Deformable Convolutional Networks for Efficient Mixed-Type Wafer Defect Pattern Recognition

Junliang Wang, *Member, IEEE*, Chuqiao Xu, *Graduate Student Member, IEEE*, Zhengliang Yang, Jie Zhang, *Member, IEEE*, and Xiaoou Li, *Senior Member, IEEE*

***Abstract*—Defect pattern recognition (DPR) of wafer maps is critical for determining the root cause of production defects, which can provide insights for the yield improvement in wafer foundries. During wafer fabrication, several types of defects can be coupled together in a piece of wafer, it is called mixed-type defects DPR. To detect mixed-type defects is much more complicated because the combination of defects may vary a lot, from the type of defects, position, angle, number of defects, etc. Deep learning methods have been a good choice for complex pattern recognition problems. In this article, we propose a deformable convolutional network (DC-Net) for mixed-type DPR (MDPR) in which several types of defects are coupled together in a piece of wafer. A deformable convolutional unit is designed to selectively sample from mixed defects, then extract high-quality features from wafer maps. A multi-label output layer is improved with a one-hot encoding mechanism, which decomposes extract mixed features into each basic single defect. The experiment results indicate that the proposed DC-Net model outperforms conventional models and other deep learning models. Further results of the interpretable analysis reveal that the proposed DC-Net can accurately pinpoint the defects areas of wafer maps with noise points, which is beneficial for mixed-type DPR problems.**

***Index Terms*—Semiconductor manufacturing, wafer defects, pattern recognition, deformable convolutional networks.**

I. Introduction

THE INTEGRATED CIRCUIT (IC) industry consists of four basic parts: IC design, wafer manufacturing, wafer test, assembly, and final test [1], as is shown in Fig. 1. After the wafer manufacturing, wafer test is conducted to evaluate the electrical function of each die by probe test [2], which consists of many test items that measure the performance of the operations of required die electronic functions. Then, these qualified dies are cut from the wafer, assumed into individual chips, and label as qualified products after the final test [3]. The testing results of all dies on a wafer are saved on a wafer map [4], which has a two-dimensional adjacent matrix. Defect pattern recognition (DPR) classifies wafer maps into several categories by analyzing the distribution of dies failed in wafer test, which provides insights for reasoning the root cause of defects in fabrication processes [5], [6]. Nowadays, DPR is conducted manually piece by piece in typical semiconductor wafer manufacturing systems (SWFSs) by experienced engineers, which is time-consuming and expensive. Moreover, the capacity of the SWFS is improved rapidly with the application of the automatic material handling system, robots, advanced planning system, etc. The manual defective pattern recognition can hardly satisfy the system needs [7]. Therefore, there is a great demand for automated DPR in the operation of SWFSs.

During the wafer fabrication, the integrated circuits were built layer-by-layer in each die on a piece of silicon wafer [8]. Directed by the Moore's law, the layer of the integrated circuits and the number of the dies on a wafer are increased greatly with the development of integrated circuits technology [9]. In different circuit layers, different defects were caused by various faults in production. As a result, different types of defects are mixed on a wafer map coming into being a mixed-type defect. As is shown in Fig. 2-a), three different defect patterns ("Edge-loc", "Scratch", and "Count") occur in a wafer map.

Compared with DPR with single-type pattern recognition tasks, the DPR of mixed-typed wafer maps remains to be tough due to the various graphic characteristics of mixed defects. In wafer fabrication, some process is conducted by rotating the wafer. For instance, the photoresist was coated on the wafer by a photoresist spinner in a photolithography stage. In this process, the wafer may have defects in different angels due to the inappropriate rotation speed and the viscosity of the photoresist. In SWFSs, the relative location of composed defects on a piece of wafer can be different from the same mixed-type defect, which results in complicated and diverse graphic characteristics. As shown in Fig. 2-b), four wafer maps have different graphic characteristics, which have the same defect mixed by "Dount", "Loc" and "Scratch". In this scenario, the graphic characteristics of four subfigure vary considerably depending on the relative locations of mixed-types defects. Moreover, single defects are easy to be covered up. For example, the defects in subfigure (2) and (4) are blend in with "Dount", "Loc" and "Scratch", where the "Loc" and "Scratch" are hard to identify.

Manuscript received June 9, 2020; revised July 27, 2020; accepted August 20, 2020. Date of publication September 2, 2020; date of current version October 29, 2020. This work was supported in part by National Natural Science Foundation of China under Grant 51905091, in part by the Fundamental Research funds for the Central Universities under Grant 2232019D3-34, and in part by Shanghai Sailing Program under Grant 19YF1401500. *(Corresponding author: Jie Zhang.)*

Junliang Wang and Jie Zhang are with the College of Mechanical Engineering, Donghua University, Shanghai 201620, China (e-mail: junliangwang@dhu.edu.cn; mezhangjie@dhu.edu.cn).

Chuqiao Xu is with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xuchuqiao@sjtu.edu.cn).

Zhengliang Yang is with the College of Mechanical Engineering, Donghua University, Shanghai 201620, China.

Xiaoou Li is with the Department of Computer Science, CINVESTAV-IPN, Mexico City 07360, Mexico (e-mail: lixo@cs.cinvestav.mx).

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.2020.3020985

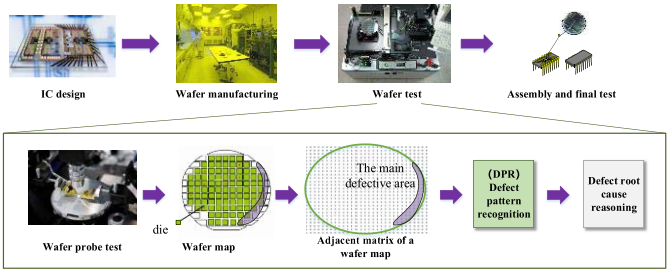


Fig. 1. DPR in semiconductor wafer manufacturing systems.

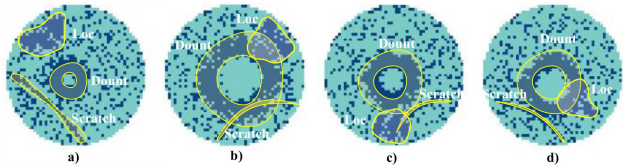
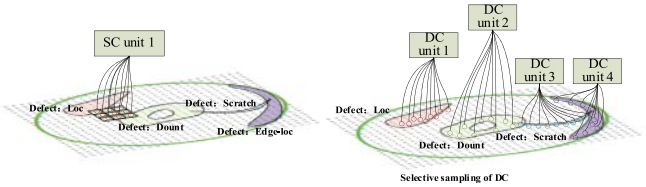


Fig. 2. Four wafer maps of the same mixed-type defect type.

The DPR of wafer maps is a computer vision task, which has received sufficient consideration [10], [11]. The previously proposed methods can be divided into two types: model-based methods and data-driven methods. the model-based methods use a predefined defective function (such as probability distribution function or spatial dependence function) for each wafer map and select the best matching defect pattern [12], [13]. Data-driven methods classify wafer maps through machine learning from massive data [14], [15], such as the RBFN, decision forest, and support vector machine [5].

