결함 탐지 시스템을 위한 반지도 학습 활용

응우옌티호아이투, 한동석*

경북대학교 전자전기공학부

thunguyen@knu.ac.kr, *dshan@knu.ac.kr

Utilization of Semi-Supervised Learning for Defect Detection Systems

Nguyen Thi Hoai Thu, Dong Seog Han

School of Electronic and Electrical Engineering, Kyungpook National University

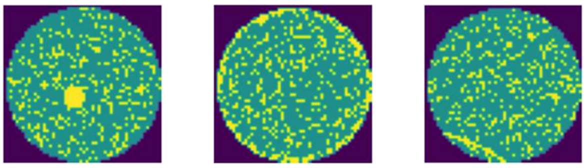
Abstract

Defect detection plays a crucial role in many automatic systems, helping to prevent system crashes, send alerts to users and assure the product quality. Common deep learning-based defect detection systems are often trained on large labeled datasets using supervised learning methods in order to make robust classification results. However, it is expensive and time-consuming to annotate big datasets. In this paper, we proposed a deep learning-based classification model for the wafer map defect detection system constructed by a wide residual network and a semi-supervised learning algorithm to utilize the vast amount of unlabeled data. Our experiment results on the MixedWM38 dataset indicate that by using both labeled and unlabeled data, the semi-supervised learning-based system outperforms the baseline systems that are trained by a supervised learning method on labeled data only.

I. Introduction

Wafer, in the manufacturing of integrated circuits (ICs) and micro-devices, is a thin slice of semiconductor material, such as monocrystalline silicon. It is used as the substrate for microelectronic circuits to be built layer by layer upon the wafer. Due to the thin characteristic and many microfabrication processes such as doping, oxidation…, wafers are extremely sensitive to defects. Although advanced cleanroom conditions are strictly required, wafers are still highly prone to defects due to the microscopic scale.

To overcome this challenge, different real-time defect detection methods have been studied, such as optical inspection and e-beam review. In these methods, a graphical representation of a wafer called wafer map is generated. It can contain physical parameters that are measured on the wafer [1] or can directly show the location and type of test results for each individual chip. Examples of the wafer map in the MixedWM38 dataset are shown in Fig. 1. As wafer maps provide important information to identify the root causes of chip failures, they are widely used in recent deep learning-based wafer defect detection systems [2]. With the great success of deep learning algorithms, these systems have achieved high detection accuracies [3]. These systems utilize deep learning networks to take wafer maps as input, extract salient features, and output the defect type and the defect area. However, most of these systems are trained by supervised learning algorithms with the requirement of large labeled datasets, which are often expensive and time-consuming to collect. Moreover, manual data annotation performed by human experts is highly subjective.



a) Center b) Edge ring c) Scratch

Fig 1. Examples of different defect patterns on wafer maps in the MixedWM38 dataset

Therefore, in this study, instead of the commonly used supervised learning approach, we apply different algorithms of the semi-supervised learning (SSL) approach to utilize the vast amount of unlabeled wafer map data that is already available during the wafer manufacturing process.

II. Methodology

The overall architecture of the proposed semi-supervised learning-based wafer map defect detection system is shown in Fig. 2. The system contains two main components: a deep learning neural network and a semi-supervised learning algorithm. We use a wide residual network [4] for extracting the features from input wafer maps and classifying the defect

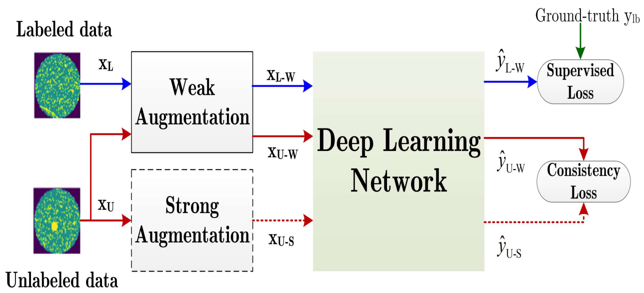


Fig 2. Overall architecture of the semi-supervised learning-based wafer map defect detection system.

pattern. The semi-supervised learning algorithm, with two loss functions: supervised loss and consistency loss, is utilized for training the network on both labeled and unlabeled data. In this study, we leveraged two of the most powerful SSL algorithms: FixMatch [5] and ReMixMatch [6] for our experiments.

