Unsupervised Wafer Defect Classification Model Based On Joint Reconstruction And Clustering

Luyang Jie [1, a, *], Yilong Guo [1, b], Yiming Yao [1, c], and Yongkang Liu [1, d]

[1] School of Microelectronics, Shanghai University, Shanghai, 201899, China

[a, *] jly@she.edu.cn, [b] wdz_gyl@shu.edu.cn, [c] yaoym@shu.edu.cn, [d] yk_liu@shu.edu.cn

Abstract. Defect detection of unpatterned wafer is very important for determining the causes of wafer defects, and it is also a significant way to improve production yield. At present, the defect detection model based on the deep learning method has been widely used and has shown promising performance. However, the labor cost of supervised learning method based on labeled samples is very expensive, so the wafer defect detection model based on unsupervised learning is a future research direction. In this paper, we propose an unsupervised wafer defect classification model(UWDDM), in which the reconstruction loss is used to train the feature extractor of wafer defects, and the classifier is trained based on the cross entropy loss of clustering results and classification results. The experimental results on the real dataset MixedWM38 show that the proposed model has higher recognition accuracy and less inference time than other unsupervised models. This unsupervised learning method has great potential for wafer defect detection. It can avoid the expensive manual labeling cost and provide an effective way to automatically detect wafer defects.

Keywords: unpatterned wafer; defect detection; unsupervised learning; clustering.

1. Introduction

The silicon wafer is used for manufacturing semiconductor chips. Wafer defect detection is a key period in the semiconductor manufacturing process, and its accuracy and efficiency have an important impact on the quality and production cost of semiconductor products [1]. With the continuous development of semiconductor technology, the manufacturing process of wafers has become more and more complicated, and the types of defects on wafers have become more diverse. Therefore, it is a very challenging task to accurately identify wafer defects. With the development of machine vision technology, the study of wafer images has become a hotspot in wafer defect detection [2,3]. Neural network model based on supervised learning has achieved excellent results in wafer defect detection [4,5]. However, it is expensive and time-consuming work to obtain enough labeled data for wafer defect detection. Therefore, unsupervised learning has become an attractive method to reduce the data marking burden of supervised learning.

In the past few years, the unsupervised learning method has made remarkable progress in the field of wafer defect detection. In order to detect the system failure modes at the wafer level, Alawieh M B et al. [6] proposed to use the binary test results of all dies on multiple wafers to cluster these wafers according to their failure feature spaces, and finally identify the potential failure system modes. Weber C et al. [7] used automatic encoders and clustering to identify and group unknown defect patterns. Geng S et al. [8] proposed a method of wafer surface defect pattern recognition based on unsupervised learning. This method fine-tuned the classification model through a small number of label images with wafer defect patterns, which can accurately identify various types of defect patterns and show its advantages on the real data set WM-811K. Fan S K S et al. [9] proposed an anomaly detection method, which used an automatic de-noising encoder to learn the main representation of normal wafers from the readings of equipment sensors and used it as a single-class classification model.

The purpose of this paper is to explore the application of unsupervised learning methods in wafer defect detection, so as to solve the problem of data labeling faced by supervised learning. We hope to improve the accuracy and efficiency of wafer defect detection by deeply studying various

unsupervised learning technologies and applying them to wafer datasets. Through this research, we are expected to provide more feasible wafer defect detection solutions for the semiconductor manufacturing industry, thus improving product quality, reducing production costs, and promoting the further development of semiconductor technology.

In this paper, we have the following contributions: (i) A wafer defect detection method based on unsupervised learning is studied, which ensures the accuracy of defect identification on the premise of saving a lot of labor costs. (ii) Compared with other unsupervised defect detection methods, the proposed defect detection method has a faster inference speed.