Recently, with the development of intelligent manufacturing technologies [16], [17], data-driven methods have been widely applied in the mixed-type DPR. According to the complexity of machine learning models, the proposed methods can be divided into two types: shallow models and deep learning models. Shallow models refer to the machine learning models with shallow structures, such as the RBFN, decision forest, and support vector machine [5]. Aiming to achieve high accuracy, shallow models should be equipped with good feature extractors (such as correlogram and Radon transform [18]), which can obtain differentiable features from raw data. However, shallow models appear to be limiting if the input features are of low quality [9]. Wang [19] integrated classifier chains and support vector machines (CC-SVM) together to identify seven mixed-type defective patterns. Particularly, the eigen-decomposition of the normalized affinity matrix was used to separate the mixed clusters. Yuan *et al.* [20] proposed a multistep defect analysis approach to recognize six mixed-type defective patterns, which provided a similarity-based clustering technique to group the local defects into clusters according to their spatial locations. Kim *et al.* [13] proposed an infinite warped mixture model (iWMM) for six mixed-type defect patterns in wafer bin maps. The iWMM was a nonparametric Bayesian model that can find nonlinearly separable clusters with complex shapes (with non-Gaussian distributions), which enabled separating mixed-type defect patterns into several clusters of different defect patterns.

With the development of artificial intelligence, deep learning methods extract features from raw data automatically



a). Sampling mode of SC b). Sampling mode of DC

Fig. 3. Sampling mode of both DC and SC.

through multi-layer neural networks [21]–[23], which shows incredible performance in computer version tasks [24]. Nakazawa and Kulkarni [25] presented a convolutional neural networks (CNN) method for wafer map pattern classification and image retrieval, and the results indicated that the CNN model is effective in pattern recognition. In Nakazawa's work, 22 defects mixed by six basic types (random, ring, edge, scratch, cluster, gross) were investigated in different locations with theoretically generated wafer maps. Kyeong and Kim applied the CNN model to classify mixed-typed defect patterns of wafer bin maps [26] by building an individual classification model for each single defect pattern to determines if the corresponding pattern exists when several defect patterns are mixed over a wafer. In this work, 16 defects mixed with four basic types: circle, ring, scratch, and zone were considered.

The mixed-type defects investigated in the current studies were theoretical and simplified, which results in research gaps in the implementation of the current CNN models in real SWFSs. According to the observation in real SWFSs, the graphic characteristic of a mixed-type defect is various since the relative position and angle of basic defects are different, which would result in a high possibility of misclassifying defect patterns from different angles. As is shown in Fig. 2, four wafer maps have the same mixed-typed defect combined the "Dount", "Scratch" and "Loc" with different relative positions and angles. In a pilot test, a validated CNN classifier incorrectly suggested "Dount+Loc" type for Fig. 2-b) and Fig. 2-c), and a "Dount" type for the defect in Fig. 2-d). The unexpected results may be caused by the following issues:

1). A softmax multi-output layer was implemented in the validated CNN, where each neuron represents a mixed-type of defect. In this way, the graphic features of mixed-type of defects are considered in a unified way. As is shown in Fig. 2, there is a huge difference from overall graphic characteristics between wafer maps with defects in various relative positions and angles. For an analogy, a model well-trained to recognize cats may be confused by a picture with a cat that has two ears on its abdomen. Hence, a decomposition strategy should be applied to identify if "ears" and "abdomen" exists separately for higher accuracy.

2). The standard convolution (SC) kernels samples data from the source pixels with a standard rectangular. In mixed-typed DPR of wafer maps, the sampled source pixels may contain information from several different defects, which is bad for data representation with a decomposition strategy. As is shown in Fig. 3-a), the sampled source pixels of SC unit 1 covers information about defect "Loc", "Dount" and noisy data, which may disturb the decision making for only one defect in a decomposition strategy for mixed-type DPR.

Motivated by the above-mentioned issues, this article proposed an improved deformable convolutional (DC) network for the mixed-type DPR of wafer maps. In the proposed method, selective sampling and data representation are performed by a deformable convolution unit. And a multi-label output layer is designed to identify if the wafer map contains each single defect pattern. The rest of this article is organized as follows: Section II reviews the related works and summaries the research gaps. Section III designs the proposed deformable convolutional networks for mixed-type DPR of wafer maps. Section IV reports numerical experiments of the defective pattern recognition based on wafer map datasets. Finally, conclusions and recommendations for future work are summarized in Section V.

II. The Deformable Convolutional Neural Network

Among the machine learning technologies, the deformable convolutional (DC) network [27] selectively samples from input data based on the position information for further displacement adjustment, which has been extensively applied in multi-angle object detection and hyperspectral image classification [28]. A one-shot DC module was designed with a predictor-net to predicts interpolation offsets for guiding the deformation of the receptive field to enhance the robustness of respiratory motion estimation in ultrasound sequences [29]. Cao *et al.* [30] proposed a multi-scaled DC object detection network to detect small, dense objects. Wang *et al.* [31] devised a DC neural network module to align different frames at the feature level to handle large motions in video restoration tasks. A novel spatiotemporal DC neural network [32] was designed for end-to-end learning to extract complex features for dynamic gesture recognition. To capture information from irregular objects, a temporal DC was presented to enable free-form deformation of temporal sampling in video captioning [33]. Instructed by the offset, DC network can freely sample from input data to extract high-quality features to achieve better performance in complex tasks.

Given that the deformable convolutional network has yielded impressive results for image classification tasks with complex location information, such descriptors are expected to also be effective in multi-angle wafer defective pattern recognition. Different from the traditional CNN model, the DC unit introduces a 2-D offset matrix to offer deformable sampling locations of each convolutional kernel. With the deformable sampling locations, deformable features can be obtained by compressing the neighboring similar structural information into fixed grids. In the DPR of mixed-typed wafer maps, DC units can extract features with different graphic shapes, which can be applied to extract features for the type of each basic defect of wafer maps. The feature of mixed-type defects can be represented by several DC units with a decomposition mechanism, and fused by subsequent normal convolutional operations. As is shown in Fig. 2-c), four different DC units can extract defects with specific graphic shapes individually. For example, "DC unit 4" in Fig. 2.-c) can extract "Edge-loc" efficiently. With the corporation of DC units, the mixed-type defect pattern can be recognized in a decomposition method. Therefore, a deformable convolutional network method is improved to cope with the various graphic characteristics in mixed-type wafer defective pattern recognition.

III. The Mixed-Type Defects of Wafer Maps

From the graphic features of wafer maps, the single defective wafer maps can be divided into eight types, which is widely known as "Center(C)", "Donut(D)", "Edge-loc(EL)", "Edge-ring(ER)", "Loc(L)", "Near-full(NF)", "Scratch(S)", and "Random(R)" [18]. "Center" (shown in Fig. 4-C2) refers to the pattern that defective dies are distributed in the center of wafer maps, which may be caused by the abnormal center temperature in implant operation. "Dount" (shown in Fig. 4-C3) means the defective dies form a ring structure, which may be caused by the inaccurate chemical mechanical polishing. "Edge-loc" (shown in Fig. 4-C4) is a regional defect pattern on the edge of a wafer map, which may result from pollution in fabrication. "Edge-ring" (shown in Fig. 4-C5) is a global defect, where defective dies are distributed around the circle of wafer maps. "Loc" (shown in Fig. 4-C6) is a frequently occurred regional defect pattern, which may be caused by particle pollution or uneven heating. "Near-full" (shown in Fig. 4-C7) is a rarely happened defect pattern, which may result from receipt failure. "Scratch" (shown in Fig. 4-C8) means the defective dies are distributed in a long narrow area, which may be caused by mechanical and foreign body injury. "Random" (shown in Fig. 4-C9) means the defective dies randomly occur in a wafer map.

