Minimizing Labeling, Maximizing Performance: A Novel Approach to Nanoscale Scanning Electron Microscope (SEM) Defect Segmentation

Yibo Qiao[1,3], Weiping Xie[1,3], Shunyuan Lou[1,3], Qian Jin[1,3], Lichao Zeng[2,3], Yining Chen[1,3*], Qi Sun[1*], and Cheng Zhuo[1*]
[1]Zhejiang University, Hangzhou, China
[2]University of Science and Technology of China, Hefei, China
[3]Zhejiang ICsprout Semiconductor Co., Ltd., Hangzhou, China
*Corresponding author email: {czhuo, yining.chen, qisunchn}@zju.edu.cn

ABSTRACT

In semiconductor manufacturing, pinpointing nanoscale wafer defects is crucial for yield and reliability. Deep learning methods for defect segmentation rely heavily on large, labor-intensive datasets and focus mainly on macroscopic wafer defects, not nanoscale morphology. Our research introduces a hybrid weakly supervised scanning electron microscope (SEM) defect segmentation system with two sub-networks: one for accurate defect localization and image cropping, another for detailed segmentation. Validated on 1,328 SEM image defects from a real facility, our model surpasses existing weakly supervised methods and equals fully supervised models in accuracy, with 10% labeling effort, providing a novel approach for high-precision defect segmentation.

1 INTRODUCTION

The semiconductor manufacturing process is composed of thousands of steps [1], such as thin film deposition, ion implantation, etching, polishing, *etc.* Even minor abnormalities can lead to wafer surface defects. These defects, far from being mere indicators of equipment malfunction or operational errors, are closely linked to the electrical performance of semiconductor devices, making their precise detection and analysis crucial for semiconductor yield and reliability enhancement [2, 3].

Historically, defect detection at the wafer level has been dominated by two approaches: spatial distribution analysis and scanning electron microscope (SEM)-based examination of microscopic morphology, as shown in Figure 1. While wafer maps provide a macroscopic view of defect distribution, their lack of detail falls short in identifying the subtle imperfections crucial for root cause analysis [3]. Conversely, SEM allows for the detailed examination of nanoscale defects, where even minuscule details are critical. This method enables engineers to capture multi-angled images of each defect, offering a comprehensive view of their shape and topography. Such detailed analysis is essential for pinpointing process-related defect causes and implementing corresponding measures to boost production quality and consistency [4].

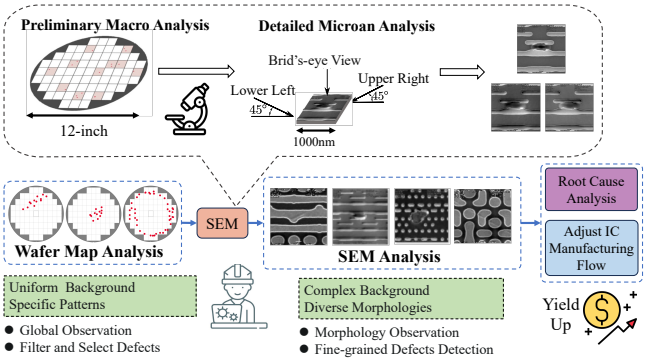


Figure 1: The application of wafer maps and SEM images in defect analysis.

Early semiconductor defect detection relied on traditional methods such as rule-based methods, statistical methods, and morphological algorithms[5]. These approaches necessitated laborious manual feature extraction and rule formulation, thus consuming significant human resources. In recent years, significant progress has been made in the research on semiconductor defect detection, owing to the emergence of deep learning technology. Deep learning harnesses the powerful learning capabilities of neural networks, enabling automatic learning and feature extraction from large volumes of training data without the need for manual intervention. The deep learning approach has demonstrated the potential to improve efficiency in the defect detection field [6]. The literature is replete with studies on the spatial distribution of defects, and such research has progressed significantly [7–9]. Conversely, the microscopic morphology analysis of SEM defects confronts a host of intrinsic challenges that have yet to be fully addressed. These challenges are twofold: firstly, the defects captured in SEM images exhibit a diverse range of shapes and structures set against a complex background, making categorization and analysis a daunting task; secondly, variability within defect categories, and similarity across different ones, complicates the detection process to the extent that even seasoned engineers are faced with considerable analytical demands [10].

