classification process, which depends on the detection model. Based on the dataset divided by Fig. 6(b), we used the wafer map with defective pattern and *None* pattern to train the model.

During training, samples were randomly cut and rotated to expand dataset. We used cross entropy as the loss function, and used Adam algorithm encapsulated in Tensorflow for optimization. Batch-size was 64, each round of iteration contained 520 batches. We chose the learning rate of 0.0001, and Fig. 7 shows the accuracy curve on training and validation set after each iteration.

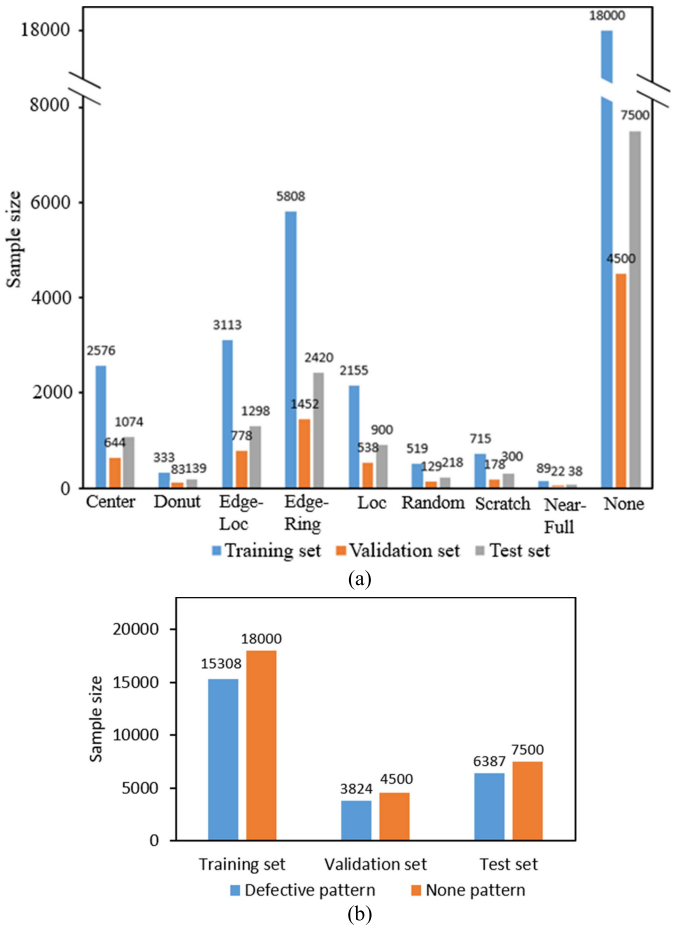

Fig. 6. (a) Sample distribution of 9 wafer map defect patterns, the proportion of training set, validation set and test set are 60%, 15% and 25%. (b) Sample distribution of defective pattern and *None* pattern.

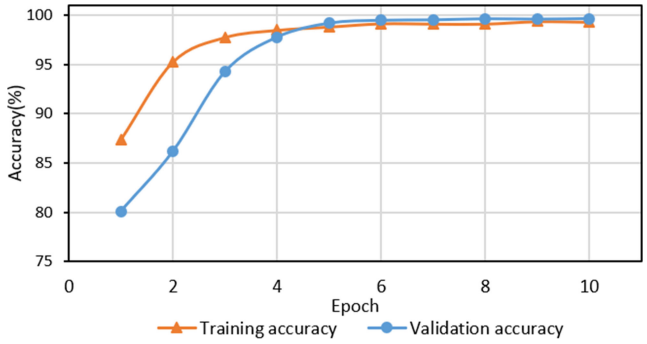

Fig. 7. The defect pattern detecting accuracy on training set and validation set during training process.

We evaluated the accuracy of the detection model on the test set. Table II lists the detection rates of each defect pattern, our model can completely distinguish *None* pattern and defective pattern. For each pattern, the model achieves a high accuracy rate, the average accuracy is 99.98%.