However, with the development of IC technology, mixed-type defects have widely occurred in fabrication, where defects in different IC layers are overlapped together. From the number of mixed-type defects, the investigated mixed-type patterns of the state-of-the-art are under 22 types. In Nakazawa and Kulkarni's work [25], 22 defects mixed by six basic types (random, ring, edge, scratch, cluster, gross) were investigated in different locations with theoretically generated wafer maps. In Kyeong and Kim's work [26], 16 defects mixed with four basic types: circle, ring, scratch, and zone were considered. However, the frequently occurred defect patterns are more than 30 types based on the observation in a SFWS in Shanghai, which indicated the previously investigated data are not sufficient for industrial practice.

According to the investigation in a real SWFS, 38 frequently occurred mixed-type defect patterns were found combined with one normal defect, eight basic defects, and 29 mixed defects, which were divided into four parts: single type, 2 mixed-type, 3 mixed-type, and 4 mixed-type (detailed in Table I). As shown in Fig. 4, there are 9 single type defect patterns (Pattern ID: C1-C9) consists of 8 single defects and the normal type. The 2 mixed-type contained 13 defect patterns (Pattern ID: C10-C22) mixed with 2 single types. There are 12 mixed-type defect patterns (Pattern ID: C23-C34) mixed with 3 single types, and 4 mixed-type defect patterns (Pattern ID: C35-C38) mixed with 4 single types. In Fig. 4, the pattern name of a mixed pattern indicates the combined single defects. For example, the pattern name of C11 is "C+ER" is mixed by the single type defects center and edge-ring together.

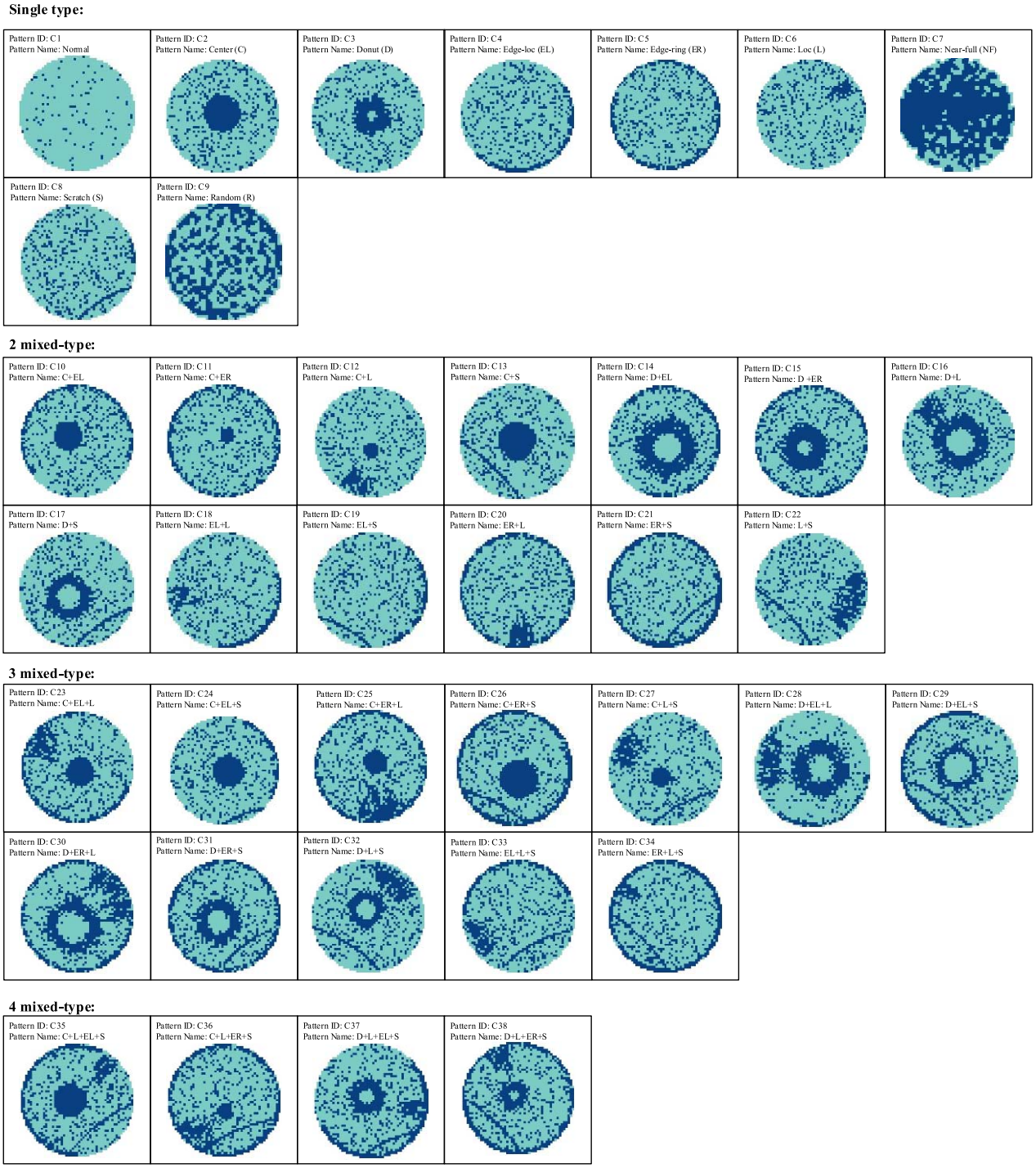


Fig. 4. The 38 types of original wafer map defects.

In this article, the above 38 mixed-type defect patterns collected from real-world fabrication were investigated. Due to the insufficient samples in some mixed-type wafer defect patterns, a simulation method [26] was employed to generate samples to keep the balance between various types of samples. We named this new dataset containing 38 types of wafer defect patterns as "MixedWM38". MixedWM38 contains one thousand wafer maps for each category. The MixedWM38 dataset in this article is available at https://github.com/Junliangwangdhu/WaferMap.

IV. Method

This section describes the framework of the proposed deformable convolutional networks for imbalanced learning, which consists of three parts: the defect pattern recognition networks structure, the deformable convolutional networks structure, and the multi-label network output layer.