The hurdles are further exacerbated by the technical limitations of existing deep learning methods in this context. The performance of these methods is intrinsically linked to the availability of extensive, fully annotated datasets. The acquisition of these datasets is hindered by the confidentiality of SEM images and the labor-intensive manual annotation process, rendering this approach impractical and unsustainable. This scarcity of labeled data is a critical bottleneck, severely impeding the progress of automated SEM defect analysis [10]. In light of these challenges, recent methodologies have sought

DAC '24, June 23–27, 2024, San Francisco, CA, USA

ACM ISBN 979-8-4007-0601-1/24/06…$15.00
https://doi.org/10.1145/10.1145/3649329.3655939

to alleviate the onerous labeling task. Techniques such as active learning, semi-supervised learning, and weakly supervised learning have been proposed as less labor-intensive alternatives [11–13]. Yet, these advanced algorithms, primarily tailored for natural images, falter when confronted with the unique complexities of industrial SEM images [11]. The rich semantic information inherent to natural images does not translate to the high variability and subtle nuances of SEM defect images, rendering these weakly supervised methods less effective when directly applied to the domain of semiconductor manufacturing.

The crux of the problem in current defect analysis methodologies for semiconductor manufacturing is multi-dimensional: (1) The intricate and often confounding backgrounds in SEM images place exacting demands on analytical models. (2) The reliance on deep learning techniques on extensive data sets poses a significant challenge in a field where such data is scarce. (3) The direct application of existing weak supervision methods is largely ineffective for the nuanced requirements of SEM defect analysis. (4) The dearth of diverse and labeled SEM datasets constrains the evolution of automated analysis techniques for these defects.

To tackle these issues, we design a novel hybrid weakly supervised network. This network effectively alleviates the rigorous demands imposed by fully supervised learning methods. By leveraging a substantial volume of data with cheap image-level classification labels and a small dataset with expensive pixel-level labels, we achieve highly accurate pixel-level segmentation results for all images. Our main contributions are summarized as follows:

- We propose a novel weakly supervised automated analysis for SEM defects, aiming to address the lack of research and knowledge in this specific domain.
- We propose a hybrid weakly supervised segmentation network to utilize limited labeled information to achieve precise SEM image segmentation results effectively.
- We introduce a versatile key area localization mechanism that significantly amplifies the efficacy of existing segmentation techniques.
- To facilitate the research of SEM defect morphology analysis, we collect and label a new nanoscale SEM defect dataset from a real semiconductor fabrication facility.

The experimental results show that our proposed network outperforms the state-of-the-art remarkably, and achieves more than fully supervised performance with only 10% of manual annotation effort.

2 PRELIMINARIES

2.1 Related Work

In the nanoscale SEM defects analysis, the cost to address SEM image segmentation issues will increase significantly due to the extensive manual workloads required for pixel-level labels by experienced engineers. Existing approaches significantly reduce the burden of manual labeling, particularly weakly supervised learning, which effectively balances the level of manual labeling required with the resulting model performance [14–16].

Figure 2 depicts the various levels of labeling efforts featured with different supervisions. Fully supervised learning requires the largest amount of labeling. Existing research on industrial defect detection focuses on fully supervised approaches, relying on a large amount of accurately labeled data to train the models [7–9]. Particularly in defect segmentation tasks, accurate pixel-level segmentation labels are required, resulting in high costs. Purely weakly supervised methods

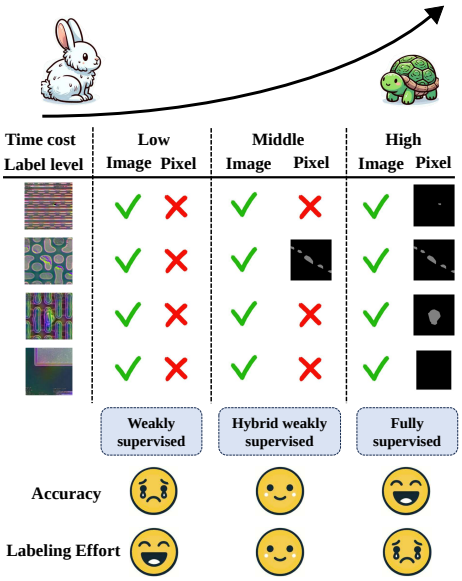


Figure 2: Different supervision methods and their required labeling effort.

have the least labeling effort, but due to the significant information gap between image-level and pixel-level labels, the segmentation performance of purely weakly supervised methods is difficult to guarantee. Hybrid weakly supervised learning, employed in this study, falls under the broader category of weakly supervised learning. Compared to purely weakly supervised methods, hybrid weakly supervised methods enhance performance by incorporating a small amount of auxiliary information.

