RA-UNet: A New Deep Learning Segmentation Method for Semiconductor Wafer Defect Analysis on Fine-grained Scanning Electron Microscope (SEM) Images

Yibo Qiao, Zhouzhouzhou Mei, Yuening Luo, and Yining Chen

***Abstract*— Given the increasing complexity in the modern semiconductor integrated circuit manufacturing process, a variety of defects may occur in each process step, which would eventually lead to loss in wafer yield. The complication and heterogeneity of defect morphologies add great challenges for in root cause analysis. In addition, because the defect data of scanning electron microscope(SEM) images is not easy to obtain, there is no public large data set, which brings difficulties to the training of the algorithm. In this study, we introduce a novel UNet architecture that integrates deep residual networks and an attention mechanism for the segmentation of wafer SEM images. The proposed methodology adopts an encoder-decoder structure, and adds an intermediate attention module (IAM) to enhance features using residual attention mask blocks (RAMBs). To validate the efficacy of the proposed RA-UNet model, a real dataset of defect SEM images in a foundry was manually collected and labeled. The results demonstrate that the proposed model achieves an Intersection over Union (IoU) of 71.11%, providing empirical evidence for the effectiveness of the segmentation approach in the analysis of wafer defect SEM images.**

***Index Terms*— Convolutional neural network(CNN), computer vision, deep learning, defect segmentation, scanning electron micro-scope (SEM), semiconductor manufacturing.**

I. Introduction

WAFER defect detection has garnered significant attention in the field of yield prediction, and has been widely explored by scholars. Integrated circuit manufacturing involves complex technological processes, such as thin film deposition, ion implantation, etching, and polishing. With the continual shrinking of chip feature sizes, circuits now contain more layers, and each chip undergoes hundreds of processing steps before delivery. Any deviation from the normal manufacturing process may cause the generation of wafer surface defects [1], [2]. Therefore, precise identification of these defects is critical to enable engineers to locate any abnormal equipment or operations, and ultimately improve production yield by making timely adjustments.

The authors are with the College of Micro and Nano Electronics, Zhejiang University, Hangzhou 310058, China (e-mail:22141056@zju.edu.cn;22241039@zju.edu.cn; 22241008@zju.edu.cn;yining.chen@zju.edu.cn)(Corresponding author: Yining Chen)

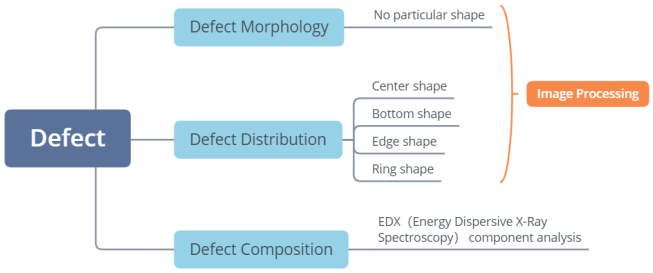

Fig. 1. Three directions of wafer defect detection.

Fig.1 illustrates the three approaches for wafer defect detection: defect distribution, defect morphology, and defect composition. Computer vision technology can be used for the automated detection of the distribution and morphology of defects. Defects typically exhibit a specific distribution pattern, such as central, bottom, edge, circular, and so on. This pattern is often indicative of a particular process flow, which can be used to identify potential sources of defects. In recent years, various machine learning and deep learning methods have been heavily applied in the detection of wafer map defect distribution [3]–[5].

Fig.2 illustrates the connection between the defect distribution and the defect morphology. Once the range of process steps that contribute to defects has been determined based on the distribution pattern, yield engineers can gradually narrow down this range by analyzing the morphology of the defects. This analysis enables the engineers to identify the specific process that is causing the defects, and optimize it to enhance the yield. As such, the morphological analysis of defects plays a crucial role in the root cause analysis of defects, and in improving the overall yield.

