Journal of Intelligent Manufacturing (2025) 36:319–330
https://doi.org/10.1007/s10845-023-02242-w

Shape prior guided defect pattern classification and segmentation in wafer bin maps

Rui Wang[1] · **Songhao Wang**[2] · **Ben Niu**[3]

Received: 16 March 2023 / Accepted: 9 October 2023 / Published online: 7 November 2023

Abstract
In semiconductor manufacturing, patterns formed by defective dies in a wafer bin map (WBM) reveal possible problems during the wafer fabrication process. Therefore, the identification of these patterns is important for root cause diagnosis and yield enhancement. Recently, as the manufacturing process becomes increasingly complicated, mixed-type defect patterns have been frequently observed on the WBMs. The joint classification and segmentation of each pattern contained in the mixed-type pattern is challenging, especially when these patterns are connected or overlapped. This study proposes a shape prior guided method in which the shape templates are deformed to match the patterns with a spatial transformer network. The deformation based method is able to give plausible results while reducing computation cost of the network. Furthermore, the method is flexible and easily extended to deal with complex shapes by defining different shape templates. The experimental results demonstrate the effectiveness of the proposed method in the identification of the pattern types, as well as the separation of each isolated pattern.

Keywords Semiconductor manufacturing · Wafer bin map · Shape prior · Defect pattern recognition

Introduction

Semiconductor manufacturing is a complex process which consists of hundreds of steps. Large amount of data are automatically generated, providing valuable information for process monitoring and yield enhancement. Among the collected data, wafer bin maps (WBMs) are widely examined, which encodes the spatial distribution of defective dies on the wafer (Kim & Behdinan, 2023). Specifically, the defective dies are represented with value 1 and the normal dies with value 0. WBMs may contain typical patterns formed by defective dies that are caused by specific manufacturing processes, and thus the detection and recognition of such patterns is of great significance for the early detection of abnormalities and the analysis of root causes (Abd Al Rahman et al., 2021; Chien et al., 2013). Figure 1 shows four typical examples of defect patterns. WBMs contain both systematic defects that form clusters, as well as random defects that are isolated and seen as noise to the patterns. These patterns can trace back to assignable causes during the manufacturing process. For example, scratches are caused by machine handling problems, while ring patterns are generated by layer-to-layer misalignment in storage node and etching problems (Kim et al., 2018).

Conventionally, the identification of defect patterns in WBMs relies on manual inspection with the human eye, which is time-consuming and highly subjective (Hsu et al., 2020). In Recent years, deep learning methods are widely applied to solve practical problems in both academia and industry. For example, Adly et al. (2015) proposed a simplified subspaced regression framework for the identification of defect patterns in WBMs. In Jin et al. (2020), features extracted from a convolutional neural network (CNN) were

✉ Ben Niu
drniuben@gmail.com

Rui Wang
r.wang@hit.edu.cn

Songhao Wang
wangsh2021@sustech.edu.cn

[1] School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China

[2] School of Business, Southern University of Science and Technology, Shenzhen, China

[3] College of Management, Shenzhen University, Shenzhen, China

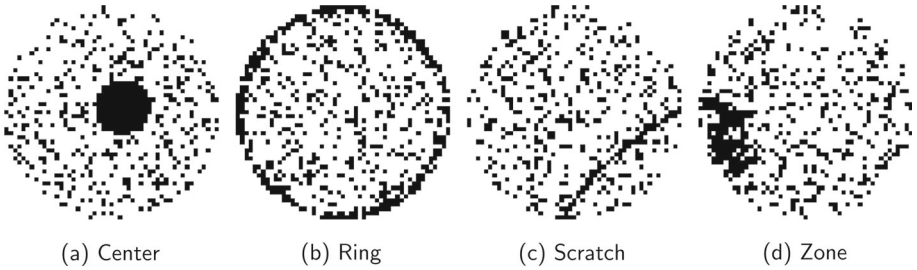

Fig. 1 Examples of WBMs with single patterns

used to train an error-correcting output codes (ECOC) model for WBM pattern classification. Yu and Liu (2020) built a two-dimensional principal component analysis-based convolutional autoencoder for the recognition of defect patterns in WBMs. The framework of ensemble convolutional neural network (ECNN) (Hsu & Chien, 2022; Piao & Jin, 2023) was also applied for WBM pattern classification, and a weighted majority was used to choose weights for the base CNN models.

Most of the existing studies focus on the identification of single defect patterns. Nowadays, mixed-type defect patterns, where multiple types of defects exist in a piece of wafer, become more common due to the evolution of the manufacturing technology. Multiple manufacturing problems can be identified from mixed-type patterns, with each defect pattern assigned to its corresponding root cause (Yuan & Kuo, 2008). The appearance of the mixed-type patterns may vary in different sizes and locations with interactions in the WBMs. For the classification of mixed-type defect patterns in WBMs, deep learning models have been widely used due to their great power in image recognition. Kyeong and Kim (2018) classified mixed-type defect patterns in WBMs with combinations of individual CNN-based classification models for each pattern. Wang et al. (2020) introduced a deformable convolutional network to extract high-quality features for the mixed-type pattern recognition problem. Shin et al. (2022) proposed a Mixup-based strategy for training CNNs to detect mixed-type defects in WBMs using samples with only single-type patterns. These models successfully recognize multiple patterns in the WBMs, but they are not able to find the location of each specific pattern, which provides information for subsequent analysis of scanning electron microscopy (SEM) and energy-dispersive x-ray (EDX) to determine the source of defect generation (O'Leary et al., 2020).

The joint identification of the class and the location of mixed-type patterns is complicated, as the combination of patterns may vary in number, types and positions, etc. The complexity of mixed-type patterns may confuse the classifiers, while increasing the difficulty of pattern separation. To solve this problem, Kim et al. (2018) adopted a clustering based method called the infinite warped mixture model (iWMM) to find different pattern clusters without the identification of number of groups. Wang and Chen (2022) developed a tensor voting based approach to extract regions and curves from the WBM patterns, which were then recognized by a decision tree with the shape information. However, the clustering methods cannot give semantically interpretable categories that are meaningful in real applications. Recently, Yan et al. (2023) proposed to use semantic segmentation based approach to segment the WBM into different pattern classes. Nag et al. (2022) presented WaferSegClassNet (WSCN), a novel network based on encoder-decoder architecture, to perform simultaneous classification and segmentation of both single and mixed-type defect patterns.

