

# Mixed-Type Wafer Failure Pattern Recognition (Invited Paper)

Hao Geng  
ShanghaiTech University

Qi Sun  
Chinese University of Hong Kong

Tinghuan Chen  
Chinese University of Hong Kong

Qi Xu  
University of Science and Technology  
of China

Tsung-Yi Ho  
Chinese University of Hong Kong

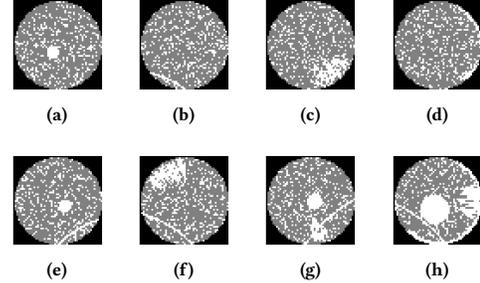
Bei Yu  
Chinese University of Hong Kong

## ABSTRACT

The ongoing evolution in process fabrication enables us to step below the 5nm technology node. Although foundries can pattern and etch smaller but more complex circuits on silicon wafers, a multitude of challenges persist. For example, defects on the surface of wafers are inevitable during manufacturing. To increase the yield rate and reduce time-to-market, it is vital to recognize these failures and identify the failure mechanisms of these defects. Recently, applying machine learning-powered methods to combat single defect pattern classification has made significant progress. However, as the processes become increasingly complicated, various single-type defect patterns may emerge and be coupled on a wafer and thus shape a mixed-type pattern. In this paper, we will survey the recent pace of progress on advanced methodologies for wafer failure pattern recognition, especially for mixed-type one. We sincerely hope this literature review can highlight the future directions and promote the advancement of the wafer failure pattern recognition.

## 1 INTRODUCTION

As the physical size of transistors continues to shrink without worsening the performance of chips, an increasing number of integrated circuit components can be patterned and then etched onto wafers, thus introducing more functionality and memory capacity. There isn't a free lunch, though. The probabilities of defects from modern fabrication process (including photolithography, etching deposition, and metallization) on the surface of the wafers increase as well. Owing to various processes, the types of defects are also diverse. What's worse, different defects are coupled on the same wafer because of recent advances in downsizing technology node and an increase in wafer size. Typically, after wafer is fabricated, several tests are performed to check the function of each die on wafer via a wafer probe. The testing results for the dies are represented as binary values on the wafer map, and defective dies on a wafer map are very likely to converge into a specific distribution pattern. Based on this prior information, experienced experts manually analyze and recognize these failure patterns in manufacturing processes to improve yield. This procedure, known as wafer defect pattern recognition, offers hints and insights into how to improve yield by reasoning the root cause of defects during fabrication. However,



**Figure 1: The nine kinds of wafer map patterns in MixedWM38: (a) Center; (b) Scratch; (c) Location; (d) Edge-Location; (e) 2 single-type defects mixed (Center, Scratch); (f) 2 single-type defects mixed (Scratch, Location); (g) 3 single-type defects mixed (Center, Scratch, and Location); (h) 4 single-type defects mixed (Center, Scratch, Location, and Edge-Location). Each wafer image has  $52 \times 52$  pixels (dies) with 3 pixel levels: 0, 127 and 255. Locations with pixel level 0 (i.e., the black pixels in wafer images) are not part of the wafer. Grey pixels with pixel level 127 represent dies with a passing label, while white pixels refer to those with a fail label.**

this human visual inspection takes a lot of time and is very subjective. Therefore, automated and artificial intelligence-powered wafer defect recognition process is in demand.

Generally, the defects can coarsely fall into three categories [1]: random kind, systematic and repeatable type, the combination of previous two defects. The first type is mainly caused by particles such as dust that scattered all over the wafer surface. Hence, there is no specific clustering pattern. This influence on yield can be weakened by improving the stability and accuracy of the fabrication. The second kind of defects have obvious clustering phenomenon. The factor of this category is mask-induced or radial variations during the photolithography [2]. The last one is most common. Almost all the defective wafer maps in the most broadly adopted academia benchmarks such as WM-811K [3] and MixedWM38 [4] belong to the third type. More specifically, WM-811K dataset contains 8 groups of defective patterns collected from industry, while MixedWM38 covers 38 classes of frequently occurred mixed-type defect patterns where 29 categories of mixed defects and 9 types of sing-type wafer patterns exist. For further explanation, we exemplify some sorts of wafer defect map visualizations from MixedWM38 in Figure 1.

