Deep Learning Wafer Defect Detection Overcomes Limited Label Data: Semi-Supervised Learning Techniques with FixMatch and Autoencoder

Seongjin Park
University of Wisconsin-Madison
Computer Science & Data Science
(seongjinpark99@gmail.com)
https://github.com/Seongjin74/Wafer-Defect-Classification

1. Introduction

With the rapid development of deep learning technology, tasks traditionally performed by humans—such as classification and prediction—are increasingly being automated. In particular, deep learning has demonstrated outstanding performance in areas like image classification, defect detection, and predictive modeling, gradually replacing conventional methods. However, efficiently utilizing limited data is a critical challenge when building high-performance deep learning models.

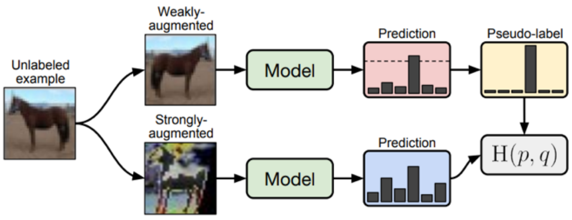
Traditional deep learning approaches primarily rely on supervised learning, which requires both input data and corresponding labels for training. This reliance poses a significant drawback due to the considerable labor and costs involved in labeling. For example, in complex environments such as semiconductor manufacturing processes, collecting a large volume of labeled data demands substantial resources and expense. To overcome these limitations, semi-supervised learning techniques—which reduce labeling costs and labor while still delivering excellent performance—have recently gained prominence.

In this study, we propose a semi-supervised learning method for wafer defect detection. Specifically, we introduce a method that effectively combines limited labeled data with a large amount of unlabeled data by leveraging the FixMatch algorithm, one of the latest semi-supervised learning techniques. Additionally, we employ an autoencoder as a strong augmentation module to generate reconstructed, transformed images from the original image, enabling the model to learn from diverse forms of augmented data while maintaining consistency with the pseudo-labels. This approach goes beyond mere image restoration with an autoencoder; it is integrated into the strong augmentation phase of FixMatch to maximize the benefits of semi-supervised learning.

This paper first reviews the theoretical background and existing research on various deep learning training techniques—including semi-supervised learning and autoencoders—and provides a detailed explanation of a wafer defect detection system that combines the FixMatch algorithm with an autoencoder-based approach. Subsequently, the performance of the proposed method is evaluated through various experiments, and the results are analyzed to explore its potential contributions to defect detection and prediction in semiconductor manufacturing processes.

2. Related Work

Recently, the FixMatch algorithm has gained atten tion in the field of semi-supervised learning as a method that demonstrates excellent performance despite its simplicity.



<Figure 1> Diagram of FixMatch Sohn et al. (2020) proposed a method in which pseudo-labels are generated based on predictions from weak data augmentations. These pseudo-labels are then applied to strongly augmented images only when they meet a certain confidence threshold (e.g., 0.90 or above) [1]. Specifically, after applying weak augmentation to the original image, the model's predicted distribution is obtained, and the class with the highest probability is selected as the pseudo-label. Subsequently, the model parameters are updated to maintain consistency with the pseudo-label for the result obtained from applying strong augmentation to the same image, thereby enabling effective learning even in scenarios with limited labeled data.

In the field of wafer defect detection, deep learning-based methods—particularly those using convolutional neural networks (CNNs)—have been actively researched, moving beyond traditional image processing techniques. Wang et al. (2018) proposed a wafer defect inspection system using CNNs and achieved higher detection rates compared to conventional statistical methods. Additionally, Kim et al. (2019) developed a wafer defect detection model that exhibits excellent performance even with limited labeled data by incorporating semi-supervised learning techniques[2][3].

Alongside these advancements, research involving autoencoders is also underway.

<Figure 2> A Schema of an Autoencoder

Autoencoders effectively extract key features from an image by compressing the input into a latent space during the encoding phase and then reconstructing the original image during the decoding phase. This method can simultaneously perform noise removal and feature reconstruction, thereby enhancing the robustness of data augmentation and defect detection. Choi et al. (2021) demonstrated that an autoencoder-based approach can effectively detect even minute defects in semiconductor manufacturing processes[4].

