Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conditional Generation of Overlapping Mixed-type Wafer Bin Map Defects via Feature-wise Linear Modulation

Jaejoon Choi
Kyung Hee University
Younghoon Kim
y.kim@khu.ac.kr

Kyung Hee University https://orcid.org/0000-0003-0265-3825

Research Article

Keywords: Wafer Bin Map, Mixed-Type Defect Patterns, Conditional Semantic Generation, Feature-wise Linear Modulation, Overlapping Defect Disentanglement

Posted Date: August 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10356216/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Conditional Generation of Overlapping Mixed-type Wafer Bin Map Defects via Feature-wise Linear Modulation

Jaejoon Choi[1] and Younghoon Kim[1*]

[1*]Department of Industrial and Management Systems Engineering, Kyung Hee University, 1732, Deogyeong-daero, Giheung-gu, Yongin-si, 17104, Gyeonggi-do, Republic of Korea.

*Corresponding author(s). E-mail(s): y.kim@khu.ac.kr;

Abstract

In semiconductor manufacturing, mixed-type wafer bin map (WBM) defects contain spatial evidence valuable for process diagnosis and yield management. Existing approaches, however, either provide image-level labels without localizing constituent defects or produce ambiguous predictions in overlapping regions, limiting their usefulness for engineering decision support. To address this gap, we propose the Wafer Conditional Generative Network (WCGN), a conditional generation framework that integrates Feature-wise Linear Modulation into a U-Net architecture. By conditioning the decoder on a target defect label, WCGN converts mixed-type WBM analysis into a label-conditioned mask generation task and disentangles concealed constituent patterns within overlapped defects. Experiments on MixedWM38 for classification and on a generated mixed-type benchmark with pixel-level masks for segmentation show that WCGN outperforms state-of-the-art baselines, achieving an Exact Match Ratio of 97.85% and a mean Intersection over Union of 0.8725. The model also improves segmentation of thin scratch patterns and maintains strong performance on unseen defect combinations excluded from training. These results show that conditional mask generation can externalize process-relevant spatial evidence from complex WBMs rather than only output black-box labels. From an engineering informatics perspective, WCGN provides an interpretable representation of overlapping defect structures that can support engineer-in-the-loop diagnosis, process monitoring, and yield management in semiconductor manufacturing.

Keywords: Wafer Bin Map, Mixed-Type Defect Patterns, Conditional Semantic Generation, Feature-wise Linear Modulation, Overlapping Defect Disentanglement

1 Introduction

Semiconductor manufacturing is a complex process comprising hundreds of steps. Following wafer fabrication, circuit probe testing is performed on each die to spatially display the pass/fail results. Therefore, the wafer map, which contains the test results for each individual die, serves as a critical source of data [1]. In the semiconductor manufacturing process, defect classification plays a crucial role by providing information to identify and rectify process-related issues. Consequently, the importance of automatic defect classification for semiconductor chips has been emphasized [2]. From this perspective, wafer maps have become a critical component for recognizing problems in the semiconductor manufacturing process. Accordingly, the analysis of wafer map data has assumed a pivotal role in process optimization and yield enhancement [3]. Research in wafer map analysis utilizes the wafer bin map (WBM) dataset, which consists of two-dimensional images where pixels are labeled as normal, defect, and background [4]. These WBMs enable the scientific identification of process factors that cause failures, thereby contributing to fundamental process improvements and yield enhancement. They have been shown to be critical data for defect detection in the semiconductor manufacturing process [5]. WBM defects are commonly categorized into random defects, systematic defects, and mixed-type defects. Random defects are scattered without a clear spatial structure and are often associated with environmental or stochastic disturbances in the manufacturing line. Systematic defects, in contrast, exhibit recognizable geometric patterns and are more directly linked to specific process issues within the fab [6]. Mixed-Type Defects are particularly important because they contain multiple defect mechanisms within a single wafer map, often preserving process-relevant characteristics of their constituent defect types. As semiconductor fabrication and integrated circuit technologies continue to advance, this problem has become more challenging: multiple defect types may co-occur on the same wafer, and their spatial regions may partially overlap. In such cases, one defect pattern can conceal or distort another, making reliable interpretation substantially more difficult [7].

As illustrated in Fig. 1, WBMs exhibit not only non-overlapping defect patterns but also patterns where single defects overlap, resulting in a concealed appearance, with these overlapping regions highlighted by a red gradient in the figure. Ultimately, this phenomenon leads to the occlusion of single defects. Additionally, the regions outlined in blue represent the WBM defect patterns, while the green text indicates the specific defect type corresponding to each outlined area. As observed in the segmentation visualization results of the WBMs presented in Fig. 1, representative segmentation methods define the overlapping regions of single defects as the single-type defect deemed most probable based on the training WBM data. In practice, Yan et al. [8] observed that among single-type WBM defects, the scratch pattern is thinner and more easily compromised compared to other defect types. Therefore, to maintain the integrity of the scratch, they employed a method that prioritizes the segmentation and preservation of scratch defects when overlaps occur. However, as shown in the segmentation results of the proposed Wafer Conditional Generative Network (WCGN) in Fig. 1, unlike the representative segmentation methods, our model makes it possible to identify the specific constituent defect types within overlapping

defects. This demonstrates that the proposed approach resolves the phenomenon of defect patterns becoming concealed due to overlapping—a limitation inherent in conventional methods—thereby achieving significantly more precise segmentation than previously possible.

Consequently, the identification and segmentation of mixed-type WBM defects are significantly more challenging compared to defects in conventional wafer manufacturing processes, and the difficulties and limitations associated with research in this area continue to be highlighted to this day. Unlike single-type defects, mixed-type defects often involve overlapping patterns, making them significantly more complex and difficult to detect [9]. From the perspective of yield enhancement, employing an interpretable classification model is essential to identify the root causes of manufacturing process issues and to develop effective solutions [10]. Kim et al. [11] explicitly detailed the appearance patterns and underlying causes associated with various WBM defect types. This suggests that root cause diagnosis for mixed-type WBMs is achievable based on the specific manifestation of each constituent defect. However, Taha [12] noted that most process engineers heavily rely on their personal expertise and experience to correlate specific characteristics of WBM defects with corresponding manufacturing process stages. Consequently, the root cause diagnosis of defects is highly dependent on the individual proficiency of the engineer, and the manual nature of this approach renders the decision-making process inefficient. This makes defect interpretation labor-intensive and inconsistent, especially when multiple defect mechanisms are superimposed. Therefore, automated WBM analysis should ideally provide not only accurate recognition performance but also interpretable spatial evidence that supports engineering decision-making.

Existing research has made substantial progress in WBM analysis, but important limitations remain. Image-level classification models [7, 13, 14] have improved mixed-type defect recognition under class imbalance, geometric variation, and limited-data conditions. However, because these methods primarily output class labels, they do not explicitly localize constituent defect regions. Segmentation-based approaches [8, 15–19] improve the interpretability of defect patterns by providing pixel-level localization. However, for the overlapped regions highlighted in red in the Original Image of the WBM in Fig. 1, these approaches apply a mutually exclusive classification scheme per pixel within a single prediction pipeline, as demonstrated by the Representative Segmentation Methods in Fig. 1. Consequently, depending on the model's architecture or training, only one of the overlapping defects is ultimately selected. This inherent limitation exists within the approach itself. As a result, a gap persists between automated defect recognition and process-relevant evidence generation: current methods do not sufficiently disentangle overlapping constituent defects in a form that engineers can directly inspect during diagnosis.

To address this gap, we propose WCGN, a conditional generation framework that integrates Feature-wise Linear Modulation (FiLM) [20] into a U-Net [21] architecture. The core novelty of WCGN is that it reformulates mixed-type WBM analysis as a label-conditioned mask generation problem rather than a conventional one-pass classification or segmentation problem. Given the same mixed-type wafer map and a queried

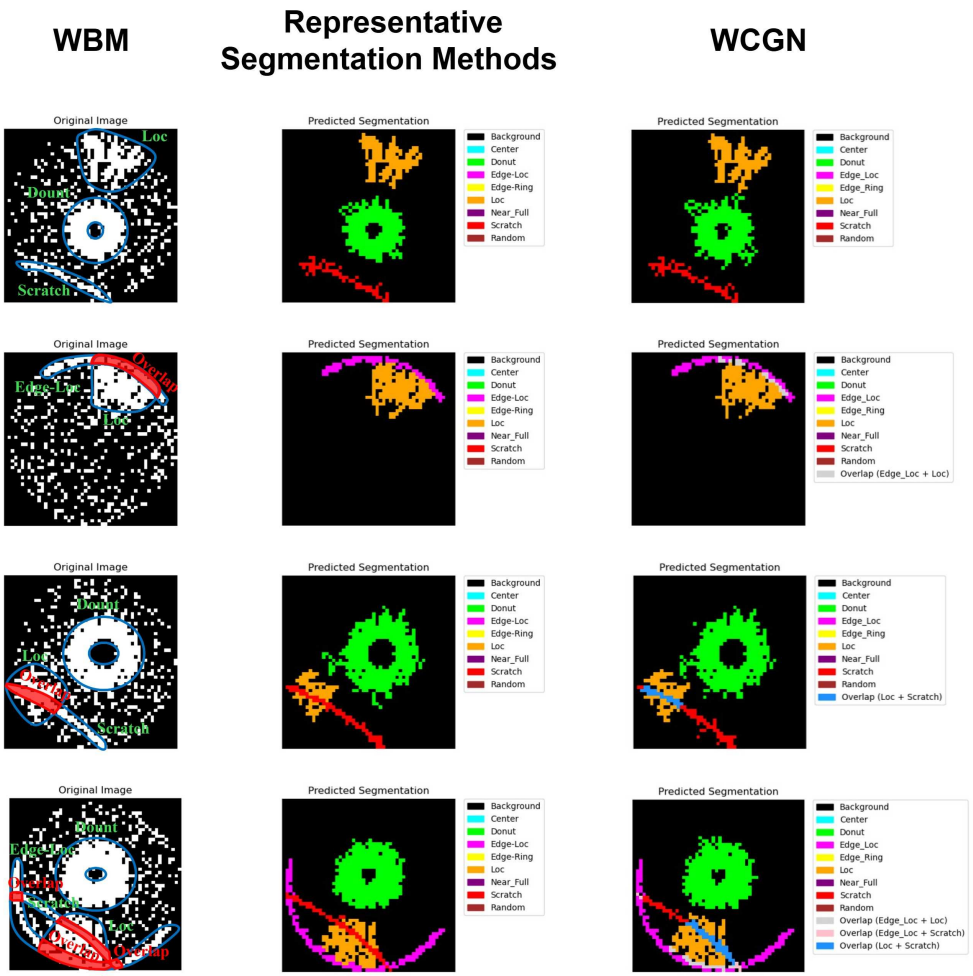


Fig. 1: Segmentation Differences between Representative Methods and WCGN

single-defect label, WCGN generates the corresponding defect-specific mask. By iteratively querying different defect labels, the model can explicitly separate constituent defects, including those concealed in overlapping regions. The final multi-label prediction is determined by the presence or absence of defect pixels within the defect-specific masks generated through these queries. Reconstructed from these generated masks, it becomes possible to identify all defect labels within the overlapped regions, as demonstrated by WCGN in Fig. 1. In this way, the proposed framework not only predicts defect labels but also resolves the limitations of prior studies by deriving and externalizing inspectable pixel-level evidence regarding defect patterns and labels for process diagnosis and engineering decision-making. Experimental results on the MixedWM38 dataset and a generated benchmark comprising pixel-level masks show that WCGN

improves both classification and segmentation performance, including in overlapped regions and unseen mixed-type settings. These results demonstrate that by providing interpretable pixel-level representations of complex defect combinations, WCGN enables automated process monitoring and diagnosis for the complex and novel defect combinations emerging in advanced wafer manufacturing processes. Consequently, it facilitates engineering decision-making and root cause analysis—challenges that were not fully resolved in previous research. Furthermore, this implies that the perspective of reformulating the segmentation problem into a label-conditioned mask generation problem inherently improves the model's performance.