Kim et al. (2022) applied the single shot detector (SSD), which is an object detection method, for the detection and localization of mixed-type patterns. Chiu and Chen (2021) utilized the instance segmentation model Mask R-CNN with data augmentation for the segmentation and classification of defect patterns. Existing solutions are capable of simultaneously obtaining the class and the position information of the mixed-type patterns, when the patterns are connected. For the unwarpping of overlapped patterns, Kong and Ni (2019) first used a CNN to classify WBMs with single or mixed-type patterns, and then located and classified the overlapping patterns through template matching method. They (Kong & Ni, 2020) also developed a U-net based boundary detection and overlapped pattern unwarpping method for complex pattern segmentation, followed by a CNN to classify the patterns. For current studies on the recognition of mixed-type defect patterns, unsupervised clustering based methods require sophisticated algorithms to form clusters, while the clusters may not be assignable to typical WBM patterns. Segmentation based methods rely on deep models that are computationally expensive, yet they fail to take the pattern overlap into account. Unlike a single end-to-end model, unwarpping methods represent a process of several steps, which appears more complicated and may need a greater amount of processing. Furthermore, although existing methods can provide plausible result, there is no guarantee that the extracted patterns satisfy the desired shape constraints. Therefore, the segmentation and classification of mixed-type pattern in WBMs is still challenging, especially when these patterns are overlapped or connected.

In this study, we propose a practical approach for the segmentation and classification of mixed-type patterns in

WBMs. Different from the conventional deep learning based segmentation models (Shelhamer et al., 2017; Ronneberger et al., 2015), which detect the objects by optimizing pixel-wise loss functions, our model adopts the shape priors during training. Since the defect patterns in WBMs are usually common types, their shape priors could be pre-determined. In the literature, the incorporation of shape priors into network training have shown improved performance for image segmentation (Lee et al., 2019; Zotti et al., 2018). In our model, the segmentation is performed with deformations of the given shape prior, which gives plausible results that adhere to specific shape constraints, while reducing the computation cost of neural networks. Spatial transformer network (STN) (Jaderberg et al., 2015), which was originally proposed to improve model invariance to image translation, scale and rotation, is used to perform affine transformations to align the shape priors to the patterns of interest. The STN could be applied for segmentation task with satisfactory performance, under the hypothesis that it requires much less precise features than a segmentation network (Pak et al., 2020). The use of STN allows the separation of overlapped and connected patterns by matching with pre-defined pattern templates separately. To be specific, each pre-defined shape template is deformed separately to match the pattern in the WBM to be examined. Consequently, each pattern type is extracted independently by the deformations of the corresponding shape template, making it possible for a pixel to be allocated to more than one pattern class, which eliminates the influence of the interactions between different patterns. Our proposed method is able to locate and classify mixed-type patterns simultaneously, and can be easily extended with additional shape templates as input to the model.

The remainder of this article is structured as follows. "Methodology" section gives a detailed introduction to the proposed defect pattern classification and segmentation method. In "Experiments" section, the experimental results are presented to demonstrate the effectiveness of the proposed model. Finally, "Conclusion" section concludes this paper.

Methodology

In this section, we propose a novel classification and segmentation model for mixed-type defect patterns in WBMs. The approach is inspired by the idea of network based registration, and it takes shape priors as the additional input to the network. The segmentation is then performed by learning the optimal deformations of the shape templates for the accurate alignment with the given image. An overview of the method is depicted in Fig. 2. The detailed description of the proposed method is presented in the following subsections.

Shape templates

The utilization of shape priors in segmentation networks can be in various forms, such as level sets (Cremers et al., 2006), statistical models (Milletari et al., 2017) and conditional random fields (Zheng et al., 2015). In this paper, we define multiple templates manually as shape priors, and they are added to the network as auxiliary inputs that are deformed to match the patterns of interest. In this way, the model estimates the transformation of the templates regarding each class instead of a point-wise segmentation map, which provides a solution to the extraction of complex and overlapped patterns. The segmentation and classification results heavily rely on the deformation of the shape templates, which should be carefully chosen in order to match most of the samples. Generally, these shape priors can be obtained with manual selection of either real or simulated patterns. The templates should be typical, clear, and preferably without noise, to improve the accuracy of deformation.

Network architecture

We present a model that takes images and templates as input, and outputs images warped to the templates. Our model enables the incorporation of any shape prior to segment and classify the mixed-type patterns, overcoming the difficulty when the patterns are connected or overlapped. In addition, the template deformation based pattern extraction reduces the chance of discretization of the extracted patterns.

Image features

In this study, as the classification and segmentation are performed at the same time, a shared network is applied for image feature extraction to reduce the computation cost. The target images and shape templates are jointly sent to the network as input, which extracts valuable information from the images under the guidance of the shape priors. Features should be carefully extracted to be representative in the tasks of both segmentation and classification. We choose to use a feed-forward network architecture to perform feature extraction in this stage. The residual network (ResNet) (He et al., 2016) has been widely applied in the field of image segmentation and object detection, outperforming other deep learning models in various tasks. It is possible to train much deeper networks without increasing the amount of parameters through skip connections between the layers (He et al., 2020). Therefore, the ResNet is applied for feature extraction in our model. The extracted features are then flattened for further processing.

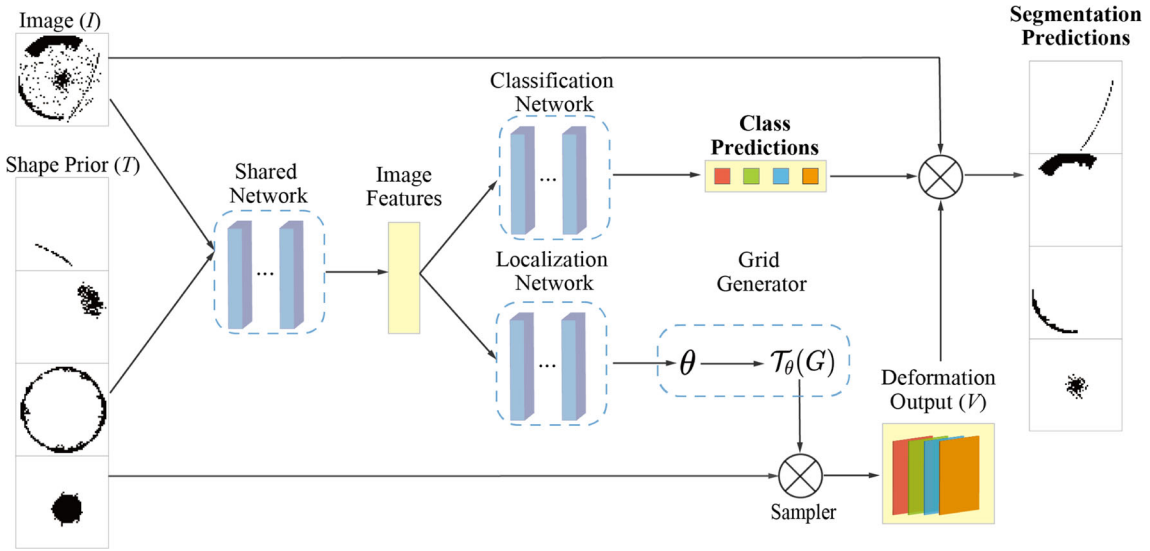

Fig. 2 Overview of the proposed model

Pattern classification

The number of types grows in an exponential way for the mixed-type defect patterns. Theoretically, 2^L defect patterns are identified considering the mixture of L type of single defect patterns. To simplify the representation of the defect types, one-hot encoding is applied to describe the mixed-type defect patterns with several single patterns (Wang et al., 2020). For a mixed-type defect pattern, the pattern type is defined as a row vector $y_i = [y_i^1, y_i^2, \ldots, y_i^L]$, where $y_i^j \in \{0, 1\}$ indicates the occurrence of the single defect pattern. For example, when $L = 4$, $y_i = [0, 1, 0, 1]$ indicates a mixture of defect pattern type two and four in a single WBM. For multi-label classification, two fully-connected layers together with the Sigmoid activation function are defined to identify the inclusion of the pattern types. The output of the network is a vector of length L, with each element c_i^j ranging from 0 to 1. The classification is conducted by applying a threshold, if the value $c_i^j \geq 0.5$, then the WBM is predicted to contain the jth defect pattern type.