Most of the current wafer defect studies are qualitative in nature, with a focus on the types of defect aggregation shapes. However, since the process of defect analysis is complex and requires gathering as much information as possible to determine the root cause, quantitative analysis of defects - such as measuring the area of a single defect - is also crucial. The scanning electron microscope(SEM) map of the defect

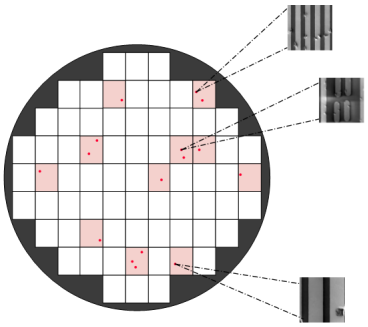

Fig. 2. Relationship between defect distribution and defect morphology. In the figure, the circle represents the entire wafer, the rectangle represents a single chip, and each point corresponds to a defect. The SEM image of each defect can be captured by photographing it separately, allowing for observation of its morphology.

can be segmented to enable quantitative analysis of defect morphology.

Despite extensive research on segmentation in the field of computer vision, segmenting SEM images of wafer defects remains challenging due to certain properties of these images. One of the main difficulties is the limited availability of datasets as the size of defect areas and back-grounds vary significantly, resulting in imbalanced distribution of positive (defective) and negative (non-defective) image samples. This issue is more pronounced in mature processes where the majority of image samples do not contain defects. In general, the challenges associated with wafer defect morphology segmentation include a small number of samples, an imbalance in front and background (defect and non-defect areas), and an imbalance of positive and negative samples.This study proposes a new method for defect morphology analysis of wafer SEM images, which is a task that has received little attention in the existing research. The proposed method, named RA-UNet, is a hybrid deep learning method based on UNet that combines a semantic segmentation network with an attention mechanism. The main contributions of this study are as follows:

1) Introducing a new task of defect morphology analysis of wafer SEM images and addressing the lack of research in this area.
2) Proposing a new network structure that combines the deep residual network and the attention mechanism with UNet. It focuses more on the features of the defective part of the image and reduce the interference of background information, thus improving the ability of UNet to extract defect morphology features.
3) Creating a new dataset of real wafer defect morphology by manually labeling the original image data collected from a wafer factory, which is used to facilitate the research of wafer defect morphology analysis.

The remainder of this paper is organized as follows.In Section II, we provide an overview of previous research on wafer defect detection and image segmentation techniques. Section III details the structure of our proposed RA-UNet network, which includes the encoder part, intermediate attention module, and decoder part. In Section IV, we present the experiments we conducted, which encompass the datasets, data augmentation and preprocessing, evaluation modalities, ablation experiments, and model fine-tuning. We also analyze the experimental results in this section. Finally, Section V provides the conclusion of this study.

II. RELATED WORKS

In this study, our related work focuses on three parts, which are wafer defect analysis, image segmentation, and wafer defect data set.

A. Wafer Defect Analysis

In the past, wafer surface inspection relied on manual visual methods due to technical limitations. Sampling inspection was also commonly used to reduce labor costs, but this method was slow and had low accuracy. With the introduction of various testing instruments, defect detection became possible using electromagnetic induction principles, such as eddy current and magnetic flux leakage detection [6], [7]. However, these methods were limited to detecting defects in conductive materials, and many steps in the wafer manufacturing process involve nonconductive materials.

In recent years, there has been a shift towards intelligent methods in wafer defect detection with the emergence of computer vision technology. Various machine learning models, including Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and autoencoder, have been applied to this field. Chen et al. (2003) used supervised learning to develop an online wafer bin map pattern recognition system, which solved the problems of product dependence and excessive derived patterns in unsupervised learning [3]. Huang et al. (2019) proposed an intelligent defect detection system for metal products based on machine learning, using generative adversarial networks (GAN) to address the issue of insufficient surface defect datasets and achieve the detection of common stains, edge defects, scratches, and more [4]. Yu developed a hybrid deep learning model by proposing a stacked convolutional sparse denoising auto-encoder (SCSDAE) that integrates CNN with SDAE, allowing for the accumulation of robustness layer by layer and effective feature learning [5].The majority of prior research on wafer defects has focused on analyzing the location distribution and classification of defect image patterns, treating each defect as a point on a map and analyzing the extent of defect aggregation. How-ever, in practical engineering applications, the process problems associated with these image patterns are often easily identified by engineers, and thus this approach has limitations in fully addressing the needs of wafer defect analysis. To address this issue, this study places emphasis on defect morphology analysis, which is the crucial step in identifying the root cause of defect formation and typically requires significant time and effort from engineers. Specifically, this study aims to analyze the topography of individual defects, which represents the next step in defect distribution analysis beyond location, and has received relatively little attention in previous research.

