Pre-trained CNN-based TransUNet Model for Mixed-Type Defects in Wafer Maps

YOUNGJAE KIM[1,2], JEE-HYONG LEE[1*], JONGPIL JEONG[1*]
[1]Department of Computer Science and Engineering,
Sungkyunkwan University,
2066 Seobu-ro Jangan-gu, Suwon, 16419,
REPUBLIC OF KOREA

[2]Device Solutions Division,
Samsung Electronics,
1, Samsung-ro, Giheung-gu, Yongin-si, Gyeonggi-do 17113,
REPUBLIC OF KOREA

**Corresponding Authors*

Abstract: - Classifying the patterns of defects in semiconductors is critical to finding the root cause of production defects. Especially as the concentration density and design complexity of semiconductor wafers increase, so do the size and severity of defects. The increased likelihood of mixed defects makes finding them more complex than traditional wafer defect detection methods. Manually inspecting wafers for defects is costly, creating a need for automated, artificial intelligence (AI)-based computer vision approaches. Previous research on defect analysis has several limitations, including low accuracy. To analyze mixed-type defects, existing research requires a separate model to be trained for each defect type, which is not scalable. In this paper, we propose a model for segmenting mixed defects by applying a pre-trained CNN-based TransUNet using N-pair contrastive loss. The proposed method allows you to extract an enhanced feature by repressing extraneous features and concentrating attention on the defects you want to discover. We evaluated the model on the Mixed-WM38 dataset with 38,015 images. The results of our experiments indicate that the suggested model performs better than previous works with an accuracy of 0.995 and an F1-Score of 0.995.

Key-Words: - classification, mixed-type wafer maps, TransUNet, N-pair contrastive loss function, multitask learning, transformer layer

Received: July 8, 2022. Revised: May 29, 2023. Accepted: June 19, 2023. Published: July 18, 2023.

1 Introduction

The process of making semiconductor wafers is broadly categorized into front-end and back-end processes and is made through eight different processes, [1]. The front-end process is the process of designing and etching semiconductor chips onto wafers, while the back-end process is the process of cutting the chips from the wafers, wrapping them in insulation, and laying wires to reliably deliver power, [2], [3]. In specific, the front-end process, sometimes referred to as the wafer process, is the process of repeating the formation and cutting of different types of materials on the face of a wafer to create electronic circuits to make a single semiconductor chip, [4]. Previous processes include photolithography, which prints patterns of semiconductor circuits on wafers, etching to cut away parts other than the circuit pattern, deposition to create insulating thin films to separate and protect the metal from the circuit for electrical signal transmission, and metalization to create wiring, [5], [6]. With such a wide variety of processes, there are many different patterns of defects that can appear on a wafer. After wafer manufacturing, we run several tests to inspect defects and display them as wafer maps of binary numbers. The result of classifying the dies on the wafer map in this way shapes a particular pattern and is visually represented, [7]. The different patterns of defects in the wafer map are associated with the fabrication process. So, exactly categorizing the pattern of defects in a wafer map can help determine the source of defects in the manufacturing process. These classifications are important because they give engineers clues for troubleshooting, [8].

A semiconductor wafer is a circular plate made by growing a column of single crystals, such as silicon (Si) or gallium arsenide (GaAs), which are the core materials of semiconductor integrated circuits. A semiconductor integrated circuit is an electronic component that integrates many devices on a single chip to perform and organize various functions. This means that, since semiconductor integrated circuits build their circuits on thin circular plates or wafers, the wafer is the foundation of the semiconductor.

Wafer mapping is commonly used for data analysis of semiconductor manufacturing processes. Wafer mapping generates a color-coded map of semiconductor device performance on the surface of a wafer based on the test results of each chip failure. The created wafer maps have one or more patterns depending on the distribution of the defect chips. In semiconductor processes, different defect patterns occur because the defect chip pattern will look different depending on the source of the irregularity. Therefore, wafer maps defect analysis of defective chip patterns provides critical information for finding anomalies in semiconductor processes and determining the cause of defects. Fig. 1 shows a wafer map.