Template deformation

The model utilizes the idea of STN that deforms the shape prior template T to match the input image I by learning the transformation parameters θ. $\mathcal{T}_\theta(G)$, where G is a regular grid $G = \{G_p\}$ of pixels $G_p = (x_p^t, y_p^t)$, is defined as the function that maps the target coordinates (x_p^t, y_p^t) to the source coordinates (x_p^s, y_p^s). Each output pixel is computed by a sampling kernel applied to the template grid, and for 2D affine transformation, the function is

$$\begin{pmatrix} x_p^s \\ y_p^s \end{pmatrix} = \mathcal{T}_\theta(G_p) = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_p^t \\ y_p^t \\ 1 \end{pmatrix}. \tag{1}$$

This transformation defines six parameters that allows cropping, translation, rotation and scale of the templates.

For WBMs, the template T, input image I and the output feature map V are all of size $H \times W$. The output feature map V is then defined by the sampling points $\mathcal{T}_\theta(G)$ and the template T. $V(p)$ is the output value of pixel p at location (x_p^t, y_p^t), determined by a sampling kernel $k()$ applied to the input coordinates:

$$V(p) = \sum_n^H \sum_l^W T(n, l) k(x_p^s - l; \Phi_x) k(y_p^s - n; \Phi_y) \quad \forall p \in [1, \ldots, HW], \tag{2}$$

where Φ_x and Φ_y are the parameters of the sampling kernel $k()$, and $T(n, l)$ is the value at location (n, l). For the kernel $k()$, bilinear image interpolation is applied, that is

$$V(p) = \sum_n^H \sum_l^W T(n, l) \max(0, 1 - |x_p^s - l|) \max(0, 1 - |y_p^s - n|). \tag{3}$$

The deformation parameters θ is learned by a neural network conditioned on the image I and the template T. For simplicity, a multi-layer perceptron (MLP) with two fully-connected layers is applied for the estimation of deformation parameters.

Pattern segmentation

The extracted pattern of each defect mode should have its original shape instead of making up new defective points. Therefore, the deformed shape priors are then multiplied with the original input image I by element-wise production, to capture the shape characteristics of the typical defect pattern. After the multiplication, the deformed shape priors may still match with the isolated defective points, which should be avoided in the process of pattern extraction. To improve the accuracy of segmentation, the classification results are combined to get the final prediction. That is, if the pattern exists in the WBM, the segmentation mask is retained as it is in the prediction result. Otherwise, the mask of 0 values is returned, if the pattern is not detected in the WBM. For a given WBM i, the proposed model is capable of jointly classifying the defect patterns $y_i \in \{0, 1\}^L$ and identifying the corresponding masks $M_i \in \{0, 1\}^{L\times H\times W}$, where L is the number of defect pattern types on the WBM. The segmentation mask of each pattern class is given as:

$$M_i^j = y_i^j \cdot I_i \circ V_i^j, \tag{4}$$

where $\circ$ denotes the Hadamard product of two matrices.

Model evaluation

For multi-label classification, Hamming loss, mixed-type accuracy and exact match accuracy are commonly used to measure the performance (Chiu & Chen, 2021; Tsoumakas & Katakis, 2007). Hamming loss (HL) computes the proportion of labels that are incorrectly predicted, and smaller value gives better performance. Mixed-type accuracy (MTA) calculates the ratio of the size of the interaction set to the union set, measured between the actual and predicted classes. Exact match accuracy (EMA) is the percentage of WBMs that the predicted classes exactly match the actual pattern. Higher values of the two accuracy metrics indicate the superiority of the model. Let y_i and $\hat{y}_i$ be the actual and predicted defect pattern class, respectively. The metrics are defined as:

$$HL = \frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L} I(\hat{y}_i^j \neq y_i^j), \tag{5}$$

$$MTA = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i \cap y_i|}{|\hat{y}_i \cup y_i|}, \tag{6}$$

$$EMA = \frac{1}{N}\sum_{i=1}^{N} I(\hat{y}_i = y_i), \tag{7}$$

where N is the number of WBMs.

To quantify the segmentation quality, the intersection over union (IoU) (Çiçek et al., 2016) that rewards a complete and total match between predicted and ground-truth area is introduced. The IoU is defined as the area of the overlapped pattern divided by the area of the union pattern.

$$IoU_i^j = \frac{Area(\hat{M}_i^j \cap M_i^j)}{Area(\hat{M}_i^j \cup M_i^j)}, \tag{8}$$

For each mixed-type pattern, the mean IoU (mIoU) is used to evaluate the model segmentation performance.

$$mIoU_i = \frac{1}{L}\sum_{j=1}^{L} IoU_i^j, \tag{9}$$

where M_i^j and $\hat{M}_i^j$ denote the actual and predicted mask of the jth pattern in the ith WBM. If pattern j is not detected in the WBM i, the IoU is not calculated.

Experiments

Data description

The technique described in Kyeong and Kim (2018) was used to generate mixed-type WBMs to evaluate the proposed method. The shape of the generated patterns was based on both real data and domain knowledge, with noise added to the simulated patterns. The mixture of four single-type patterns were considered, i.e. Scratch, Zone, Ring and Center, as illustrated in Fig. 1. The size of each WBM was 56×53, i.e., each WBM contained a total of 2968 dies. The percentage of defective dies ranged from 3% to 27%. Sixteen classes of mixed-type patterns were generated with combination of the four identified patterns. A total of 8000 samples were generated, 500 for each pattern. The data was separated into training, validation and testing datasets in a ratio of 8:1:1.

Model training

The shape priors used in our model are shown in Fig. 2. The shape templates were chosen from generated patterns manually after noise removal, which should be representative of the pattern types with clear and perfect shapes on the WBM. Zero padding were performed for the WBMs to be a perfect square with size 56×56 for easier computation.