The main contributions of this study are summarized as follows:

1. We reformulate mixed-type WBM analysis from direct classification/segmentation into a label-conditioned defect-mask generation problem. Through this, we propose a unified and interpretable model capable of performing both conditional mask generation and the classification of WBM patterns and labels, including overlapping defects.
2. As a significant contribution to engineering informatics, the proposed conditional generation framework provides actionable intelligence for process decision. By enabling the visual analysis and disentanglement of concealed single defects within overlapping patterns, it directly assists process engineers in diagnosing complex root causes arising from increasingly sophisticated semiconductor manufacturing processes.
3. By conditioning on single defect type labels to generate individual mask images even from overlapping regions, our model successfully identifies each constituent defect concealed by occlusion. This demonstrates validated capabilities in clear defect generation, localization, and visualization, moving beyond standard vision-based anomaly detection to provide interpretable evidence for root cause analysis.
4. The proposed model demonstrated superior classification performance not only on patterns included in the training data but also in experiments involving unseen mixed-type WBMs that were excluded during the training phase. These results suggest that this framework possesses significant practical potential and robustness for yield management and process optimization in modern semiconductor manufacturing, where unknown defect combinations frequently emerge due to the continuous advancement of IC and fabrication technologies.

The remainder of this paper is organized as follows: Section 2 introduces and reviews the research topic and related studies. Section 3 describes the proposed conditional generation framework and network architecture. Section 4 explains the experimental datasets, measurement metrics, and presents the results to validate the effectiveness of our model. Section 5 discusses the implications of the findings and the limitations of the study. Finally, Section 6 summarizes the conclusions and suggests directions for future research.

2 Related Works

In the early stages of research, semiconductor manufacturing technology was not as advanced; thus, defect analysis and classification could be adequately performed primarily through preprocessing techniques. Representative approaches included methods fusing the Radon transform with similarity ranking [22], decision tree ensemble [23], techniques utilizing density- and geometry-based clustering and feature extraction [24], and finally, advanced methods based on linear discriminant analysis, such as joint local and nonlocal linear discriminant analysis [25].

However, with the increasing sophistication of ICs and the growing complexity of semiconductor processes, the emergence of mixed-type defects has necessitated the development of more advanced AI classification models, including machine learning and deep learning. Currently, representative analyses of defects are broadly divided into two categories. The first is Defect Pattern Classification, which functions as an image-level classification. The second is Defect Pattern Segmentation and Segmentation-based Classification, which segments defect regions and identifies the defect type for each segmented area.

2.1 Defect Pattern Classification

To address the complexities of WBM defect analysis, various network architectures and strategies have been extensively investigated. Early approaches focused on classifying mixed-type defects using individual classifiers or pre-trained autoencoders to handle scalability and single-type defect data constraints [26, 27]. Subsequently, research evolved to handle geometric variations and multi-scale issues; for instance, Wang et al. [7] introduced deformable convolutional networks for irregular shapes, while Wei and Wang [13] combined convolutional neural networks (CNNs) with transformers to enhance feature extraction. To mitigate data limitations such as class imbalance and distribution mismatch, techniques involving synthetic data generation, similarity searching, and generative adversarial network-based augmentation have been employed [28–30]. Furthermore, research was conducted to address the inherent limitations of WBM datasets. Advanced frameworks utilizing self-supervised learning for unlabeled data, lightweight architectures for efficiency, and contrastive learning for interpretability have further pushed the boundaries of defect classification performance [31–34]. Recently, to recognize fine wafer defects, methods utilizing multi-scale feature fusion have been employed to combine the robust semantic information of high-level features with the high-resolution details of low-level features. This approach offers the advantage of enhancing recognition efficiency while avoiding ultra-high computational costs [14].

Collectively, these studies have successfully proposed models robust to the geometric variations and data imbalances inherent in the sophisticated semiconductor manufacturing process. They have demonstrated effectiveness in enhancing detection accuracy for multi-scale and mixed-type defects through methods like data augmentation and advanced neural architectures. However, a critical limitation persists across these reviewed studies: they primarily rely on image-level classification. This approach fails to provide spatial location information for defects, hindering precise root cause

analysis. Consequently, the inability to localize specific defect pixels or interpret the spatial combination of defect types on the WBM remains a significant drawback, highlighting the need for models capable of pixel-level interpretation.

2.2 Defect Pattern Segmentation and Segmentation-based Classification

Research within the domain of defect pattern recognition and segmentation has evolved to address the need for precise localization and unit separation. Early efforts by Kim et al. [9] focused on clustering techniques to separate patterns, while subsequent deep learning approaches aimed at simultaneous classification and segmentation gained prominence. Chiu and Chen [15] as well as Cha and Jeong [17] explored instance segmentation using Mask R-CNN and attention-enhanced U-Nets, respectively, though challenges with overlapping pixels remained. Nag et al. [16] proposed WaferSegClassNet, a lightweight network that simultaneously performs classification and segmentation of mixed-type defects via a shared encoder architecture. By maximizing the representational power of the latent space through pre-training with N-pair contrastive loss, the model achieved high efficiency, enabling real-time processing with significantly fewer parameters compared to existing models. Yan et al. [8] presented a U-Net-based framework that applies semantic segmentation techniques to decompose and identify complex mixed-type defect patterns at the single-defect level. Through connected region analysis, the method demonstrated superior generalizability, effectively handling unseen mixed-type defect patterns. Yin et al. [18] proposed a self-supervised contrastive learning framework that integrates global and local representations for the classification and segmentation of mixed-type WBM defect patterns. By jointly learning pixel-level details alongside image-level context, the model enables the precise identification of complex mixed-type defect patterns even in environments with limited labeled data. Additionally, for micro-defect analysis on wafers, Qiao et al. [19] utilize a ResNet encoder for effective feature extraction, while employing VGG blocks for the decoder. Furthermore, they introduce an attention module between the encoder and decoder to emphasize key features associated with defect regions. Collectively, these studies primarily leverage Autoencoder architectures, such as U-Net, modifying encoder-decoder structures with various modules like convolution, ResNet, and VGG blocks to achieve clear pixel-level segmentation.

These segmentation-based classification models offer a significant advantage over conventional image-level classification by providing superior interpretability regarding pixel-level details and the localization of defect combinations. By segmenting regions at the pixel-level, they allow for a more granular analysis of defect types. However, a critical limitation persists in the accurate segmentation and classification of pixel regions where mixed-type defects overlap within mixed-type patterns. This inability to effectively disentangle overlapping defects is a primary cause of suboptimal performance in the current landscape, particularly given the increasing complexity of semiconductor manufacturing which frequently produces such mixed flaws. Consequently, further research is essential to overcome these limitations and advance towards models that preserve high interpretability even in the presence of complex,

overlapping defect combinations. Therefore, this study proposes a conditional generation framework that generates a mask for a specific defect by taking its label as a conditional input. When comparing the proposed model with the representative prior studies presented in Table 1, WCGN can decompose the mixed-type defects based on the conditioned labels, enabling the reconstruction of individual defect patterns. This allows for the implementation of a highly interpretable model for WBM defect analysis. Furthermore, it overcomes the existing limitation by generating the appropriate mask image precisely according to the conditioned label, even in pixel regions where mixed-type defects overlap.

2.3 Data Augmentation and Generative Modeling

To alleviate the limitations of WBM datasets, recent studies have actively introduced data augmentation and generative-model-based strategies to improve defect classification performance. In real wafer defect inspection settings, the major problems are data scarcity, insufficient labeling, class imbalance, and data dimensional inconsistency. To address these issues, one study utilized generation and augmentation through an autoencoder together with adversarial domain adaptation [35].

Afterward, Generative Adversarial Networks (GANs) and Diffusion Models were mainly employed. First, GAN is a new generative modeling framework that can generate sophisticated data only through backpropagation, without complex indirect probabilistic inference or a Markov chain [36]. Byun and Baek [37] proposed a deep convolutional generative adversarial network for wafer map synthesis, which generates mixed-type patterns by combining thresholding techniques and pixel-wise summation for labeling, because existing deep-learning-based studies require a large amount of labeling data. They stated that this model could contribute to solving the problem of insufficient labeled data in single- and mixed-type defect pattern classification models. Aung et al. [38] found that because mixed-type WBMs have more complex pattern structures than single-type WBMs, the generator model must be searched multiple times so that GAN-based data generation does not produce images biased toward specific defects. Nevertheless, they showed that GANs can generate diverse defects even in challenging environments with limited data. Yu et al. [39] stated that generative augmentation techniques using GANs suffer from training instability and image blurring, making effective data augmentation difficult. Therefore, they reported that performing data augmentation by adding random Gaussian noise to a Convolutional Autoencoder could achieve higher classification performance while reducing the number of parameters and computational cost. Li et al. [30] proposed a methodology that performs stable data augmentation using a Jacobian regularized generative adversarial network, followed by final defect classification using a ResNet-18-based classification model. Through this, they improved the generation quality of minority-class wafer patterns by addressing the training instability and low diversity of generated images that conventional GAN-based models exhibit. However, they also pointed out disadvantages in high-dimensional large-scale training datasets due to increased computational cost and complexity, and they noted that additional research is needed for the Scratch pattern in particular because its morphology is discontinuous and fragmented.

Another approach, the Diffusion Model, is mainly used for denoising. Denoising Diffusion Probabilistic Models [40] introduced diffusion probabilistic models and showed that they are a type of latent variable model and are effective for high-quality image synthesis. Yang et al. [41] showed that Diffusion Models have emerged as a powerful new family of deep generative models that achieve record-breaking performance in various application fields such as image synthesis, video generation, and molecular design, and that they are being broadly applied from computer vision, natural language generation, and time-series data modeling to interdisciplinary applications in other scientific fields. Long et al. [42] stated that as wafer manufacturing processes become more sophisticated, mixed-type WBMs become increasingly complex and vulnerable to noise interference. To address this, the study proposed an end-to-end model based on a denoising diffusion model. However, it also had limitations in that the computational complexity was very large and it could not detect unidentified defect patterns. Li et al. [43] proposed a denoising diffusion probabilistic model equipped with an auxiliary classifier to improve the number of imbalanced WBMs. In that study, the trained model gradually removes predicted noise from initial random noise to generate new wafer maps. Through this, the classification problem caused by sample imbalance was addressed by generating new samples with the same defect characteristics from random noise, but there was also a limitation in that the data generation process required a very large amount of time. Moon and Kim [34] pointed out that existing WBM defect pattern classification studies cannot respond to new defect patterns that were absent from the training data and therefore misclassify them. In addition, there is the problem that it takes a great deal of time and cost for expert engineers in real semiconductor sites to manually label large amounts of data. That study proposed an Input-guidance Diffusion Model and showed robustness in detection performance, including for subtle defects and complex mixed-type defect patterns that were difficult for existing models to detect. However, it also reported that relatively high training cost and inference time are limitations.

In summary, studies based on Autoencoders, GANs, and Diffusion Models have evolved as auxiliary strategies. They have shown that they can model complex defect distributions and improve robustness to patterns that are not sufficiently included in the training data. Although they can augment the training data or improve recognition performance, they do not explicitly determine which constituent defects compose a mixed-type WBM or how those defects overlap spatially. Therefore, it is appropriate to understand this line of research as a practical strategy for data preprocessing and augmentation in WBM analysis. In other words, although it is effective in alleviating data imbalance and distribution sparsity, it still has limitations in interpreting the internal constituent defects of mixed-type WBMs at the pixel level and in providing the explainability needed for process root-cause analysis.