B. Image Segmentation

In recent years, there have been significant advancements in image segmentation technology based on deep learning. Long et al. proposed the first method, FCN, to use a classification network for segmentation tasks and demonstrated that the problem of image segmentation can be trained end-to-end. FCN replaced the fully connected layer in the classification network with a convolutional layer and a pooling layer, allowing the network to adapt to pixel-level segmentation tasks [8]. Since then, many new structures in the field of image semantic segmentation have been based on improved versions of FCN. In 2015, UNet, which modified and extended the FCN architecture, was proposed. This network has an encoder-decoder structure and can achieve very accurate segmentation results with very few training datasets [9]. Following this, many studies have improved the UNet architecture, such as UNET++ [10], UNet3+ [11], U2Net [12], and CE-Net [13].The improvements made to UNet have been successful in achieving satisfactory segmentation results in medical images. However, industrial defects share similar characteristics with medical images, such as low contrast and blurred boundaries. Nastaran et al. proposed a fully convolutional neural network based on UNet for automatic defect detection of industrial surfaces [14]. However, compared to other industrial defects, wafer defects are more complex in their causes and have various defect types that are multiscale. A new network is needed to address these characteristics and obtain more accurate defect segmentation for wafers.

C. Wafer Defect Data Set

At present, the only public large data set related to wafer defect analysis is WM-811k, which contains 811,457 crystal circle distribution patterns. Almost all the research on wafer defect distribution is verified on this data set [15], [16]. However, this data set can only be used to study the location distribution of defects in the entire wafer map. More specific defect information, such as size and shape, cannot be obtained from this data set. There is no common data set for the SEM image of wafer defects, which makes the study of defect morphology much more difficult than that of defect distribution.

In order to facilitate the analysis of defect morphology and shorten the time to find the root cause of defects, we collected and manually labeled images to construct the corresponding SEM image data set of wafers. Having a dedicated data set for SEM images of wafer defects can also help to standardize the evaluation of different defect detection models.

III. METHODOLOGY

Due to the complexity of the causes of defects in the IC manufacturing process, there is limited SEM image data available for each specific defect cause. To maximize the use of the available data set, various image segmentation methods were reviewed, and the jump connection and U-shaped structure in the UNet network were chosen as the basis for improvement. This modified approach was de-signed to fully extract the semantic information of the defect area and improve the segmentation accuracy.

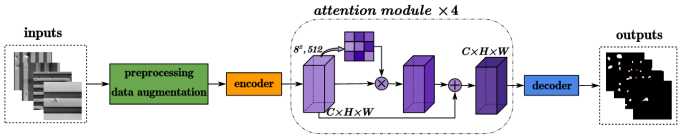

Fig. 3. System diagram of wafer defect segmentation: data augmentation, preprocessing, encoding, intermediate attention module, decoding.The output map has two labels: foreground (filled in white) and background (filled in black).

TABLE I
RA-UNET CONFIGURATION("CONV" AND "CONCAT" DENOTE CONVOLUTION AND CONCATENATION, RESPECTIVELY)

Layer	Outputsize	Operation
Conv0_1	128 ×128 × 64	Conv 3×3, BN, $RELU$
Conv1_1	64 ×64 × 64	Res block ×3
Conv2_1	32 ×32 × 128	Res block ×4
Conv3_1	16 ×16 × 256	Res block ×6
Conv4_1	8 ×8 × 512	Res block ×3
Conv4_2	8 ×8 × 512	Attention block ×4
Conv3_2	16 ×16 × 256	VGG block ×1
Conv2_2	32 ×32 × 128	VGG block ×1
Conv1_2	64 ×64 × 64	VGG block ×1
Conv0_2	128 ×128 × 64	VGG block ×1
Conv0_3	128 ×128 × 1	Conv 1 ×1

Fig.3 shows the flow chart of the system. The pro-posed RA-UNet network in this study adopts an encoder-decoder architecture, which can be divided into three parts: encoder, intermediate attention module, and decoder. The architecture of the system is shown in Fig. 4. The encoder is responsible for feature extraction, the intermediate attention module(IAM) containing the residual attention mask blocks(RAMBs) recalibrates the extracted features to focus more on the defect part, and the decoder restores the image size for the fusion of deep and shallow features. The output is a single-channel image, where each pixel block represents the probability that the pixel belongs to either the defect or background category, allowing for the segmentation of these two classes. Table I shows the configuration of RA-UNet. The following sections explain the details of each step.