Thus, semi-supervised learning and autoencoder-based techniques are contributing to overcoming the limitations of labeled data and enhancing model generalization in wafer defect detection. In this study, we propose an approach that combines the FixMatch algorithm with an autoencoder, aiming to validate its performance and practicality in encompassing and extending previous research.

3. Proposed Methodology

In this study, we propose a novel methodology that combines the semi-supervised FixMatch algorithm with an autoencoder-based data augmentation technique to address the wafer defect detection problem. The proposed method consists of data preprocessing, data augmentation, model design, and training procedures, with a particular focus on the combination of weak and strong augmentations and their integration into the FixMatch framework.

First, the wafer map data is transformed into a normalized 26×26 three-channel image using the process_wafer_map function. During this preprocessing step, each pixel value of the 2D wafer map is mapped according to a predefined colormap, effectively accentuating the defect patterns in the original data.

Regarding data augmentation, the proposed method combines two augmentation strategies. Weak augmentation applies resizing and minor rotational transformations to the original image to preserve its basic characteristics, while strong augmentation generates a variety of transformed images through more aggressive geometric modifications. Notably, the autoencoder-based augmentation technique plays a key role in the strong augmentation stage. The trained autoencoder extracts the latent representation of the input image, adds noise to the latent space, and then generates a reconstructed image via the decoder. This reconstruction process allows the autoencoder to learn the intrinsic features of the wafer image, and the reconstructed output is further processed into discrete values. In addition, during the strong augmentation phase, a random 90-degree rotation is applied to the autoencoder-augmented result to generate even more diverse transformed images, enabling the model to robustly learn from various augmentations.

The FixMatch algorithm first generates pseudo-labels based on the model's predictions from the weakly augmented image. Pseudo-labels are considered reliable only if the prediction probability exceeds a predefined threshold (e.g., 0.95). The model is then trained to maintain consistency between the pseudo-label and the result obtained from applying strong augmentation to the same original image. This process contributes to enhancing the model's generalization performance by effectively utilizing a large amount of unlabeled data even when labeled data is limited.

For model design, a simple yet efficient CNN architecture, SimpleCNN, was adopted. This network extracts low-level features from the image through three convolutional layers and two max pooling layers, and then predicts the final class through fully connected layers. Owing to its concise structure, SimpleCNN has low computational complexity and converges quickly in a semi-supervised learning environment, making it suitable for both supervised and semi-supervised experiments.

The overall training process begins with applying both weak and strong augmentations to each dataset. In the supervised learning experiments, the model was trained using limited labeled data with a conventional CrossEntropyLoss, whereas in the semi-supervised experiments, the model was trained using both labeled and unlabeled data according to the FixMatch algorithm. For the unlabeled data, an additional loss is computed to maintain consistency between the pseudo-label generated from the weakly augmented image and the strongly augmented image, with weights assigned according to the confidence of the pseudo-label. The proposed method is a comprehensive approach that integrates data preprocessing, augmentation, semi-supervised learning, and autoencoder-based reconstruction for wafer defect detection, aiming to achieve high detection rates and generalization performance while effectively utilizing limited labeled data.

4. Experiments

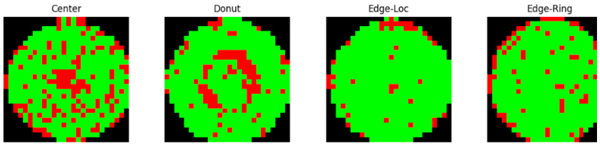
4.1 Dataset

this study, the performance of the proposed methodology was evaluated using the WM-811K wafer map dataset, which was collected from semiconductor manufacturing processes. The WM-811K dataset, introduced by Wu et al. (2015), is a large-scale wafer map dataset containing a total of 811,457 wafer maps collected from 46,393 LOTs in an actual manufacturing environment. The defect types are categorized as Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, Near-full, none, among others[5]. This dataset is released under the CC0 (Public Domain) license and is widely used in research on wafer defect detection and similarity assessment.