2.2 Collection and Processing of Dataset

Currently, there is a lack of a real and well-labeled SEM image dataset for nanoscale surface defects. In this work, we collected defects from various key manufacturing steps at a 12-inch wafer fabrication facility, using KLA corporation-produced puma9550 and EDR7380 machines.

We gathered six typical types of defects including holes, particles, infilm, puddles, scratches, and pattern deformations (abbreviated as "deform"). Figure 3 exemplifies some defective images. To obtain more comprehensive defect information, each defect was captured from three different incident angles, as demonstrated in Figure 1, resulting in 3 SEM images per defect.

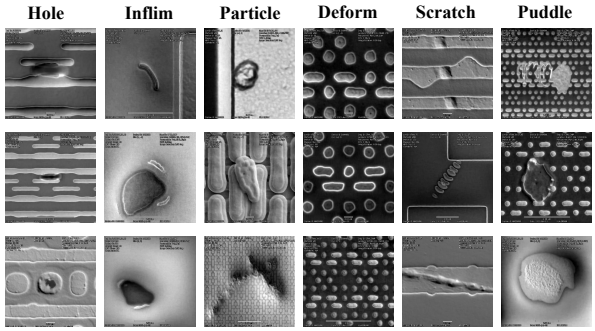


Figure 3: The six classes of defects.

The captured images are synthesized to aggregate information, as shown in Figure 4. Specifically, each defect had three SEM images formed from bottom-left, top-right, and bird's-eye view, each visualized as a single-channel grayscale image. Consequently, the three images were combined into a three-channel image. This is conducive to a more comprehensive utilization of defect information. After this preprocessing, we obtained 1,328 combined images of defects and 100 combined images of non-defect samples.

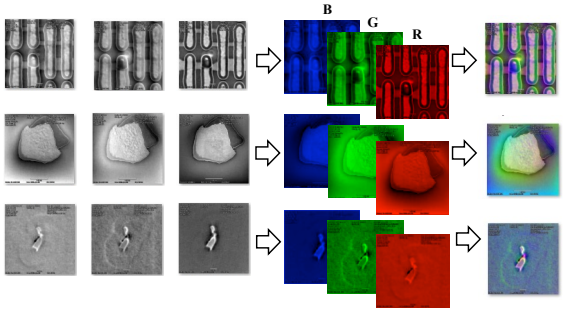


Figure 4: Three single-channel grayscale SEM images are synthesized into one three-channel RGB image.

We used the annotation software Labelme to make careful image-level classification annotation and pixel-level segmentation annotation for SEM images after fusion. Each SEM defect image is marked with the presence or absence of defects, the category of defects and the pixel part belonging to defects in the map. The annotated dataset can be used for accurate benchmarking of nanoscale defect classification and defect segmentation methods in chip manufacturing.

2.3 Problem Definition

Problem 1 (Hybrid Weakly Supervised SEM Defect Segmentation)**.** Given a dataset of defective SEM images, labeled with image-level labels and pixel-level labels, the objective of the hybrid weakly supervised SEM defect segmentation system is to train a network that can generate high-precision pixel-level segmentation results for all SEM defect images using minimal label information.

3 PROPOSED FRAMEWORK

3.1 Framework Overview

Our hybrid weakly supervision framework consists of two stages, corresponding to the key area localization network (KALNet) and the weakly supervised segmentation network (WSSNet), as shown in Figure 5. It includes sub-modules such as regions of interest (RoI) acquisition, cropping and augmentation module, and attention feedback module. In the first stage, SEM images are processed through a region of interest acquisition network to identify key regions where defects are located. These key regions are then used for image cropping and augmentation to expand the dataset. In the second stage, the encoder extracts features from the enhanced images and their corresponding key areas. After obtaining the segmentation results from the decoder, a gradient attention feedback is sent back to the network, encouraging the network to focus more on the defect areas in the SEM images. This iterative process enhances the feature extraction of the defect regions and assists the network in effectively capturing defect features.