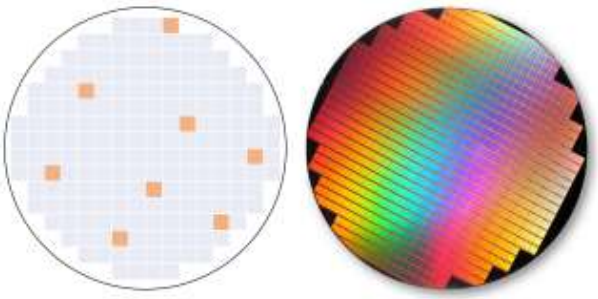

Fig. 1: Example of wafer map.

Modern developments in micronization techniques and increases in wafer size have increased the likelihood of creating more than two defect patterns, [9], [10]. Recent advances in Fig. 1. Example of wafer map. AI technology has spurred research focused on deep learning. The study, [11] suggested using CNN to classify mixed defect patterns. In [12], the authors suggested a deformable CNN. In [13], the authors suggested an infinite warped mixture model for clustering mixed-type defects. In [14], the authors applied augmentation techniques for the segmentation of mixed defect patterns and proposed a masked R-CNN. In [15], the authors proposed an Improved U-Net with a Residual Attention Block for mixed-defect wafer maps.

Taking classification one step further, segmentation is used as a way to detect defects. Classification simply categorizes the input image, but segmentation can provide inferences about the data on a pixel level. This can greatly help users make decisions by providing them with additional information. There is a lot of segmentation research going on, especially in the medical field, to find diseases. In [16], the authors applied a model that applies attention gates on U-Net to perform medical image segmentation. With attention gates, the model is trained to maintain target structures of different patterns and scales in focus automatically. A recent study suggested a cascaded neutralization dual attention U-Net for achieving enhanced tumor segmentation, [17]. In [18], the authors suggested a new U-Net architecture that uses aggregated residual blocks and a soft attention mechanism to segment COVID-19-infected areas. They proposed a cascading structure to scale low-resolution quality prediction and a dual-attention module to enhance the feature representation of tumor segmentation. In [19], the authors proposed TransUNet for medical image segmentation

Mixed defects are harder to recognize because as the defects are mixed, the pattern becomes more complex, and different defects overlap. So, we use a pre-trained CNN-based TransUNet to deliver classification and segmentation results targeted at defect areas to engineers. We specifically contribute as follows:

1. N-pair contrastive loss-based pre-training to generate good Feature maps that focus on defect details you want to find.
2. Applying multi-task learning to TransUNet for wafer defects.
3. Reduce unnecessary labor and time by creating pseudo-label data required to train segmentation models with automatic defect masking techniques.

The structure of the paper is as follows. Section 2 describes the background and related work. Section 3 explains in detail the architecture and characteristics of the proposed model. Section 4 explains the experimental setup and results. Finally, Section 5 sets out our conclusions and suggestions for future research.

2 Background & Related Work

2.1 TransUNet

Hybrid CNN-Transformer as Encoder. Instead of taking a pure transformer for an encoder, TransUNet uses a CNN transformer. It is a hybrid model that

initially uses a Convolutional Neural Network as a feature extractor to build a feature map for the input. Patch embedding is performed on a 1×1 patch extracted from the Convolutional Neural Network feature map instead of the raw image. Cascaded Upsampler (CUP). The decoder consists of several upsampling stages that decode hidden features for the resulting segmentation mask output. Together with the hybrid encoder, the CUP forms a U-shaped architecture, with skip-connections that enable feature aggregation at different resolution levels. Fig. 2 shows the architecture of TransUNet

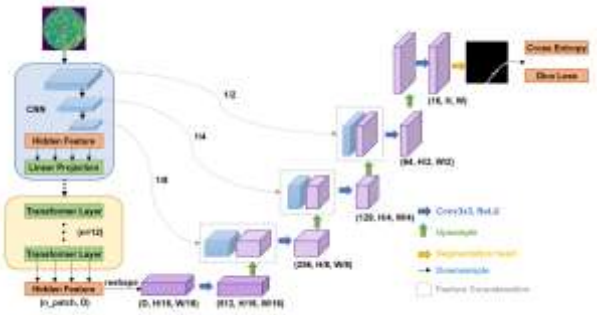

Fig. 2: Overall architecture of TransUNet.