B. Wafer Defect Pattern Classification Accuracy

After the wafer map pattern is determined to be defective, the specific defect pattern should be classified. WM-811K wafer dataset contains 8 defect patterns, namely *Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-Full*. Based on the samples which divided according to Fig. 6(a), we trained and tested the wafer map defect pattern classification model. Due to the non-uniform distribution of sample quantity among these patterns, it is easy to overfit, so we randomly cut and rotate the training samples, and employ regular technique as described below.

TABLE II
ACCURACY OF EACH DEFECT PATTERN ON TEST SET

Defect patterns	Accuracy
Center	99.88%
Donut	100%
Edge-Loc	100%
Edge-Ring	99.95%
Loc	100%
Random	100%
Scratch	100%
Near-Full	100%
None	100%
Average	99.98%

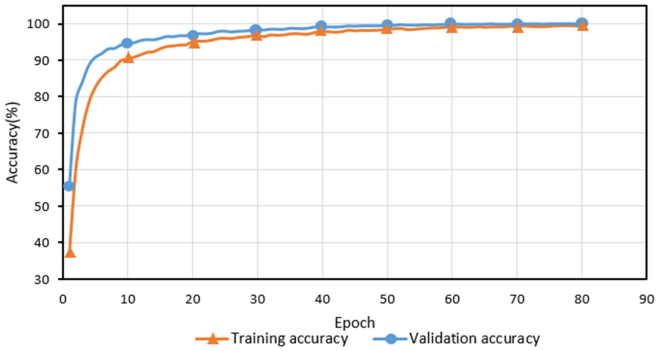

Fig. 8. The classification accuracy of defect patterns on training set and validation set during training process.

Adam optimization algorithm and cross entropy were still used in training, and we selected L2 regularization to limit parameter variation according to

$$L = E + \lambda \sum_j \omega_j^2, \tag{4}$$

where E is the original cost function, λ is the regularization coefficient and valued of 1×10^{-6} in our method. ω is the weight parameter value of each iteration. The sample batch-size was 128, each iteration contained 120 batches. Fig. 8 shows the classification accuracy of training and validation set while learning rate is 0.0001.

Fig. 9 is the confusion matrix for the test set. The average recognition rate for the 8 defect patterns is 93.25%. From the confusion matrix we can find that the model easily misclassified samples into *Loc* and *Scratch* patterns. The model's ability to recognize *Donut* and *Loc* patterns needs to be improved. In order to compare the results of SVM, Decision Tree and JLNDA [5] algorithms, we experimented on the test set. Since there are too few samples of *Donut* and *Near-Full* patterns, which is not conducive to the experiment, we first expand the

TABLE III
COMPARISION WITH JLNDA, SVM AND C4.5 ON TEST SET

Classifier	Proposed method	JLNDA[5]	SVM	C4.5
Center	91%	**93%**	89%	87%
Donut	86%	**98%**	98%	96%
Edge-Loc	**92%**	77%	76%	64%
Edge-Ring	**100%**	100%	98%	95%
Loc	**88%**	81%	69%	65%
Random	**95%**	94%	90%	83%
Scratch	**93%**	-	82%	71%
Near-Full	**100%**	-	100%	100%
Average	**93.13%**	90.5%	87.75%	82.63%

Prediction

Annotation	Center	Donut	Edge-Loc	Edge-Ring	Loc	Random	Scratch	Near-Full
Center	92.08% (989)	0.72% (1)	0.46% (6)	0	2.56% (23)	0.46% (1)	0	0
Donut	0.28% (3)	87.05% (121)	0	0	0.11% (1)	0	0	0
Edge-Loc	0.93% (10)	0	92.06% (1195)	0.58% (14)	6% (54)	0.92% (2)	0.67% (2)	0
Edge-Ring	0	0	0.15% (2)	95.87% (2320)	0	0	0	0
Loc	5.51% (59)	12.23% (17)	6.63% (86)	0	89.89% (809)	0.46% (1)	4.33% (13)	0
Random	0	0	0.08% (1)	0.17% (4)	0	94.03% (205)	0	0
Scratch	1.12% (12)	0	0.39% (5)	3.39% (82)	1.44% (13)	4.13% (9)	95% (285)	0
Near-Full	0.09% (1)	0	0.23% (3)	0	0	0	0	100% (34)

Fig. 9. Confusion matrix for the test set, the average accuracy is 93.25%.