3 Methodology

Semantic segmentation aims to assign a categorical label to every pixel in an image. It plays a crucial role in terms of understanding image data and within various systems [44]. Such image segmentation techniques have been utilized across a wide range of

Table 1: Comparison of Representative Related Works

Category	Author	Key Features	Pros	Cons
Defect Pattern Classification	Wang et al. (2020) [7]	Deformable convolutional networks. Capable of transforming sampling locations using 2D offset matrices.	Accurately identifies defect regions in noisy WBMs.	Low precision for thin patterns such as scratch defects.
Defect Pattern Recognition Segmentation-based Classification	Cha and Jeong (2022) [17]	U-Net composed of ResNet Blocks and CBAM. Focuses on target defect patterns utilizing Residual Attention Blocks.	High capability in focusing on defect pixels and suppressing noise.	Frequently misidentifies overlapping defect pixels during classification.
Defect Pattern Classification	Wei and Wang (2022) [13]	CNN and Transformer. Utilizes local perception of CNNs and global perception of transformers.	Superior performance in recognizing edge-located and mixed-type defects.	Simple classification model; precise interpretation is not possible.
Defect Pattern Recognition Segmentation-based Classification	Yan et al. (2023) [8]	U-Net composed of convolution blocks. Utilizes pixel connectivity and U-Net-based semantic segmentation.	Decomposes mixed-type defects into single defects via segmentation.	Unable to visualize individual defect patterns within mixed-type WBMs.
Defect Pattern Classification	Zhao et al. (2024) [14]	Feature pyramid network and convolution block attention module. Utilizes multi-scale feature fusion and attention mechanisms.	Enhances focus on core WBM information compared to existing methods.	Unable to identify spatial location information for defects in mixed-type WBMs.
Data Augmentation and Generative Modeling	Li et al. (2025) [30]	ResNet-18 based classification model following data augmentation using a Jacobi regularized Generative Adversarial Network.	Resolves the training instability and low diversity of generated images observed in conventional GAN models, and improves the generation quality of minority-class wafer patterns.	Increases computational cost and complexity.
Data Augmentation and Generative Modeling	Moon and Kim et al. (2025) [34]	Input-guidance Diffusion Model utilizing reconstruction-based Unsupervised Out-of-Distribution detection that can be trained without labels.	Demonstrates robustness even against subtle defect patterns that are difficult to detect and complex mixed-type patterns.	High training costs and long inference times.
Defect Pattern Recognition Segmentation-based Classification	Qiao et al. (2025) [19]	U-Net composed of ResNet Blocks and VGG Blocks. Focuses on spatial information using FRelu and residual connections.	Resolves data imbalance through a hybrid loss function.	Lacks defect pixel label classification capabilities due to its focus on defect shape analysis.

fields. In the domain of general imagery, models such as fully convolutional networks [45], Mask R-CNN [46], and the segment anything model [47] are prominent. Image segmentation models are also actively employed in diverse specialized fields, including U-Net [21] in biomedical imaging and SegNet [48] in autonomous driving. Among these, U-Net has been modified and evolved in various ways. Representative examples include 3D U-Net [49], V-Net [50], Attention U-Net [51], R2U-Net [52], UNet++ [53], TransUNet [54], EfficientNet [55], and UNet– [56]. Furthermore, there are numerous representative studies [8, 15–18, 57] that have utilized U-Net-related models specifically for WBM datasets. This demonstrates that the U-Net architecture is not only widely adopted for image segmentation across various disciplines but is also extensively utilized for analyzing WBM datasets. Consequently, this study selected the U-Net architecture as the backbone for performing image segmentation. However, to perform complex visual reasoning required for data such as mixed-type WBMs, an additional methodology is necessary. To address this, we applied FiLM [20], a general conditioning method that modulates the computation process of the neural network according to the WBM label, thereby enabling the model to handle such complex tasks. Specifically, the FiLM layers condition the network on single-type defect labels derived as pixel-level labels during the data preprocessing stage—to segment a mixed-type WBM image into distinct WBM images corresponding to each specific single-type defect label. The framework is designed to determine the final label of the mixed-type WBM based on these segmented WBM images.

3.1 Framework

In this study, as illustrated in Fig. 2, we propose a method that generates semantic WBMs via a conditional generation approach utilizing WBMs with pixel-level labels, and subsequently classifies defect patterns based on these generated outputs. Initially, we extract single-type WBMs from the MixedWM38 dataset and generate pixel-level labels using pixel connectivity to compensate for the lack of inherent segmentation data. These are superimposed to create a benchmark mixed-type dataset with pixel-level labels. Using this synthesized dataset, our model is trained to segment images according to a FiLM layer conditioned on single-type defect labels. The trained model is then applied to the actual MixedWM38 dataset, where it assigns final one-hot encoded labels based on the detection of defect pixels in the segmented results, allowing for both accurate classification and visual verification.

3.2 Conditional Generative Network Model

In this study, we propose a conditional generation network to effectively separate and classify complex mixed-type defect patterns within WBMs. The core of the proposed methodology is a U-Net architecture integrated with FiLM, which selectively activates specific defect regions by receiving a single defect label as a condition. The overall framework is illustrated in Fig. 3.

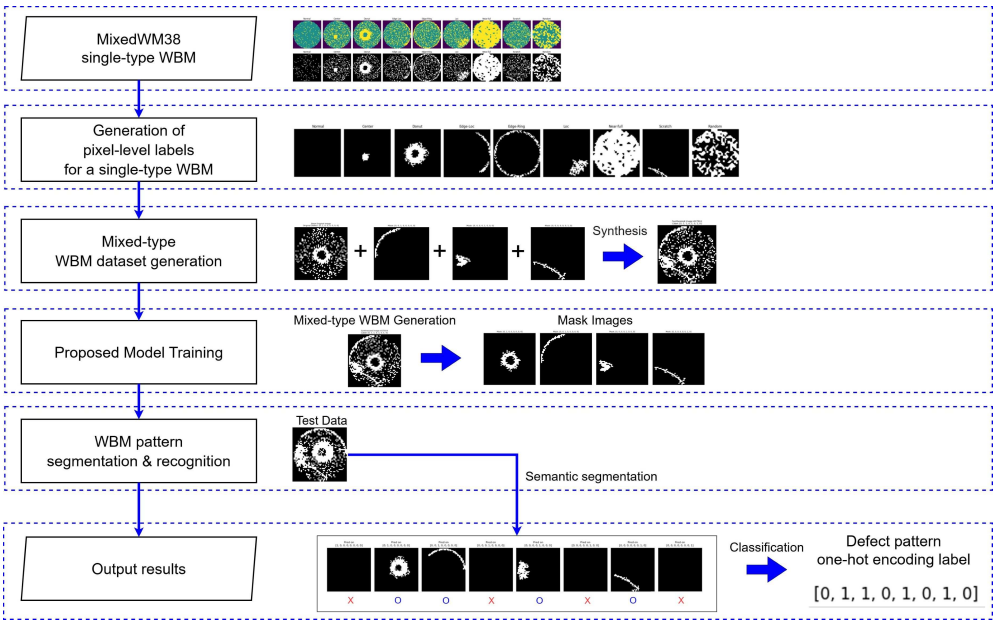


Fig. 2: Proposed Framework

3.3 Conditional Generation U-Net Architecture

The WCGN model, denoted as F takes an input wafer image $X \in \mathbb{R}^{H\times W\times 1}$ and a target defect label $l \in \{0,1\}^C$ as inputs, and outputs a binary segmentation mask $\hat{M} \in \mathbb{R}^{H\times W\times 1}$ for the corresponding defect:

$$\hat{M} = F(X, l; \theta) \tag{1}$$

where θ denotes the trainable parameters. The model consists of a contracting path for extracting features from the input image and an expansive path for restoring the original resolution. Notably, to address dimension mismatch issues arising from the input image size ($H \times W$), we applied a dynamic padding strategy. The detailed process of dimensional transformation is as follows.

3.3.1 Contracting Path

The encoder consists of four downsampling blocks. The size of the input image X is $H \times W \times 1$. In the first block, convolution is performed using 64 filters of size 3×3, resulting in a feature map of size $H \times W \times 64$. Subsequently, the resolution is reduced to $\frac{H}{2} \times \frac{W}{2} \times 64$ via 2×2 max pooling. Downsampling proceeds in the same manner while doubling the number of channels. After the second block, a feature map of $\frac{H}{4} \times \frac{W}{4} \times 128$ is generated. In the third block, this is further downsampled to $\frac{H}{8} \times \frac{H}{8} \times 256$ feature map. Finally, after passing through the fourth block, latent features of size $\frac{H}{16} \times \frac{H}{16} \times 1024$ are obtained.

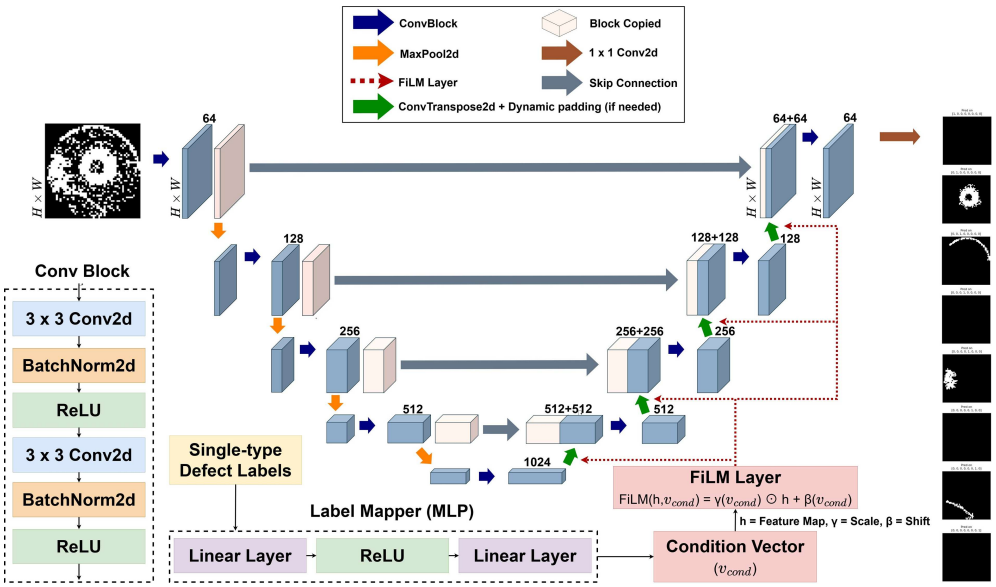


Fig. 3: Architecture of the Proposed WCGN. The network extracts spatial features from a mixed-type WBM input, while the label mapper generates a condition vector (v_{cond}) from target single-type defect labels. FiLM layers dynamically modulate the decoder's feature maps using this condition vector to generate the corresponding single-defect mask.

3.3.2 Expansive Path with Dynamic Padding

The decoder performs up-sampling in the reverse order of the encoder to restore the spatial resolution. The deepest feature map of size $\frac{H}{16} \times \frac{W}{16} \times 1024$ is first expanded to $\frac{H}{8} \times \frac{W}{8} \times 512$ via a 2×2 transposed convolution. A critical challenge arises during the subsequent up-sampling stages where dimension mismatches may occur due to the fractional loss from previous floor-processing in the encoder. When the up-sampled decoder features do not perfectly match the corresponding encoder feature maps retrieved for skip connections, the framework dynamically computes the spatial difference $(\Delta H, \Delta W)$. To resolve this dimension mismatch, this study introduces a dynamic padding mechanism that appends the required pixels to the right and bottom boundaries of the decoder feature map, successfully reconstructing it to match the encoder's dimensions. Following the concatenation of these aligned maps, the resolution is sequentially restored to $\frac{H}{4} \times \frac{W}{4}$, $\frac{H}{2} \times \frac{W}{2}$ and eventually to the original size of $H \times W$. Finally, a binary mask of size $H \times W \times 1$ is output via a 1×1 convolution.

3.4 Feature-wise Linear Modulation Layer (FiLM)

Since a standard U-Net attempts to predict all defect classes simultaneously, ambiguity arises between pixels in regions where defects overlap. To overcome this, we introduced the FiLM mechanism into each stage of the decoder. The FiLM layer takes

the target defect label as a condition and linearly modulates each channel of the feature map, thereby selectively emphasizing only the features related to that defect.

3.4.1 Condition Vector Generation via Label Mapper

To perform conditional modulation, the target defect label (one-hot vector) l must first be projected into a continuous latent space. For this purpose, we designed a label mapper composed of a two-layer multi-layer perceptron (MLP). The condition vector v_{cond} is calculated as follows:

$$v_{\text{cond}} = (W_2 \cdot \text{ReLU}(W_1 \cdot l + b_1)) + b_2 \tag{2}$$

where W_1, W_2 are weight matrices and b_1, b_2 are biases. The generated condition vector v_{cond} contains the unique semantic information of the defect and is subsequently used as input for the FiLM generator.

3.4.2 Feature Modulation Mechanism

The FiLM layer infers channel-wise scale parameters γ and shift parameters β from the condition vector v_{cond}. For an intermediate feature map $h \in \mathbb{R}^{H \times W \times Ch}$ of the decoder, the FiLM operation is defined as a channel-wise affine transformation:

$$\text{FiLM}(h, v_{\text{cond}}) = \gamma(v_{\text{cond}}) \odot h + \beta(v_{\text{cond}}) \tag{3}$$

Here, $\odot$ denotes element-wise multiplication, and γ and β are broadcasted across the spatial dimensions ($H \times W$) of the feature map. This process guides the model to amplify ($\gamma > 1$) activation patterns related to the target defect (l) across the entire input image while suppressing ($\gamma < 1$) irrelevant patterns or noise. Through this, the model can effectively disentangle the complex mixed-type defect image X into a single defect mask M_{pred}.