A. Data Preprocessing

During the wafer test, a wafer map is generated by taking two probes that touche every die on a wafer to evaluate the

TABLE I
THE COMPARISON OF MIXED-TYPE DPR CATEGORIES

	Nakazawa's work [26]	*Kyeong's work[27]*	*ours*
Single type	18 (6 types with different locations)	5	9
2 mixed-type	4	6	13
3 mixed-type	0	4	12
4 mixed-type	0	1	4
Total	22	16	38

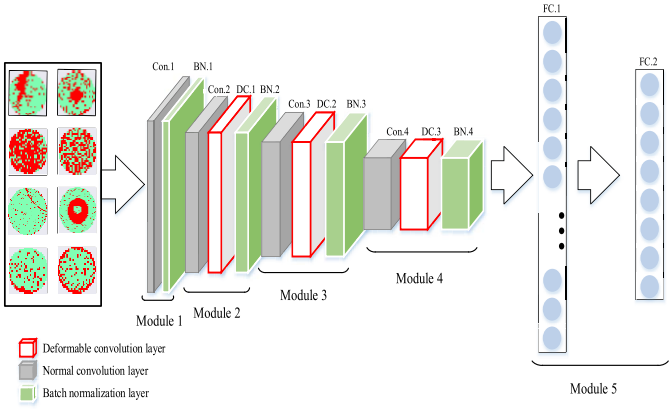


Fig. 5. The structure of the proposed DC-Net.

electrical property. The results of all dies on the i^{th} wafer were recorded in a wafer map, which is mathematically an adjacent matrix m_i. A unit x_{ij} in m_i denotes the results of the probe test of a die at j^{th} column and i^{th} row. The value of x_{ij} can be selected from the set $\{0, 1, 2\}$, where 0 means there is no die corresponding to x_{ij}, 1 means the die corresponding to x_{ij} is qualified, and 2 means the die corresponding to x_{ij} get functional faults. In the SWFSs, the layout of wafers specifically differs from the design of circuits. As a result, the size of the wafer maps can be 45*48, 53*58, 26*26, 60*40, or any other sizes. Therefore, all the wafer maps are standardized into a unified size in data preprocessing. In this method, the toolkit of the cubic interpolation method in 'OpenCV Computer Vision' [34] is applied to resize all raw wafer maps. The i^{th} unified wafer map is defined to be $M_r{}^i$, which is taken as the input of DPR model.

B. The Defect Pattern Recognition Networks Structure

The network model designed in this section is composed of five modules (blocks). The specific network structure is shown in Fig. 5.

The first module contains a normal convolution layer (Con. 1), where the input wafer maps are convoluted with 32 convolution kernels of size 5×5 pixels. Furthermore, a batch-normalization layer (BN.1) is designed in this module for better stability during model training. The second module equipped with a normal convolution layer (Con. 2), a deformable convolutional layer (DC.1), and a batch-normalization layer (BN.2). In this module, the input data are convoluted with 64 convolution kernels of size 3×3 pixels. The third module equipped with a normal convolution layer (Con. 3), a deformable convolutional layer (DC.2), and

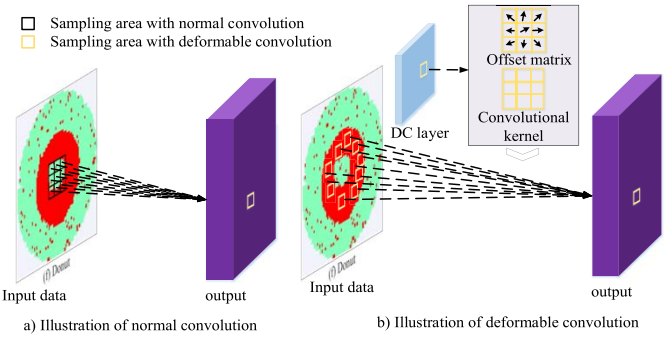


Fig. 6. Illustration of the deformed sampling locations.

a batch-normalization layer (BN.3). In this module, 128 convolutional kernels are contained in each layer. The fourth module contains a normal convolution layer (Con. 4), a deformable convolutional layer (DC.3), and a batch-normalization layer (BN.4). To keep a large data representation space, 128 convolutional kernels are contained in each layer of this module. The fifth module is made up of two full connection layers (FC.1 and FC.2). The FC.1 is made up of 128 neurons, and the FC.2 contains 8 neurons. In this model, the activation functions of input and hidden layers are Relu. The sigmoid function is adopted in the output layer to generate the membership in this multi-label classification problem. The training loss of the proposed DC-Net is described as follows:

$$loss = -\left[y * ln(p) + (1 - y)\ln(1 - p)\right] \tag{1}$$

$$p = sigmoid(\hat{y}) = \frac{1}{1+e^{-\hat{y}}} \tag{2}$$

where y is the actual label, $\hat{y}$ denotes the predicted values through the networks.

1) Deformable Convolutional Layer: Different from the normal convolution model, the deformable convolutional layer adds a 2D offset matrix for a convolution kernel to samples from input data selectively. As is shown in Fig. 6. DC consists of twin-channel feature preprocessing. The upper one learns the offset from a convolution layer to obtain the overall offset domain. And another one convolutes the input data according to the offset for each convolution kernel. As illustrated in Fig. 6, the offsets are obtained by applying a convolutional layer over the same input feature map. The convolution kernel is of the same spatial resolution and dilation as those of the current convolutional layer. The output offset fields have the same spatial resolution with the input feature map, which can be learned during back propagation. Different from the normal convolution network, DC can detect the key features with irregular shapes effectively. For example, the sampling area of the DC unit in Fig. 6-b). is distributed along the ring of defect dies in the input wafer map.

The DC works in this article as follows:

$$y_{pq} = \sum_{ij\in\{1,2\cdots l\}} w_{dc}^{ij} x_{p+w_{offx}^{ij}, q+w_{offy}^{ij}} \tag{3}$$

W_{dc} is a deformable convolutional kernel with the size of [l, l], W_{offx} is the offset matrix in X-direction, and W_{offy} is the offset matrix in Y-direction. w_{dc}^{ij}, w_{offx}^{ij} and w_{offy}^{ij} are elements of the matrix W_{dc}, W_{offx}, and W_{offy}, where $i, j \in \{1, 2\cdots l\}$.

$x_{p,q}$ is the input data of the DC from input matrix $X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,m} \end{bmatrix}$, y_{pq} is the output of the DC process.

The offset matrixes are determined during the back-propagation. Usually, the offset value w^{ij}_{offx}, and w^{ij}_{offy} are decimals. To implement the obtained offset values in index determination, a bilinear interpolation is performed in deformable convolution to legitimize the sampling area.

2) Decomposition Strategy for Mixed-Type Defects: With the increased complexity of ICs, the defects in different IC layers are mixed in a wafer map. There are the 2^8 types of mixed defects in theory, which results in a huge target space. To simplify the representation of target space, a decomposition strategy with one-hot encoding is performed to describe the mixed-type defect patterns of a wafer map with several containing single patterns, since eight single defect patterns are graphically independent. The estimated one-hot code by the DC-Net is a row vector $P_{oh} = [p^o_1, p^o_2 \cdots p^o_8]$, where $p^o_i \in [0, 1]$ means the probability belonging to the i^{th} single defect category. And the real value of a one-hot code is $\hat{P}_{oh} = [\hat{p}^o_1, \hat{p}^o_2 \cdots \hat{p}^o_8]$, where $\hat{p}^o_i \in \{0, 1\}$ indicates if the wafer map contains the i^{th} single defect pattern. In this model, the encoded single defect patterns are ordered from C2 to C9 in Fig. 4. For example, $[0, 0, 1, 0, 0, 0, 1, 0]$ indicates that this wafer map contains third and seventh single defects: (Dount and Scratch).