3.2 Key Area Location Network

Weakly supervised approaches typically focus on improving Class Activation Maps (CAMs), often leading to incomplete attention maps. Our approach, however, relaxes the requirement for CAM accuracy. By identifying crucial regions first, we then apply cropping enhancement to these areas, offering more usable data for later weakly supervised segmentation networks.

RoI Acquisition Module: This module maximizes the use of classification labels in SEM images to identify key areas in images with and without defects. We implement a convolutional neural network for binary classification to differentiate SEM images that contain defects from those that don't. The network utilizes ResNet34 for feature extraction, where each Res-Block improves gradient propagation through residual connections, facilitating easier training and optimization of the network. The gradient information corresponding to the convolution layer contains information corresponding to the weights of each feature channel, which reflects the importance of different feature channels for the classification results [17]. In order to obtain the RoI in the images that the network focuses on, we perform weighted summation of the weights and the channels of the last convolutional feature map. This process associates the output features of the classification network with the spatial positions of the input SEM images, thereby obtaining the most crucial regions, which we refer to as RoI, that enable the model to distinguish between the presence and absence of defects in SEM images.

The methodology for RoI acquisition module can be articulated through the following formula:

$$L^c_{\text{RoI}} = \text{Acti}\left(\sum_k w^c_k F^k\right), \tag{1}$$

where Acti represents the activation function, and c represents the corresponding class c. F^k represents the feature map of the k-th channel. w^c_k represents the weight for the k-th channel with respect to class c.

The calculation formula for w^c_k is as follows:

$$w^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial F^k_{ij}}, \tag{2}$$

where y^c represents the score for class c, which is the value that comes out of the network before the softmax operation is applied to compute probabilities for each class. F^k_{ij} represents the element at the i-th row and j-th column in the k-th channel of the feature map. Z represents the number of elements in the feature map.

Cropping and Augmentation Module: As demonstrated in Figure 6, the SEM images were cropped and enhanced based on the identified RoI. To preserve the complete information of defects, the cropping window was set to include the key areas. After cropping, the size of the processed images was adjusted to that of the original images to ensure consistency in image scales. The cropped images were then fed back to the key area localization network for repeated training, which enhanced the network's detection of defects in SEM image sections. Through cropping, the available data was expanded, and to some extent, the problem of imbalance between defects and backgrounds was mitigated.

3.3 Weakly Supervised Segmentation Network

The weakly supervised segmentation network adopts an encoder-decoder architecture and utilizes two types of information for attention extraction. Additionally, to maximize the use of available data, we implemented transfer learning, which involved transferring the weights from KALNet, after their initial training, to serve as the foundational weights for our segmentation network.