3.5 Hybrid Loss Function

In WBM segmentation tasks, defect regions occupy a very small proportion of the total wafer area. To address this class imbalance problem and secure precise segmentation performance for small defect patterns, we propose a hybrid loss function combining focal loss [58] and dice loss [50]. The final objective function $\mathcal{L}_{\text{total}}$ is defined as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Focal}} + \mathcal{L}_{\text{Dice}} \tag{4}$$

3.5.1 Focal Loss

Focal Loss is designed to down-weight easy examples and focus on hard examples. In a binary classification problem, letting $\hat{m}$ be the model's predicted probability for a pixel and m be the ground truth pixel value, Focal Loss is defined as:

$$\mathcal{L}_{\text{Focal}}(\hat{m}, m) = -\alpha_{\text{focal}}(1 - \hat{m}_t)^{\gamma_{\text{focal}}} \log(\hat{m}_t) \tag{5}$$

where $\hat{m}_t$ is $\hat{m}$ when $m = 1$ and $(1 - \hat{m})$ when $m = 0$. Focal Loss necessitates the tuning of two hyperparameters: the balancing factor α and the focusing parameter

γ. This increases learning efficiency for sparse defect pixels even in situations where background pixels are overwhelmingly dominant.

3.5.2 Dice Loss

Dice Loss is based on the Dice Coefficient, which measures the overlap between the predicted mask and the actual mask, optimizing the similarity of the region itself.

$$\mathcal{L}_{\text{Dice}}(\hat{M}, M) = 1 - \frac{2\sum_i(\hat{m}_i m_i) + \epsilon}{\sum_i \hat{m}_i + \sum_i m_i + \epsilon} \tag{6}$$

Here, $\hat{M}$ and M represent the predicted mask and the actual ground truth mask, respectively, and $\hat{m}_i$ and m_i denote their corresponding i-th pixel values. ϵ is a smoothing constant (10^{-6}). Since dice loss directly maximizes the agreement between regions regardless of defect size, it is effective in preserving boundary details.

3.6 Inference Algorithms

Algorithm 1 outlines the procedure for generating a mask for a specific defect during the inference phase. The model extracts features via the encoder and then dynamically modulates the feature maps of the decoder according to the condition vector to generate a binary mask. Specifically, this section details the procedure for generating a defect-specific mask during the inference phase, taking the input image X and the target label l as inputs. This process comprises three main stages: feature encoding, condition embedding, and conditional decoding. First, in the feature encoding phase, the encoder $\mathcal{E}$ extracts multi-scale feature maps h_{enc} from the input image X. This process follows the contracting path described in Section 3.3.1, capturing features ranging from low-level boundary details to high-level morphological information. Simultaneously, in the condition embedding phase, the target defect label l is passed through the label mapper Φ to be transformed into a latent condition vector v_{cond}. As defined in Eq. 2, this process projects the semantic information of the defect into a continuous vector space. The core of the algorithm lies in the iterative processing of the conditional decoding phase. At each stage k, the decoder $\mathcal{D}$ upsamples the feature map and concatenates it with the corresponding feature map h_{enc}^k from the encoder. To address subtle dimensional mismatches between the encoder and decoder, the dynamic padding operation is applied beforehand, ensuring precise pixel-level alignment. The concatenated feature maps undergo convolution and are subsequently modulated by the FiLM mechanism. The scale (γ_k) and shift (β_k) parameters derived from the condition vector v_{cond} are applied to the feature maps according to Eq. 3, enhancing activation patterns relevant to the target defect while suppressing irrelevant features. Finally, the modulated highest-level feature map is transformed into the final binary segmentation mask $\hat{M}$ via a 1×1 convolution and a sigmoid activation function. Through this procedure, the model can precisely isolate only the specific defect regions corresponding to the given condition, even within complex mixed-type defect patterns.

Since WCGN functions as a conditional generation model rather than a direct classifier, it requires a one-hot label vector for a specific single defect to generate a

Algorithm 1: Conditional Mask Generation Procedure

Input: Mixed-Type WBM Image X, Target Defect Label l
Output: Generated Final Binary Segmentation Mask $\hat{M}$

Parameters: Encoder $\mathcal{E}$, Decoder $\mathcal{D}$, Label Mapper Φ
Feature Encoding:
$h_{enc} \leftarrow \mathcal{E}(X)$ // Extract multi-scale feature maps (Sec. 3.3.1)
Condition Embedding:
$v_{\text{cond}} \leftarrow \Phi(l)$ // Generate condition vector using Eq. (2)
Conditional Decoding:
Initialize h_{dec} with the bottleneck features.
for each decoder block k **do**
  Compute Modulation Parameters: $\gamma_k, \beta_k \leftarrow \text{mathrm}Proj_k(v_{\text{cond}})$
  Upsample Features: $h_{\text{up}} \leftarrow \text{ConvTranspose}(h_{\text{dec}})$
  Align Dimensions: $h_{\text{up}} \leftarrow \text{DynamicPad}(h_{\text{up}}, h_{\text{enc}}^k)$
  Concatenate: $h_{\text{cat}} \leftarrow \text{Concat}(h_{\text{up}}, h_{\text{enc}}^k)$
  Apply Convolution: $h_{\text{conv}} \leftarrow \text{ConvBlock}(h_{\text{cat}})$
  Apply FiLM: $h_{\text{dec}} \leftarrow \gamma_k \odot h_{\text{conv}} + \beta_k$ // Use Eq. (3)
end
Final Prediction:
$\hat{M} \leftarrow \text{Sigmoid}(\text{Conv}_{1\times1}(h_{\text{dec}}))$
return $\hat{M}$

mask. To perform multi-label classification on test images where ground truth labels are unknown, we devised an iterative inference strategy, as detailed in Algorithm 2. The procedure iterates through the set of all pre-defined defect classes $S_{\text{class}} = \{c_1, \ldots, c_C\}$, where C denotes the number of single-type defect classes. For each class c_i, the model constructs a corresponding one-hot query vector $l^{(i)} \in \{0,1\}^C$ and feeds it into the network alongside the input image X. The network then generates a conditional predicted mask $\hat{M}^{(i)}$, and its activated pixels are counted according to the indicator function $\mathbb{I}$. To determine the presence of the defect c_i, we calculate the number of activated pixels N_{pixels} within $\hat{M}^{(i)}$. If N_{pixels} exceeds a predefined threshold τ (set to 0 in this study), the class is marked as present ($\hat{y}_i \leftarrow 1$). By aggregating these results across all classes, the model explicitly reconstructs the final multi-hot classification vector $\hat{y}$. This approach allows WCGN to utilize its precise pixel-level segmentation capabilities to achieve robust image-level classification.

4 Experiment

4.1 Experimental Dataset and Settings

In this study, we utilized the MixedWM38 dataset [7]. As this dataset is derived from actual semiconductor wafer manufacturing processes, it serves as a widely used public benchmark in related research. As shown in Fig. 4, the MixedWM38 dataset comprises one normal type, 8 types of single-type defects, 13 types of 2 mixed-type defects, 12 types of 3 mixed-type defects, and 4 types of 4 mixed-type defects. Specifically, the mixed-type defect patterns consist of combinations of single-type defects, excluding C7 (Near-Full) and C9 (Random). The original MixedWM38 dataset comprises a total of 38,015 WBMs. As detailed in Fig. 4, the distribution varies for certain classes: C7 contains 149 samples, C9 includes 866 samples, and C24 (C+EL+S) consists of 2,000 samples, while all remaining categories contain exactly 1,000 samples each. To ensure

Algorithm 2: Inference Strategy for Classification

Input: Mixed-Type WBM Image X, Set of Defect Classes $S_{\text{class}} = \{c_1, \ldots, c_C\}$, Threshold τ (e.g., 0)

Output: Predicted Multi-Hot Label Vector $\hat{y} \in \{0,1\}^C$

Initialize prediction vector: $\hat{y} \leftarrow 0$

Iterate through all defined single defect types:

for $i \leftarrow 1$ **to** C **do**

  1. Construct Target Label Vector for Class c_i:

  Let $l^{(i)} \in \{0,1\}^C$ be the one-hot vector for class c_i.

  2. Conditional Generation (Algorithm 1):

  $\hat{M}^{(i)} \leftarrow F(X, l^{(i)})$

  3. Pixel-level Activation Check:

  $N_{\text{pixels}} \leftarrow \sum_{x,y} \mathbb{I}(\hat{M}^{(i)}(x,y) > 0.5)$

  4. Classification Decision:

  if $N_{\text{pixels}} > \tau$ **then**

    $\hat{y}_i \leftarrow 1$ // Defect c_i is present

  end

end

return $\hat{\mathbf{y}}$

data balance among single-type WBMs while preserving the inherent characteristics of actual defect types, we augmented the C7 and C9 classes to 1,000 samples using rotation and flipping techniques. Each sample consists of a visual image of the WBM and its corresponding label. The images have a resolution of 52×52 pixels and consist of three discrete pixel values: 0 represents a blank spot, 1 represents a normal die that passed the electrical test, and 2 represents a broken die that failed the electrical test. The label of WBM X is expressed as $y \in \{0,1\}^C$, a multi-hot encoded vector, where C denotes the number of single-type defect classes. Since the primary objective of this study is to segment and classify defective dies, we performed a binarization process where pixels with values 0 (blank spot) and 1 (normal die) were converted to 0, and pixels with value 2 (broken die) were converted to 1. This transformation was designed to enable the model to focus exclusively on the broken dies within the WBM.

All models proposed in this study were implemented in Python 3.10.18 and Pytorch 2.5.1. The experiments were conducted on a workstation equipped with a 13$^{\text{th}}$ Gen Intel® Core™ i7-13700K CPU and an NVIDIA GeForce RTX 4080 GPU to accelerate model training and inference. WCGN was trained using the Adam [59] optimizer with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. We employed a cosine annealing learning rate scheduler [60] ($\eta_{min} = 1 \times 10^{-6}$) to adjust the learning rate dynamically over 200 epochs. The batch size was set to 64. For the objective function, a hybrid loss combining focal Loss ($\alpha = 0.25, \gamma = 1.5$) and dice Loss was utilized to address the class imbalance problem inherent in wafer defect datasets. For the model training dataset, we utilized single-type defect WBMs from the MixedWM38 dataset and mixed-type defect WBMs generated using the method described in Sec. 4.2. This training dataset comprises 1,000 samples for each defect type. To ensure consistent data partitioning, the training dataset was split into training and validation sets at a ratio of 8:2. To guarantee a clear performance comparison of the models developed in this study, the test datasets were organized as follows: For the evaluation of classification performance, we exclusively employed the mixed-type defect WBMs from the

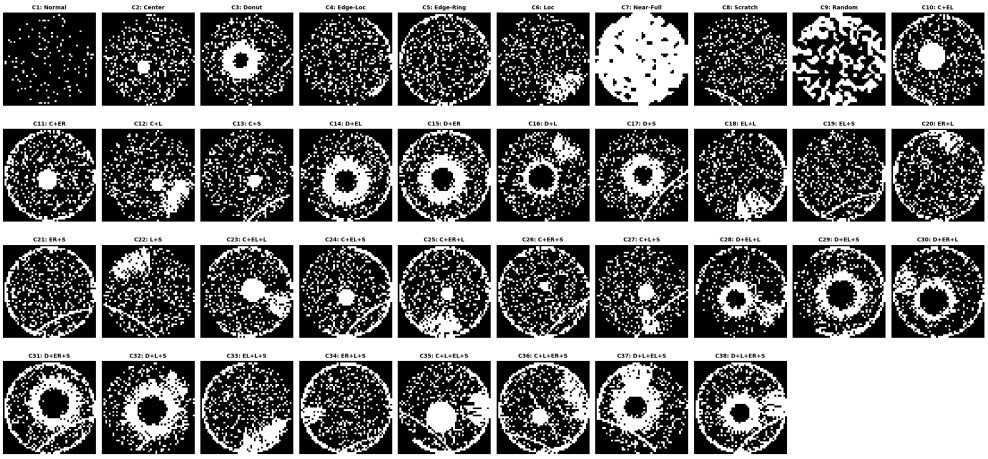


Fig. 4: Overview Map of the MixedWM38 Dataset

original MixedWM38 dataset, which represents defect patterns occurring in real-world semiconductor manufacturing processes. Conversely, for the evaluation of segmentation performance, we utilized the mixed-type defect WBMs generated via the method proposed in Sec. 4.2.