Aiming to deal with the multi-label classification problem, a multi-label network output layer is designed in this section. In DC-Net, a fully connected layer (fc.1) is adopted to learn the feature extracted by the convolutional layers. Then, another fully connected layer (FC.2) determines the including defects of the wafer map. In FC.2, the Sigmoid activation function is adopted in the output layer, and the 8 neuron nodes in the output layer are activated with the Sigmoid function independently. After activation, the value of each neuron will fall independently from the range of 0 to 1, indicating the membership of a defect mode. The discriminative criterion is formulated with a threshold: if $x^o_i \geq t_{sh}$, this wafer map contains the i^{th} defect category, otherwise, the wafer map doesn't contain the i^{th} defect category. To achieve sensitive discrimination performance, t_{sh} is set to be 0.5 in this model, where the slope of the sigmoid function is maximized at 0.5.

V. Results and Analysis

In this section, the performance of the proposed approach is evaluated with a dataset named "MixedWM38" containing 38 thousand wafer maps as described in Section II. In each category of wafer maps, 80% of the records are taken as the training data, 20% are used for validation. Before each experiment, wafer maps are randomly selected from MixedWM38 to form the training and validating set independently. The experiments are conducted in Python compiler, TensorFlow 1.0, and CUDA 8.0, the computer with option: Linux system, Intel Xeon CPU E5-2630 V4 @2.20GHz and GeForce RTX 2080 Ti. Compared results with typical multi-label classification methods.

TABLE II
Experimental Parameter Settings of the Proposed DC-Net

Description	Value
Learning rate	0.0001
Optimizer	SGD
Batch size	1024
Decay	1e-6
Momentum	0.9

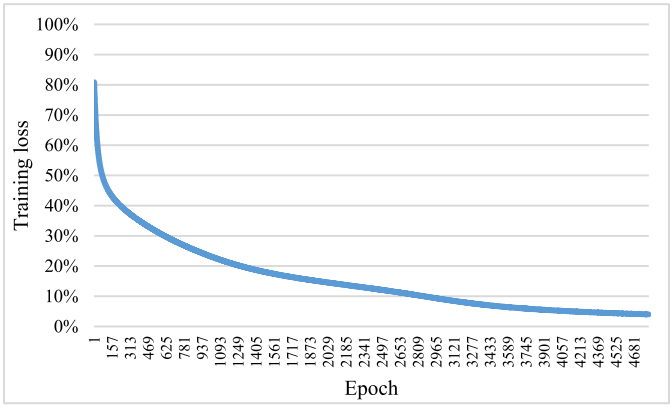


Fig. 7. The training loss curve of DC-Net.

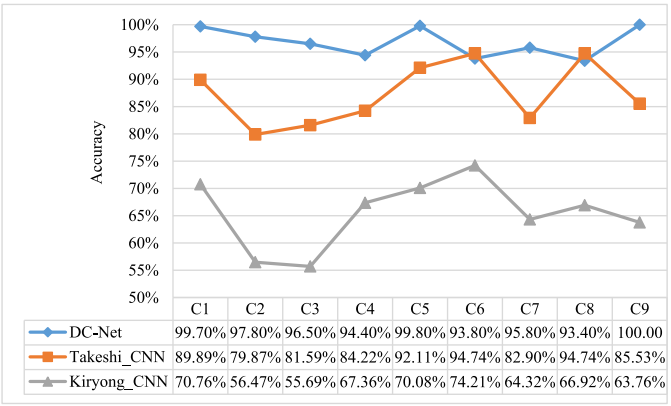
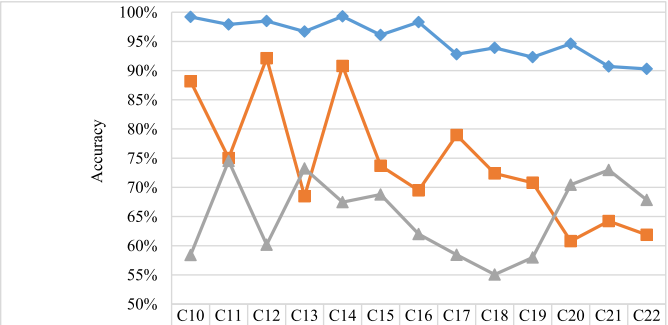


Fig. 8. The accuracy of comparable methods in single type DPR.

The performances of the proposed DC-Net were first compared with Takeshi_CNN [25], Kiryong_CNN [26]. Takeshi_CNN and Kiryong_CNN are the state-of-the-art methods proposed to cope with mixed-type DPR of wafer maps. In the experiments, the parameter settings of the proposed model that offer the best forecasting accuracy are listed in Table II.

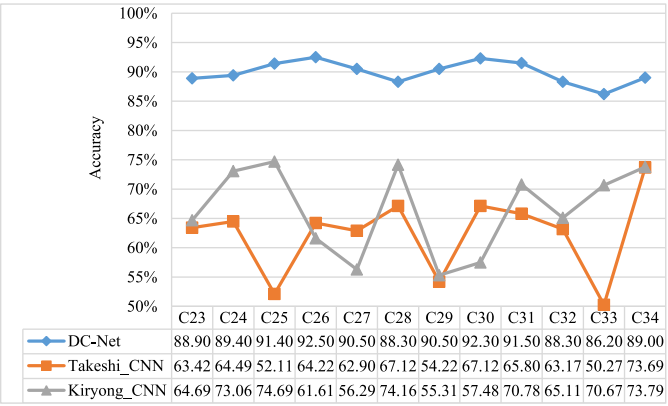
In the training phase, the training loss curve of DC-Net is reported in Fig. 7. As can been seen, the training error is less than 0.05 after 4800 iterations, which demonstrates that the DC-Net model converges nicely.

As is shown in Fig. 8 to Fig. 11, the accuracy obtained by the validation dataset of DPR is competitive among the evaluated methods. The average accuracy of "C1-C38" of the proposed DC-Net was 93.2%, which was 20.68% higher than Takeshi_CNN, and 27.91% higher than Kiryong_CNN. From



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
DC-Net	99.20	97.90	98.50	96.70	99.30	96.10	98.30	92.80	93.90	92.30	94.60	90.70	90.30
Takeshi_CNN	88.16	75.01	92.11	68.48	90.79	73.69	69.48	78.96	72.38	70.79	60.79	64.22	61.86
Kiryong_CNN	58.41	74.54	60.18	73.24	67.46	68.78	61.99	58.44	55.06	57.99	70.45	72.98	67.85

Fig. 9. The accuracy of comparable methods in 2 mixed-type DPR.