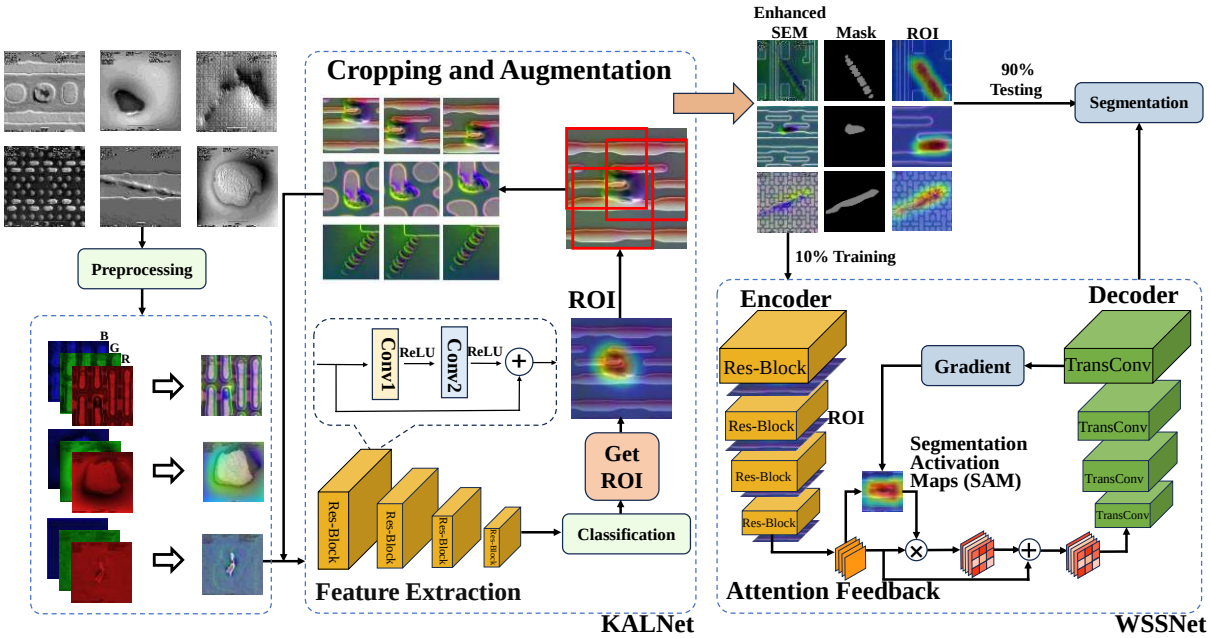


Figure 5: The proposed hybrid weakly supervised SEM image segmentation framework.

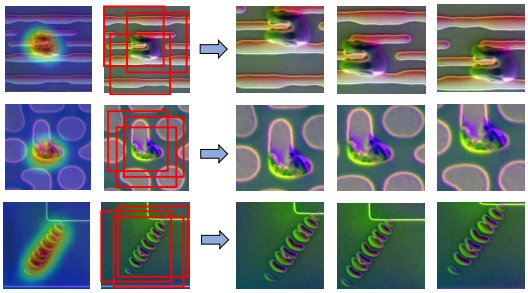


Figure 6: RoI cropping and data augmentation.

Encoder: The encoder utilizes ResNet34 as the backbone network and incorporates the RoIs obtained from KALNet into each convolutional block through residual connections. By applying weighted summation of the feature maps with RoIs, the encoder strengthens its ability to extract defect features by emphasizing attention to them. The methodology for encoder can be expressed as:

$$F \otimes K(j,k) = \sum_{i=1}^{c} \sum_{m_0=1}^{m} \sum_{n_0=1}^{m} F(i, j-m_0, k-n_0) K(m_0, n_0) + H(j,k), \tag{3}$$

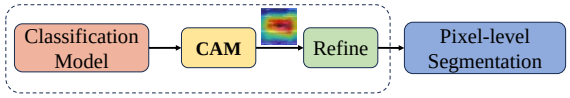
where $K(j,k)$ represents the value at position (j,k) in the output feature map of the convolution operation. $F(i, j-m_0, k-n_0)$ represents the value at position $(i, j-m_0, k-n_0)$ in the input feature map, $K(m_0, n_0)$ represents the weight value of the convolution kernel at position (m_0, n_0), $H(j,k)$ represents the value of the residual connection, specifically referring to the value of the RoI image at position (j,k).

Attention Feedback Module: In the mainstream weakly supervised approaches, based on the information learned from the classification model, the CAM is utilized to generate seed areas for each training image [18]. Subsequently, the semantic information is propagated from these seed areas to the entire image to generate pixel-level labels. However, solely relying on the information from the classification network is insufficient to bridge the gap between image-level and pixel-level labels in our task.

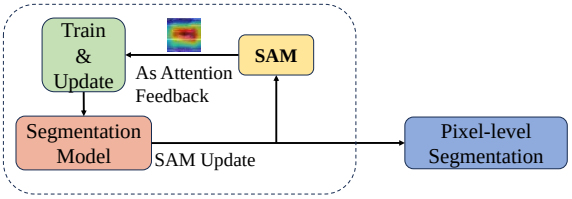
We have bridged this gap by leveraging a small number of pixel-level segmentation labels. Through the utilization of this sparse dataset, we have obtained segmentation activation maps (SAM) based on the gradient information obtained through backpropagation in the segmentation network. Figure 7 illustrates a comparison between the mainstream image-level weakly supervised frameworks and our framework. The most significant difference from mainstream frameworks is that we incorporate SAM into the training process, enabling the activation regions to evolve with network iterations. In each training iteration, SAM is employed as an attention weight, which is multiplied by and added to the last feature map of the encoder. Through this approach, we feed back these activation regions to the network, providing the network with a self-explanation of its own behavior. This process forms a "self-adjusting" mechanism to enhance the network's focus on the actual defects.

Despite the limited usage of segmentation labels during training, these labeled data play a crucial role in correcting the attention regions of the activation responses. By effectively utilizing the scarce segmentation labels, reliable supervision with semantic affinity can be achieved, leading to accurate segmentation results even in scenarios with extremely limited data and weak supervision.