2. Methodology

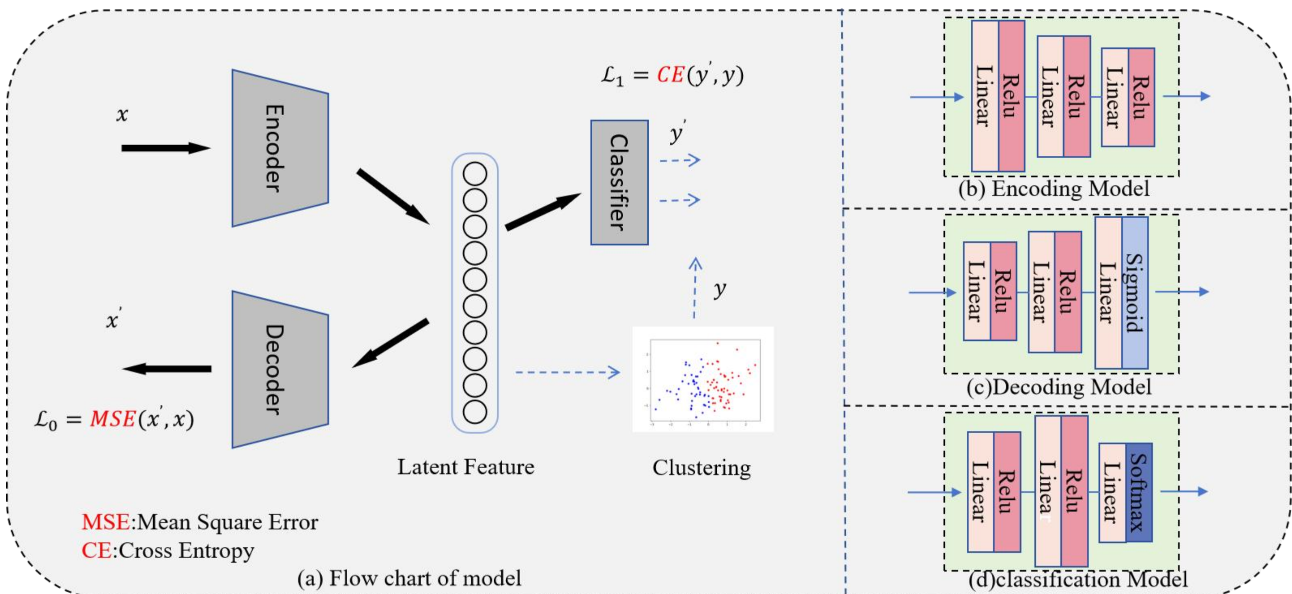

Fig. 1 Overall framework of model. (a) is the overall process of model training, and (b) to (c) are the concrete realization of the encoder, decoder, and classifier modules respectively.

This section describes the UWDDM, as shown in Fig. 1, which mainly consists of three parts: feature encoder, feature decoder, and classifier. Mean square error loss and cross-entropy loss are used to optimize the weight parameters of the model.

2.1 Data Preprocessing

Because some wafer defects are too small to be easily found, we resize all the data to 224×224, which has another advantage that it is convenient for us to use other research methods for experiments, because in other research methods, the feature map may disappear after many convolution operations because the output samples are too small. In addition, on the premise of ensuring the training data is as real as possible, in order to enhance the generalization ability of the model, we use horizontal flipping and vertical flipping to enrich the data of the sample. Through data augmentation, we can expand the training dataset and improve the robustness of the model for different transformations and perspectives. Finally, data regularization is used to make the data more consistent with the independent and identically distributed conditions, make the data better adapted to the saturation region of the activation function, and speed up the training of the model.