We implement two types of augmentation: weak and strong. The weak augmentation applies some light transformations, including random crop, random horizontal/vertical flip, whereas the strong augmentation has some additional transformations, including RandAugment, CutOut methods. The labeled wafer map $\boldsymbol{x}_L$ indicated in blue color is input to the weak augmentation before being forwarded to the deep learning network for defect classification. A cross-entropy loss function is applied to the predicted output $\hat{y}_L$ and the ground-truth label y_L to calculate the supervised loss. On the other hand, the unlabeled data $\boldsymbol{x}_U$ is input to both weak and strong augmentation modules to create two different versions of the input wafer map. The core idea here is to use a consistency loss to enforce the model to output the same defect class. By combining both the supervised loss of the labeled data and the consistency loss of the unlabeled data, the deep learning network is capable of learning important information from the unlabeled data.

We conduct our experiments on the MixedWM38 dataset [3]. The dataset contains more than 38000 wafer maps of 1 normal pattern and 37 defect patterns, including 8 single defect patterns (center, donut, edge-loc, edge-ring, loc, near full, scratch, random) and 29 mixed defect patterns. Some of the defect patterns are shown in Fig. 1. The dataset is split into a training set and a test set with the ratio of 80% and 20%, respectively. From the training set, we randomly select 5 subsets to be used as labeled data with sizes of 100, 200, 400, 600 and 800 samples. The rest of the training data is used as unlabeled data.

Results of the experiments are shown in Fig. 3. It can be clearly seen that models trained by the two SLL algorithms outperformed the one trained by a common supervised learning method. The ReMixMatch method gains the best performance compared to the other two methods, especially, when there are only 200 labeled samples which are about 5 samples per class, it achieves an accuracy of 92.53%.

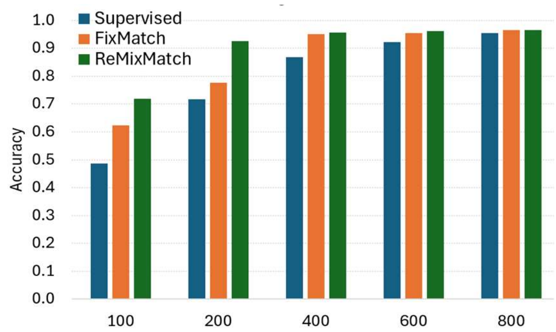


Fig 3. Comparison results of different learning methods on the deep learning-based wafer map defect detection system.

III. Conclusion

In this study, we proposed a deep learning-based wafer map defect detection using a wide residual network and two semi-supervised learning algorithms to utilize both labeled and unlabeled data. The semi-supervised learning-based systems outperform the common supervised learning method, thus helping to reduce the cost of data annotation. In future work, we will consider different SLL algorithms and apply them to the wafer map defect detection systems.

ACKNOWLEDGMENT

This research was supported by the Core Research Institute Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2021-NR060127).

References

[1] T. Yuan, W. Kuo, and S. J. Bae, "Detection of spatial defect patterns generated in semiconductor fabrication processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, no. 3, pp. 392–403, 2011.

[2] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Transactions on Semiconductor Manufacturing*, vol. 31, no. 2, pp. 309–314, 2018.

[3] J. Wang, C. Xu, Z. Yang, J. Zhang, and X. Li, "Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition," *IEEE Transactions on Semiconductor Manufacturing*, vol. 33, no. 4, pp. 587–596, 2020.

[4] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

[5] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.

[6] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel, "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring," *arXiv preprint arXiv:1911.09785*, 2019.