Decoder: The decoder employs the method of transposed convolution (TransConv) to upsample and gradually restore the resolution of the image. Compared to other common upsampling methods, transposed convolution upsampling can learn more complex feature representations. After the upsampling, feature maps of the same size from the encoder are directly concatenated to the corresponding transposed convolutional layers in the decoder, in order to convey more detailed and comprehensive information through the decoder. This skip-connection approach enhances feature extraction in the



(a) The linear structure of mainstream weakly supervised networks



(b) The cyclical structure of our network

Figure 7: The distinctions between our network and mainstream weakly supervised networks.

network. Upon concatenation, the new feature maps continue transposed convolution to further restore the resolution by a factor of two. This process is repeated multiple times until the image is restored to its original resolution. Finally, the number of channels is reduced to 1 and the probability of each pixel being a defect is determined using the sigmoid function.

4 EXPERIMENTAL RESULTS

In this study, our hybrid weakly supervised SEM image segmentation network is evaluated on the newly collected dataset, which is further split into training and testing sets. The partition ratios of training or testing sets are varied with the supervision manners. Models are trained using PyTorch 2.0.1 with an Nvidia GeForce RTX4090.

4.1 Evaluation Metrics

IoU (Intersection over Union) is a metric evaluating the overlap between predicted and ground truth bounding boxes in object segmentation. Precision measures the proportion of true results, both *true positives* (TP) and *true negatives* (TN), among the total number of pixels examined. Recall measures the proportion of actual positive pixels that were correctly identified by the model. The F1 score is the harmonic mean of precision and recall, which effectively processes category imbalance issues arising from the disproportionate representation of defect and background areas. The formulations are shown in Equation (4).

$$\begin{aligned} Recall &= \frac{TP}{TP+FN}, \\ Precision &= \frac{TP}{TP+FP}, \\ F1 &= 2 \times \frac{Precision \times Recall}{Precision + Recall}, \end{aligned} \tag{4}$$

where FP and FN denote *false positives* and *false negatives*.

4.2 Ablation Studies

In semiconductor manufacturing, there are numerous types of defects that are challenging to identify due to their extremely small size, irregular shapes, or low contrast with the background. For instance, the hole defects we have collected closely resemble particle defects, making them visually indistinguishable, thus demanding a higher level of segmentation performance from the model. To profile the impacts of the two sub-networks on the SEM defect segmentation performance, we conducted ablation studies, as shown in Table 1.

Table 1: Ablation studies on network modules under fully supervised learning.

Block	IoU (%)	Recall (%)	Precision (%)	F1 (%)
Backbone	77.89	86.50	88.96	86.71
KAL	88.74	93.16	94.99	93.81
WSS	80.81	87.43	91.48	89.10
KAL+WSS	**91.08**	**94.98**	**95.63**	**95.25**

*Primary comparison is conducted based on IoU values.
All metrics range between 1 to 100, with higher values indicating better performance.

Since the standalone KALNet does not possess segmentation capabilities, we concatenated a mainstream segmentation network UNet after KALNet. To mitigate the influence of imbalanced datasets, the ablative experiments were conducted in a fully supervised manner.

The table demonstrates that both sub-networks have a positive effect on the SEM defect segmentation performance, particularly KALNet, which contributes to a performance improvement of 10.85%. It demonstrates that our KALNet can accurately locate the regions of defects, while the attention mechanism of the WSSNet can selectively direct the network's focus towards the defects. The best segmentation performance of the model is achieved when both sub-networks are used together, resulting in a 13.19% improvement. Figure 8 presents the convergence characteristics of IoU and Loss when both sub-networks are jointly utilized. It can be observed that our model demonstrates rapid convergence during the training phase, converging at approximately 150 epochs. Additionally, both the loss function and IoU curve demonstrate stable characteristics. Our model possesses the ability to quickly learn defect features and adjust corresponding parameters in actual applications of semiconductor manufacturing. The demonstrated convergence shows that our model can effectively handle complex and varying nanoscale SEM defects in practical applications.

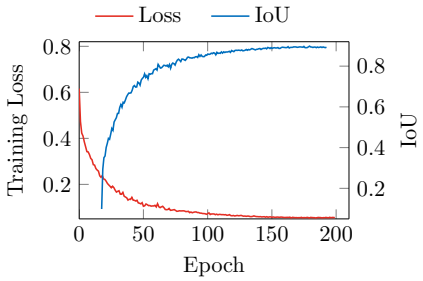


Figure 8: The convergence of IoU and Loss.