4.2 Mixed-Type WBM Data Generation

4.2.1 Generation of Pixel-Level Labels for Single-Type WBM

The data generation process is based on single-type WBMs from the MixedWM38 dataset, retaining the original resolution of 52×52. Fig. 5 illustrates the process of generating mixed-type WBM data employed in this study. The original WBMs in MixedWM38 consist of pixels labeled as 0 (blank spot), 1 (normal die), and 2 (defect die). As this study focuses primarily on the defect dies (pixels label 2), a binarization process was conducted wherein pixels labeled 0 and 1 were set to 0, and pixels labeled 2 were set to 1. This design enables the WCGN to concentrate more intensively on defect pixels. Subsequently, preprocessing was performed to derive WBMs corresponding to pixel-level labels, serving as mask images. For this purpose, the pixel connectivity method [61] was utilized. Pixel connectivity is broadly categorized into two types: 4-connected component and 8-connected component. 4-connected component considers pixels to be connected if they share an edge with one of the four adjacent pixels (up, down, left, right), treating them as part of the same label. 8-connected component considers pixels connected if they are adjacent in any of the eight directions, including the four diagonals. This concept of pixel connectivity introduces the notion of Simple-Connectedness in digital images and has been proven valid [61]. Indeed, prior WBM research has utilized this method to derive WBM images corresponding to single-type pixel-level labels [8]. The filtering strategy utilizing pixel connectivity employed in this study to generate mask images is detailed in Table 2.

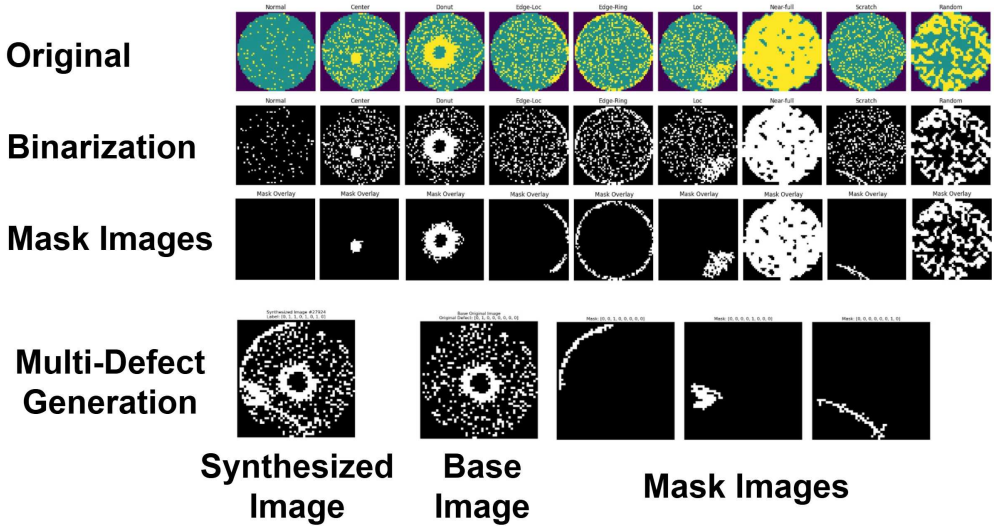


Fig. 5: Mixed-Type WBM Data Generation Method

4.2.2 Mixed-Type WBM Generation Method

The generation of mixed-type WBM data employs a Base-anchored synthesis approach to maintain realistic background noise distributions while securing pixel-level labels. Let $S_{\text{mix}} = \{c_1, c_2, \ldots, c_k\}$ be the set of defect types to be mixed. First, one defect type $c_{\text{base}} \in S_{\text{mix}}$ is randomly selected as the anchor. A real single-type WBM image corresponding to c_{base} is chosen as the background canvas, denoted as I_{base}. This ensures that the synthesized image retains the inherent noise characteristics of a real wafer map. For the remaining defect types $c_i \in S_{\text{mix}} \setminus \{c_{\text{base}}\}$, we randomly select a processed mask M_{c_i} from the pre-generated library of pixel-level labels for single-type WBMs. The final mixed WBM, I_{mixed}, is generated using a pixel-wise maximum operation:

$$I_{\text{mixed}} = \max\left(I_{\text{base}}, \bigcup_{i \neq \text{base}} M_{c_i}\right) \tag{7}$$

This operation effectively simulates the superposition of defects where a defect pixel (value 1) overrides a background pixel (value 0). Fig. 5 illustrates the overall procedure of this generation method. Crucially, the mask images corresponding to each constituent single defect—including the mask of the base image—are preserved during this process. These masks serve as the ground truth labels for separating individual defect patterns during the training of the proposed conditional generation model. Consequently, this generation approach not only preserves distinct defect patterns and random noise but also establishes a precise supervision signal for overlapping defect disentanglement.

Using this methodology, we generated 2,000 mixed-type WBMs for each mixed-type category present in the MixedWM38 dataset. These generated samples were

Table 2: Single-Type Defect WBMs Filtering Strategy and Rationale

Defect Type	Connectivity	Rationale	Filtering Strategy
Center	4-connected component	Maintain cohesiveness of dense, blob-like shapes.	Preserve only the largest connected component.
Donut	8-connected component	Maintain seamless continuity of curved ring shapes.	Preserve only the largest connected component.
Loc	4-connected component	Preserve the density (compactness) of local clusters.	Preserve only the largest connected component.
Scratch	8-connected component	Prevent fragmentation (breaking) of thin diagonal patterns.	Preserve only the largest connected component.
Edge-Ring	-	Restrict distribution to the wafer edge area only.	Apply ring mask (Outer $R = 26$, Width=4px).
Edge-Loc	8-connected component	Remove edge local patterns and noise.	Ring mask + Area threshold (≥ 10).
Random	All	Maintain noise characteristics distributed across the wafer.	No filtering.
Near-Full	All	Maintain high-density distribution characteristics across the wafer.	No filtering.

subsequently partitioned into two subsets: 1,000 samples were allocated for the training and validation set, while the remaining 1,000 samples were designated as the test set. This test set was specifically utilized to evaluate the segmentation performance of mixed-type WBMs in the experiments described in the subsequent sections.

4.3 Measurement

This study employs performance metrics tailored for multi-label classification tasks [62]. Specifically, the exact match ratio (EMR) imposes the strictest criterion, counting a prediction as correct only if the predicted label set is identical to the ground truth label set; partial matches are treated as failures. In contrast, metrics such as precision and recall are utilized to account for partially correct predictions.

Let $\mathrm{Pat}_{\mathrm{set}} = \{\mathrm{pat}_1, \ldots, \mathrm{pat}_m\}$ be the set of unique label patterns in the dataset. For a specific pattern pat_k, let S_k denote the set of sample indices belonging to this

pattern, and $N_k = |S_k|$ be its support. We evaluate pattern-wise performance based on element-wise matches within S_k. Let TP_k, FP_k, and FN_k denote the total count of true positives, false positives, and false negatives for pattern pat_k, respectively. These are calculated by aggregating bit-wise comparisons across all samples in S_k.

The pattern-wise accuracy, precision, recall, and F1-score are defined as:

$$\mathrm{Precision}_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FP}_k} \tag{8}$$

$$\mathrm{Recall}_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FN}_k} \tag{9}$$

$$\text{F1-score}_k = 2 \cdot \frac{\mathrm{Precision}_k \cdot \mathrm{Recall}_k}{\mathrm{Precision}_k + \mathrm{Recall}_k} \tag{10}$$

Additionally, the exact match ratio (EMR) imposes a strictly perfect match criterion:

$$\mathrm{EMR}_k = \frac{1}{N_k} \sum_{n \in S_k} \mathbb{I}(y_n = \hat{y}_n) \tag{11}$$

where y_n and $\hat{y}_n$ are the ground truth label and predicted label vectors for the n-th sample, and $\mathbb{I}(\cdot)$ is the indicator function.

To quantify the overall error rate at the label level, we employ Hamming Loss (HL). Unlike EMR, which penalizes the entire instance for a single mismatch, HL evaluates the fraction of incorrectly predicted labels across all samples. For a pattern group S_k, the Hamming Loss is defined as:

$$\mathrm{HL} = \frac{1}{N_k \cdot C} \sum_{n \in S_k} \sum_{j=1}^{C} \mathbb{I}(y_{n,j} \neq \hat{y}_{n,j}) \tag{12}$$

where C represents the number of single defect classes, and $\mathbb{I}(y_{n,j} \neq \hat{y}_{n,j})$ indicates a mismatch between the ground truth and predicted label for the j-th class of the n-th sample. A lower Hamming Loss indicates better performance.

To aggregate these metrics, we use macro and weighted averages. Weighted averages were employed to address the uneven distribution in the MixedWM38 dataset, specifically the presence of 2,000 samples in the C24 class versus 1,000 samples in others, thereby ensuring that the evaluation metrics properly consider the sample count of each class. Let Met_k represent any metric ($\mathrm{Precision}_k, \ldots, \mathrm{EMR}_k$) for pattern pat_k.

$$\text{Macro Avg} = \frac{1}{m} \sum_{k=1}^{m} \mathrm{Met}_k, \quad \text{Weighted Avg} = \frac{\sum_{k=1}^{m} (\mathrm{Met}_k \cdot N_k)}{\sum_{k=1}^{m} N_k} \tag{13}$$

For segmentation performance, we utilize the Mean Intersection over Union (mIoU) [63]. For each class c_i, let $\hat{M}^{(i)}$ and $M^{(i)}$ denote the predicted and ground truth masks. The Intersection over Union (IoU) for class c_i is defined as:

$$\mathrm{IoU}_{c_i} = \frac{|\hat{M}^{(i)} \cap M^{(i)}| + \epsilon}{|\hat{M}^{(i)} \cup M^{(i)}| + \epsilon} \tag{14}$$

Table 3: Experiments on Focal Loss for WCGN

Hyperparameters		Macro Avg		Weighted Avg		HL	mIoU
$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$	EMR	F1-Score	EMR	F1-Score		
0.50	1.5	0.9722	0.9937	0.9723	0.9937	0.0044	0.8551
0.50	2.0	0.9755	0.9946	0.9754	0.9945	0.0039	0.8641
0.25	2.0	0.9762	0.9941	0.9763	0.9941	0.0040	0.8689
0.25	**1.5**	**0.9786**	**0.9948**	**0.9785**	**0.9948**	**0.0036**	**0.8725**

where ϵ is a smoothing factor to prevent division by zero. The mIoU is computed by averaging IoU scores across valid classes:

$$\text{mIoU} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{C} \mathbb{I}(S_{c_i} > 0) \cdot \text{IoU}_{c_i} \tag{15}$$

where S_{c_i} is the support for class c_i, and N_{test} is the number of classes present in the test set. Here, $\mathbb{I}(\cdot)$ is the indicator function, outputting 1 if the defect class c_i is present in the test set ($S_{c_i} > 0$) and 0 otherwise.

4.4 Results

4.4.1 Experimental Results for Focal Loss Tuning

WCGN employs focal loss as its objective function. Lin et al. [58] previously investigated performance variations based on these parameters, identifying $\alpha = 0.25$ and $\gamma = 2.0$ as the optimal settings. Guided by these findings, we evaluated our WCGN using combinations of $\alpha \in \{0.25, 0.5\}$ and $\gamma \in \{1.5, 2.0\}$. The resulting performance metrics—EMR, F1-score, and mIoU—are summarized in Table 3. Among the tested combinations, $\alpha = 0.25$ and $\gamma = 1.5$ demonstrated superior performance, achieving a weighted average EMR of 0.9786, a weighted average F1-score of 0.9948, and an mIoU of 0.8725. The HL also exhibited its lowest value, demonstrating that this specific hyperparameter configuration yielded the fewest misclassifications in the prediction of WBM defect types. Consequently, we adopted this configuration ($\alpha = 0.25, \gamma = 1.5$) for the Focal Loss in the final implementation of WCGN.