A. Encoder

This study employs ResNet34 as the encoder. The network has skip connections to address gradient issues as it deepens [17]. The encoder has two types of residual blocks with short-circuit connections: Identity Block (same input and output channels) and Conv Block (different input and output channels). The Identity Block is used for feature extraction when the number of channels is matched. The Conv Block adds a convolutional layer to modify the channel count before using the Identity Block.

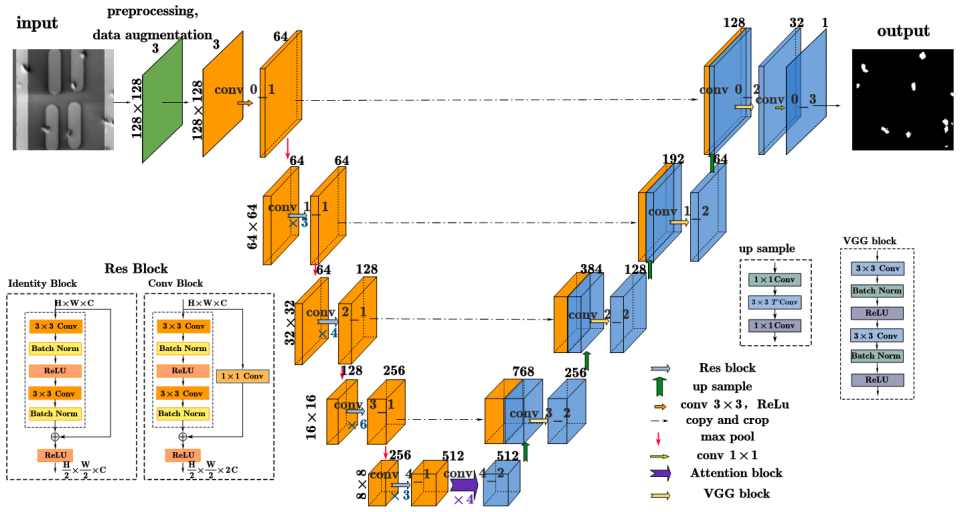

Fig. 4. Structural diagram of the proposed RA-UNet model.The network uses ResNet for feature extraction, and the attention module is added between the encoder and decoder of UNet, with the activation function using FRelu.

B. Intermediate Attention Module

The convolution operation aggregates spatial and channel information of the features without distinction, but the relevance of these types of information is not necessarily equal. To recalibrate the importance of the channel information, a Squeeze-and-Excitation Block (SE) was proposed in the literature [18]. As depicted in Fig. 5-(a), this block first generates a $1\times1\times$C feature channel descriptor and multiplies it with the original feature map channel-wise to recalibrate the channel. This allows the amplification of important channel information. However, this study prioritizes the spatial information of features in SEM defect segmentation. Therefore, the spatial weight of each channel of the feature map is recalibrated to amplify important spatial information. The proposed RA-UNet structure integrates the Residual Attention Network method from [19] as the Residual Attention Mask Block (RAMB) in the intermediate attention module between the encoder and decoder, building on the idea of attention. The specifics of this integration are illustrated in Fig. 5-(b).