It is recognized that distinct categories of defects have varied distribution patterns. For instance, the defects in the Center are concentrated around or near the center of the wafer in circular or

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
ASPDAC '23, January 16–19, 2023, Tokyo, Japan  
© 2023 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9783-4/23/01.  
<https://doi.org/10.1145/3566097.3568363>

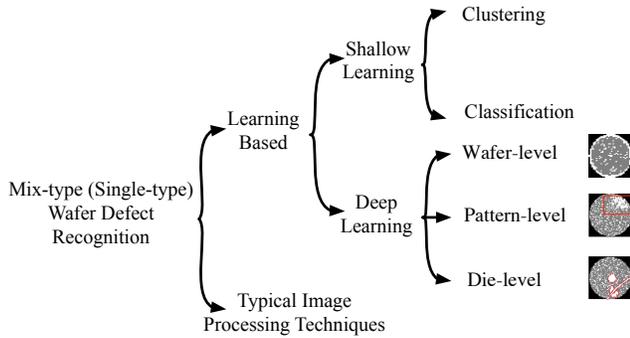


Figure 2: The taxonomy of prior arts.

ring-like patterns, and the contaminants in the Scratch display linear or curvy distribution from the edge towards the center of the wafer, and the defects in the Edge-Location are clustered on the edge of the wafer map. It is worth mentioning that these sorts of defects have determined possible root cause. The Center category is generated due to abnormality in liquid flow or in liquid pressure, while the Edge-Location class is caused by pollution in fabrication. When it turns to mixed-type defects, a fundamental factor is the challenge brought by the advance of technology node. To meet the requirements, the integrated circuits are built in a layer-by-layer style in each die on a piece of silicon wafer [5]. Researchers believe that different types of defects which come from different layers are mixed and eventually being a mixed-type one [4].

There have been numerous academic attempts to automatically recognize wafer map defect patterns. For a better illustration, we visualize the taxonomy of these methods in Figure 2. We can observe that before machine learning-based approaches are developed, some traditional digital image processing techniques like typical image alignment are utilized in automated defect classification (ADC) system [6]. Over past decades, machine learning approaches including the shallow learning paradigm and the deep learning paradigm have been broadly accepted in our community. They are anticipated to play an ever-more-important role in upgrading the quality of chip designs that come from the execution of complete CAD flows and subroutines, in addition to improving the standard models used in CAD tools [7–19]. Due to their exceptional performance, the learning-based defect recognition approaches are highlighted in our work, particularly the deep learning-driven ones. In the following, we will summarize a few illustrative artworks based on the taxonomy displayed in Figure 2.

The rest of the paper is organized as follows. Section 2 will give an overview of the wafer failure pattern recognition problem. Section 3 will summaries the shallow learning techniques-based works. Section 4 will introduce the existing deep learning-based arts including the advanced wafer-level frameworks, pattern-level and die-level approaches. Section 5 will conclude the whole paper and have some discussions.

## 2 PROBLEM DESCRIPTIONS

Most wafer failure pattern recognition frameworks typically take wafer maps as inputs, and their output should be accurate recognition results. One significant benefit of viewing the wafer maps

as images is that by retaining the wafer maps in their image representation, the defect information may be preserved the best. Wafer failure patterns, which are by nature spatial patterns, can thus be categorized visually.

It is essential to note that the single-type defect problem is the special instance of mixed-type one. As a result, the corresponding single-type recognition frameworks can be thought of as the prototypical solution to the mixed-type issue. We will provide some representative existing methods for single-type and mixed-type defect issues.

## 3 SHALLOW LEARNING-BASED RECOGNITION PARADIGM

A large amount of pioneered artworks [3, 5, 20–25] have been put forth based on shallow learning-based recognition paradigm.