4.3 Segmentation Performance

To evaluate our proposed hybrid weakly supervised SEM image segmentation network, we compared it on two aspects. The first aspect of our evaluation involved a performance comparison with existing weakly supervised networks, CAM [19] and SC-CAM [20]. The second aspect is the comparison with fully supervised networks, where we selected SegNet [21], FCN [22], UNet [23] mentioned in wafer-related papers [24] and DeepLab v3+ [25], which performs state-of-the-art on segmentation tasks. Table 2 presents comprehensive experimental results. Our Hybrid weakly supervised network achieves an IoU of 80.70%, significantly surpassing other weakly supervised networks. Moreover, even with only 10% of the training set, it outperforms all the fully supervised models listed in the table.

Apart from the IoU, our model outperforms all fully supervised and weakly supervised models in terms of recall, precision, and F1 scores. The recall is 88.44%, which demonstrates our model exhibits

Table 2: Comparison of our model with existing fully supervised and weakly supervised models.

Model	Supervision	IoU (%)	Recall (%)	Precision (%)	F1 (%)
UNet [23]	Full	77.89	86.50	88.96	86.71
DeepLab v3+ [25]	Full	71.58	80.44	86.71	82.83
FCN-ResNet50 [22]	Full	67.09	75.28	85.62	79.57
CAM [19]	Weak	54.25	76.41	66.61	68.44
SC-CAM [20]	Weak	46.91	63.29	67.94	62.30
KAL+WSS (Ours)	Weak	**80.70**	**88.44**	**90.13**	**88.86**

Table 3: The relationship between training set proportion and the performance of our model (from 5% to 40%).

Training Set proportion	IoU (%)	Recall (%)	Precision (%)	F1 (%)
5%	75.35	84.54	87.38	85.33
10%	80.70	88.44	90.13	88.86
20%	84.34	91.48	91.45	91.16
30%	86.17	91.35	**93.76**	92.31
40%	**87.16**	**92.61**	93.61	**92.90**

a low omission rate, enabling comprehensive and sensitive detection of defects. The precision, which is 90.13%, signifies that our model demonstrates the capability to precisely separate pixels that pertain to defects in SEM images. It also implies that our model exhibits a lower rate of false alarms, which allows engineers to focus more attentively on genuine defects in practical applications. Additionally, with an F1 score of 88.86%, our model has achieved a high level of comprehensive performance in reducing the false negative rate and improving sensitivity.

To examine how the model's performance improves with increasing training data, Table 3 demonstrates the segmentation performance of our model under different training set ratios. Our segmentation performance demonstrates significant and rapid improvement, even with a relatively small dataset. This observation underscores the efficiency and robustness of our model in learning from limited labeled data, which suggests that the labor and time-intensive process of data annotation can be drastically reduced without compromising on the accuracy of defect segmentation. This will save engineers a significant amount of time in analyzing defects, thus improving the efficiency of semiconductor manufacturing, reducing production costs, and ultimately increasing yield.

4.4 Flexibility

Table 4 presents the performance improvement of existing segmentation networks achieved by KALNet. All networks in the table exhibit an IoU improvement exceeding 10%, highlighting the impressive performance of KALNet. This capability enhances the segmentation performance of existing networks for intricate and challenging defects, further emphasizing KALNet's exceptional defect localization ability. The ability of KALNet to freely combine with other networks allows for enhanced flexibility and scalability by leveraging the distinctive features and advantages of different networks in defect feature extraction. This capability becomes particularly valuable when dealing with various types of defects or addressing different application scenarios in semiconductor defect analysis. Through the integration of multiple models, KALNet can cater to specific defect analysis requirements.

5 CONCLUSIONS

In this study, we introduce a novel hybrid weakly supervised SEM defect segmentation system, leveraging deep learning for unparalleled pixel-level accuracy with strategic utilization of limited labelled data. This approach significantly reduces manual labeling,

Table 4: The integration of our KALNet with existing fully supervised models.