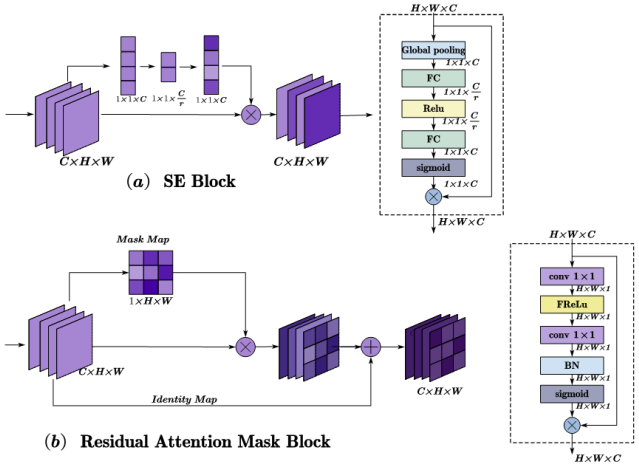

Fig. 5. Squeeze-and-Excitation Block (SE block) vs. RAMB block. (SE focused on weight recalibration of different channels, RAMB focused on spatial weight recalibration of the same channel.)

1) Coefficient of Attention: As illustrated in Fig.5- (b), the attention module initially applies a 1×1 convolution to reduce the number of feature channels in the attention branch to 1. Then, the convolution is used with FRelu activation to extract information and obtain the Mask map of the position information in the feature map. The Sigmoid function is used to obtain the attention coefficient, which has a value range of 0 to 1. Finally, the attention coefficient from the Mask map is multiplied with the original feature map to eliminate unimportant information, and the new feature is obtained by enhancing the constraint according to the feature spatial location information. This method effectively improves the information expression of the defect area and suppresses the irrelevant background area.

2) Residual Connection: Since the parameters in the feature map are weighted from 0 to 1, this operation will lead to the reduction of feature parameter values. In order to avoid it, it is important to consider adding an identity mapping branch to the attention module. By doing so, the introduction of the attention module will not compromise the accuracy of the network.

3) Activation function—Frelu: In order to effectively leverage the global context of an image, the activation function of the attention module is FRelu (Funnel activation function), which extends ReLU and PReLU to 2D activation functions.

Compared to other methods, such as extended convolution [20], STN [21], active convolution [22], demorphable convolution [23], FRelu incurs minimal additional memory overhead. FRelu can be expressed as:

$$f(x_{c,i,j}) = max(x_{c,i,j}, T(x_{c,i,j})) \tag{1}$$

$$T(x_{c,i,j}) = x^{w}_{c,i,j} \cdot p^{w}_{c} \tag{2}$$

C. Decoder

In this study, deconvolution is used to gradually restore the image resolution. Compared to other common upsampling methods such as bilinear [24] and Unpooling [25], deconvolution upsampling [26] can learn more complex feature representations. During the restoration process, channel concatenating

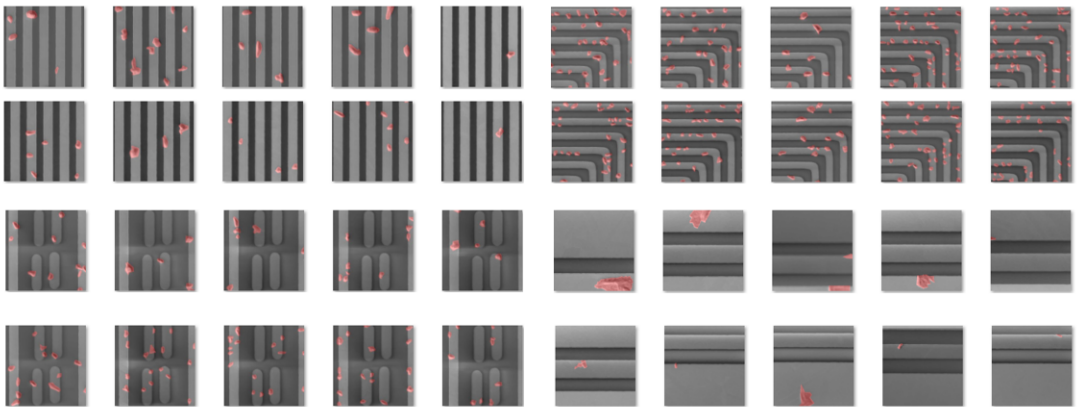

Fig. 6. Example wafer defect SEM plot dataset. (The images were manually annotated using "Labelme". The red area in the image represents the foreground, which corresponds to the undesirable part in the manufacturing process and serves as the defect part in our study.)