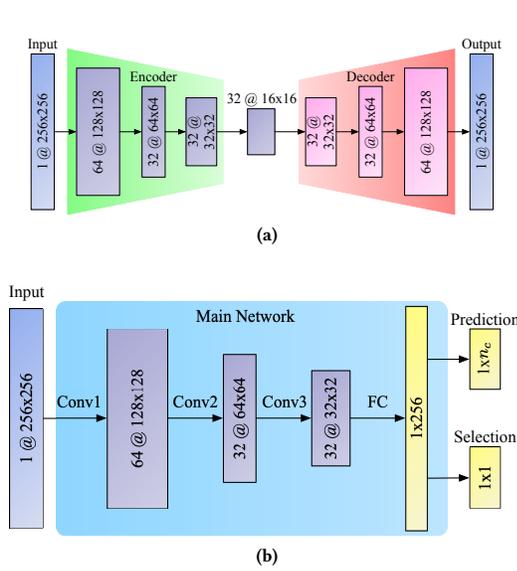
In the manner of clustering paradigm, some studies take into account creating a pre-defined probability distribution function for each defective pattern. A comparison of distribution models using the information criterion is then used to evaluate which statistical model is most suited. These methods assume that wafer maps will behave a mixture of defect clusters, with each cluster being represented by a particular distribution model. As a result, the mixed-type defect recognition can be resolved. For example, [22] put up with a multi-step approach via a statistical model-based clustering. The first step acts like denoising, which divides all local (or systematic) failures from the global (or random) failures. Second, a clustering technique is used to group the local defective dies into clusters. Finally, assigning the appropriate statistical model to each defective pattern. It is worth noticing that the clustering technique works on dies. On the contrary, in the other clustering-based works like [25], clusters of wafer-level maps are first built, and then the failure pattern recognition work is assigned to experienced engineers. Briefly, the clustering-based methods are significantly reliant on the expertise of the professionals because of building the distribution model or labeling work.

In the taste of classification paradigm, the feature vector of a wafer map is usually manually-crafted. For example, Wu *et al.* [3] trickily design radon-based and geometry-based features to feed the support vector machine classifier, while Yu *et al.* [24] exploit a manifold learning algorithm, joint local and non-local linear discriminant analysis, to obtain the features for the consequent Fisher discriminant classifier. It is in clear view that the domain knowledge of expert engineers is once again required when manually designing the feature. Such methods do not fully automate feature design or labeling for failure wafer patterns. Even worse, the feature design component and the follow-up classifier separate, causing the convergence of the entire failure classification framework to sub-optimality.

Briefly stated, when faced with complicated and varied multi-patterns, the performance of the aforementioned shallow learning-based approaches is typically inadequate.

## 4 DEEP LEARNING-BASED RECOGNITION PARADIGM

In the last ten years, deep learning approaches have seen substantial progress in a variety of vision-related tasks. As a result of

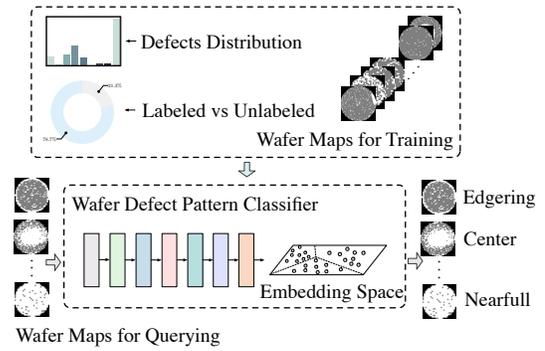


**Figure 3: The illustration of the data augmentation and selective wafer defect recognition network (reproduced from [26]).**

the success, the deep learning-based recognition paradigm has been extensively investigated and explored, and has become the mainstream in the academia. As mentioned, the mixed-type defect problem can be narrowed down to the single-type issue, and the associated single-type frameworks can serve as the prototype for mixed-type defect recognition. Thus, we review a wide range of deep learning-based studies looking into the solutions to sing-type defect recognition problem as well as the mixed-type one. Figure 2 shows three parallel defect recognition branches formed by all the techniques at three distinct levels: wafer-level, defect pattern-level, and die-level (i.e., pixel-level).

#### 4.1 Wafer-level Frameworks

Due to the superior performance and the robustness to random noise, convolutional neural network (CNN) has been frequently applied to classify wafer failure patterns. [27] develops a CNN model made up of 8 convolutional layers and 2 fully connected layers for defect recognition. Compared with manual feature design in shallow learning, the proposed methods effectively extracts valuable information from wafer map through the convolution operation. Besides, data augmentation techniques like vertically or horizontally flipping, slightly rotating wafer maps in a pure computer vision (CV) fashion is exploited in the training stage. Since the semiconductor manufacturing process becomes more complex and very sophisticated, it is challenging to collect enough various defect patterns. To address the issue, Ji *et al.* [28] apply generative adversarial networks (GANs) [29] to supplement the lack of training data, and finally improve the performance of the CNN’s classifier. [30] designs an improved GAN model integrated with CNN and other deep learning models called adaptive balancing generative adversarial network for identification of defective patterns in wafer maps. [31] brings up some insightful data augmentation methods including chip reverse, translate, and combine.