4.4.2 Classification Results of Single and Mixed-Type WBMs on the MixedWM38 Dataset

This section presents the classification results obtained by applying WCGN to mixed-type WBMs within the MixedWM38 dataset. Among the comparative models, those [8, 17, 19] discussed in Section 2.2 retained their original architectures as presented in their respective studies, while incorporating the "scratch priority selection" condition. This condition, proposed by Yan et al. [8], aims to preserve the integrity of scratch defects, which are notoriously difficult to segment and classify due to their thin morphology. As shown in Table 4, which compares the proposed model against these baselines under the specified condition, our method achieved EMR scores of 97.86%

and 97.85%, demonstrating superior capability in accurately identifying WBM labels. Furthermore, the model achieved a Precision of 99.62%, indicating exceptional classification performance relative to prior studies. In comparison to the baseline models, the proposed method exhibited an average improvement of 0.8% in EMR for 2 mixed and 3 mixed-type defects. Notably, for 4 mixed-type defect patterns—the most complex defect type in WBMs—it showed an improvement of approximately 0.4%. These results confirm that WCGN outperforms existing state-of-the-art methods in classifying mixed-type WBM defects. As shown in Table 5, the results demonstrate that WCGN achieves the highest performance in terms of both macro and weighted averages for the EMR—which indicates the accuracy of WBM label classification—and the F1-score, which comprehensively evaluates precision and recall. This signifies that WCGN exhibits superior classification capabilities compared to both major categories of WBM analysis approaches: those focused solely on classification and those utilizing segmentation-based classification.

In this study, single-type WBMs from the MixedWM38 dataset were utilized both for the data generation process detailed in Section 4.2 and as training data for the model. Consequently, single-type WBMs were excluded from the comparative analysis presented in Table 4. To address this, we conducted a separate classification performance experiment utilizing exclusively the single-type WBMs from the MixedWM38 dataset. This experiment focused solely on the single-type categories (C1–C9) depicted in Fig. 4, each consisting of 1,000 samples. The test dataset was constructed by selecting 200 samples for each single-type category, while the remaining data were split into training and validation sets at an 8:2 ratio. To appropriately evaluate the single-type datasets used in this experiment, standard confusion matrix-based metrics were utilized: accuracy, precision, recall, and f1-score [64]. The results are summarized in Table 6 and Table 7. These findings demonstrate that while the proposed model is optimized for multi-label tasks, its capabilities are not limited solely to multi-label classification. Specifically, the results indicate that even when trained exclusively on single-type WBMs, the model achieves high accuracy without misclassifying inputs as different single-type defects or erroneously predicting them as multi-label defects, outperforming the comparative models.

Table 4: Classification Results for Mixed-Type WBMs on MixedWM38 Dataset

Defect Pattern	EMR (%)				Precision (%)				Recall (%)			
	Improved U-Net	Semantic U-Net	RA-Unet	WCGN	Improved U-Net	Semantic U-Net	RA-Unet	WCGN	Improved U-Net	Semantic U-Net	RA-Unet	WCGN
C10: C+EL	96.90%	97.40%	98.30%	98.20%	98.71%	98.90%	99.35%	99.45%	99.10%	99.10%	99.30%	99.20%
C11: C+ER	98.60%	99.00%	98.80%	99.50%	99.30%	99.50%	99.40%	99.75%	100.00%	99.95%	99.80%	99.75%
C12: C+L	98.60%	97.00%	96.50%	98.80%	99.50%	98.71%	98.47%	99.65%	99.80%	99.80%	99.80%	99.75%
C13: C+S	98.30%	98.70%	99.50%	99.50%	99.16%	99.35%	99.75%	99.85%	99.95%	99.85%	100.00%	99.90%
C14: D+EL	96.50%	97.30%	97.80%	98.10%	98.46%	99.00%	99.25%	99.60%	99.30%	99.15%	99.40%	99.15%
C15: D+ER	97.90%	97.60%	98.00%	98.70%	98.91%	98.81%	99.01%	99.35%	99.90%	99.75%	99.85%	99.55%
C16: D+L	97.30%	95.30%	96.50%	97.80%	99.10%	98.13%	98.90%	99.50%	99.55%	99.45%	99.25%	99.40%
C17: D+S	98.80%	98.30%	98.70%	99.40%	99.60%	99.25%	99.55%	99.85%	99.80%	99.90%	99.80%	99.85%
C18: EL+L	97.80%	97.10%	96.60%	97.60%	99.40%	99.00%	98.90%	99.25%	99.40%	99.50%	99.35%	99.55%
C19: EL+S	97.40%	98.10%	98.30%	98.40%	98.81%	99.15%	99.25%	99.60%	99.25%	99.25%	99.45%	99.20%
C20: ER+L	96.80%	97.20%	95.00%	97.20%	98.76%	98.96%	98.31%	99.00%	99.55%	99.50%	98.75%	99.00%
C21: ER+S	98.90%	99.20%	99.40%	99.60%	99.55%	99.60%	99.75%	99.85%	99.85%	99.90%	99.90%	99.80%
C22: L+S	95.00%	96.90%	97.60%	98.50%	99.09%	99.25%	99.75%	100.00%	98.30%	99.15%	99.05%	99.25%
C23: C+EL+L	94.60%	94.20%	95.40%	96.50%	98.80%	98.48%	99.00%	99.36%	99.00%	99.17%	99.03%	99.03%
C24: C+EL+S	97.40%	97.30%	97.70%	97.60%	99.37%	99.33%	99.53%	99.57%	99.47%	99.37%	99.30%	99.20%
C25: C+ER+L	96.40%	97.30%	96.30%	98.30%	99.20%	99.37%	99.20%	99.63%	99.57%	99.70%	99.43%	99.57%
C26: C+ER+S	98.00%	99.20%	98.90%	99.30%	99.50%	99.77%	99.67%	99.80%	99.83%	99.93%	99.90%	99.77%
C27: C+L+S	95.90%	96.80%	97.50%	98.30%	99.63%	99.37%	99.77%	100.00%	98.97%	99.57%	99.37%	99.43%
C28: D+EL+L	96.20%	96.30%	96.40%	96.80%	99.17%	99.27%	99.40%	99.56%	99.37%	99.33%	99.17%	99.13%
C29: D+EL+S	95.40%	96.10%	96.80%	97.20%	98.80%	98.93%	99.17%	99.43%	98.80%	98.97%	99.03%	99.10%
C30: D+ER+L	89.50%	89.70%	95.20%	95.70%	96.81%	96.79%	98.87%	98.77%	99.20%	98.43%	99.17%	98.77%
C31: D+ER+S	97.30%	98.10%	98.40%	99.00%	99.24%	99.40%	99.57%	99.77%	99.77%	99.87%	99.77%	99.80%
C32: D+L+S	95.90%	96.50%	96.40%	97.90%	99.76%	99.47%	99.66%	100.00%	98.87%	99.37%	99.10%	99.30%
C33: EL+L+S	94.10%	95.90%	94.90%	96.00%	99.13%	99.36%	99.29%	99.50%	98.33%	98.77%	98.47%	98.67%
C34: ER+L+S	95.80%	96.80%	96.80%	97.70%	99.70%	99.80%	99.87%	99.90%	98.87%	99.07%	99.03%	99.23%
C35: C+L+EL+S	94.30%	95.40%	95.80%	95.90%	99.32%	99.25%	99.37%	99.57%	98.78%	99.08%	99.15%	98.95%
C36: C+L+ER+S	96.30%	97.50%	97.40%	98.10%	99.80%	99.82%	99.77%	99.85%	99.22%	99.48%	99.55%	99.58%
C37: D+L+EL+S	93.30%	95.60%	94.80%	95.40%	99.35%	99.52%	99.55%	99.72%	98.65%	99.10%	98.78%	98.80%
C38: D+L+ER+S	95.60%	96.20%	96.00%	96.90%	99.85%	99.67%	99.77%	99.82%	99.02%	99.32%	99.15%	99.20%
Macro Avg	96.37%	96.83%	97.09%	**97.86%**	99.16%	99.15%	99.35%	**99.62%**	99.29%	**99.41%**	99.35%	99.34%
Weighted Avg	96.41%	96.84%	97.11%	**97.85%**	99.17%	99.15%	99.35%	**99.62%**	99.30%	**99.41%**	99.35%	99.34%

Table 5: Expanded Classification Results for Mixed-Type WBMs on MixedWM38 Dataset

Method	Overlap Disentanglement	EMR	Hamming Loss	Macro F1-Score	Weighted F1-Score
DC-Net [7]	X	0.8903	0.0165	0.9765	0.9764
MSF-Trans [13]	X	0.9486	0.0080	0.9882	0.9885
Improved U-Net [17]	X	0.9641	0.0053	0.9923	0.9923
Semantic U-Net [8]	X	0.9684	0.0049	0.9927	0.9928
FPRN-CBAM [14]	X	0.9616	0.0060	0.9915	0.9914
RA-UNet [19]	X	0.9711	0.0044	0.9935	0.9935
WCGN	**O**	**0.9785**	**0.0036**	**0.9948**	**0.9948**

Table 6: Classification Results for Single-Type WBMs on MixedWM38 Dataset

Defect Pattern	Accuracy (%)				Precision (%)				Recall (%)			
	Improved U-Net	Semantic U-Net	RA-Unet	WCGN	Improved U-Net	Semantic U-Net	RA-Unet	WCGN	Improved U-Net	Semantic U-Net	RA-Unet	WCGN
C1: Normal	100.00%	100.00%	99.94%	100.00%	100.00%	100.00%	99.50%	100.00%	100.00%	100.00%	100.00%	100.00%
C2: Center	99.94%	99.89%	99.89%	99.94%	100.00%	100.00%	100.00%	100.00%	99.50%	99.00%	99.00%	99.50%
C3: Donut	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
C4: Edge-Loc	99.17%	99.22%	99.56%	99.44%	100.00%	100.00%	100.00%	100.00%	92.50%	93.00%	96.00%	95.00%
C5: Edge-Ring	99.39%	99.39%	99.61%	99.56%	97.01%	94.79%	96.62%	96.60%	97.50%	100.00%	100.00%	99.50%
C6: Loc	99.83%	99.78%	99.83%	99.94%	100.00%	100.00%	100.00%	100.00%	98.50%	98.00%	98.50%	99.50%
C7: Near-Full	99.83%	99.78%	99.78%	99.72%	99.00%	99.00%	98.51%	98.03%	99.50%	99.00%	99.50%	99.50%
C8: Scratch	99.94%	99.94%	99.83%	100.00%	100.00%	99.50%	100.00%	100.00%	99.50%	100.00%	98.50%	100.00%
C9: Random	99.56%	99.72%	99.67%	99.72%	100.00%	99.49%	100.00%	99.49%	96.00%	98.00%	97.00%	98.00%
Avg	99.74%	99.75%	99.79%	**99.81%**	**99.56%**	99.20%	99.40%	99.35%	98.11%	98.56%	98.72%	**99.00%**

Table 7: Expanded Classification Results for Single-Type WBMs on MixedWM38 Dataset

Method	Accuracy	Precision	Recall	F1-Score
DC-Net [7]	0.9923	0.9719	0.9622	0.9654
MSF-Trans [13]	0.9965	0.9846	0.9839	0.9841
Improved U-Net [17]	0.9974	**0.9956**	0.9811	0.9881
Semantic U-Net [8]	0.9975	0.9920	0.9856	0.9885
FPRN-CBAM [14]	0.9975	0.9893	0.9889	0.9889
RA-UNet [19]	0.9979	0.9940	0.9872	0.9905
WCGN	**0.9981**	0.9935	**0.9900**	**0.9916**

4.4.3 Semantic Segmentation Results of Mixed-Type WBMs on WBM Generation Method

The methodology employed to generate mixed-type WBMs from single-type defect WBMs was previously detailed in Section 4.2. Since the MixedWM38 dataset consists solely of WBM images and their corresponding one-hot encoded 8-dimensional label vectors, it inherently lacks the pixel-level annotations required for semantic segmentation. Consequently, we evaluated the semantic segmentation performance of WCGN using the mixed-type dataset created via the aforementioned generation method. The results, summarized in Table 8, demonstrate that WCGN achieves superior performance in terms of overall mIoU. Notably, we observed a significant improvement in the segmentation of scratch defects. Given that scratches are widely recognized as the most challenging defect type to detect and segment within WBM analysis, this performance enhancement represents a primary contribution of this study to the field of WBM semantic segmentation.