Mode	IoU (%)	Recall (%)	Precision (%)	F1 (%)	ΔIoU(%)
UNet [23]	77.89	86.50	88.96	86.71	10.85
KAL+UNet	**88.74**	93.16	94.99	**93.81**	
DeepLab v3+ [25]	71.58	80.44	86.71	82.83	13.80
KAL+DeepLab v3+	85.38	**94.92**	89.43	91.87	
FCN-ResNet50 [22]	67.09	75.28	85.62	79.57	16.81
KAL+FCN-ResNet50	83.90	87.87	94.72	90.96	
FCN-ResNet101 [22]	66.43	74.22	87.23	78.46	17.63
KAL+FCN-ResNet101	84.06	87.60	**95.22**	90.96	
SegNet [21]	64.95	75.44	82.02	77.18	**17.70**
KAL+SegNet	82.65	88.69	91.99	90.07	

improving efficiency in semiconductor defect analysis. Future work will expand our dataset and enhance the framework, aiming for minimum manual intervention, marking a shift toward automated defect analysis in semiconductor manufacturing.

ACKNOWLEDGMENTS

This work was supported in part by Zhejiang Provincial R&D Key Project (Grant No. 2024SJCZX0031) and NSFC (Grant No. 62034007).

REFERENCES

[1] Michael Quirk et al. *Semiconductor manufacturing technology*, volume 1. Prentice Hall Upper Saddle River, NJ, 2001.
[2] Shu-Kai S Fan et al. Key parameter identification and defective wafer detection of semiconductor manufacturing processes using image processing techniques. *IEEE TSM*, 2019.
[3] Edmund G Seebauer et al. Trends in semiconductor defect engineering at the nanoscale. *Materials Science and Engineering: R: Reports*, 70(3-6):151–168, 2010.
[4] Dan Malinaric et al. Case study for root cause analysis of yield problems. In *2000 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop. ASMC 2000 (Cat. No. 00CH37072)*, pages 8–13. IEEE, 2000.
[5] Thibault Lechien et al. Automated semiconductor defect inspection in scanning electron microscope images: a systematic review. *arXiv:2308.08376*, 2023.
[6] Hao Geng et al. Mixed-type wafer failure pattern recognition. In *Proc. ASPDAC*, 2023.
[7] FL Chen et al. Logic product yield analysis by wafer bin map pattern recognition supervised neural network. In *IEEE Int. Symp. Semicond. Manuf. Conf. Proc.*, 2003.
[8] Chung-Chi Huang et al. Study on machine learning based intelligent defect detection system. In *MATEC Web Conf.*, 2018.
[9] Jianbo Yu et al. Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map. *Comput. Ind.*, 2019.
[10] Tim Houben et al. Depth estimation from a single sem image using pixel-wise fine-tuning with multimodal data. *Machine Vision and Applications*, 2022.
[11] Wei Shen et al. A survey on label-efficient deep image segmentation: Bridging the gap between weak supervision and dense prediction. *IEEE TPAMI*, 2023.
[12] Shuanlong Niu et al. Defectgan: Weakly-supervised defect detection using generative adversarial network. In *Proc. CASE*, pages 127–132. IEEE, 2019.
[13] Qijing Wang et al. Waferhsl: Wafer failure pattern classification with efficient human-like staged learning. In *Proc. ICCAD*, pages 1–8, 2022.
[14] Deepak Pathak et al. Constrained convolutional neural networks for weakly supervised segmentation. In *Proc. ICCV*, pages 1796–1804, 2015.
[15] Emily Denton et al. Semi-supervised learning with context-conditional generative adversarial networks. *arXiv preprint arXiv:1611.06430*, 2016.
[16] Xinyu Fu et al. Region-guided pixel-level label generation for weakly supervised semantic segmentation. In *Proc. ICCV*, pages 1–6, 2021.
[17] Ramprasaath R Selvaraju et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proc. ICCV*, 2017.
[18] Bolei Zhou et al. Learning deep features for discriminative localization. In *Proc. CVPR*, 2016.
[19] Bolei Zhou et al. Learning deep features for discriminative localization. In *Proc. CVPR*, 2016.
[20] Yunchao Wei et al. Revisiting dilated convolution: A simple approach for weakly- and semi-supervised semantic segmentation. In *Proc. CVPR*, 2018.
[21] Vijay Badrinarayanan et al. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE TPAMI*, 2017.
[22] Patrick Brandao et al. Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks. *Journal of Medical Robotics Research*, 2018.
[23] Olaf Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *Proc. MICCAI*, 2015.
[24] Hao Geng et al. Mixed-type wafer failure pattern recognition. In *Proc. ASPDAC*, 2023.
[25] Liang-Chieh Chen et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proc. ECCV*, 2018.