**Figure 4: The proposed wafer failure pattern classifier (reproduced from [32]).**

Based on the foundation laid by the prior research, Alawieh *et al.* [26] present a defect pattern classification framework built upon deep selective learning (showed in Figure 3(b)), integrating a reject option which can further increase model accuracy by not making predictions for samples with a high probability of misclassification. Moreover, the convolutional auto-encoder network (seen in Figure 3(a)) is harnessed as the core to synthesize wafer maps for data augmentation. It provides the chance to incorporate the data generation and recognition parts into a unified framework. The highlight of this method is considering the combination of deep learning-enabled recognition and manual inspection so that the possibility of method grounding increases. Another spotlight work is [32]. Different from the prior arts, Geng *et al.* [32] observe and attempt to tackle two major problems existing in previous works and the widely used benchmark (e.g., WM-811K). One is although a few works [26, 27, 33] exploit some wafer synthesis techniques to alleviate the imbalance issue, the generation process and the consequent classifier calibration are isolated to each other. Additionally, synthesis may lead to the label perturbation. Another is most earlier arts disregard unlabeled wafer maps. Hence, in [32], the few-shot learning and self-supervised learning paradigms are combined, and an end-to-end wafer failure pattern classifier based on a deeper and wider backbone (i.e., Inception-v4 [34]) is built. The working mechanism is depicted in Figure 4. The majority of the wafer-level frameworks have similar architectures with those in Figures 3 and 4.

Mentioning to mixed-type issue, the following is a summary of a number of noteworthy works. As one of the pioneer works offering CNN-based recognition method, [33] utilizes a simple network consisting of 3 convolutional layer and 2 fully connected layers. In addition, the authors point out that the experimental benchmark is plagued by the imbalanced distribution problem. In order to feed enough training data, 28600 synthetic wafer maps of 22 classes include sing-type and mixed type defects are generated via numerical simulation. Then, 1191 real wafer maps are adopted for CNN performance evaluation. Kyeong *et al.* in [36] adopt several CNN models to classify the mixed-type wafer defect patterns. Specifically, an individual classification model is built for each single defect pattern. When classifying mixed-type defect patterns, each model checks whether the relevant pattern exists. In this work, 16 defects mixed with four basic types: circle, ring, scratch, and zone are considered. However, when the category number of

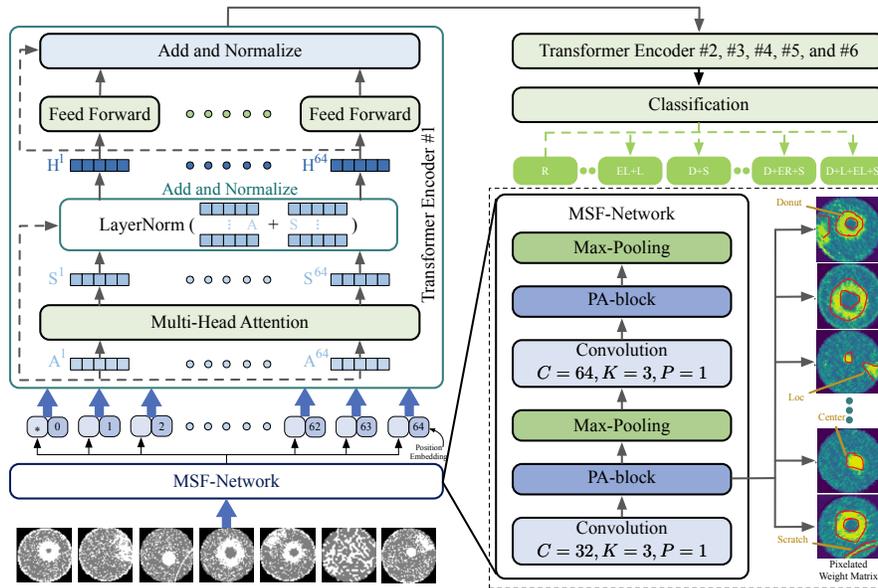


Figure 5: The architecture of the transformer-based defect recognition (reproduced from [35]).