2.2 Feature Extraction

The feature extraction method proposed in this paper is based on the reconstruction loss and learns the low-dimensional representation of data by minimizing the reconstruction loss. Specifically, we first build a self-encoder network, as shown in Fig. 1(a). Among them, the encoder (as shown in Fig. 1(b)) maps the original data x into a low-dimensional space, and the decoder (as

shown in Fig. 1(c)) reconstructs the low-dimensional x_{latent} representation into the original data $x^{'}$. The formula is described as follows:

$$x_{latent} = f(Wx + b)\#(1)$$

$$x^{'} = g\left(f\left(W^{'}x_{latent} + b^{'}\right)\right)\#(2)$$

Where W and $W^{'}$ are the weight matrices of encoder and decoder, b and $b^{'}$ are offset vectors, and g and f are activation functions.

It is worth emphasizing that we choose the three-layer fully connected layer as the middle hidden layer of the encoder and decoder. We found through experiments that compared with the convolution module, multilayer perceptron has a stronger ability in feature extraction. The reason why we choose the three-layer fully connected layer is that too few layers will lead to the failure of the model to effectively extract sample defect features, while too many layers will increase the parameters of the model and cause unnecessary waste of resources. Next, we define the reconstruction loss function, which is used to measure the difference between the original data and the reconstructed data. In order to prevent gradient explosion and gradient disappearance during model training, we introduce the ReLU activation function into the encoder and decoder. In addition, the ReLU activation function also has the advantage of keeping the model sparse. Finally, we use optimization algorithms such as gradient descent to minimize the reconstruction loss $\mathcal{L}_0$, so as to learn the low-dimensional representation of data. Through an iterative optimization process, we can gradually improve the performance of the feature extraction model. Feature extraction based on reconstruction loss is described as follows:

$$\mathcal{L}_0\left(x, x^{'}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(x - x^{'}\right)^2 \#(3)$$

2.3 Classifier

Although the sample label can be obtained by clustering the low-dimensional features of the samples extracted by the encoder, the clustering-based method has certain requirements for the number of samples and its time consumption is too high. So we designed a classifier to learn the clustering function, hoping to reduce the inference time of the model while ensuring the clustering accuracy. Since the classifier is too simple to learn the clustering function well, it is necessary to map the low-dimensional vector to the high-dimensional space first, and then to the cluster space to output the cluster to which it belongs, as shown in Fig. 1(d). The potential features output by the encoder are simultaneously input into the classifier and the clustering module. The output of the clustering module is used as the corresponding label of the output of the classifier to calculate its cross-entropy loss $\mathcal{L}_1$. The gradient descent optimization algorithm is adopted to optimize the classifier with the cross-entropy loss function as the objective function. The description of optimizing the classifier according to the clustering results is as follows:

$$y^{'} = Classifier(x_{latent})\#(4)$$

$$y = Cluster(x_{latent})\#(5)$$

$$\mathcal{L}_1\left(y, y^{'}\right) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\left(y_{ij}\log y^{'}_{ij}\right)\#(6)$$

In the above formula, $Classifier$ and $Cluster$ are classifier and clustering operation respectively, and $y^{'}$ and y are their corresponding output results respectively.

3. Experimental and Analysis

3.1 Dataset Description

The dataset used in this study is the public dataset MixedWM38 [10], which consists of one kind of normal sample, eight kinds of single defect samples, and 29 kinds of mixed defect samples. In

this study, there are 1600 normal and defective samples in the training set and 400 normal and defective samples in the test set. The resolution of each sample is 52×52, and there are only three values of 0, 1, and 2 in each sample. For the convenience of demonstration, we map the pixel value of the sample to the range of 0~255 and convert it into a color map, as shown in Fig. 2.