The entire dataset was first preprocessed to extract wafer map images and label information, and then partitioned into a balanced dataset for each class. In the supervised learning experiments, up to 100 training samples and 50 test samples per class were used, while in the semi-supervised learning experiments, 10,000 unlabeled samples were additionally utilized along with the same 100 labeled samples.

During the preprocessing stage, the input wafer map was converted into a normalized three-channel image using the process_wafer_map function. This function maps the values within the wafer map (0, 1, 2) to a predefined colormap (0 → black, 1 →

green, 2 → red) and then resizes the image to the specified dimensions (26×26), producing an image normalized to the [0, 1] range.



<Figure 3> Wafer data image after preprocessing

Subsequently, the Pandas library was used to filter the labeled data, and a training/test dataset was constructed for each class in an 80:20 ratio. For the supervised learning experiments, limiting the number of samples per class helped mitigate data imbalance and provided a baseline for comparison.

4.2 Deep Learning Model

A simple CNN architecture, SimpleCNN, was employed. This model was designed to extract features through three convolutional layers and two max pooling layers, followed by fully connected layers for final class prediction. Training was performed using the Adam optimizer (LR=1e-3) with CrossEntropyLoss. A scheduler that reduced the learning rate by half every 10 epochs was applied, and training proceeded for a total of 50 epochs.

4.3 Supervised Learning

In the supervised learning experiments, the model was trained using a Supervised DataLoader. During the training process, the loss value gradually decreased over epochs, and the final classification accuracy on the test dataset was evaluated at approximately 73.72%. In addition, precision, recall, and f1-score for each class were computed to assess the model's detailed performance, and a confusion matrix was used to visually analyze the classification results among the classes.

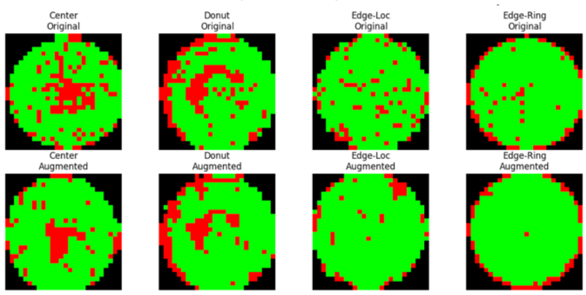
Test Accuracy: 73.72%

	precision	recall	f1-score	support
Center	0.88	0.84	0.86	50
Donut	0.90	0.92	0.91	50
Edge-Loc	0.70	0.64	0.67	50
Edge-Ring	0.87	0.90	0.88	50
Loc	0.59	0.48	0.53	50
Near-full	0.97	1.00	0.98	30
Random	0.85	0.92	0.88	50
Scratch	0.46	0.42	0.44	50
none	0.51	0.62	0.56	50
accuracy			0.74	430
macro avg	0.75	0.75	0.75	430
weighted avg	0.73	0.74	0.73	430

<Figure 4> Supervised learning results table

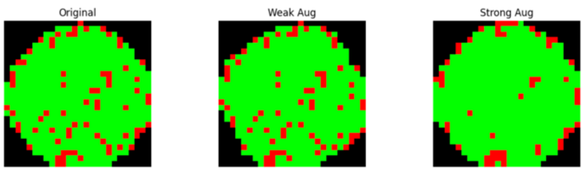
4.4 Semi-Supervised Learning

In the semi-supervised learning experiments, pseudo-labels were first generated based on the model's prediction results from images with weak augmentation. At this stage, only predictions with a probability exceeding a predefined threshold (e.g., 0.95) were considered reliable enough to serve as pseudo-labels. For the same original image, strong augmentation was then applied so that the model could be trained to maintain consistency with the pseudo-label. The strong augmentation not only applied geometric transformations but also incorporated an autoencoder-based augmentation module. Specifically, the trained autoencoder transformed the input image (26×26×3) into a latent representation of size 13×13×64 via the encoder, added a small amount of noise (noise_std=0.05) in the latent space, and then reconstructed the image through the decoder.