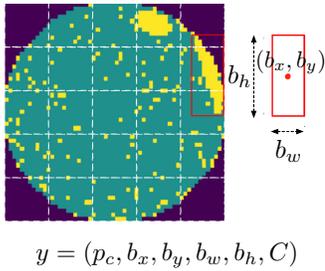
defect patterns rises, more CNN model is required. Wang *et al.* [37] present a tensor voting method to extract and separate mixed-type patterns from wafer maps. It firstly partition mixed-type defect patterns into clusters, and then a simple decision tree is constructed to extract region and curve patterns by tensor voting process. The experimental results suggests the effectiveness of the proposed method in identifying both single and mixed-type defect patterns. [4] exploits a deformable convolutional network for mixed-type issue. To selectively sample from mixed defects and then extract features from wafer maps, an improved deformable convolutional unit is devised. Additionally, a multi-label output layer is enhanced with a one-hot encoding strategy, which decomposes extracted mixed features into each fundamental single defect. It can be inferred that the computational cost is much higher than vanilla CNN model. More importantly, this research has significantly advanced the research of mixed-type defect recognition by releasing a benchmark, *MixedWM38*. Very recently, [35] firstly present a multi-scale information fusion transformer for mixed-type wafer recognition issue. The authors argue that due to the working paradigm (e.g., padding and down-sampling) and the limitations of the convolution operator (e.g., fixed-size kernels), CNN models are more likely to ignore crucial information of some defects and lack the global view of wafer maps. Observing the proposed framework in Figure 5, the MSF-network is comprised of two convolutional layers and a pixelated attention block (i.e., PA-Block), while the transformer encoder encodes the positional information provided by the MSF-network. The architecture makes use of CNN’s local perception and transformer’s global perception to fuse the corresponding information.

## 4.2 Pattern-level and Die-level Frameworks

In light of the fact that each die has the label indicating pass or fail in the test, the pattern-level and pixel-level operations can be carried out after wafer maps translating into images. In addition

to reporting the defect classes, the defect pattern-level and die-level methods show more thorough information covering pattern locations, shapes, *etc.*, which would help with further analysis and provide a better reasoning.

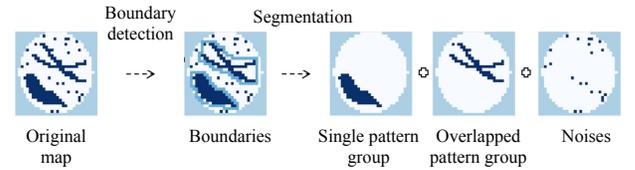
Pattern-level techniques can predict some bounding boxes which would contain certain patterns. As illustrated in Figure 6, the output results may include the prediction ( $p_c$ ), the center coordinates, widths, and heights of the boxes and the confidence ( $C$ ) that each bounding box contains a defect pattern. [38] applies the YOLOv3 [39], a popular object detection model in computer vision, to detect defects. To collect the training data, [38] uses the image-capturing system to scan hundreds of industrial dies. The GAN model is subsequently adopted to generate synthetic data to supplement the dataset. Credited to YOLOv3, this method outperforms Faster R-CNN [40] and SSD [41] which are another well-known object detection models. Despite the excellent performance of the method in [38], the inference cost is much higher than the baselines, which makes practical usage more difficult. [42] uses the single shot detector (SSD) [41] to detect defects quickly and deliver superior recognition performance. To train the default boxes to effectively represent the patterns’ positions, a matching strategy that identifies the default boxes that are significantly related to the ground truth bounding boxes is required. As a matching strategy, the Jaccard overlap coefficient between the ground truth boxes and the default boxes is calculated. The ground truth bounding box is then matched to the default boxes only if the Jaccard overlap score is higher than 0.5. Further, researchers found that the category dependencies among defects are clearly distinct from those among typical objects in CV. It is not possible to directly utilize the object detection methods in CV for recognizing defects. To resolve this issue, [43] develop a category-related non-maximum suppression (CR-NMS) method which employs the Cover Percent (CoP) instead



**Figure 6: An example of pattern-level recognition (reproduced from [44]).**

of Intersection over Union (IoU) to guide the bounding box regression. Moreover, a two-stage bounding box regression algorithm is proposed to remove the duplicate boxes.