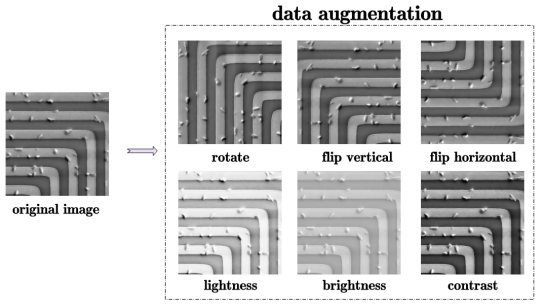

Fig. 7. Example image of data augmentation.

is performed to connect feature channels of the same size from the encoder part. This skip-level connection refines the features extracted by the network [27]. After the concatenating, the new feature map undergoes a VGG convolution operation [28], and deconvolution continues to restore double the resolution. This process is repeated multiple times until the image is restored to the initial resolution. The number of channels is then compressed to 1, and the sigmoid function of formula (2-3) is used to determine the probability of each pixel being a defect.

$$Sigmoid(x) = \frac{1}{(1+e^{-x})} \tag{3}$$

D. Loss Function

For defect segmentation, the problem of unbalanced foreground and background, and unbalanced negative and positive samples needs to be addressed. Therefore, this study uses a combined loss function of binary cross-entropy (BCE) and Dice Loss to tackle the imbalance issue.

$$BCELoss = -(ylog(p(x) + (1-y)log(1-p(x)) \tag{4}$$

$$DiceLoss = 1 - \frac{2TP}{(FP+2TP+FN)} \tag{5}$$

$$BCE\_DiceLoss = \lambda BCELoss + \mu DiceLoss \tag{6}$$

TABLE II
CONFUSION MATRIX FOR DEFECT SEGMENTATION

Prediction/reality	Positive example (defect)	Negative example (background)
Positive example (defect)	TP	FP
Negative example (background)	FN	TN

The combined loss function is represented in (6), where BCE loss, which is represented in (4), avoids the gradient disappearance by making the gradient independent of the sigmoid function during backpropagation.The formula for Dice Loss is shown in (5), where the meanings of TP, FPand FN are shown in Table II. Dice loss is a loss function that was originally proposed for the imbalance of positive and negative samples in semantic segmentation [29].

This study combines BCE Loss and Dice Loss by adjusting the combination ratio of the two, which is the ratio of λ to μ, as discussed in Section IV.

IV. Experimental Analysis

In this section, we trained the RA-UNet network structure (shown in Fig.3) on the SEM data set of metal layers collected from a wafer factory, selected the number of RAMBs through experiments,and compared our proposed method with other existing methods, and then conducted ablation experiments, verified the effectiveness of the ResNet encoder and intermediate attention module proposed in this study. Finally, we adjusted the model parameters, improved the accuracy of defect segmentation.

A. Data Set

We collected 1000 metal layer defect maps from a 12-inch 55nm wafer factory, and manually labeled them. Each image has a resolution of 480 $\times$ 480, and the ratio of training set to test set is 8:2. We ensured that the test set does not overlap with the training set to guarantee accurate test results and better evaluation of the model's generalization ability. As shown in

TABLE III
COMPARISON OF RESULTS WITH AND WITHOUT PREPROCESSING

Method	Preprocessing	F1	IoU	Difference Value
UNet	×	80.80%	67.91%	0.13%
	✓	80.92%	68.04%	
RA-UNet	×	81.87%	69.37%	0.26%
	✓	82.05%	69.63%	

Fig.6, in each SEM image, the regular rectangular strip part represents the metal line, while the irregular part under the raised or concave region represents the defect.

B. Data Augmentation and Preprocessing

To augment our data set and improve the robustness of our neural network training model, we followed a three-step process. Firstly, we randomly rotated the images by 90°, 180°, or 270°. Next, we applied a random flip to the images, either horizontally or vertically. Finally, we randomly selected one operation from lightness, brightness, or contrast, according to the normalized probability. These operations were chosen to preserve the global texture of the original image and generate a large number of new SEM graphs. In the preprocessing stage, we performed median filtering on the images obtained after data augmentation to reduce image noise and facilitate subsequent feature extraction. An example image of data augmentation is shown in Fig.7.