For shared network that aims to extract valuable information from the images, ResNet was adopted for the network architecture. Commonly used number of layers in ResNet are 18, 34, 50 and 101. They gave similar results for the WBM classification and segmentation task identified in this study. Considering the computation cost and model performance, a simplified ResNet was chosen and the details of the network are shown in Fig. 3. The WBM samples together with

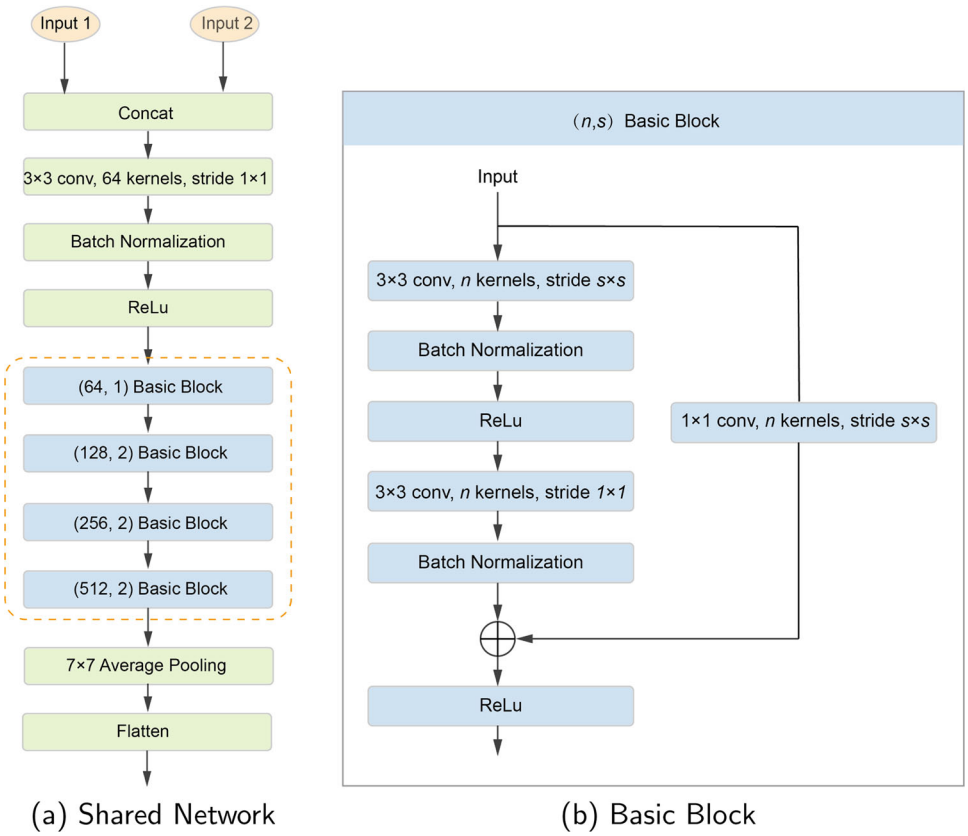

Fig. 3 Schematic architecture of the shared network and the basic blocks used to construct the model

shape templates were first concatenated and sent to a convolutional layer, and the model was then connected with four residual blocks, before average pooling and flattened for further processing. The output size for the shared network was 512. The classification and localization networks were chosen to have two fully-connected layers, with 256 neurons for the first layer, before they were connected to the output layer with the corresponding number of neurons for classification and deformation. The Sigmoid activation function was adopted for the identification of the pattern classes and the transformation parameters. The transformation learned by the segmentation was applied to the shape templates to match the WBMs, and the result of the deformation was then combined with the predicted classes to obtain the segmentation results.

The proposed model accomplishes both classification and segmentation in the same model, so the loss function consists of two parts, i.e., the classification loss and mask loss. Classification loss calculates the cross entropy of the predicted class and actual class. Mask loss is a weighted cross entropy function of the predicted and ground truth mask defined as

$$Loss_M = -w \cdot p \log(q) - (1-p)\log(1-q), \tag{10}$$

where w is the weighting factor, p and q are target and predicted distributions, respectively. Larger values of $w > 1$ penalize false negatives, which avoids predicting small pattern areas as background noise due to class imbalance of defective and normal pixels. Compared with the area in a whole WBM, the appearance of defect patterns is sparse, especially for scratches. Therefore, different weights were chosen for these patterns, and the weight of [16, 4, 1, 1] was adopted for the patterns Scratch, Zone, Ring and Center, considering the size and the segmentation difficulty of the patterns. The weights of the loss components can be adjusted, according to the significance of the tasks.

$$Loss = Loss_y + \lambda * Loss_M,\ \lambda \in \mathbb{R}^+. \tag{11}$$

We chose the combination $\lambda = 1$ for this study, based on the results on the validation set. The model were trained for 2000 epoches, and the batch size was set to 64. The Adam optimizer (Kingma & Ba, 2014) was applied with an exponential learning rate decay with initial learning rate of 0.01. The model was coded using Python 3.7, TensorFlow 1.15 environment on a Windows 10, and executed on a computer with a Core i9-10885 H CPU@2.40 GHz and a NVIDIA Quadro RTX 5000 GPU.

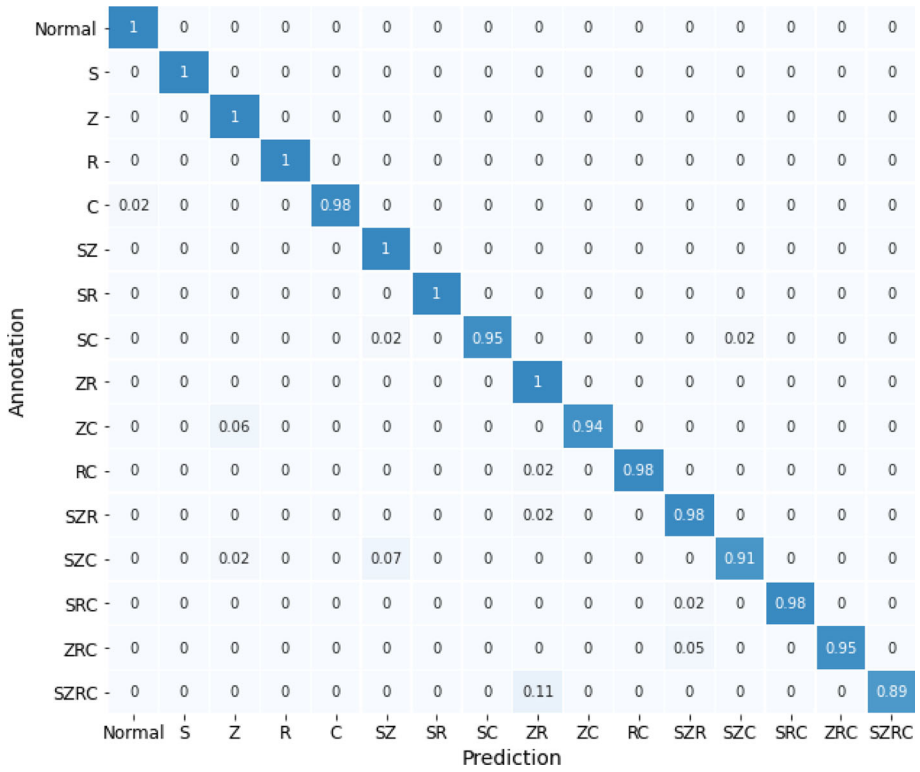

Fig. 4 Confusion matrix result of the mixed-type pattern classification

Results of pattern classification

We first examined the classification performance of the proposed method. Figure 4 presents the confusion matrix of the mixed-type pattern classification. The percentage of WBMs that was accurately classified was shown in the matrix diagonal, and mixed-type patterns were denoted with the combination of the pattern initials. The overall accuracy reaches 97.3% for the 16 pattern classes. The model gave satisfactory results ($\geq$ 95%) for most of the classes, especially for patterns with single type and the mixture of two defect types. It was more difficult to recognize patterns when three or more defect types exist on the same WBM, as their interactions may increase the probability of misclassification. The recognition rates were 91% and 89% for the pattern SZC and SZRC, respectively. This may be caused by the connectivity of the scratch and zone patterns at the edge, as well as the scratch and center patterns at the middle part, while the scratch patterns appear long and thin that were hard to be detected.