The die-level methods assign a label to each die (or pixel) in a wafer map image such that pixels with the same label share certain characteristics, as shown in Figure 7. [45] applies semantic segmentation models like SegNet [46], U-Net [47], and FCN [48] into the wafer defect recognition. The inputs to the network are wafer maps with defect clusters and random defects, while the target outputs have only the defect clusters. In the training dataset, the basis defect patterns are randomly combined to generate the mixed-type wafer map. Though the testing performance on 1191 real wafers is good, the testing set is small and non-public. Besides, in [45], the segmentation algorithms are directly employed without any specific customization for the wafer tasks. [49] proposes a qualitative and quantitative analysis approach for the mixed-type wafers. The boundary detection method via U-Net and an overlapped pattern unwrapping approach are utilized to segment pattern groups and overlapped patterns, respectively. The mixed-type wafers are then transformed into multiple single-type wafers. A CNN classifier is calibrated to predict classes of single-type defects. After segmentation and recognition, the defects are remapped to the original mixed-type wafers to calculate the sizes of the defects, which reflects the impacts of the defects. The mixed-type wafer dataset is generated by the superimposition of single-type wafers, which would be impractical in real industrial environments. The experimental results of [49] show its performance advantages over classical methods such as Gaussian mixture model, infinite warped mixture model, and support vector clustering. [50] harnesses the Mask R-CNN as the segmentation model cooperating with a data pre-processing method for mixed-type wafer defect recognition. During data pre-processing, mask labeling, rotational and copy-paste data augmentation are used to gather sufficient mask-annotated mixed-type wafer maps. Then the Mask R-CNN model is trained with the constructed dataset to classify and locate distinct defect patterns. The output layers of the network have two branches: the first utilizes a fully connected layer to predict the object class and the position of the bounding box, and the second employs convolutional layers to obtain the detailed pattern segmentation masks. [51] exploits a residual attention block that combines an attention mechanism with a residual block to improve the U-Net to segment a mixed defect. The residual block [52] accepts the input as it is and adds it to the learned function to solve the gradient vanishing problem as the network gets deeper. The residual blocks are added to each step of the contracting path and



**Figure 7: The segmentation example for the wafer defect detection (reproduced from [49]).**

the expanding path in the U-Net. The attention method follows the [53] and sequentially applies channel attention and spatial attention in the residual block. Channel attention aggregates the spatial information by max-pooling and average pooling operations. Spatial attention firstly concatenates two aggregated values generated by performing max pooling and average pooling along the channel axis of the feature map, and then multiplies the feature map with the input channel attention map element-wisely. Similar to [50], [54] studies a lightweight encoder-decoder-based model, WaferSegClassNet, which has two branches for segmentation and classification respectively. The encoder uses a series of convolution blocks with pooling layers to extract multi-scale defective pattern information. The decoder block serves a dual purpose of performing classification and producing the semantic segmentation masks by recovering the spatial information.

In a nutshell, even though these wafer recognition algorithms have positive outcomes in their experiments, the typical object detection and segmentation methods are almost directly applied without being tailored to mixed-type wafer defects problems.

## 5 CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have surveyed the recent line of arts in techniques of wafer failure pattern recognition, especially for mixed-type issue. These arts contribute to promoting the advancement of defect recognition. In the future, we believe the following aspects are worth studying.

(1) Confidential wafer data in each foundry may not be insufficient to calibrate a reliable model. On the other hand, learning-based techniques by their very nature rely on absorbing enormous amounts of data in order to train and test algorithms. The conflict between the security of wafer information and the development of a learning-based wafer recognition method emerges. One potential solution to this dilemma is using the federated learning paradigm [55]. Instead of storing all data on one server, in federated learning, each foundry trains the model locally with its wafer map data and then uploads the local model to the central server. The central server aggregates local models and then sends back the model updates to each foundry. During the procedure, all local data is still kept confidential, while the robust defect recognition model is built.

(2) Although deep learning-based algorithms have demonstrated superior performance, these algorithms are susceptible to perturbations (e.g., random defects). It poses a great challenge to the deployment of these methods in real semiconductor wafer manufacturing systems. Studying the concepts of adversarial attacks and defenses [56] come into being a possible prospect.

(3) As wafer defect patterns get increasingly complicated, effectively extracting and fusing multi-level features from wafer-level to die-level becomes substantial. Moreover, how to incorporate

the prior information like the statistical distribution of each defect pattern is worthy of further thought.

## ACKNOWLEDGMENT

This work is sponsored by Shanghai Pujiang Program (Project No. 22PJ1410400).