2.2 Related Works on Defect Classification

In [11], the authors suggested using CNN to classify mixed defect patterns. They train distinct classification models to classify single defects and then use all models to find the occurrence of mixed-type defects. However, this method is not efficient, as it requires a significant increase in both the storage and computing power overhead, making it less scalable. In addition, separate training is required for every model. They also only study four single defect patterns: 'zone', 'circle', 'scratch', and 'ring'. In contrast, we test our model on 8 single-type and 13 mixed-type defect patterns. For defect classification, the study, [13], proposed a deformable convolutional network (DCNet). For extracting a feature representation of a defect, DCNet uses transform convolution (DC) to focus on a sampling region of the defective dies. The output layer has multi-labels and one-hot encoding. This converts mixed types of defects into separate, single defects so that you can effectively identify each defect.

3 Pre-trained CNN-based TransUNet

3.1 Pre-training TransUNet using N-pair Contrastive Loss

In this study, we use an N-pair Contrastive Loss Function to pre-train a CNN on the encoder of the TransUNet and then trained the entire TransUNet using cross-entropy loss and dice loss. The cross-entropy loss is the most commonly used loss function for supervised learning. But, we use N-pair contrastive loss in pre-train to ensure that features of the same class are closer than features of different classes. In our experiments, using this loss outperform supervised learning using cross-entropy loss. Encoder training based on N-pair contrast losses can provide a better representation of the latent dimension of wafer maps. It increases the accuracy of the whole network.

By pre-training the encoder with N-pair contrast loss, the distance between similar embeddings is reduced. This allows for better feature learning in the encoder stage and helps with the segmentation of the decoder. To produce a feature representation of the input image, we pre-trained the encoder on 150 epochs using N-pair contrast loss. Fig. 3 shows the proposed pre-training structure.

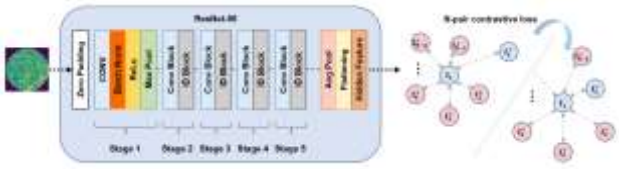

Fig. 3: Overall architecture of proposed pre-training.

N-pair loss is a generalized version of triplet loss, consisting of one anchor, one positive sample, and $(\mathrm{N}-1)$ negative samples. If $\mathrm{N}=2$, this is equivalent to a triplet loss. This is optimized for identifying a positive sample from multiple negative samples. Consider the training data $\{\mathrm{x}, x_+, x_1, \cdots, x_{n-1}\}$, where x_+ is a positive sample of x and $x_+, x_1, \cdots, x_{n-1}$ are negative. The $(\mathrm{N}+1)$-tuple loss is defined as follows, where f is the embedding kernel defined by the deep neural network.

$$\mathrm{L}\left(\mathrm{x}, x^+, \{x_i\}_{i=1}^{N-1}; f\right) = \log(1+\sum_{i=1}^{N-1}\exp(f^T f_i - f^T f^+)) \quad (1)$$

The multi-class N-pair loss (N-pair-mc) is defined as follows:

$$L_{N-pair-mc}\left(\{(x_i, x_i^+)\}_{i=1}^{N}; f\right) = \frac{1}{N}\sum_{i=1}^{N}\log(1+\sum_{j\neq i}\exp(f_i^T f_j^+ - f_i^T F_i^+)) \quad (2)$$

3.2 Pseudo-Defect Masking

To train the segmentation model, we need images and masks that serve as a label. Generally, labeling is performed by humans using an image masking program. These traditional approaches are time-consuming and need a lot of labor. To overcome that problem, we use a technique that automatically masks defects. A defect represented by a wafer map consists of a set of defects on a die. So, we perform the masking in a way that separated the connected pixels. We used the measure. label feature of Scikit-image, and Fig. 4 shows the connectivity option. We use 2-connectivity based on what we did with the connectivity option, as shown in Fig. 5.