<Figure 5> Example wafer images before and after autoencoder augmentation for each class

The reconstructed output underwent a postprocessing step involving pixel-level discrete transformation, followed by an additional random rotation in 90-degree increments. These strongly augmented images preserved the core defect patterns of the original image while generating various forms of transformed images, thus enforcing consistency with the pseudo-labels generated by the FixMatch algorithm.



<Figure 6> Comparison example of weakly and strongly augmented unlabeled wafer images used in semi-supervised

learning

In the early stages of the experiment, the utilization rate of pseudo-labels from the unlabeled data was extremely low. However, by epoch 50, approximately 16% (e.g., 1002 out of 6272) of the unlabeled samples had pseudo-labels applied. The semi-supervised learning experiments achieved an overall test accuracy of 76.05%.

[Semi-supervised Experiment Results]
Test Accuracy: 76.05%

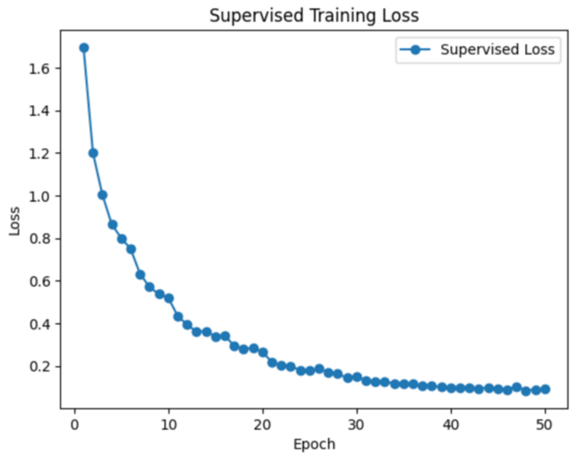
	precision	recall	f1-score	support
Center	0.90	0.90	0.90	50
Donut	0.90	0.92	0.91	50
Edge-Loc	0.76	0.62	0.68	50
Edge-Ring	0.96	0.88	0.92	50
Loc	0.65	0.48	0.55	50
Near-full	0.97	1.00	0.98	30
Random	0.81	0.92	0.86	50
Scratch	0.50	0.52	0.51	50
none	0.54	0.70	0.61	50
accuracy			0.76	430
macro avg	0.78	0.77	0.77	430
weighted avg	0.77	0.76	0.76	430

<Figure 7> Semi-supervised learning results table

In summary, the FixMatch algorithm effectively leveraged both limited labeled data and a large amount of unlabeled data by combining pseudo-label generation from weak augmentation with autoencoder-based strong augmentation (involving the addition of slight noise, reconstruction, and additional random rotations) to enhance the model's robustness to diverse transformed data. Moreover, the proposed autoencoder-based augmentation technique significantly contributed to improving the model's generalization performance in the semi-supervised learning process by effectively preserving the inherent features of the original image while maximizing data diversity through additional geometric transformations.

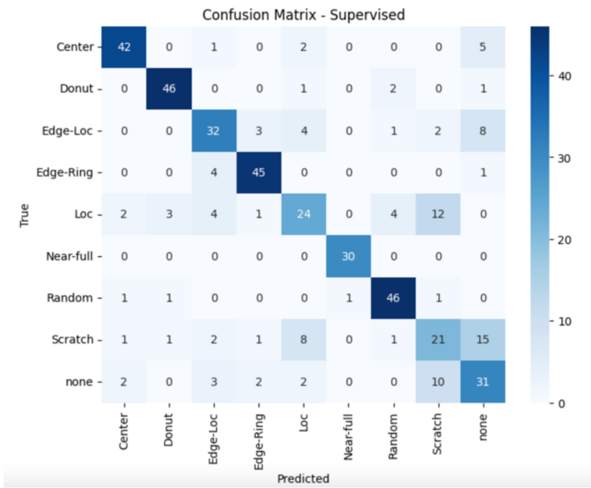
5. Experimental Evaluation

In this study, the performance and training processes of models trained using both supervised and semi-supervised methods were compared and analyzed quantitatively. In the supervised learning experiment, the SimpleCNN model was trained for a total of 50 epochs, with the initial training loss starting at approximately 1.73 and decreasing to 0.0920 by epoch 50.