Table 8: Semantic Segmentation Results (IoU, mIoU) on Mixed-Type WBMs from the Proposed Generation Method

	Improved U-Net	Semantic U-Net	RA-UNet	WCGN
Center	**0.9507**	0.9496	0.9467	0.9505
Donut	0.8947	0.8960	**0.8966**	0.8930
Edge-Loc	0.8940	0.8877	0.8973	**0.9061**
Edge-Ring	0.9117	0.9085	0.9124	**0.9278**
Loc	0.8563	0.8549	**0.8606**	0.8473
Scratch	0.6970	0.7011	0.7080	**0.7104**
mIoU	0.8674	0.8663	0.8703	**0.8725**

A key contribution of WCGN is its ability to segment overlapping regions within mixed-type WBMs, effectively disentangling superimposed defect patterns. To validate this capability, we constructed a specialized dataset by extracting only those WBMs exhibiting overlapping defect pixels from the mixed-type WBM test set generated in Section 4.2. Subsequently, performance evaluations were conducted exclusively on this dataset, with the results summarized in Table 9. The results indicate that WCGN achieved the highest classification performance, demonstrating an EMR improvement of 0.2%. For mIoU, measurements were restricted to segmentation-based models. Unlike existing baseline models, which cannot identify the specific constituent defect types within an overlapping region, WCGN can explicitly determine which defect types are overlapping due to its conditional U-Net architecture. In the performance evaluation process detailed in this section, WCGN operates under more stringent conditions compared to existing baseline models. Since WCGN performs both segmentation and classification to identify the specific types of overlapping

defects, incorrect predictions regarding the constituent defects could lead to a significant degradation in performance metrics, a risk not present in models that do not attempt such disentanglement. Despite these challenges, WCGN demonstrated superior performance across all evaluation metrics: classification on the real-world MixedWM38 dataset, segmentation on mixed-type WBMs from the proposed generation method, and both classification and segmentation specifically on the subset of overlapping mixed-type WBMs. This superior performance confirms the model's ability to accurately and precisely perform segmentation and classification even for overlapping defects. Ultimately, these findings suggest that WCGN is capable of decomposing complex defect combinations resulting from advanced semiconductor manufacturing processes, enabling the precise segmentation and classification of constituent defect types, even in cases of overlap.

Table 9: Classification and Segmentation Results on Overlapping Mixed-Type WBMs from the Proposed Generation Method

Method	Overlap Disentanglement	EMR	Hamming Loss	Macro F1-Score	Weighted F1-Score	mIoU
DC-Net [7]	X	0.9546	0.0067	0.9900	0.9912	-
MSF-Trans [13]	X	0.9722	0.0040	0.9927	0.9946	-
Improved U-Net [17]	X	0.9742	0.0035	0.9936	0.9952	0.8513
Semantic U-Net [8]	X	0.9774	0.0033	0.9941	0.9954	0.8506
FPRN-CBAM [14]	X	0.9800	0.0030	0.9960	0.9961	-
RA-UNet [19]	X	0.9836	0.0023	0.9962	0.9969	0.8555
WCGN	**O**	**0.9859**	**0.0022**	**0.9972**	**0.9970**	**0.8602**

4.5 Ablation Study

A distinguishing feature of WCGN's inference mechanism is its ability to perform semantic segmentation of WBM defects tailored to specific single-type defects via the FiLM layer. According to Perez et al. [20], the primary advantages of the FiLM layer include robust performance and superior generalization capabilities, even when trained on limited data. Their research demonstrated consistent performance across various ablation studies, suggesting that the mechanism can be flexibly adapted to a wide range of model architectures. Furthermore, the layer enables effective learning with minimal data and exhibits exceptional generalization to complex or substantially different data distributions not encountered during training. This implies a high degree of adaptability in unseen mixed-type scenarios. Motivated by these findings, we conducted two specific ablation studies to evaluate these capabilities. In these experiments, we evaluated classification performance using the 2 mixed, 3 mixed, and 4 mixed-type WBMs from the MixedWM38 dataset as the test set. In the first ablation study, the training data consisted exclusively of single-type WBMs from the MixedWM38 dataset and the generated 2 mixed and 3 mixed-type WBMs; notably, 4 mixed-type WBMs were excluded from the training set. In the second ablation

study, the training set was further restricted to include only single-type WBMs and generated 2 mixed-type WBMs, thereby excluding both 3 mixed and 4 mixed-type WBMs. The results of the first ablation study revealed that WCGN achieved an EMR improvement of 0.9% and a performance gain of 0.2% compared to baseline methods. Ultimately, WCGN demonstrated superior performance, recording an EMR of 0.9762 and an F1-score of 0.9946. Similarly, in the second ablation study, WCGN maintained the highest performance among the compared methods. These findings confirm that the WBM inference mechanism facilitated by the FiLM layer in our architecture delivers excellent performance in the designed ablation studies, demonstrating robust classification capabilities even in unseen mixed-type scenarios.

Table 10: Classification Results for Ablation Study (Training Set: Single, 2 Mixed, and 3 Mixed-Types)

Method	EMR	Hamming Loss	Macro F1-Score	Weighted F1-Score
Improved U-Net [17]	0.9575	0.0063	0.9905	0.9906
Semantic U-Net [8]	0.9670	0.0050	0.9928	0.9928
RA-UNet [19]	0.9675	0.0047	0.9927	0.9928
WCGN	**0.9762**	**0.0036**	**0.9946**	**0.9945**

Table 11: Classification Results for Ablation Study (Training Set: Single and 2 Mixed-Types)

Method	EMR	Hamming Loss	Macro F1-Score	Weighted F1-Score
Improved U-Net [17]	0.9427	0.0082	0.9888	0.9889
Semantic U-Net [8]	0.9523	0.0070	0.9904	0.9905
RA-UNet [19]	0.9534	0.0069	0.9907	0.9907
WCGN	**0.9554**	**0.0066**	**0.9913**	**0.9914**

5 Discussion

This study shows that mixed-type WBM analysis can be improved by reformulating overlapping-defect interpretation as a label-conditioned mask generation problem rather than as a conventional multi-class recognition task. Across the reported experiments, WCGN achieved the best overall performance in classification on the MixedWM38 benchmark and in segmentation on the generated mixed-type dataset with pixel-level annotations. These results are important not only because the proposed model improved aggregate metrics such as EMR, F1-score, and mIoU, but also because they clarify how a conditional segmentation framework can address a core weakness of prior approaches: the inability to explicitly separate constituent defect patterns when multiple defect types overlap in the same wafer region. One of the

most notable findings is that WCGN produced clearer separation with less damage in overlapping defect regions and showed particularly strong segmentation and detection performance for scratch defects. This result can be interpreted from a structural perspective. The proposed model dynamically applies query-label-specific affine transformations, represented by the FiLM parameters (γ, β), to the intermediate feature maps. Accordingly, even when large and spatially dense defect pixels strongly dominate the image representation, the model can suppress such interfering responses once the scratch condition is queried, while selectively amplifying the fine linear activation patterns associated with scratch defects. In other words, the model does not treat all defect evidence uniformly. Instead, it adaptively reweights the feature space according to the queried defect identity, which is particularly beneficial for thin and low-area patterns that are easily concealed by larger overlapping defects. This provides a plausible explanation for the improved scratch performance observed in the segmentation experiments.

The same mechanism also helps explain the strong performance of WCGN in overlapping mixed-type cases. Existing models tend to learn mixed defects as holistic visual combinations, meaning that they implicitly memorize a closed set of composite patterns. Under such a learning regime, when an unseen combination appears, the model can lose a stable mapping boundary in latent space because the combined visual pattern no longer matches previously learned prototypes. In contrast, WCGN decomposes the problem into two coordinated but more generalizable steps: the label mapper projects multi-class relationships into a continuous latent space, and the FiLM module extracts class-specific modulation parameters independently for each queried defect. As a result, the model does not primarily memorize the entire geometry of a defect combination. Rather, it learns an independent mechanism for filtering out a target defect component from a complex image. This distinction is important for generalization. Even when a new defect combination is encountered, the model can still operate robustly if the target label condition is provided, because the task becomes conditional extraction of a constituent pattern rather than closed-set recognition of a predefined mixture. This interpretation is consistent with the ablation results. WCGN maintained the best performance even when some mixed-type combinations were excluded from training, suggesting that the proposed framework captures transferable class-wise representations rather than depending solely on direct exposure to every possible composite case. From an engineering informatics standpoint, this is a meaningful advantage because real semiconductor processes continuously generate new, irregular, and partially overlapping defect combinations. A model that depends on enumerating all mixture patterns in advance is difficult to sustain in practice, whereas a model that can isolate defect constituents conditionally is more compatible with evolving manufacturing environments.

The practical benefit of this framework is therefore not limited to higher benchmark accuracy. In real manufacturing scenarios, process engineers are not only interested in whether a wafer contains a mixed defect, but also in which constituent defect types are present, where they overlap, and which one is likely to be dominant in the failure mechanism. By explicitly generating label-conditioned masks, WCGN provides interpretable spatial evidence that can support faster root-cause analysis, failure

triage, and yield-improvement actions. For example, the ability to preserve the morphology of scratch defects while separating them from dense neighboring defects can help prevent diagnostically important but visually weak signals from being overlooked. Likewise, robust operation in unseen mixed-type settings can reduce the maintenance burden associated with frequent model retraining whenever new composite defect combinations emerge in advanced process nodes.

Despite these advantages, several limitations remain. First, the mixed-type segmentation benchmark was constructed synthetically because the original MixedWM38 dataset does not provide pixel-level ground truth for overlapping defects. Although the generation strategy was designed to preserve realistic connectivity characteristics, further validation on real annotated mixed-type defects is still necessary. Second, because the model conditions segmentation on label embeddings for all defect classes, the training and inference pipeline may be more computationally demanding than simpler discriminative baselines. Third, while the current results support the interpretability and robustness of the proposed framework, further analysis using industrial case studies would strengthen claims regarding decision-support value in live production settings. Future work should therefore focus on large-scale real-world validation, computational optimization, and the integration of additional manufacturing context beyond label information alone.

6 Conclusion

This study proposed WCGN as a conditional generation framework for mixed-type WBM analysis, with the specific goal of overcoming the limitations of conventional models in overlapping-defect scenarios. By integrating FiLM into a U-Net-based architecture, the proposed framework transformed mixed-defect analysis from a closed-form recognition problem into a label-conditioned mask generation problem. Experimental results showed that WCGN consistently outperformed strong baseline models in both classification and segmentation tasks, including the challenging subset of overlapping mixed-type defects. The contribution of the proposed method is not limited to metric improvement. The results suggest that WCGN is especially effective in two technically important situations: preserving clear segmentation with minimal interference in overlapping defect regions, and accurately detecting and segmenting fine scratch patterns that are easily obscured by stronger surrounding defect signals. A key structural reason for this behavior is that FiLM enables the model to dynamically modulate intermediate features according to the queried label. Consequently, even when dominant defect regions strongly disturb the representation, the model can suppress irrelevant responses and selectively amplify subtle linear activations corresponding to scratches or other target defect types. This mechanism provides a more defensible explanation for the observed performance gains than a purely empirical comparison alone.

Another major outcome of this study is the demonstrated robustness of WCGN in unseen mixed-type scenarios. Rather than memorizing the entire appearance of specific defect combinations, the proposed framework learns to extract constituent defect elements through class-wise conditional modulation. Because the label mapper organizes defect relations in a continuous latent space and FiLM produces independent

modulation factors for each queried class, the model remains effective even when new defect combinations appear that were not explicitly observed during training. This characteristic is particularly valuable in semiconductor manufacturing, where emerging process complexity can continuously generate novel and irregular mixed-failure patterns. From a practical perspective, WCGN offers a meaningful benefit for industrial diagnosis and yield management. The model produces interpretable pixel-level evidence for each queried defect class, allowing engineers to inspect not only whether a wafer is abnormal, but also how different defect patterns coexist, overlap, and contribute to the final failure signature. Such information can improve process monitoring, support more precise root-cause analysis, and reduce the risk of overlooking thin but critical defect patterns such as scratches. In this sense, the proposed framework functions not only as a vision model but also as a decision-support mechanism for manufacturing intelligence.

This study is still limited by its dependence on synthetically generated mixed-type masks and by the computational overhead associated with label-conditioned inference. Future research should validate the framework on real industrial datasets with pixel-level annotations for overlapping defects, examine lightweight implementations for scalable deployment, and extend the conditioning mechanism to incorporate richer process-related information. Nevertheless, the present findings indicate that WCGN provides a technically and practically meaningful direction for robust, interpretable, and generalizable mixed-type defect analysis in semiconductor manufacturing.

Despite these significant findings, this study is limited by its reliance on synthesized mixed-type datasets due to the scarcity of pixel-level ground truth data for real overlapping defects. Future research will focus on validating the framework using real-world industrial datasets containing complex mixed patterns and exploring scalability for real-time inference in large-scale manufacturing environments. Additionally, while the proposed WCGN utilized WBM label embeddings to achieve precise segmentation of overlapping defects, we believe it is crucial for future studies to enhance model extensibility by integrating not only label embeddings but also various other forms of WBM-related information.