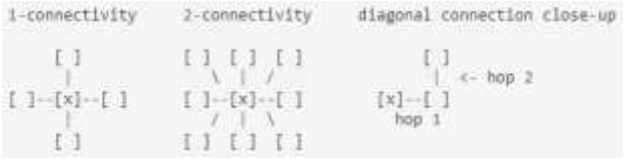

Fig. 4: Connectivity option description.

Fig. 5: Execution result by option.

4 Experiment and Results

4.1 Experiment Environments

To check our proposed model's performance, we conduct classification and segmentation with wafer maps including both single and mixed defects. Every experiment was run on four V100 GPUs with 16 CPU cores, 528 GB of MEM, and 32 GB of RAM, using the torch open-source library. Table 1 summarizes the system specification.

Table 1. System specification

Hardware Environment	Software Environment
CPU: 16Cores MEM: 528GB GPU: V100 x 4	Linux Torch 1.4.0 Python3.7

4.2 Datasets

Mixed-Type Defects. We are aiming to identify mixed defects. Therefore, we chose to use a mixed wafer defect dataset offered by the Intelligent Manufacturing Institute and Donghua University. It contains both single defects and mixed defects. Single defect classes are organized as follows Center, Donut, Edge-Loc, Edge-Ring, Loc, Scratch, Random, and Near-full. The center is defects clustered in the center. Donut is a ring formed by defects in the center. Edge-Loc is a cluster localized at the edge. Edge-Ring is Ringed clusters around boundaries. Loc is localized clusters that occur regularly. Scratch is a distribution of defects in long, narrow areas. Near-full is abnormal failure patterns. Random is random defects with no pattern. Fig. 6 shows a Single wafer map defect. Fig. 7 shows a mixed-type wafer map defect. For mixed-type Defects, we used 10,400 of the two-type mixed defects images as training data and 2,600 as evaluation data.

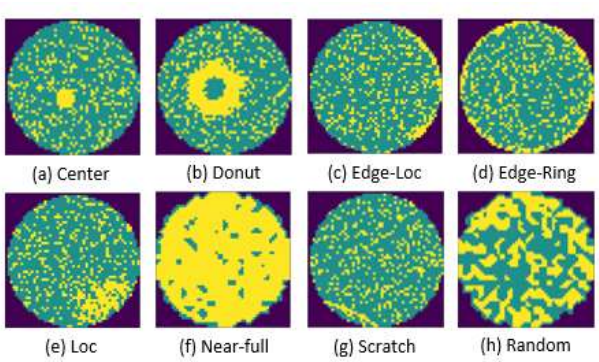

Fig. 6: Single wafer map defect: (a) Center (C); (b) Donut (D); (c) Edge-Loc (EL); (d) Edge-Ring (ER); (e) Loc (L); (f) Near-full; (g) Scratch (S); (h) Random.

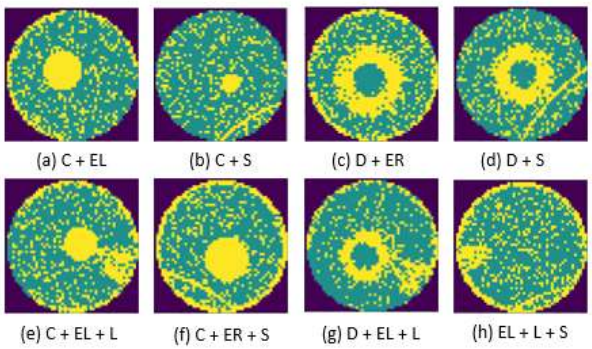

Fig. 7: Mixed-type wafer map defect: (a) Center + Edge-Loc; (b) Center + Scratch; (c) Donut + Edge-Ring; (d) Donut + Scratch; (e) Center + Edge-Loc + Loc; (f) Center + Edge-Ring + Scratch; (g) Donut + Edge-Loc + Loc; (h) Edge-Loc + Loc + Scratch.