The classification performance of the proposed method was compared with two CNN based models renamed as CNN_N (Nakazawa & Kulkarni, 2018) and CNN_T (Kyeong & Kim, 2018), which were used to recognize mixed-type defect patterns on WBMs. The comparison results are shown in Table 1. The Hamming loss, mixed-type accuray and exact match accuracy were computed for these three methods. The values of MTA were higher than that of EMA, as EMA is more strict which measures the percentage of WBMs that were perfectly recognized. Overall, our method outperformed the two CNN based models for all the three indices. For single-type defect patterns, the Normal, Scratch, Zone and Ring patterns were successfully identified, and the recognition rate of center patterns reached 98.2%, which was competitive with the CNN models. For mixed-type defect patterns, our method gave better performance for most of the pattern types, especially for the two-type mixture patterns SZ, SR and ZR, which were perfectly classified in the testing results. For complex defect patterns that contains three or more types of defect patterns, the performance of our method was superior for SZR, SRC and SZRC patterns. Though our method performed relatively worse on the classes SZC and ZRC than CNN_N, it still obtained acceptable results, which demonstrated the effectiveness of the proposed method.

Results of pattern segmentation

For defect pattern segmentation, the IoU was calculated for each pattern type, and the results are presented in Table 2. For single type defect patterns, zone patterns gave the best performance, with the average IoU reaches 0.936. The scratch patterns were relatively difficult to be located, as they can appear at any positions in the WBMs and can be mixed up with the isolated noise due to their long and thin appearances. The results of center patterns were worse than that of zone and ring patterns, which may be explained by the irregular-

Table 1 Comparison results of pattern classification

Pattern type	HL			EMA			MTA		
	Our method	CNN_N	CNN_K	Our method	CNN_N	CNN_K	Our method	CNN_N	CNN_K
Normal	**0.0000**	0.0045	0.0133	**1.000**	0.982	0.946	**1.000**	0.982	0.946
S	**0.0000**	**0.0000**	**0.0000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Z	**0.0000**	0.0068	0.0068	**1.000**	0.973	0.973	**1.000**	0.987	0.987
R	**0.0000**	0.0144	0.0048	**1.000**	0.942	0.981	**1.000**	0.952	0.981
C	0.0045	**0.0000**	**0.0000**	0.982	**1.000**	**1.000**	0.982	**1.000**	**1.000**
SZ	**0.0000**	0.0330	0.0425	**1.000**	0.887	0.849	**1.000**	0.943	0.934
SR	**0.0000**	0.0100	0.0300	**1.000**	0.960	0.880	**1.000**	0.980	0.943
SC	0.0188	**0.0000**	0.0204	0.950	**1.000**	0.918	0.975	**1.000**	0.969
ZR	**0.0000**	0.0204	0.0255	**1.000**	0.918	0.898	**1.000**	0.963	0.959
ZC	0.0147	0.0227	**0.0045**	0.941	0.909	**0.982**	0.971	0.967	**0.994**
RC	0.0088	0.0102	**0.0000**	0.982	0.959	**1.000**	0.988	0.983	**1.000**
SZR	**0.0047**	0.0268	0.0268	**0.981**	0.911	0.893	**0.993**	0.964	0.964
SZC	0.0278	0.0532	**0.0160**	0.911	**0.979**	0.936	0.963	**0.993**	0.981
SRC	**0.0108**	0.0217	0.0326	**0.978**	0.935	0.913	**0.989**	0.976	0.964
ZRC	0.0127	**0.0000**	0.0048	0.949	**1.000**	0.981	0.988	**1.000**	0.995
SZRC	**0.0266**	0.0389	0.0667	**0.894**	0.844	0.733	**0.964**	0.961	0.933
Overall	**0.0081**	0.0134	0.0184	**0.973**	0.950	0.930	**0.988**	0.978	0.972

It shows the comparison results of different methods. The best results are shown in bold to highlight the superiority

Table 2 Pattern segmentation results of different defect patterns

Pattern type	Extracted patterns			
	Scratch	Zone	Ring	Center
S	0.825	–	–	–
Z	–	0.936	–	–
R	–	–	0.907	–
C	–	–	–	0.851
SZ	0.729	0.918	–	–
SR	0.649	–	0.899	–
SC	0.800	–	–	0.852
ZR	–	0.854	0.844	–
ZC	–	0.936	–	0.852
RC	–	–	0.906	0.841
SZR	0.553	0.829	0.838	–
SZC	0.702	0.918	–	0.839
SRC	0.630	–	0.896	0.844
ZRC	–	0.850	0.852	0.844
SZRC	0.541	0.817	0.832	0.833
Overall	0.678	0.882	0.872	0.844

ity and ambiguity of the region edges. For mixed-type defect patterns, the difficulty of pattern segmentation increased with the mixture of more defect pattern types, and therefore the accuracy decreased with the complexity of mixed-type patterns. For the extraction of scratch patterns, the co-occurrence of ring patterns lowered the values of the IoU indice, as they were both curves and their interactions at the wafer edge were not easy to be resolved. The segmentation of ring patterns and the zone patterns were largely influenced by each other, which were connected at the wafer edge. The center patterns, however, were not significantly affected by other patterns, because the chances of their interactions were relatively lower.

The segmentation performance of connected patterns is illustrated in Fig. 5. Three overlapped examples were given, in which the complex patterns were hard to be separated into clusters of single patterns. The four columns illustrated the separated patterns matched by the four corresponding templates, respectively. In all the three WBMs, the zone and ring patterns were connected, which was often the case when they both appeared in the same WBM. In the first WBM, the scratch was an arc with a large radius, which was connected with the center region. For the second WBM, it was hard to determine the location of the scratch pattern, as it was connected with the ring and the zone patterns. In the third example, the scratch and the zone regions were connected and close to each other, making it difficult to differentiate between the two patterns. It was observed from Fig. 5 that our method performed well in the extraction of connected and overlapped patterns of different shapes. Our method was also robust when dealing with isolated points, which were seen as noise to the defect patterns. The overall results from representative examples showed that our method was able to extract single defect patterns that were clear and complete, from noisy WBMs with complex patterns of arbitrary shapes.