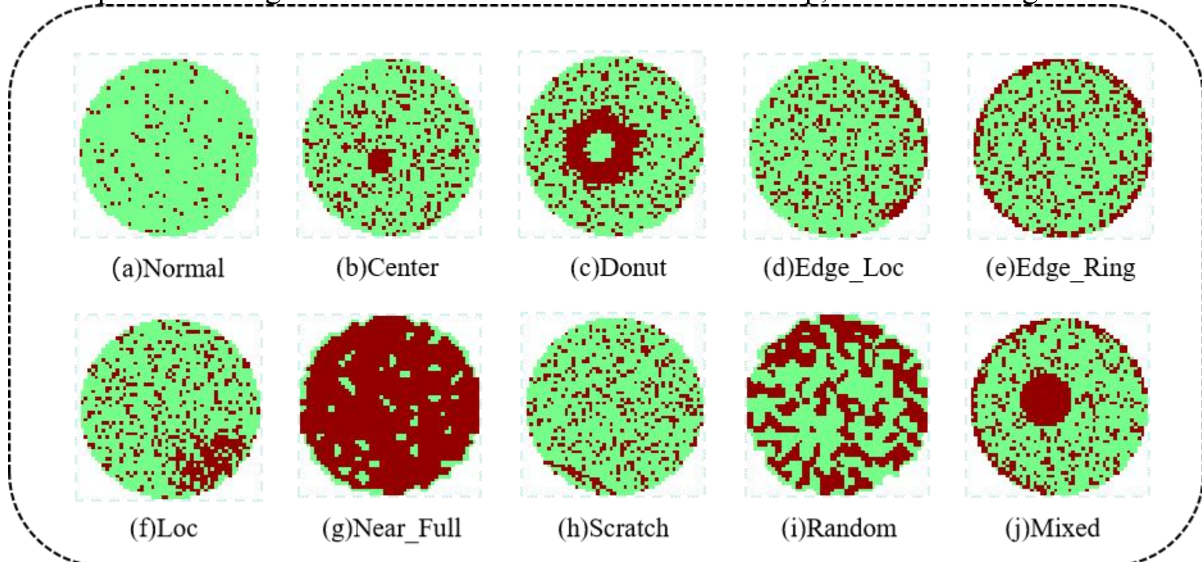

Fig. 2 Sample of wafer defect data. (a) is a normal sample, (b)~(i) is a single defect sample, and (j) is a mixed defect sample.

3.2 Result and Analysis

In order to explore a defect detection method based on unsupervised learning, we mainly compare three common mainstream unsupervised learning methods based on autoencoder, contrast learning, and deep clustering. The dimension of the low-dimensional vector output by the above model is set to 10, and then the clustering of the output low-dimensional vector is obtained by K-means clustering. After tuning each model, the learning rate and optimizer selection of the final model are shown in Table 1.

Table 1. Learning rate(LR) and optimizer of different models.

Model	AE [11]	CAE [12]	SAE [13]	DAE [14]	simCLR [15]	Deepcluster [16]	UWDDM
LR	0.00001	0.001	0.01	0.0001	0.0003	0.05	0.01
Optimizer	Adam	Adam	Adam	Adam	Adam	SGD	Adam

The performance of the model is evaluated by cluster analysis indicators, such as Adjusted Rand Index(ARI), Adjusted Mutual Information(AMI), V-measure, and Fowlker-Mallows Index(FMI). The results are shown in Table 2.

Table 2. Comparison of ARI, AMI, FMI and V-measure of different unsupervised methods.

Model	ARI	AMI	FMI	V-measure
AE	1.0000	1.0000	1.0000	1.0000
CAE	0.6716	0.6433	0.8381	0.6442
DAE	0.9701	0.9436	0.9850	0.9437
SAE	0.9214	0.8779	0.9606	0.8782
simCLR	0.8927	0.8194	0.9462	0.8197
Deepcluster	0.7915	0.7440	0.8961	0.7445
UWDDM	1.0000	1.0000	1.0000	1.0000

As can be seen from the above table, because there are only three types of pixel values in the wafer defect image, the performance indexes of each model are still relatively high. It is worth noting that although the unsupervised learning method based on contrast learning and deep clustering has achieved remarkable results in other tasks, the method based on autoencoder is more effective in unsupervised wafer defect detect tasks. We believe that the training method based on

reconstruction loss can capture the deep feature information of wafer defects more effectively, which is also an important reason why our method optimizes the feature extractor with reconstruction loss as the objective function. After adjusting the parameters of various models, AE and UWDDM can achieve the best performance.