4.3 Results

Mixed-Type Defect Result. We compared our performance to previous works in Table 2. Our model was better than previous studies. After pre-training the CNN among TransUNet's encoders using the N-pair contrastive loss function, the self-attention-based transformer layer concentrated on the defect and treated the neighboring defective dies as noise, resulting in improved results. We can also see better performance compared to previous studies. For accuracy or F1-Score, our model shows improvement. Fig. 8 illustrates the confusion matrix of mixed-type defects. In problems such as statistical classification in machine learning, a confusion matrix is a table that allows us to visualize the performance of a classification algorithm trained with supervised learning. Each column of the matrix represents an instance of the predicted class and each row represents an instance of the true class (or vice versa). Precision, indicated by the yellow diagonal line, was able to achieve a result of 0.995. Table 3 details the model's performance in detecting two types of mixed defects. The F1-Socre for two-type mixed defects reached 0.995.

Table 2. Comparison with other models

Model	Accuracy	F1-Score
Our Model	0.995	0.995
	0.826	0.824
[12]	0.962	0.962
[13]	0.977	0.977
[14]	0.980	0.974
[15]		

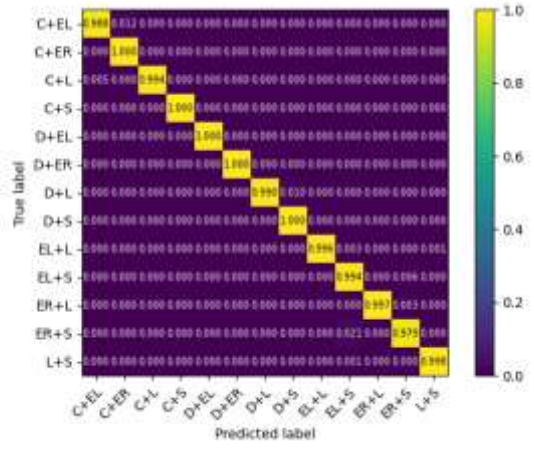

Fig. 8: Testing results of mixed-type defects: normalized confusion matrix (C + EL, C + ER, C + L, C + S, D + EL, D + ER, D + L, D + S, EL + L, EL + S, EL + L, ER + S, L + S).

Table 3. Mixed-type testing result

Defect Type	F1-Score		
	baseline	w/multitask	w/pretrain+multitask
C + EL	0.903	0.895	0.990
C + ER	0.955	0.959	0.993
	0.954	0.953	0.997
C + L	0.981	0.984	1.000
	0.930	0.941	1.000
C + S	0.968	0.972	1.000
	0.924	0.942	0.995
D + EL	0.975	0.979	0.995
	0.875	0.887	0.998
D + ER	0.890	0.921	0.984
	0.852	0.875	0.999
D + L	0.917	0.918	0.986
	0.965	0.969	0.999
D + S			
EL + L			
EL + S			
ER + L			
ER + S			
L + S			

Fig. 9 shows some samples of inference results for two-type mixed defects. It is composed of the original image, pseudo-label, and inference results. When benchmarked against the pseudo-label mask, we can check that it correctly predicts a single defect. Since the pseudo label data serves as the label, segmentation results are produced as close to the pseudo label. Since the masking was done programmatically, there were mismatches in the defects, but there was no problem detecting the defects. It could even correctly predict the majority of defects that were a mix of both types, but it was difficult to recognize when two defects overlapped. By more than a certain percentage, such as when Local and Edge-Loc defects overlapped, or when Scratches overlapped with other defects.