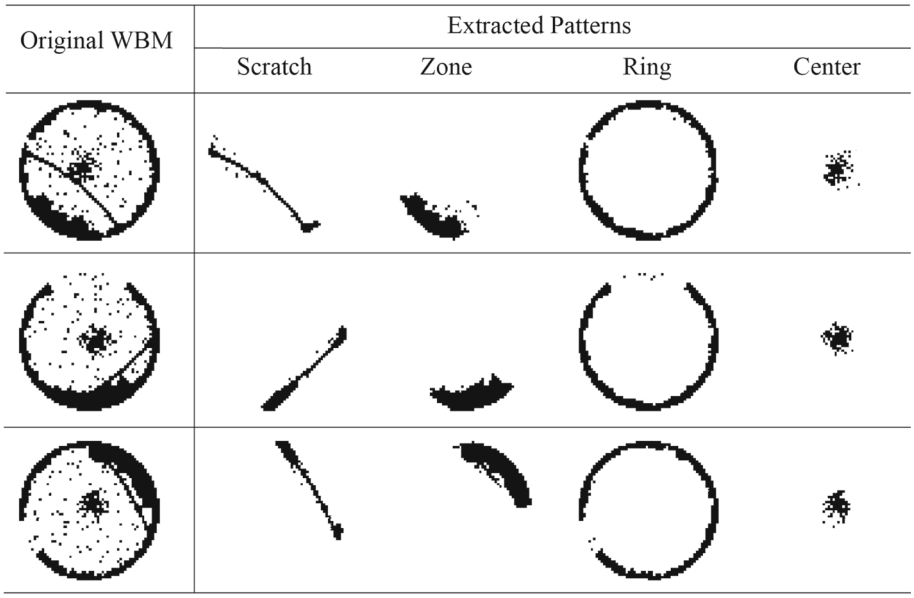

Fig. 5 Extraction of connected patterns in the WBMs

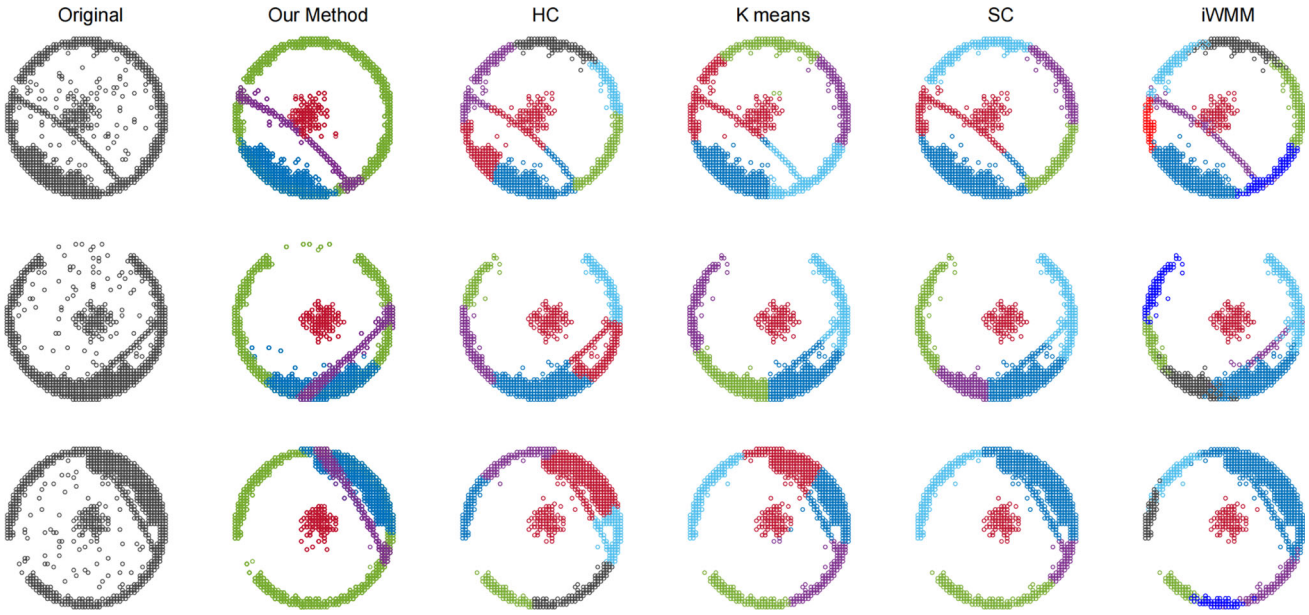

Fig. 6 Comparison of pattern segmentation results of the hierarchical clustering (HC), K-means, spectral clustering (SC), iWMM and our method

To further evaluate the ability of our method for pattern segmentation, we compared our method with the conventional unsupervised clustering methods that was widely applied in the literature, namely, hierarchical clustering (HC), K-means, spectral clustering (SC) and iWMM (Kim et al., 2018). Our method was inherently robust to remove noisy points from the WBMs, so no additional denoising step was required. For these clustering based methods to be compared, we applied the denoising method mentioned in Wang and Chen (2022) in order to compare the pattern separation performance. For hierarchical clustering, we followed the bottom-up approach. Average linkage and Euclidean distance were chosen for the formation of clusters. To use K-means for clustering, the number of clusters K is required. Here we choose $K = 5$, i.e., five clusters are constructed based on the defect points. Spectral clustering performed clustering task by first constructing an affinity matrix, and eigenvectors were then computed to form the new data points. Classic clustering algorithms were applied on the new data points to find the final clusters. We construct the affinity matrix with the Euclidean distance and use K-means for the clustering of new data points. For iWMM, the same parameter setting of Kim et al. (2018) was applied.

Table 3 mIoU results for different numbers of mixed pattern types

Method	Number of mixed pattern types					Overall
	0	1	2	3	4	
Our method	**1.000**	**0.879**	**0.839**	0.800	0.775	**0.845**
U-Net	0.982	0.577	0.627	0.746	**0.877**	0.682
WSCN	0.801	0.785	0.797	**0.817**	0.817	0.800
FCN	1.000	0.799	0.732	0.667	0.594	0.741
SegNet	0.806	0.852	0.779	0.751	0.728	0.789

Figure 6 compares the clustering performance of our method and the four competitive methods mentioned above. The first column represents the three WBMs to be examined, and rows 2 to 6 show the results obtained using our method, HC, K-means, SC and iWMM, respectively. The colors indicate different clusters for each of the WBM. The results of our method was shown with superposition of the extracted patterns following the sequence of ring, zone, scratch and center for better visualization. Overall, our method was capable of removing noise and insignificant shapes, which was comparable with the existing denoising method in Wang and Chen (2022). In addition, our method shows better performance compared with traditional clustering based methods, which may group connected patterns into the same clusters or separate the same pattern into many fractions instead. For the first WBM, HC, K-means and SC cannot separate the scratches from the cluster at the wafer center due to the connectivity of the two patterns. The iWMM was able to isolate the scratch correctly, but it cannot automatically combine connected ring patterns at the edge. For the second WBM, the ring and zone patterns at the edge cannot be differentiated using HC, K-means and SC, while iWMM shows good separation results; however, it divided the edge pattern into two local clusters. All the methods gave similar clustering results for the last WBM, which failed to separate scratch and zone patterns when they were really close to each other.