	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34
DC-Net	88.90	89.40	91.40	92.50	90.50	88.30	90.50	92.30	91.50	88.30	86.20	89.00
Takeshi_CNN	63.42	64.49	52.11	64.22	62.90	67.12	54.22	67.12	65.80	63.17	50.27	73.69
Kiryong_CNN	64.69	73.06	74.69	61.61	56.29	74.16	55.31	57.48	70.78	65.11	70.67	73.79

Fig. 10. The accuracy of comparable methods in 3 mixed-type DPR.

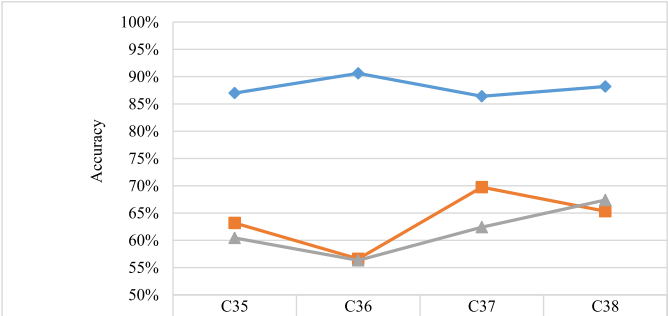
the view of single defect pattern recognition, as shown in Fig. 8, the average accuracy of the proposed DC-Net from C1 to C9 was 96.80%, which was 9.51% higher than Takeshi_CNN, and 31.29% higher than Kiryong_CNN. The accuracy of the proposed method is only slightly lower than that of Takeshi_CNN method in two single types. Furthermore, with the results on mixed-type of defects "C10-C38", as shown in Fig. 9 to Fig. 11, the DC-Net was extraordinarily better than other compared methods. The average accuracy of the proposed DC-Net from C10 to C38 was 92.12%, which was 24.19% higher than Takeshi_CNN, and 26.90% higher than Kiryong_CNN. It demonstrated the DC-Net was more effective for complex mixed-types. From the view of industrial practice, the detection of normal wafers is critical, since the misclassification would result in product quality problems. The accuracy of the proposed method with "C1" is 99.7%, which demonstrated the effectiveness of DC-Net in industrial practice.

To evaluate the performance in the false recognition and missing recognition of wafer defects, the statistical parameters of Precision (P) and Recall (R) are employed in our experiments, which are defined as follow:

True Positive (TP): predicting positive, the actual is positive.

False Positive (FP): predicting positive, the actual is negative.

False Negative (FN): predicting negative, the actual is positive.



	C35	C36	C37	C38
DC-Net	87.00%	90.60%	86.40%	88.20%
Takeshi_CNN	63.17%	56.60%	69.75%	65.33%
Kiryong_CNN	60.43%	56.32%	62.40%	67.39%

Fig. 11. The accuracy of comparable methods in 4 mixed-type DPR.

True Negative (TN): predicting negative, the actual is negative.

$$\mathrm{A} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{5}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{6}$$

Note that a higher P value indicates less false recognition that the actual normal patterns are recognized as the defect. A higher R value indicates less missing recognition that the actual defects are recognized as the normal patterns. The performance of DC-Net in precision and recall are reported in Table III.

As can be seen, the average precision and recall of DC-Net are very close to 1, which demonstrates that the proposed method has minuscule false recognition missing recognition for wafer map defects. As for the performance in each category, the average precision and recall for C1-C38 are greater than 0.9, except for C8 (scratch defect). The lower precision of C8 indicates that other defect patterns (especially C19, EL+S) are misrecognized as scratch defect pattern. The C19 wafer defect pattern is mixed with edge-loc and scratch defects. In this wafer defect pattern, the edge-loc defect is much less apparent compared with scratch defect, which is easy to be filtered out as noise by DC-Net. In future work, we will be committed to balance the recognition of noise and weak features in DC-Net.

A. Improvements Analysis of Proposed DC-Net

The proposed method is customized from CNN with the DC module and the one-hot encoding. To evaluate the performance gains from the DC module and one-hot output, the DC-Net model is further compared with the basic CNN (BCNN) and CNN with one-hot encoding (OH-CNN). The BCNN shares a similar network structure with DC-Net, but the DC module is replaced to be a normal convolution module. The OH-CNN has the same input layer and hidden layers with DC-Net, adopts an output layer with Softmax activation function and 38 neurons transformed by label powerset.

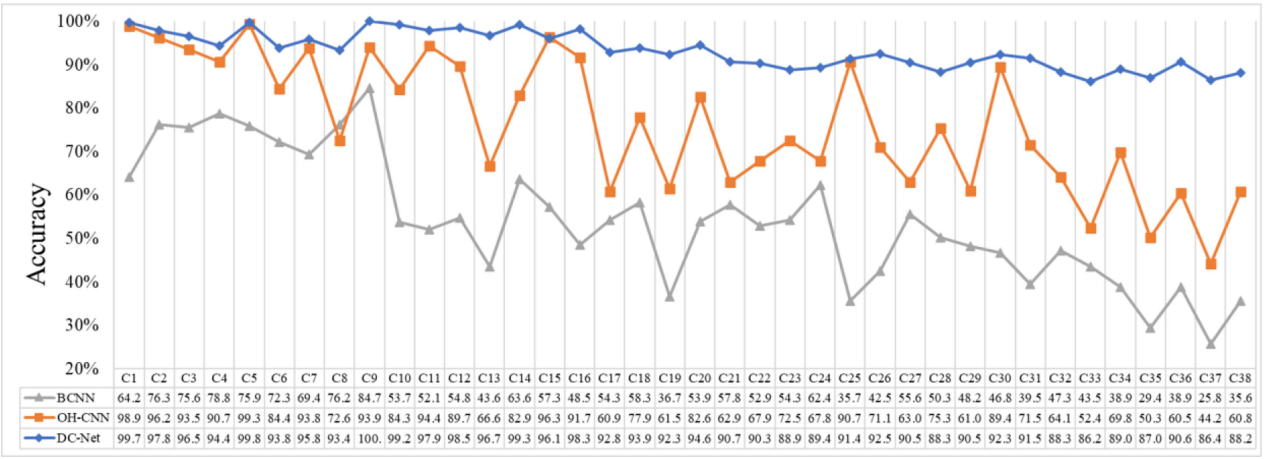


Fig. 12. The performance of DC-Net OH-CNN and BCNN.