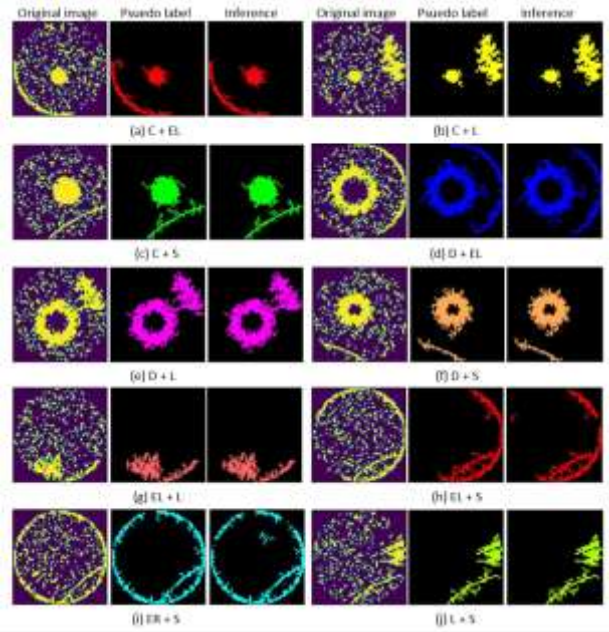

Fig. 9: Inference results in two-types mixed defects: (a) Center + Edge-Ring defect; (b) Center + Loc defect; (c) Center + Scratch defect; (d) Donut + Edge-Loc defect; (e) Donut + Loc defect; (f) Donut + Scratch defect; (g) Edge-Loc + Loc defect; (h) Edge-Loc + Scratch defect; (i) Edge-Ring + Scratch defect; (j) Loc + Scratch defect.

5 Conclusion

In this study, we used an N-pair Contrastive Loss Function to pre-train a CNN on the encoders of the TransUNet and then learn the entire TransUNet, resulting in good feature maps that emphasize the details of defects we were looking for. Human defect labeling is a very inefficient method because it is subjective and labor-intensive. In this study, we used automatic masking to solve the inefficiency problem. Our suggested approach delivered better results than the previous approach. Accuracy was 0.995, and F1-Score was 0.995. These results provide engineers with the exact location of the error, which helps them determine the cause of the problem. The study allowed us to detect mixed defects, and the automatic masking technique saved us unnecessary labor and time. This saves labor for existing workers and provides accurate defect detection.

We tested the model in public large wafer map datasets, but further validation on real-world datasets can be considered in future work. We can also use methods like transfer learning. In the future, we will concentrate our research on decreasing the size of the model and increasing performance.

Acknowledgment:
This research was supported by the SungKyunKwan University and the BK21 FOUR(Graduate School Innovation) funded by the Ministry of Education(MOE, Korea) and the National Research Foundation of Korea(NRF). And this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2021R1F1A1060054).

References:

[1] Shu-Kai S. Fan, Chia-Yu Hsu, Du-Ming Tsai, Fei He, Chun-Chung Cheng, Data-driven approach for fault detection and diagnostic in semiconductor manufacturing, *IEEE Transactions on Automation Science and Engineering*, Vol.17, Issue.4, 2020, pp. 1925–1936.

[2] Erik H.M. Heijne, Future semiconductor detectors using advanced microelectronics with post-processing, hybridization and packaging technology. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol.541, Issues.1-2, 2020, pp. 274–285.

[3] Chih-Hung Chang, Brian K. Paul, Vincent T. Remcho, Sundar Atre, James E. Hutchison, Synthesis and post-processing of nanomaterials using microreaction technology, *Journal of Nanoparticle Research*, Vol.10, 2008, pp. 965–980.

[4] P. Grybos, Front-End Electronics for Multichannel Semiconductor Detector Systems, Institute of Electronic Systems, Warsaw University of Technology: Warsaw, Poland, 2020, pp. 132–135.