We also compared our model with four segmentation based baseline models: the U-net (Yan et al., 2023), WSCN (Nag et al., 2022), FCN (Shelhamer et al., 2017) and SegNet (Badrinarayanan et al., 2017). The cross entropy loss was utilized to compute the differentce between the predicted mask and the ground truth, and hyperparameters were chosen with the grid search. Table 3 illustrates the comparison of mIoU results considering different numbers of pattern types included in the WBMs. For the overall performance, our method was significantly better than the other methods. When the WBM has less than three patterns, which is typically the case in practical applications, our method performed the best. This results demonstrate the effectiveness of our proposed method. When four patterns co-exist in the same WBM, U-Net is better than our method, but the results in other classes are far less satisfactory. Figure 7 further plots the mIoU values over different mixed-type pattern classes. Overall, the compared segmentation based methods show similar patterns, among which WSCN seems to provide better performance. Our method outperformed WSCN in almost all the classes, with the exception of patterns that contains ZR, where results were also comparable. U-Net provided marginally higher performance in S, SZ, SZR and SZRC, but the mIoU results of other classes were significantly worse, notably for R, C, ZC and RC. Furthermore, compared to other methods, which experienced significant fluctuations, our method produced more stable results across the various pattern classes.

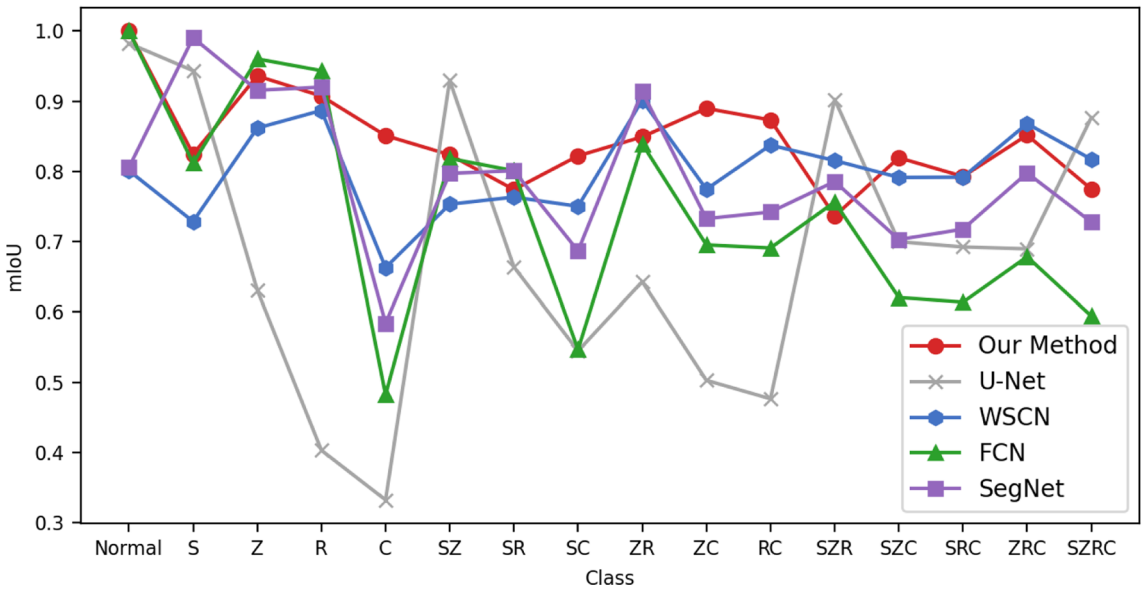

Fig. 7 The segmentation performance of our method and other segmentation models

Conclusion

This paper proposes a shape prior guided defect pattern classification and segmentation method in WBMs. The spatial transformer network is applied to deform the shape priors in order to match with the patterns for extraction. The deformation is conducted by estimating a transformation field while optimizing the alignment between the template and the target pattern. The registration based segmentation is able to extract single patterns from complex patterns composed of connected and overlapped patterns. Experiments show that the proposed method is effective for both classification and separation of mixed-type defect patterns in WBMs. The performance of the method is competitive with state-of-the-art segmentation algorithms, while it takes the connectivity and interaction of the patterns into account.

The proposed method could be easily extended by defining different representative shape priors, both real and hand-crafted ones are applicable. The model could also be separately defined, by training a network for each pattern type individually and combining the results together. In addition, more than one templates can be determined for a pattern type, improving the capability of extracting different shapes of the same pattern. Our model could be applied in various fields to solve computer vision problems, such as object detection and image segmentation, especially when the targets are connected or overlapped. In other applications, the shape priors are not restricted to shape templates, flexible forms of shape priors, including probabilistic and geometric representations, could also be incorporated into the model.

Author Contributions All authors contributed to the idea of the study. The first draft of the manuscript was written by RW and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The study is supported by Key Program of National Natural Science Foundation of China (No. 72334004), General Program of National Natural Science Foundation of China (No. 71971143), Youth Program of National Natural Science Foundation of China (No. 72101065, No. 72101106), Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515110336), Guangdong Provincial Philosophy and Social Sciences Planning Project (No. GD22CGL35), Special Projects in Key Fields of Ordinary Colleges and Universities in Guangdong Province (No. 2022ZDZX2054), University Innovation Team Project of Guangdong Province (No. 2021WCXTD002), and Shenzhen Science and Technology Program (No. RCBS20210609103119020, No. RCBS20221008093124063).

Data availability The data that support the findings of this study are available from the corresponding author upon request.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abd Al Rahman, M., Danishvar, S., & Mousavi, A. (2021). An improved capsule network (wafercaps) for wafer bin map classification based on dcgan data upsampling. *IEEE Transactions on Semiconductor Manufacturing, 35*(1), 50–59. https://doi.org/10.1109/TSM.2021.3134625

Adly, F., Alhussein, O., Yoo, P. D., Al-Hammadi, Y., Taha, K., Muhaidat, S., Jeong, Y.-S., Lee, U., & Ismail, M. (2015). Simplified subspaced regression network for identification of defect patterns in semiconductor wafer maps. *IEEE Transactions on Industrial Informatics, 11*(6), 1267–1276. https://doi.org/10.1109/TSM.2018.2841416

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Chien, C.-F., Hsu, C.-Y., & Chen, P.-N. (2013). Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence. *Flexible Services and Manufacturing Journal, 25*(3), 367–388. https://doi.org/10.1007/s10696-012-9161-4

Chiu, M.-C., & Chen, T.-M. (2021). Applying data augmentation and mask R-CNN-based instance segmentation method for mixed-type wafer maps defect patterns classification. *IEEE Transactions on Semiconductor Manufacturing, 34*(4), 455–463. https://doi.org/10.1109/TSM.2021.3118922

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424-432). https://doi.org/10.1007/978-3-319-46723-8_49

Cremers, D., Osher, S. J., & Soatto, S. (2006). Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision, 69*(3), 335–351. https://doi.org/10.1007/s11263-006-7533-5