TABLE III
THE PERFORMANCE OF DC-NET IN PRECISION AND RECALL

	Accuracy	*Precision*	*Recall*
C1	99.70%	94.00%	91.00%
C2	97.80%	93.00%	97.00%
C3	96.50%	95.00%	93.00%
C4	94.40%	96.00%	91.00%
C5	99.80%	93.00%	97.00%
C6	93.80%	99.00%	100.00%
C7	95.80%	90.00%	94.00%
C8	93.40%	60.00%	88.00%
C9	100.00%	97.00%	93.00%
C10	99.20%	94.00%	94.00%
C11	97.90%	92.00%	99.00%
C12	98.50%	92.00%	96.00%
C13	96.70%	97.00%	89.00%
C14	99.30%	96.00%	92.00%
C15	96.10%	91.00%	98.00%
C16	98.30%	94.00%	97.00%
C17	92.80%	96.00%	94.00%
C18	93.90%	98.00%	89.00%
C19	92.30%	94.00%	91.00%
C20	94.60%	95.00%	91.00%
C21	90.70%	96.00%	92.00%
C22	90.30%	98.00%	88.00%
C23	88.90%	99.00%	96.00%
C24	89.40%	92.00%	100.00%
C25	91.40%	93.00%	91.00%
C26	92.50%	97.00%	97.00%
C27	90.50%	97.00%	93.00%
C28	88.30%	95.00%	91.00%
C29	90.50%	98.00%	97.00%
C30	92.30%	89.00%	100.00%
C31	91.50%	90.00%	94.00%
C32	88.30%	99.00%	88.00%
C33	86.20%	97.00%	93.00%
C34	89.00%	98.00%	94.00%
C35	87.00%	96.00%	99.00%
C36	90.60%	99.00%	96.00%
C37	86.40%	95.00%	89.00%
C38	88.20%	92.00%	92.00%
Average	93.20%	94.00%	95.00%

1) *Compared With BCNN:* As a deep learning method, BCNN failed to live up to expectations in almost all categories, which indicates that handling represented feature in 38 types of defects of wafer maps remains to be a tough task.

2) *Compared With OH-CNN:* With the improved one-hot encoding, 38 types of defects types are decomposed into 8 basic types. The feature space is significantly reduced in OH-CNN, since the feature of each mixed-type defect is complicated. As a result, the average accuracy of OH-CNN achieved 76%, which was 22% higher than BCNN (results based on validation were shown in Fig. 12).

The results showed the one-hot encoding is effective by decomposing each mixed-type defect into several single basic patterns. Furthermore, the average accuracy of DC-Net was 93%, which was remarkable superior to the compared methods. Especially for the pattern recognition of mixed-type categories, the average accuracy among C10 to C38 was 92%, which indicated that the selective sampling from DC module is effective to recognize represented features from mixed-type defects of wafer maps.

B. Analysis and Discussion

To evaluate the performance of the DC unit, the area of interest by the DC-Net is visualized in this section. To analyze the interesting area of this model, a feature map for each wafer map is constructed to visualize the represented feature by DC-Net, which is obtained by inverse transforming the sampling position offset of the wafer map through the deformable convolution operation at the first layer. As is shown in Fig. 13, it is the visualization of sampling strength after the DC-Net. In each feature map, the gradient from blue to red indicates the importance of the sampling points from low to high in our proposed model. As can be seen in each defect pattern, the key points in defect locations appear red, which illustrates the critical areas for abnormal pattern recognition are focused by DC unit selective sampling. The noise points in the wafer maps appear blue, which illustrates the insignificant areas hindering abnormal pattern recognition are suppressed by DC unit selective sampling. For example, in C38 mixed of four defect patterns, the appearance of the four defect patterns (“Donut”, “Loc”, “Edge-ring”, and “Scratch”) are displayed in red and the surrounding noise points are faded. This provides valid

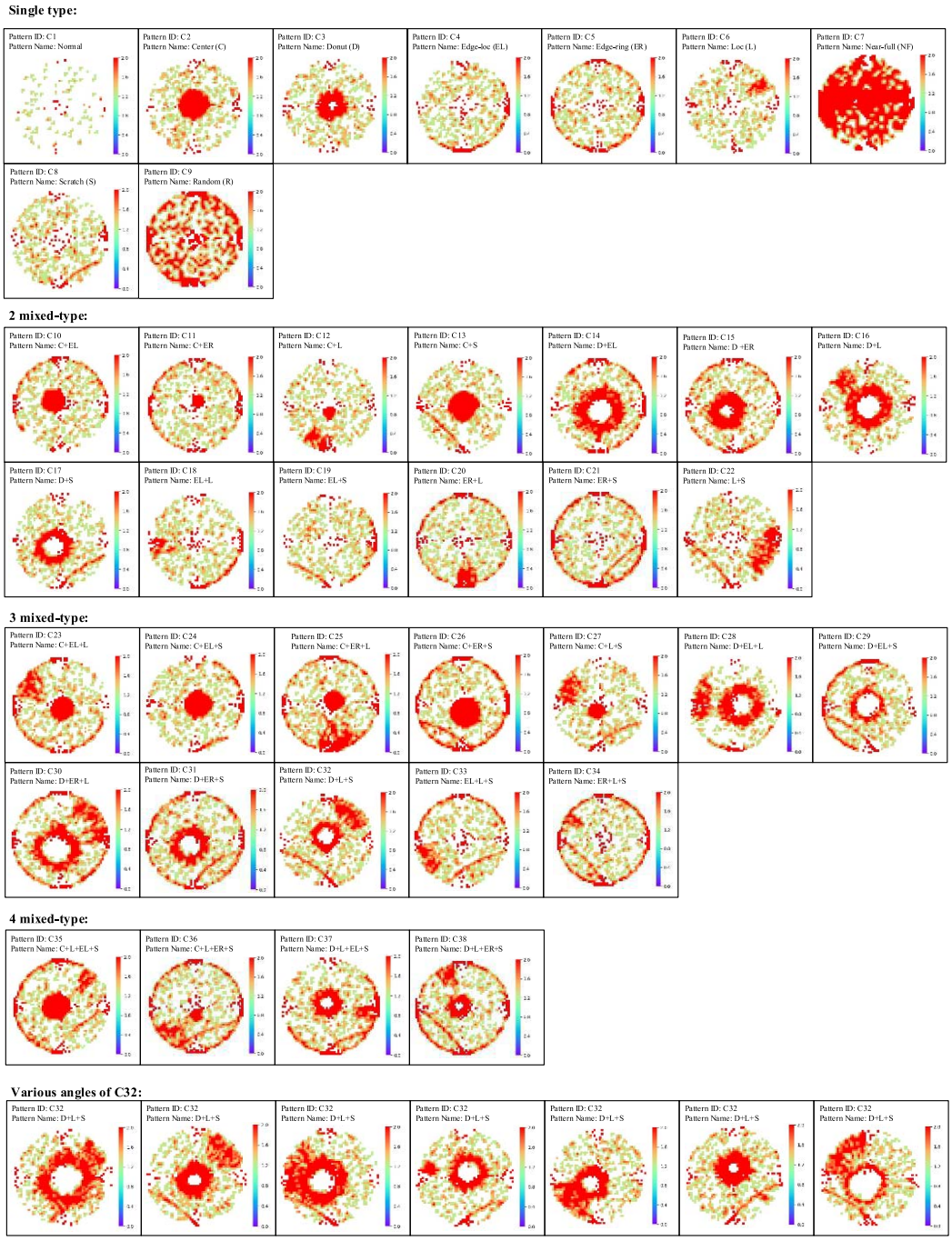


Fig. 13. The deformable convolution sampling visualization of defect wafer maps.

evidence that the DC unit greatly improves the ability to distinguish mixed features from DPR of wafer maps. Furthermore, multiple samples with various angles of C32 defect types are visualized at the bottom of Fig. 13. As can be seen, seven wafer maps have the same mixed-typed defect combined the "Dount", "Loc" and "Scratch" with different relative positions and angles, which have different graphic characteristics. However, no matter how the angle of the defect area changes, variable convolution can always find key areas for selective sampling. These illustrate that the proposed method significantly suppresses the graphic characteristic of mixed-type defect During the DPR for complicated wafer maps from real SWFSs.