TABLE IV
PRECISION OF EACH DEFECT PATTERN (%)

Defect	50	100	200	300	400	500	AP
Center	95.02	94.61	94.34	93.88	92.99	90.62	93.58
Donut	81.34	82.69	81.23	80.09	79.54	79.26	80.69
Edge-Loc	89.52	88.67	87.71	87.18	86.58	84.52	87.36
Edge-Ring	100	100	100	100	100	99.05	99.84
Loc	77.44	78.05	79.45	80.34	80.8	78.89	79.16
Random	97.38	97.59	97.87	97.85	97.47	96.13	97.38
Scratch	94.8	94.7	94.47	93.98	92.64	87.83	93.07
Near-Full	100	100	98.87	98.75	98.63	97.94	99.03

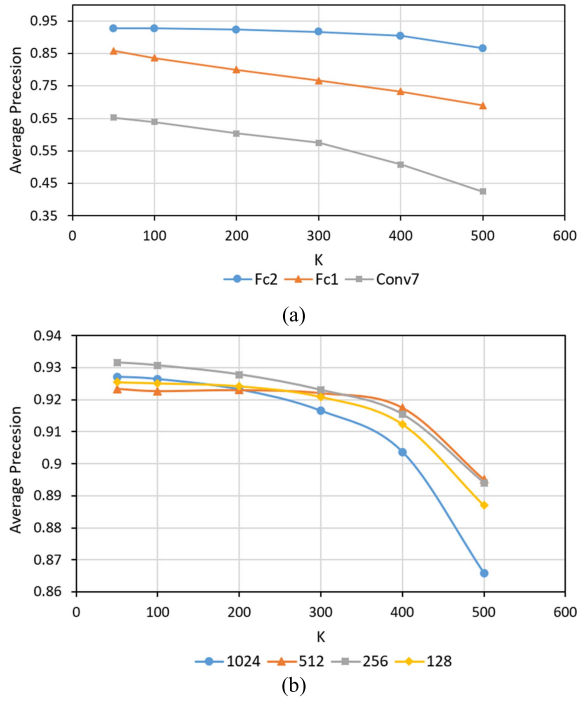

Fig. 10. (a) The accuracy curve of different layers' features. (b) The accuracy curve of different dimensions.

dataset of these two patterns by means of random rotation, scaling and adding gaussian noise. For each defect patterns, we randomly selected 100 samples from each defect pattern category in test set for testing. As can be seen from Table III, our model has a better ability to classify wafer map defect patterns. But the classification ability of *Donut* and *Loc* of our model have a weak performance, which is what we need to improve in the future.

C. Result of Defect Root Causes Analysis

We determined the feature layer and feature dimension through experiments, and carried out interval mapping for feature extracted, then ranked similarity according to hamming distance. The experimental process is as follows.

In the experiment of selecting feature layer and feature dimension, we selected 500 samples from each pattern category as a query database, therefore there was a total of 4000 samples in the database. Then, we randomly selected 300 samples from the remaining samples as the retrieved samples. The precision rate was token as the index to evaluate the performance of retrieval model, we divided the retrieval results into K grades (top K samples returned) and calculate the precision of each grade. Firstly, we evaluated the different layers of the classification model. Fig. 10 (a) shows the precision curves of Conv3_3, Fc1 and Fc2 features, and the feature expression ability of Fc2 layer was significantly better than other two layers. Then we reduced the dimensions of Fc2 layer's feature based on PCA algorithm, as shown in Fig. 10 (b). The features after dimension reduction are better than the original features, and the retrieval performance reach the best performance when the dimension was reduced to 256. We determined the appropriate dimension through comparative experiment, which is more conducive to comparing the experimental effects of different dimensions. However, cross-validation [22] is also a recommended method for PCA dimension determination.