## REFERENCES

- [1] U. Kaempf, "The binomial test: a simple tool to identify process problems," *IEEE TSM*, vol. 8, no. 2, pp. 160–166, 1995.
- [2] J.-C. Chien, M.-T. Wu, and J.-D. Lee, "Inspection and classification of semiconductor wafer surface defects using cnn deep learning networks," *Applied Sciences*, vol. 10, no. 15, p. 5340, 2020.
- [3] 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 TSM*, vol. 28, no. 1, pp. 1–12, 2015.
- [4] J. Wang, C. Xu, Z. Yang, J. Zhang, and X. Li, "Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition," *IEEE TSM*, vol. 33, no. 4, pp. 587–596, 2020.
- [5] J. Zhang, W. Qin, L. Wu, and W. Zhai, "Fuzzy neural network-based rescheduling decision mechanism for semiconductor manufacturing," *Computers in Industry*, vol. 65, no. 8, pp. 1115–1125, 2014.
- [6] P. B. Chou, A. R. Rao, M. C. Sturzenbecker, F. Y. Wu, and V. H. Brecher, "Automatic defect classification for semiconductor manufacturing," *Machine vision and applications*, vol. 9, no. 4, pp. 201–214, 1997.
- [7] H. Geng, H. Yang, B. Yu, X. Li, and X. Zeng, "Sparse VLSI layout feature extraction: A dictionary learning approach," in *Proc. ISVLSI*, 2018, pp. 488–493.
- [8] T. Chen, B. Lin, H. Geng, S. Hu, and B. Yu, "Leveraging spatial correlation for sensor drift calibration in smart building," *IEEE TCAD*, 2020.
- [9] H. Geng, W. Zhong, H. Yang, Y. Ma, J. Mitra, and B. Yu, "SRAF insertion via supervised dictionary learning," *IEEE TCAD*, vol. 39, no. 10, pp. 2849–2859, 2020.
- [10] R. Chen, W. Zhong, H. Yang, H. Geng, F. Yang, X. Zeng, and B. Yu, "Faster region-based hotspot detection," *IEEE TCAD*, 2020.
- [11] H. Yang, W. Zhong, Y. Ma, H. Geng, R. Chen, W. Chen, and B. Yu, "VLSI mask optimization: From shallow to deep learning," *Integration, the VLSI Journal*, vol. 77, pp. 96–103, 2021.
- [12] H. Geng, H. Yang, L. Zhang, J. Miao, F. Yang, X. Zeng, and B. Yu, "Hotspot detection via attention-based deep layout metric learning," *IEEE TCAD*, 2021.
- [13] T. Chen, Q. Sun, C. Zhan, C. Liu, H. Yu, and B. Yu, "Deep H-GCN: Fast analog IC aging-induced degradation estimation," *IEEE TCAD*, 2021.
- [14] Q. Sun, C. Bai, H. Geng, and B. Yu, "Deep neural network hardware deployment optimization via advanced active learning," in *Proc. DATE*, 2021, pp. 1510–1515.
- [15] Q. Sun, C. Bai, T. Chen, H. Geng, X. Zhang, Y. Bai, and B. Yu, "Fast and efficient DNN deployment via deep Gaussian transfer learning," in *Proc. ICCV*, 2021, pp. 5380–5390.
- [16] R. Chen, S. Hu, Z. Chen, S. Zhu, B. Yu, P. Li, C. Chen, Y. Huang, and J. Hao, "A unified framework for layout pattern analysis with deep causal estimation," in *Proc. ICCAD*, 2021, pp. 1–9.
- [17] Q. Xu, H. Geng, S. Chen, B. Yuan, C. Zhuo, Y. Kang, and X. Wen, "GoodFloorplan: Graph convolutional network and reinforcement learning based floorplanning," *IEEE TCAD*, 2021.
- [18] H. Geng, Y. Ma, Q. Xu, J. Miao, S. Roy, and B. Yu, "High-speed adder design space exploration via graph neural processes," *IEEE TCAD*, 2021.
- [19] H. Geng, T. Chen, Y. Ma, B. Zhu, and B. Yu, "PTPT: physical design tool parameter tuning via multi-objective Bayesian optimization," *IEEE TCAD*, 2022.
- [20] F.-L. Chen and S.-F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE TSM*, vol. 13, no. 3, pp. 366–373, 2000.
- [21] J. Y. Hwang and W. Kuo, "Model-based clustering for integrated circuit yield enhancement," *European Journal of Operational Research*, vol. 178, no. 1, pp. 143–153, 2007.
- [22] T. Yuan, W. Kuo, and S. J. Bae, "Detection of spatial defect patterns generated in semiconductor fabrication processes," *IEEE TSM*, vol. 24, no. 3, pp. 392–403, 2011.
- [23] C.-W. Chang, T.-M. Chao, J.-T. Horng, C.-F. Lu, and R.-H. Yeh, "Development pattern recognition model for the classification of circuit probe wafer maps on semiconductors," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 2, no. 12, pp. 2089–2097, 2012.
- [24] J. Yu and X. Lu, "Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis," *IEEE TSM*, vol. 29, no. 1, pp. 33–43, 2016.
- [25] M. B. Alawieh, F. Wang, and X. Li, "Identifying wafer-level systematic failure patterns via unsupervised learning," *IEEE TCAD*, vol. 37, no. 4, pp. 832–844, 2017.