[5] Ofer Sneh, Robert B. Clark-Phelps, Ana R. Londergan, Jereld Winkler, Thomas E. Seidel, Thin film atomic layer deposition equipment for semiconductor processing, *Thin Solid Film*, Vol.403, Issues.1-2, 2020, pp. 248–261.

[6] S. Joe Qin, Gregory Cherry, Richard Good, Jin Wang, Christopher A. Harrison, Semiconductor manufacturing process control and monitoring: A fab-wide framework, *Journal of Process Control*, Vol.16, Issue.3, 2006, pp. 179–191.

[7] Seokho Kang, Sungzoon Cho, Daewoong An, Jaeyoung Rim, Using wafer map features to better predict die-level failures in final test, *IEEE Transactions on Semiconductor Manufacturing*, Vol.28, Issue.3, 2015, pp. 431–437.

[8] Minghao Piao, Cheng Hao Jin, Jong Yun Lee, Jeong-Yong Byun, Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features. *IEEE Transactions on Semiconductor Manufacturing*, Vol.31, Issue.2, 2018, pp. 250–257.

[9] Cheng Hao Jin, Hyuk Jun Na, Minghao Piao, Gouchol Pok, Keun Ho Ryu, A novel DBSCAN-based defect pattern detection and classification framework for wafer bin map, *IEEE Transactions on Semiconductor Manufacturing*, Vol.32, Issue.3 2019, pp. 286–292.

[10] Hyuck Lee, Heeyoung Kim, Semi-supervised multi-label learning for classification of wafer bin maps with mixed-type defect patterns, *IEEE Transactions on Semiconductor Manufacturing*, Vol.33, Issue.4, 2020, pp. 653–662.

[11] Kiryong Kyeong, Heeyoung Kim, Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks, *IEEE Transactions on Semiconductor Manufacturing*, Vol.31, Issue.3, 2018, pp. 395–402.

[12] Junliang Wang, Chuqiao Xu, Zhengliang Yang, Jie Zhang, Xiaoou Li, Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition, *IEEE Transactions on Semiconductor Manufacturing*, Vol.33, Issue.4, 2020, pp. 587–596.

[13] Jinho Kim, Youngmin Lee, Heeyoung Kim, Detection and clustering of mixed-type defect patterns in wafer bin maps, *IISE Transactions*, Vol.50, Issue.2, 2018, pp. 99–111.

[14] Ming-Chuan Chiu, Tao-Ming Chen, Applying Data Augmentation and Mask R-CNN-Based Instance Segmentation Method for Mixed-Type Wafer Maps Defect Patterns Classification, *IEEE Transactions on Semiconductor Manufacturing*, Vol.34, Issue.4, 2021, pp. 455–463.

[15] Jaegyeong Cha, Jongpil Jeong, Improved U-Net with Residual Attention Block for Mixed-Defect Wafer Maps. *MDPI Applied Sciences Journal*, Vol.12, Issue.4, 2022.

[16] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert, Attention U-Net: Learning Where to Look for the Pancreas, arXiv 2018, arXiv:1804.03999.

[17] Yu-Cheng Liu, Mohammad Shahid, Wannaporn Sarapugdi, Yong-Xiang Lin, Jyh-Cheng Chen, Kai-Lung Hua, Cascaded atrous dual attention U-Net for tumor segmentation, *Multimedia Tools and Applications*, Vol.80, 2021, pp.30007-30031.

[18] Xiaocong Chen, Lina Yao, Yu Zhang, Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images, arXiv.2004, arXiv:2004.05645.

[19] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, Yuyin Zhou, TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation, arXiv 2021, arXiv:2102.04306.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

-Youngjae Kim set the research topic and goals, developed the software, conducted the experiments, validated, and wrote the paper.

-Jee-Hyong Lee and Jongpil Jeong conducted the review and the editing.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research was supported by the SungKyunKwan University and the BK21 FOUR(Graduate School Innovation) funded by the Ministry of Education(MOE, Korea) and the National Research Foundation of Korea(NRF). And this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2021R1F1A1060054).

Conflict of Interest

The authors have no conflicts of interest to declare.