To further verify the retrieval capability of the model, we randomly selected a sample from each pattern and calculated the precision. The results are shown in Table IV, the first row represents top K samples. With the value of K increases, the

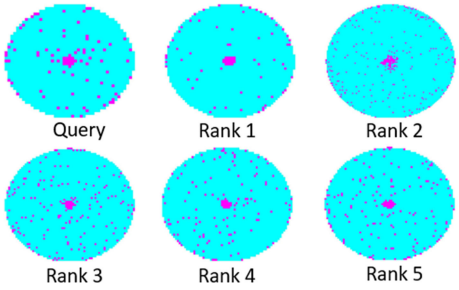

Fig. 11. Query results and sorted by similarity.

precision decreases, which indicates that the correct search results are concentrated in the range with a small value of K, our method can give priority to retrieving the correct samples, which also proves the reliability of the proposed retrieval method. AP measures the *Average Precision* of the top 500 query results for each category, we got the AP according to

$$AP = \frac{\sum_{k=1}^{500} \frac{n}{k} \times rel(k)}{\mathrm{m}}, \tag{5}$$

where k represents top k returned samples. n is the number of related samples, $rel(k)$ indicates whether the sample at position k is relevant, 0 means irrelevant, otherwise it is 1. m represents the total number of relevant samples in top 500. In the process of querying, the precision will be calculated when $rel(k) = 1$, and AP will be obtained by taking the average values of precision. As can be seen from the Table IV, the position k of the relevant samples k concentrated in the range with a small value of K grades. Our model performs worse on *Donut* and *Loc* patterns and greatly in others. We used mean average precision (MAP) to evaluate the overall performance and got the value of MAP is 91.26%, it expressed as follows:

$$MAP = \frac{\sum_{i=1}^{6} AP^{(i)}}{6}, \tag{6}$$

By measuring and ranking the similarity of features after dimension reduction, we selected the first three largest similarity samples as the queried result, and took the defect causes of queried samples as root causes of the tested sample. Each category of the dataset was taken as a database, and the tested sample was retrieved in its category database. In the case of *Center* which contains 4293 samples, Fig. 11 shows the top 5 sorted similar samples retrieved from the database, the retrieval time is 0.0098s.

D. Effectiveness of Detection-Classification Strategy

We apply the *Detection-Classification* strategy to identify the wafer map defect patterns, this is because the occurrence probability of defect pattern is low in the actual manufacturing process, most wafer maps are *None* pattern. However, the detection model proposed by us is simpler in structure and faster in calculation speed than the classification model. When the tested samples are *None* pattern, there is no need to carry out model classification, which reduces the recognition time of large quantities of wafer maps. To highlight this feature, we designed a complete recognition process based on a *None* pattern, a *Center* pattern and a *Loc* pattern respectively. We use *Detection-Classification* strategy to inspect *None* and *Center* patterns, while *Loc* pattern is directly inspected by classification model. Table V shows the time of testing three samples. The computation time of classification model is nearly twice that of detection model. If we input samples directly into the classification model like *Loc* pattern without defect detection, it will take about 0.0038s for each sample, which is not suitable for large-scale wafer map inspection with low defect rate. Therefore, our scheme is more applicable to actual wafer map inspection.