C. Evaluation index

We used Recall, F1, Accuracy, IoU to measure the performance of the proposed model. Recall indicates the probability of pixels belonging to defects be correctly segmented; Accuracy represents the probability of defect pixels be correctly classified among the pixels predicted as defects, that is, the proportion of all correctly segmented defect pixels to all pixels predicted as defects. Due to the contradiction between Recall and Accuracy sometimes, in order to better evaluate the segmentation effect, we hope to take into account both Accuracy and Recall, so we add F1 evaluation index, F1 is the harmonic mean of recall and Accuracy. IoU is a measure of the overlap between the predicted defect region and the true defect region, calcu-lated as the ratio of their intersection to their union.. The corresponding definitions are as follows:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Accuracy = \frac{TP}{TP + FP} \tag{8}$$

$$F1 = \frac{2Recall \times Accuracy}{Recall + Accuracy} \tag{9}$$

$$IoU = \frac{TP}{TP + FN + FP} \tag{10}$$

D. Experimental Results

This experiment used NVIDIA GeForce GTX1650, Windows 11 operating system, and the network model was coded in Python 3.8, using PyTorch as our deep learning framework. The network used the SGD optimizer with 300 iterations, an initial learning rate of 0.001, weight decay of 0.0001, and momentum of 0.9.

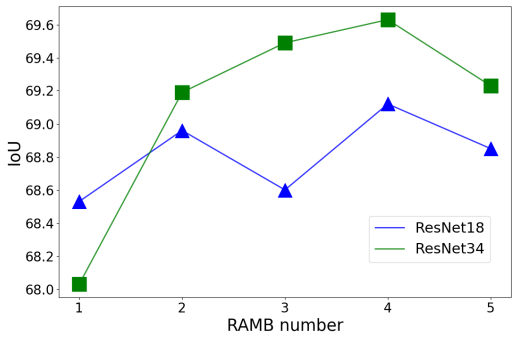

Fig. 8. IoU comparison of different ResNets with different numbers of RAMBs.

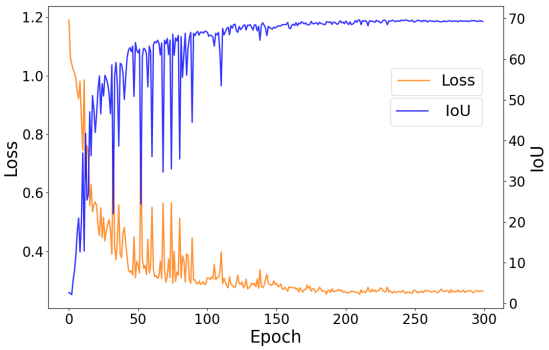

Fig. 9. Variation of Loss and IoU on the test set for RA-UNet.

To test the efficiency of the preprocessing step in improving the segmentation accuracy of metal layer defects, we compared the accuracy results of the UNet model and the proposed RA-UNet with and without preprocessing. The results are shown in Table III. Both the UNet model and the RA-UNet model perform better after preprocessing, achieving a 0.13% improvement and 0.26% improvement, respectively, compared to the results without preprocessing. In other words, the preprocessing step can enhance the metal layer defect segmentation results of the two models studied.

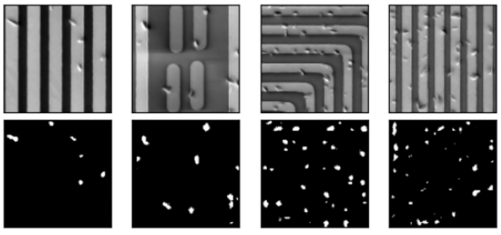

Fig. 10. Segmentation result: the first row is the original image to be seg-mented, the second row is the segmentation defect map. (white part is the defect, black part is the normal background.)