Table 3. Comparison of inference time of different unsupervised methods.

Model	Single sample	Batch samples
AE	246	1513
CAE	254	2435
DAE	258	1510
SAE	244	1497
simCLR	241	2686
Deepcluster	240	2268
UWDDM	3	1244

Although the performance of the model has reached our expected goal, the inference time of clustering low-dimensional features is too high. As shown in Table 3, for each method, the inference time of a single sample and batch sample is tested separately, the time unit is milliseconds, and the average of the three test results is taken as the final result. Because of the particularity of the clustering method, the test of a single sample is based on the known data of 200 samples (of course, the inference time of a single sample will be different because of the different number of known samples). The data volume of batch samples is 400. From the above table, it can be seen that UWDDM is about 250 milliseconds faster than other models in the inference of batch samples, and UWDDM is far faster than other models in the inference of single samples.

4. Conclusion

This study explores the feasibility of unsupervised learning methods for unsupervised wafer defect detection and proposes an end-to-end unsupervised wafer defect detection model based on reconstruction loss and clustering methods. Compared with other unsupervised methods, the model proposed in this paper not only ensures its defect detection accuracy but also reduce the model inference time. In future research, we will further explore the generalization ability and robustness of the model and apply it to the actual wafer defect detection task.

References

[1] Shankar N G, Zhong Z W. Defect detection on semiconductor wafer surfaces[J]. Microelectronic engineering, 2005, 77(3-4): 337-346.

[2] Kim Y, Cho D, Lee J H. Wafer map classifier using deep learning for detecting out-of-distribution failure patterns[C]//2020 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA). IEEE, 2020: 1-5.

[3] Saqlain M, Jargalsaikhan B, Lee J Y. A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing[J]. IEEE Transactions on Semiconductor Manufacturing, 2019, 32(2): 171-182.

[4] Wang J, Xu C, Yang Z, et al. Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition[J]. IEEE Transactions on Semiconductor Manufacturing, 2020, 33(4): 587-596.

[5] Cheon S, Lee H, Kim C O, et al. Convolutional neural network for wafer surface defect classification and the detection of unknown defect class[J]. IEEE Transactions on Semiconductor Manufacturing, 2019, 32(2): 163-170.

[6] Alawieh M B, Wang F, Li X. Identifying wafer-level systematic failure patterns via unsupervised learning[J]. IEEE transactions on computer-aided design of integrated circuits and systems, 2017, 37(4): 832-844.

[7] Weber C, Tripuramallu A, Czerner P, et al. Clustering wafer defect patterns within the semiconductor industry based on wafer Maps, using an agile unsupervised deep learning approach[C]//2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2021: 1913-1918.

[8] Geng S, Liu H, Wang F, et al. Unsupervised Learning for Wafer Surface Defect Pattern Recognition[C]//Proceedings of 2021 Chinese Intelligent Automation Conference. Springer Singapore, 2022: 275-285.

[9] Fan S K S, Hsu C Y, Jen C H, et al. Defective wafer detection using a denoising autoencoder for semiconductor manufacturing processes[J]. Advanced Engineering Informatics, 2020, 46: 101166.

[10] Wang J, Xu C, Yang Z, et al. Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition[J]. IEEE Transactions on Semiconductor Manufacturing, 2020, 33(4): 587-596.

[11] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks[J]. Advances in neural information processing systems, 2006, 19.

[12] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]//Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer Berlin Heidelberg, 2011: 52-59.

[13] Ng A. Sparse autoencoder[J]. CS294A Lecture notes, 2011, 72(2011): 1-19.

[14] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. Journal of machine learning research, 2010, 11(12).

[15] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.

[16] Caron M, Bojanowski P, Joulin A, et al. Deep clustering for unsupervised learning of visual features[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 132-149.