TABLE V
COMPARISION OF CALCULATION SPEED

Recognition process	None	Center	Loc
Detection	0.0021s	0.002s	-
Classification	-	0.0037s	0.0038s
Total time	0.0021s	0.0057s	0.0038s

V. CONCLUSION

In this paper, we propose a method for wafer map defect pattern recognition and analysis based on convolutional neural networks. We built defect pattern detection model and classification model based on CNN, and trained on real wafer map dataset. Experiments show that our model has better recognition ability than other methods and the *Detection-Classification* scheme is more suitable for large-scale wafer map inspection. In addition, this paper studies the feature expression ability of different layers and different dimensions for retrieval, and realizes the root causes analysis of wafer defect pattern based on the middle layer features of classification model and similarity ranking.

For the future research, we will focus on the ability of the model to learn new pattern. We will judge whether the defect is a new pattern or not according to the confidence of the classification model, and combine the few-shot learning algorithm to learn the new defect category.

REFERENCES

[1] S.-C. Hsu and C.-F. Chien, "Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing," *Int. J. Prod. Econ.*, vol. 107, no. 1, pp. 88–103, May 2007.
[2] Y. Han and P. Shi, "An adaptive level-selecting wavelet transform for texture defect detection," *Image Vis. Comput.*, vol. 25, no. 8, pp. 1239–1248, Aug. 2007.
[3] J. Y. Hwang and W. Kuo, "Model-based clustering for integrated circuit yield enhancement," *Eur. J. Oper. Res.*, vol. 178, no. 1, pp. 143–153, Apr. 2007.
[4] C.-H. Wang, W. Kuo, and B. Halima, "Detection and classification of defect patterns on semiconductor wafers," *IIE Trans.*, vol. 38, no. 12, pp. 1059–1068, 2006.
[5] J. Yu and X. Lu, "Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis," *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 1, pp. 33–43, Feb. 2016.
[6] F.-L. Chen and S.-F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Trans. Semicond. Manuf.*, vol. 13, no. 3, pp. 366–373, Aug. 2000.
[7] C.-F. Chien, S.-C. Hsu, and Y.-J. Chen, "A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence," *Int. J. Prod. Res.*, vol. 51, no. 8, pp. 2324–2338, Apr. 2013.

[8] L. Mika and Y. Hiltunen, "Recognition of systematic spatial patterns in silicon wafers based on SOM and K-means," *IFAC PapersOnLine*, vol. 51, no. 2, pp. 439–444, 2018.

[9] S. Lee and D. Kim, "Distributed-based hierarchical clustering system for large-scale semiconductor wafers," in *Proc. Int. Conf. Ind. Eng. Eng. Manag. (IEEM)*, Dec. 2018, pp. 1528–1532.

[10] S. J. Bae, J. Y. Hwang, and W. Kuo, "Yield prediction via spatial modeling of clustered defect counts across a wafer map," *IIE Trans.*, vol. 39, no. 12, pp. 1073–1083, Dec. 2007.

[11] 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.

[12] M. Piao, C. H. Jin, J. Y. Lee, and J.-Y. Byun, "Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 250–257, May 2018.

[13] 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, pp. 309–314, May 2018.

[14] F. Adly, P. D. Yoo, S. Muhaidat, Y. Al-Hammadi, U. Lee, and M. Ismail, "Randomized general regression network for identification of defect patterns in semiconductor wafer maps," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 2, pp. 145–152, May 2015.

[15] M. Fan, Q. Wang, and B. V. D. Waal, "Wafer defect patterns recognition based on OPTICS and multi-label classification," in *Proc. Adv. Inf. Manag. Commun. Electron. Autom. Control Conf.*, 2017, pp. 912–915.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.

[17] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704–1716, Sep. 2012.

[18] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 8689. Cham, Switzerland: Springer, 2014, pp. 584–599.

[19] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[20] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Boston, MA, USA, Jun. 2015, pp. 27–35.

[21] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 496–506, Feb. 2015.

[22] J. Camacho and A. Ferrer, "Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: Practical aspects," *Chemometr. Intell. Lab. Syst.*, vol. 131, pp. 37–50, Feb. 2014.