Statements and Declarations

Funding

This work was supported by the Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (No. RS-2023-00244330) and the National Research Foundation of Korea (NRF), South Korea grant funded by the Korea government (MSIT) (No. NRF RS-2024-00352587)

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jaejoon Choi. The first draft of the

manuscript was written by Jaejoon Choi and Younghoon Kim, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

All data are fully available without restriction. The dataset used in this study is available from the following website:
https://www.kaggle.com/datasets/co1d7era/mixedtype-wafer-defect-datasets/data

Declarations

Conflict of interest/Competing interests The authors declare no competing interests.
Ethics approval Ethical approval is not required for this study.
Consent to participate Not applicable.
Consent for publication Manuscript is approved by all authors for publication.

References

[1] Chien, C.-F., Hsu, S.-C., Chen, Y.-J.: A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence. International Journal of Production Research **51**(8), 2324–2338 (2013) https://doi.org/10.1080/00207543.2012.737943

[2] Chou, P.B., Rao, A.R., Sturzenbecker, M.C., Wu, F.Y., Brecher, V.H.: Automatic defect classification for semiconductor manufacturing. Machine Vision and Applications **9**(4), 201–214 (1997) https://doi.org/10.1007/s001380050041

[3] Jeong, Y.-S., Kim, S.-J., Jeong, M.K.: Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping. IEEE Transactions on Semiconductor manufacturing **21**(4), 625–637 (2008) https://doi.org/10.1109/TSM.2008.2005375

[4] Yuan, T., Kuo, W., Bae, S.J.: Detection of spatial defect patterns generated in semiconductor fabrication processes. IEEE Transactions on Semiconductor Manufacturing **24**(3), 392–403 (2011) https://doi.org/10.1109/TSM.2011.2154870

[5] Taam, W., Hamada, M.: Detecting spatial effects ffrom factorial experiments: An application from integrated-circuit manufacturing. Technometrics **35**(2), 149–160 (1993) https://doi.org/10.1080/00401706.1993.10485037

[6] Hsu, S.-C., Chien, C.-F.: Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing. International Journal of Production Economics **107**(1), 88–103 (2007) https://doi.org/10.1016/j.ijpe.2006.05.015

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

[8] Yan, J., Sheng, Y., Piao, M.: Semantic segmentation-based wafer map mixed-type defect pattern recognition. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **42**(11), 4065–4074 (2023) https://doi.org/10.1109/TCAD.2023.3274958

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

[10] Joo, J., Kim, C.O.: Global feature identification layer for mixed-type wafer bin map classification. Expert Systems with Applications **271**, 126709 (2025) https://doi.org/10.1016/j.eswa.2025.126709

[11] Kim, H.K., Shim, J.: Generalized zero-shot learning for classifying unseen wafer map patterns. Engineering Applications of Artificial Intelligence **133**, 108476 (2024) https://doi.org/10.1016/j.engappai.2024.108476

[12] Taha, K.: CDID: A system for identifying the root cause of a defect in semiconductor wafer fabrication. IEEE Transactions on Semiconductor Manufacturing **31**(2), 221–231 (2018) https://doi.org/10.1109/TSM.2018.2808703

[13] Wei, Y., Wang, H.: Mixed-type wafer defect recognition with multi-scale information fusion transformer. IEEE Transactions on Semiconductor Manufacturing **35**(2), 341–352 (2022) https://doi.org/10.1109/TSM.2022.3156583

[14] Zhao, D., Lei, Z., Feng, R., Gu, S., Su, Y., Wen, G.: A feature pyramid residual network with attention mechanism for wafer map mixed defect detection. In: 2024 Global Reliability and Prognostics and Health Management Conference (PHM-Beijing), pp. 1–6 (2024). https://doi.org/10.1109/PHM-Beijing63284.2024.10874728

[15] Chiu, M.-C., Chen, T.-M.: 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 (2021) https://doi.org/10.1109/TSM.2021.3118922

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

[17] Cha, J., Jeong, J.: Improved U-Net with residual attention block for mixed-defect wafer maps. Applied Sciences **12**(4), 2209 (2022) https://doi.org/10.3390/app12042209

[18] Yin, S., Zhang, Y., Wang, R.: Contrastive learning with global and local representation for mixed-type wafer defect recognition. Sensors **25**(4), 1272 (2025) https://doi.org/10.3390/s25041272

[19] Qiao, Y., Chen, Y., Liu, F., Mei, Z., Luo, Y., Chen, Y., Liao, Y., Wu, B., Deng, Y.: RA-UNet: a new deep learning segmentation method for semiconductor wafer defect analysis on fine-grained scanning electron microscope (sem) images. IEEE Transactions on Semiconductor Manufacturing, 185–193 (2025) https://doi.org/10.1109/TSM.2025.3546296

[20] Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018). https://doi.org/10.1609/aaai.v32i1.11671

[21] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, pp. 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28

[22] Wu, M.-J., Jang, J.-S.R., Chen, J.-L.: Wafer map failure pattern recognition and similarity ranking for large-scale data sets. IEEE Transactions on Semiconductor Manufacturing **28**(1), 1–12 (2014) https://doi.org/10.1109/TSM.2014.2364237

[23] Piao, M., Jin, C.H., Lee, J.Y., Byun, J.-Y.: Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features. IEEE Transactions on Semiconductor Manufacturing **31**(2), 250–257 (2018) https://doi.org/10.1109/TSM.2018.2806931

[24] Fan, M., Wang, Q., Van Der Waal, B.: Wafer defect patterns recognition based on OPTICS and multi-label classification. In: 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), pp. 912–915. IEEE, Xian (2016). https://doi.org/10.1109/IMCEC.2016.7867343

[25] Yu, J., Lu, X.: Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis. IEEE Transactions on Semiconductor Manufacturing **29**(1), 33–43 (2015) https://doi.org/10.1109/TSM.2015.2497264

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

[27] Byun, Y., Baek, J.-G.: Mixed pattern recognition methodology on wafer maps with pre-trained convolutional neural networks. In: ICAART (2), pp. 974–979 (2020). https://doi.org/10.5220/0009177909740979

[28] Shim, J., Kang, S.: Learning from single-defect wafer maps to classify mixed-defect wafer maps. Expert Systems with Applications **233**, 120923 (2023) https://doi.org/10.1016/j.eswa.2023.120923

[29] Wang, R., Wang, S.: Similarity searching for fault diagnosis of defect patterns in wafer bin maps. Computers & Industrial Engineering **185**, 109679 (2023) https://doi.org/10.1016/j.cie.2023.109679

[30] Li, J., Tao, R., Li, S., Li, Y., Huang, X.: Sample-imbalanced wafer map defects classification based on Jacobian regularized generative adversarial network. Measurement Science and Technology **36**(3), 036112 (2025) https://doi.org/10.1088/1361-6501/adb327

[31] Wang, Y., Ni, D., Huang, Z., Chen, P.: A self-supervised learning framework based on masked autoencoder for complex wafer bin map classification. Expert Systems with Applications **249**, 123601 (2024) https://doi.org/10.1016/j.eswa.2024.123601

[32] Choi, J., Suh, D.: A depthwise convolutional neural network model based on active contour for multi-defect wafer map pattern classification. Engineering Applications of Artificial Intelligence **139**, 109707 (2025) https://doi.org/10.1016/j.engappai.2024.109707

[33] Hwang, J., Ha, S., Kim, D.: Wafer defect classification algorithm with label embedding using contrastive learning. IEEE Access **13**, 9708–9717 (2025) https://doi.org/10.1109/ACCESS.2025.3527491

[34] Moon, S., Kim, S.B.: Input-guidance diffusion model for unknown defect patterns detection in wafer bin map. Advanced Engineering Informatics **64**, 103078 (2025) https://doi.org/10.1016/j.aei.2024.103078

[35] Li, A.S., Bertino, E., Wu, R.-T., Wu, T.-Y.: Building manufacturing deep learning models with minimal and imbalanced training data using domain adaptation and data augmentation. In: 2023 IEEE International Conference on Industrial Technology (ICIT), pp. 1–8. IEEE, Orlando (2023)

[36] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020) https://doi.org/10.1145/3422622

[37] Byun, Y., Baek, J.-G.: Image synthesis with single-type patterns for mixed-type pattern recognition on wafer bin maps. In: 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 039–043.

IEEE, Jeju Island (2022). https://doi.org/10.1109/ICAIIC54071.2022.9722634

[38] Aung, A.P.P., Wang, X., Wang, R., Chan, H., An, B., Li, X., Senthilnath, J.: Double oracle neural architecture search for game theoretic deep learning models. IEEE Transactions on Image Processing **34**, 2463–2472 (2025) https://doi.org/10.1109/TIP.2025.3558420

[39] Yu, N., Chen, H., Xu, Q., Hasan, M.M., Sie, O.: Wafer map defect patterns classification based on a lightweight network and data augmentation. CAAI Transactions on Intelligence Technology **8**(3), 1029–1042 (2023) https://doi.org/10.1049/cit2.12126

[40] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020) https://doi.org/10.48550/arXiv.2006.11239

[41] Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.-H.: Diffusion Models: A Comprehensive Survey of Methods and Applications. ACM Computing Surveys **56**(4), 1–39 (2024) https://doi.org/10.1145/3626235

[42] Long, Z., Yan, J., Piao, M.: Multi-scale guidance diffusion network for wafer map defect recognition. Expert Systems with Applications **267**, 126134 (2025) https://doi.org/10.1016/j.eswa.2024.126134

[43] Li, J., Tao, R., Chen, R., Chen, Y., Zhao, C., Huang, X.: Sample-imbalanced wafer map defects classification based on auxiliary classifier denoising diffusion probability model. Computers & Industrial Engineering **192**, 110209 (2024)

[44] Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G.: Understanding convolution for semantic segmentation. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1451–1460 (2018). https://doi.org/10.1109/WACV.2018.00163

[45] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015). https://doi.org/10.1109/CVPR.2015.7298965

[46] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017). IEEE

[47] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollar, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026 (2023). https://doi.org/10.1109/ICCV51070.2023.00371

[48] Siam, M., Gamal, M., Abdel-Razek, M., Yogamani, S., Jagersand, M., Zhang, H.: A comparative study of real-time semantic segmentation for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 587–597 (2018). IEEE

[49] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016, vol. 9901, pp. 424–432 (2016). https://doi.org/10.1007/978-3-319-46723-8_49 . Springer

[50] Milletari, F., Navab, N., Ahmadi, S.-A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016). https://doi.org/10.1109/3DV.2016.79 . IEEE

[51] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018) https://doi.org/10.48550/arXiv.1804.03999

[52] Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018) https://doi.org/10.48550/arXiv.1802.06955

[53] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11 (2018). https://doi.org/10.1007/978-3-030-00889-5_1 . Springer

[54] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021) https://doi.org/10.48550/arXiv.2102.04306

[55] Wang, J., Li, X., Cheng, Y.: Towards an extended EfficientNet-based U-Net framework for joint optic disc and cup segmentation in the fundus image. Biomedical Signal Processing and Control **85**, 104906 (2023) https://doi.org/10.1016/j.bspc.2023.104906

[56] Yin, L., Tao, W., Zhao, D., Ito, T., Osa, K., Kato, M., Chen, T.-W.: UNet–: Memory-Efficient and Feature-Enhanced Network Architecture based on U-Net with Reduced Skip-Connections. In: Proceedings of the Asian Conference on Computer Vision, pp. 4085–4099 (2024)

[57] Sheng, H., Cheng, K., Jin, X., Jiang, X., Dong, C., Han, T.: An efficient deep learning framework for mixed-type wafer map defect pattern recognition. AIP

Advances **14**(4) (2024) https://doi.org/10.1063/5.0190985

[58] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)

[59] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2017) https://doi.org/10.48550/arXiv.1412.6980

[60] Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2017) https://doi.org/10.48550/arXiv.1608.03983

[61] Rosenfeld, A.: Connectivity in digital pictures. Journal of the ACM **17**(1), 146–160 (1970) https://doi.org/10.1145/321556.321570

[62] Sorower, M.S.: A literature survey on algorithms for multi-label learning. Oregon State University, Corvallis **18**(1), 25 (2010)

[63] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658–666 (2019)

[64] Tharwat, A.: Classification assessment methods. Applied Computing and Informatics **17**(1), 168–192 (2021) https://doi.org/10.1016/j.aci.2018.08.003