<Figure 8> Supervised Learning Loss Graph

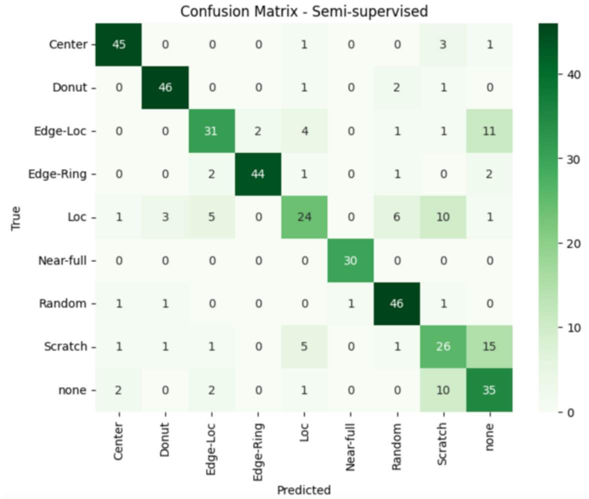
During the testing phase, the supervised learning approach achieved an overall test accuracy of 73.72%.



<Figure 9> Supervised Learning Confusion Matrix

A closer look at the detailed performance for each class revealed that the Center class had a precision of 0.88, recall of 0.84, and an f1-score of 0.86; Donut achieved 0.90/0.92/0.91; Edge-Loc scored 0.70/0.64/0.67; Edge-Ring reached 0.87/0.90/0.88; Loc obtained 0.59/0.48/0.53; Near-full recorded 0.97/1.00/0.98; Random resulted in 0.85/0.92/0.88; Scratch showed 0.46/0.42/0.44; and none achieved 0.51/0.62/0.56. These results suggest that while the supervised approach was able to achieve stable training based on the limited labeled data, certain classes (particularly Scratch and none) demonstrated relatively lower classification performance.

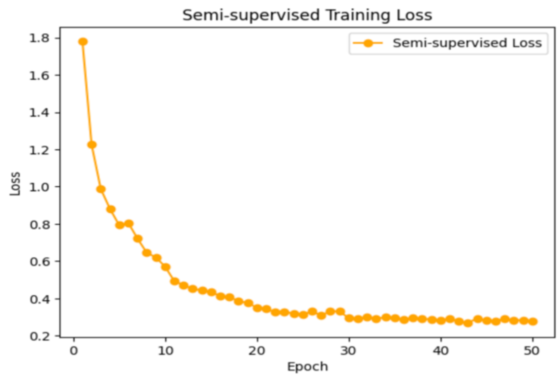
In the semi-supervised learning experiment, the FixMatch algorithm was applied using the same 900 labeled samples along with an additional 10,000 unlabeled samples. Initially, the utilization rate of pseudo-labels from the unlabeled data was nearly 0; however, as training progressed, approximately 16% of the unlabeled samples (e.g., 1002 out of 6272) were assigned pseudo-labels by epoch 50. Ultimately, the semi-supervised learning approach achieved an overall test accuracy of 76.05%, with the class-wise performance as follows:



<Figure 10> *Semi-Supervised Learning Confusion Matrix*

Center: 0.90/0.90/0.90 - Donut: 0.90/0.92/0.91 - Edge-Loc: 0.76/0.62/0.68 - Edge-Ring: 0.96/0.88/0.92 - Loc: 0.65/0.48/0.55 - Near-full: 0.97/1.00/0.98 - Random: 0.81/0.92/0.86 - Scratch: 0.50/0.52/0.51 - none: 0.54/0.70/0.61

Comparing supervised and semi-supervised learning directly, the semi-supervised approach improved the overall test accuracy by approximately 2.3%. Notably, for the Center and Edge-Ring classes, the f1-score improved by about 4% (from 0.86 to 0.90, and from 0.88 to 0.92, respectively), and for classes that initially showed relatively low performance—such as the Scratch class (improving from 0.44 to 0.51, approximately 7%) and the none class (improving from 0.56 to 0.61, approximately 5%)—significant enhancements were observed. Although the Random class experienced a slight decrease in precision, its overall f1-score remained unchanged, and for Edge-Loc, while precision improved, a slight decrease in recall resulted in only a marginal change in the f1-score.