He, F., Liu, T., & Tao, D. (2020). Why resnet works? Residuals generalize. *IEEE Transactions on neural Networks and Learning Systems, 31*(12), 5349–5362. https://doi.org/10.1109/TNNLS.2020.2966319

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Hsu, C.-Y., Chen, W.-J., & Chien, J.-C. (2020). Similarity matching of wafer bin maps for manufacturing intelligence to empower industry 3.5 for semiconductor manufacturing. *Computers & Industrial Engineering, 142*, 106358. https://doi.org/10.1016/j.cie.2020.106358

Hsu, C.-Y., & Chien, J.-C. (2022). Ensemble convolutional neural networks with weighted majority for wafer bin map pattern classification. *Journal of Intelligent Manufacturing, 33*(3), 831–844. https://doi.org/10.1007/s10845-020-01687-7

Jaderberg, M., Simonyan, K., Zisserman, A., & kavukcuoglu, K. (2015). Spatial transformer networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., & Garnett, R. (Eds.), *Advances in neural information processing systems* (Vol. 28). https://doi.org/10.48550/arXiv.1506.02025

Jin, C. H., Kim, H.-J., Piao, Y., Li, M., & Piao, M. (2020). Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes. *Journal of Intelligent Manufacturing, 31*(8), 1861–1875. https://doi.org/10.1007/s10845-020-01540-x

Kim, J., Lee, Y., & Kim, H. (2018). Detection and clustering of mixed-type defect patterns in wafer bin maps. *IISE Transactions, 50*(2), 99–111. https://doi.org/10.1080/24725854.2017.1386337

Kim, T., & Behdinan, K. (2023). Advances in machine learning and deep learning applications towards wafer map defect recognition and classification: A review. *Journal of Intelligent Manufacturing, 34*, 3215–3247. https://doi.org/10.1007/s10845-022-01994-1

Kim, T. S., Lee, J. W., Lee, W. K., & Sohn, S. Y. (2022). Novel method for detection of mixed-type defect patterns in wafer maps based on a single shot detector algorithm. *Journal of Intelligent Manufacturing, 33*(6), 1715–1724. https://doi.org/10.1007/s10845-021-01755-6

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980

Kong, Y., & Ni, D. (2019). Recognition and location of mixed-type patterns in wafer bin maps. In *2019 IEEE international conference on smart manufacturing, industrial & logistics engineering (SMILE)* (pp. 4-8). https://doi.org/10.1109/SMILE45626.2019.8965309

Kong, Y., & Ni, D. (2020). Qualitative and quantitative analysis of multi-pattern wafer bin maps. *IEEE Transactions on Semiconductor Manufacturing, 33*(4), 578–586. https://doi.org/10.1109/TSM.2020.3022431

Kyeong, K., & Kim, H. (2018). Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks. *IEEE Transactions on Semiconductor Manufacturing, 31*(3), 395–402. https://doi.org/10.1109/TSM.2018.2841416

Lee, M. C. H., Petersen, K., Pawlowski, N., Glocker, B., & Schaap, M. (2019). TETRIS: Template transformer networks for image segmentation with shape priors. *IEEE Transactions on Medical Imaging, 38*(11), 2596–2606. https://doi.org/10.1109/TMI.2019.2905990

Milletari, F., Rothberg, A., Jia, J., & Sofka, M. (2017). Integrating statistical prior knowledge into convolutional neural networks. In *International conference on medical image computing and computer-assisted intervention* (pp. 161-168). https://doi.org/10.1007/978-3-319-66182-7_19

Nag, S., Makwana, D., & R, S. C. T., Mittal, S., & Mohan, C. K. (2022). Wafersegclassnet–A light-weight network for classification and segmentation of semiconductor wafer defects. *Computers in Industry, 142*, 103720. https://doi.org/10.1016/j.compind.2022.103720

Nakazawa, T., & Kulkarni, D. V. (2018). Wafer map defect pattern classification and image retrieval using convolutional neural network. *IEEE Transactions on Semiconductor Manufacturing, 31*(2), 309–314. https://doi.org/10.1109/TSM.2018.2795466

O'Leary, J., Sawlani, K., & Mesbah, A. (2020). Deep learning for classification of the chemical composition of particle defects on semiconductor wafers. *IEEE Transactions on Semiconductor Manufacturing, 33*(1), 72–85. https://doi.org/10.1109/TSM.2019.2963656

Pak, D.H., Caballero, A., Sun, W., & Duncan, J.S. (2020). Efficient aortic valve multilabel segmentation using a spatial transformer network. In *2020 IEEE 17th international symposium on biomedical imaging (ISBI)* (pp. 1738-1742). https://doi.org/10.1109/TMI.2019.2905990

Piao, M., & Jin, C. H. (2023). CNN and ensemble learning based wafer map failure pattern recognition based on local property based features. *Journal of Intelligent Manufacturing, 34*, 3599–3621. https://doi.org/10.1007/s10845-022-02023-x

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany*, October 5–9, 2015, Proceedings, Part III 18 (pp. 234-241). https://doi.org/10.1007/978-3-319-24574-4_28

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

Shin, W., Kahng, H., & Kim, S. B. (2022). Mixup-based classification of mixed-type defect patterns in wafer bin maps. *Computers & Industrial Engineering, 167*, 107996. https://doi.org/10.1016/j.cie.2022.107996

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM), 3*(3), 1–13. https://doi.org/10.4018/jdwm.2007070101

Wang, J., Xu, C., Yang, Z., Zhang, J., & Li, X. (2020). Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition. *IEEE Transactions on Semiconductor Manufacturing, 33*(4), 587–596. https://doi.org/10.1109/TSM.2020.3020985

Wang, R., & Chen, N. (2022). Detection and recognition of mixed-type defect patterns in wafer bin maps via tensor voting. *IEEE Transactions on Semiconductor Manufacturing, 35*(3), 485–494. https://doi.org/10.1109/TSM.2022.3183008

Yan, J., Sheng, Y., & Piao, M. (2023). Semantic segmentation based wafer map mixed-type defect pattern recognition. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (Early Access)*. https://doi.org/10.1109/TCAD.2023.3274958

Yuan, T., & Kuo, W. (2008). Spatial defect pattern recognition on semiconductor wafers using model-based clustering and bayesian inference. *European Journal of Operational Research, 190*(1), 228–240. https://doi.org/10.1016/j.ejor.2007.06.007

Yu, J., & Liu, J. (2020). Two-dimensional principal component analysis-based convolutional autoencoder for wafer map defect detection. *IEEE Transactions on Industrial Electronics, 68*(9), 8789–8797. https://doi.org/10.1109/TIE.2020.3013492

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., & Torr, P.H. (2015). Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 1529-1537). https://doi.org/10.1109/ICCV.2015.179

Zotti, C., Luo, Z., Lalande, A., & Jodoin, P.-M. (2018). Convolutional neural network with shape prior applied to cardiac MRI segmentation. *IEEE Journal of Biomedical and Health Informatics, 23*(3), 1119–1128. https://doi.org/10.1109/TMI.2019.2905990