- [26] M. B. Alawieh, D. Boming, and D. Z. Pan, "Wafer map defect patterns classification using deep selective learning," in *Proc. DAC*, 2020, pp. 1–6.
- [27] M. Saqlain, Q. Abbas, and J. Y. Lee, "A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes," *IEEE TSM*, vol. 33, no. 3, pp. 436–444, 2020.
- [28] Y. Ji and J.-H. Lee, "Using GAN to improve cnn performance of wafer map defect type classification: Yield enhancement," in *2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 2020, pp. 1–6.
- [29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.
- [30] J. Wang, Z. Yang, J. Zhang, Q. Zhang, and W.-T. K. Chien, "AdaBalGAN: an improved generative adversarial network with imbalanced learning for wafer defective pattern recognition," *IEEE TSM*, vol. 32, no. 3, pp. 310–319, 2019.
- [31] Q. Zhang, Y. Zhang, J. Li, and Y. Li, "Wdp-bnn: Efficient wafer defect pattern classification via binarized neural network," *Integration, the VLSI Journal*, vol. 85, pp. 76–86, 2022.
- [32] H. Geng, F. Yang, X. Zeng, and B. Yu, "When wafer failure pattern classification meets few-shot learning and self-supervised learning," in *Proc. ICCAD*, 2021, pp. 1–8.
- [33] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE TSM*, vol. 31, no. 2, pp. 309–314, 2018.
- [34] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*, 2017.
- [35] Y. Wei and H. Wang, "Mixed-type wafer defect recognition with multi-scale information fusion transformer," *IEEE TSM*, vol. 35, no. 2, pp. 341–352, 2022.
- [36] K. Kyeong and H. Kim, "Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks," *IEEE TSM*, vol. 31, no. 3, pp. 395–402, 2018.
- [37] R. Wang and N. Chen, "Detection and recognition of mixed-type defect patterns in wafer bin maps via tensor voting," *IEEE TSM*, vol. 35, no. 3, pp. 485–494, 2022.
- [38] S.-H. Chen, C.-H. Kang, and D.-B. Perng, "Detecting and measuring defects in wafer die using GAN and YOLOv3," *Applied Sciences*, vol. 10, no. 23, p. 8725, 2020.
- [39] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015, pp. 91–99.
- [41] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. ECCV*, 2016, pp. 21–37.
- [42] T. S. Kim, J. W. Lee, W. K. Lee, and S. Y. Sohn, "Novel method for detection of mixed-type defect patterns in wafer maps based on a single shot detector algorithm," *Journal of Intelligent Manufacturing*, vol. 33, no. 6, pp. 1715–1724, 2022.
- [43] X. Wang, X. Jia, C. Jiang, and S. Jiang, "A wafer surface defect detection method built on generic object detection network," *Digital Signal Processing*, vol. 130, p. 103718, 2022.
- [44] P. P. Shinde, P. P. Pai, and S. P. Adiga, "Wafer defect localization and classification using deep learning techniques," *IEEE Access*, vol. 10, pp. 39 969–39 974, 2022.
- [45] T. Nakazawa and D. V. Kulkarni, "Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder-decoder neural network architectures in semiconductor manufacturing," *IEEE TSM*, vol. 32, no. 2, pp. 250–256, 2019.
- [46] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE TPAMI*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [48] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE TPAMI*, vol. 39, no. 4, pp. 640–651, 2017.
- [49] Y. Kong and D. Ni, "Qualitative and quantitative analysis of multi-pattern wafer bin maps," *IEEE TSM*, vol. 33, no. 4, pp. 578–586, 2020.
- [50] M.-C. Chiu and T.-M. Chen, "Applying data augmentation and mask R-CNN-based instance segmentation method for mixed-type wafer maps defect patterns classification," *IEEE TSM*, vol. 34, no. 4, pp. 455–463, 2021.
- [51] J. Cha and J. Jeong, "Improved U-Net with residual attention block for mixed-defect wafer maps," *Applied Sciences*, vol. 12, no. 4, p. 2209, 2022.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.
- [53] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: convolutional block attention module," in *eccv*, 2018, pp. 3–19.
- [54] S. Nag, D. Makwana, S. Mittal, C. K. Mohan et al., "WaferSegClassNet-a lightweight network for classification and segmentation of semiconductor wafer defects," *Computers in Industry*, vol. 142, p. 103720, 2022.
- [55] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.
- [56] N. Liu, M. Du, R. Guo, H. Liu, and X. Hu, "Adversarial attacks and defenses: An interpretation perspective," *ACM SIGKDD Explorations Newsletter*, vol. 23, no. 1, pp. 86–99, 2021.