<Figure 11> *Semi-Supervised Learning Loss Graph*

Furthermore, monitoring the training loss curve and pseudo-label utilization rate indicated that the additional learning effect derived from the unlabeled data gradually improved the model's convergence speed and generalization performance in the semi-supervised setting. The autoencoder-based strong augmentation effectively reconstructed the latent representation of the input image and, with the addition of random rotations, generated diverse transformed images. This helped the FixMatch algorithm maintain consistency with the pseudo-labels and ultimately maximized the benefits of the unlabeled data even in a limited-label environment.

In summary, the semi-supervised learning approach demonstrated improvements in overall accuracy and various class performance metrics compared to supervised learning. In particular, the combination of autoencoder-based strong augmentation with the FixMatch algorithm contributed significantly to enhancing the model's generalization capability. These results suggest that the proposed method provides a practical improvement in effectively leveraging limited labeled data for the wafer defect detection problem.

6. Conclusion and Future Work

In this study, a semi-supervised learning approach that effectively utilizes limited labeled data was applied for wafer defect detection in semiconductor manufacturing processes. The proposed method combined the FixMatch algorithm with an autoencoder-based data augmentation technique, aiming to enhance the model's generalization performance even when labeled data is scarce. Experimental results showed that while the supervised learning experiment achieved a test accuracy of 73.72%, the application of the semi-supervised learning technique reached a test accuracy of 76.05%, confirming the effective use of unlabeled data. Moreover, class-wise evaluation metrics and confusion matrix analysis indicated that the benefits of semi-supervised learning were particularly pronounced in classes with subtle defect patterns.

Future research should introduce several improvements to maximize the combined effect of the autoencoder-based strong augmentation technique and the FixMatch semi-supervised learning approach. First, instead of simply applying a fixed noise_std value in the latent space to add noise, it is necessary to dynamically adjust the noise distribution and magnitude through hyperparameter tuning, or to apply different levels of noise and combine them using an ensemble approach. This would further enhance the diversity and quality of the reconstructed images. In addition, expanding the autoencoder module to a deeper and more complex network architecture should be considered in order to extract and reconstruct the intrinsic features of wafer images more precisely.

Furthermore, improving the method for calculating the confidence of pseudo-labels in semi-supervised learning is an important task. It is worth exploring dynamic adjustment of the current fixed threshold (0.95) based on differences in model confidence between the early and later stages of training, or introducing a weighting scheme that takes predictive uncertainty into account. This would maximize the contribution of unlabeled data to model training.

Finally, while simple models like SimpleCNN have demonstrated significant performance improvements, future work should consider integrating deeper and more complex network architectures, such as ResNet or DenseNet, to further maximize the synergistic effect of semi-supervised learning and autoencoder-based strong augmentation. By systematically introducing hyperparameter tuning and various improvement strategies, a comprehensive framework can be established to maximize the accuracy and generalization capability of wafer defect detection even in environments with limited labeled data.

Reference

[1] Sohn, K., Berthelot, D., Li, C.-L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., & Raffel, C. (2020). FixMatch: Simplifying Semi-Supervised
[2] Learning with Consistency and Confidence. Wang, Y., et al. (2018). Wafer Defect Inspection
[3] Using Deep Convolutional Neural Networks. Kim, S., et al. (2019). Semi-supervised Learning for
[4] Wafer Defect Detection. Choi, J., et al. (2021). Autoencoder-Based Defect Detection in Semiconductor Manufacturing.
Wu, Ming-Ju, Jyh-Shing R. Jang, and Jui-Long Chen.
[5] "Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets." IEEE Transactions on Semiconductor Manufacturing 28, no. 1 (February 2015): 1–12.