VI. Conclusion

This study recognized the mixed-type defective patterns of the wafer maps produced by the electric property test, which can provide effective insights for the operation of the SWFSs, especially during the productivity improvement phase. The contribution of this article can be summarized as follows:

1) Compare with conventional DPR works, this article introduced deformable convolution into a CNN model, which can effectively extract defect features by concentrating sampling area on defect dies. Results clearly show that deformable convolution benefits the accuracy of mixed-type DPR of wafer maps.

2) Aiming to deal with multi-patterns of defect dies, a multi-label output layer with a one-hot encoding mechanism is designed to decompose a mixed-type defect pattern into each single defect pattern. The comparison among the OH-CNN and BCNN model indicates that the one-hot encoding mechanism can effectively significantly improve the average accuracy in multi-patterns recognition.

Taking all points into consideration, future work will be paid on applying the proposed method in defects root cause detection and reasoning of wafer fabrication. Besides, the modeling and analysis of the wafer defect to improve the wafer fabrication quality will be investigated in further research.

REFERENCES

[1] L. Mönch, R. Uzsoy, and J. W. Fowler, "A survey of semiconductor supply chain models part I: Semiconductor supply chains, strategic network design, and supply chain simulation," *Int. J. Prod. Res.*, vol. 56, no. 13, pp. 4524–4545, 2017, doi: 10.1080/00207543.2017.1401233.

[2] K. B. Lee, S. Cheon, and O. K. Chang, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.

[3] W.-T. K. Chen and C. H.-J. Huang, "Practical 'building-in reliability' approaches for semiconductor manufacturing," *IEEE Trans. Rel.*, vol. 51, no. 4, pp. 469–481, Dec. 2002.

[4] S. F. Yang and W.-T. K. Chien, "Electromigration lifetime optimization by uniform designs and a new lifetime index," *IEEE Trans. Rel.*, vol. 64, no. 4, pp. 1158–1163, Dec. 2015.

[5] F. Adly *et al.*, "Simplified subspaced regression network for identification of defect patterns in semiconductor wafer maps," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1267–1276, Dec. 2015.

[6] C.-F. Chien, C.-Y. Hsu, and P.-N. Chen, "Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence," *Flexible Services Manuf. J.*, vol. 25, no. 3, pp. 367–388, 2013.

[7] S. Kang, S. Cho, D. An, and J. Rim, "Using wafer map features to better predict die-level failures in final test," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 3, pp. 431–437, Aug. 2015.

[8] J. Zhang, W. Qin, L. H. Wu, and W. B. Zhai, "Fuzzy neural network-based rescheduling decision mechanism for semiconductor manufacturing," *Comput. Ind.*, vol. 65, no. 8, pp. 1115–1125, 2014.

[9] J. Wang, J. Zhang, and X. Wang, "A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 173–182, Feb. 2018.

[10] J. Wang, Z. Yang, J. Zhang, Q. Zhang, and W. K. Chien, "AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 310–319, Aug. 2019.

[11] T.-S. Li and C.-L. Huang, "Defect spatial pattern recognition using a hybrid SOM–SVM approach in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 374–385, 2009.

[12] W. Taam and M. Hamada, "Detecting spatial effects from factorial experiments: An application from integrated-circuit manufacturing," *Technometrics*, vol. 35, pp. 149–160, May 1993.

[13] J. Kim, Y. Lee, and H. Kim, "Detection and clustering of mixed-type defect patterns in wafer bin maps," *IISE Trans.*, vol. 50, no. 2, pp. 99–111, 2018.

[14] L.-C. Chao and L.-I. Tong, "Wafer defect pattern recognition by multi-class support vector machines by using a novel defect cluster index," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 10158–10167, 2009.

[15] G. Choi, S.-H. Kim, C. Ha, and S. J. Bae, "Multi-step ART1 algorithm for recognition of defect patterns on semiconductor wafers," *Int. J. Prod. Res.*, vol. 50, no. 12, pp. 3274–3287, 2012.

[16] J. Wang, P. Zheng, Y. Lv, J. Bao, and J. Zhang, "Fog-IBDIS: Industrial big data integration and sharing with fog computing for manufacturing systems," *Engineering*, vol. 5, no. 4, pp. 662–670, 2019, doi: 10.1016/j.eng.2018.12.013.

[17] W. Junliang, X. Chuqiao, Z. Jie, B. Jingsong, and Z. Ray, "A collaborative architecture of the industrial Internet platform for manufacturing systems," *Robot. Comput. Integr. Manuf.*, vol. 61, Feb. 2020, Art. no. 101854, doi: 10.1016/J.RCIM.2019.101854.

[18] 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.

[19] C.-H. Wang, "Separation of composite defect patterns on wafer bin map using support vector clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2554–2561, 2009.

[20] T. Yuan, W. Kuo, and S. J. Bae, "Detection of spatial defect patterns generated in semiconductor fabrication processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 392–403, Aug. 2011.

[21] Y. Xia, H. Yu, and F. Wang, "Accurate and robust eye center localization via fully convolutional networks," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 5, pp. 1127–1138, Sep. 2019.

[22] P. M. Kebria, A. Khosravi, S. M. Salaken, and S. Nahavandi, "Deep imitation learning for autonomous vehicles based on convolutional neural networks," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 1, pp. 82–95, Jan. 2020.

[23] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2017.

[24] M. Ghahramani, Y. Qiao, M. C. Zhou, A. O'Hagan, and J. Sweeney, "AI-based modeling and data-driven evaluation for smart manufacturing processes," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 4, pp. 1026–1037, Jul. 2020.

[25] 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.

[26] K. Kyeong and H. Kim, "Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 3, pp. 395–402, Aug. 2018.

[27] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 764–773.

[28] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.

[29] F. Liu, D. Liu, J. Tian, X. Xie, X. Yang, and K. Wang, "Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101793, doi: 10.1016/j.media.2020.101793.

[30] D. Cao, Z. Chen, and L. Gao, "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks," *Human Centric Comput. Inf. Sci.*, vol. 10, no. 1, p. 14, 2020, doi: 10.1186/s13673-020-00219-9.

[31] X. Wang, K. C. K. Chan, K. Yu, C. Dong, and C. C. Loy, "EDVR: Video restoration with enhanced deformable convolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work*, Jun. 2019, pp. 1954–1963, doi: 10.1109/CVPRW.2019.00247.

[32] L. Shi, Y. Zhang, J. Hu, J. Cheng, and H. Lu, "Gesture recognition using spatiotemporal deformable convolutional representation," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 1900–1904.

[33] J. Chen, Y. Pan, Y. Li, T. Yao, H. Chao, and T. Mei, "Temporal deformable convolutional encoder-decoder networks for video captioning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 8167–8174, doi: 10.1609/aaai.v33i01.33018167.

[34] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, "A brief introduction to OpenCV," in *Proc. 35th Int. Conv. MIPRO*, 2012, pp. 1725–1730.