The proposed model used either ResNet18 or ResNet34 as the encoder, and the number of RAMBs was varied from 1 to 5.The overall trend of the results is shown in Fig.8. Although

TABLE IV
IOU COMPARISON OF DIFFERENT RESNETS WITH DIFFERENT NUMBERS OF RAMBS

Method	Encoder	Numbers of RAMBs	IoU
RA-UNet	ResNet18	1	68.53%
RA-UNet	ResNet18	2	68.96%
RA-UNet	ResNet18	3	68.60%
RA-UNet	ResNet18	4	69.12%
RA-UNet	ResNet18	5	68.85%
RA-UNet	ResNet34	1	68.03%
RA-UNet	ResNet34	2	69.19%
RA-UNet	ResNet34	3	69.49%
RA-UNet	**ResNet34**	**4**	**69.63%**
RA-UNet	ResNet34	5	69.23%

TABLE V
COMPARISON RESULTS BETWEEN RA-UNET AND OTHER METHODS

Method	Tra_IoU	Val_IoU	Accuracy
SegNet	53.40%	52.64%	69.67%
Dlinknet	63.83%	49.77%	76.64%
Deeplab-v3+	69.34%	63.64%	79.01%
RA-UNet	**73.85%**	**69.63%**	**82.38%**

the IoU was lower for ResNet34 with 1 RAMB compared to ResNet18, the overall performance trend of ResNet34 is better than that of ResNet18. Therefore, ResNet34 was selected as the final encoder feature extraction model in this study. Both ResNet18 and ResNet34 had the highest IoU with 4 RAMBs, so the number of RAMBs adopted in this study was 4. The specific IoU values are presented in Table IV, with the best results highlighted in bold.

Fig.9 shows the changes of Loss and IoU of RA-UNet on the test set. The experiment shows that the evaluation loss of RA-UNet converges after about 150 epochs, and the visualization results of model evaluation are shown in Fig.10. The proposed model can accurately identify whether there is a defect on the surface of the object and segment the defect region out.

Table V presents the comparison results between RA-UNet and other methods, and our proposed model performs the best in each criterion. It is evident that RA-UNet has superior feature learning capacity compared to other methods.

E. Ablation Experiment

In this study, we proposed two improved ideas to improve the UNet model, and in order to demonstrate the effectiveness of these two ideas in the application of wafer SEM graphs segmentation, we used ablation experiments. We compared the performance of five models: original UNet (using VGG encoder), Res-UNet (UNet with ResNet encoder), A-UNet (UNet with attention module), RA-UNet (UNet with ResNet encoder), and RES-UNet (UNet with ResNet encoder). The results, depicted in Fig.11, demonstrate the effectiveness of the RA-UNet, as its IoU was very stable and had the highest value.

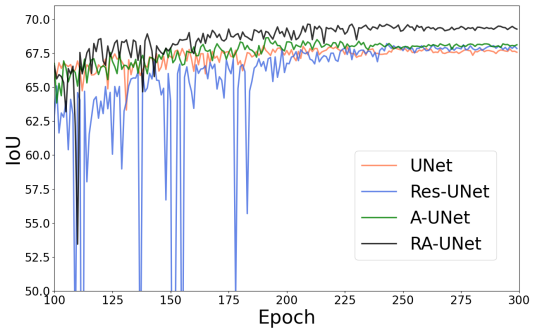

Fig. 11. IoU curves of different models in ablation experiment.

TABLE VI
COMPARISON OF ABLATION EXPERIMENT EVALUATION INDEX RESULTS

Method	ResNet Encoder	Attention Module	Accuracy	F1	Tra_IoU	Val_IoU	Difference Value
UNet	×	×	80.19%	80.92%	75.46%	68.04%	7.42%
Res-UNet	✓	×	80.60%	81.07%	70.59%	68.20%	2.39%
A-UNet	×	✓	82.33%	81.36%	74.73%	68.63%	6.10%
RA-UNet	✓	✓	82.38%	82.05%	73.85%	69.63%	4.22%

TABLE VII
FRELU ABLATION EXPERIMENTS

Activation Function	Accuracy	F1	IoU
FRelu	82.38%	82.05%	69.63%
Relu	81.94%	81.85%	69.40%

The results of the specific evaluation index are shown in Table VI. In comparison to UNet, Res-UNet showed an increase in IoU from 68.04% to 68.20%. This is because the ResNet has deeper layers, providing better feature extraction compared to UNet's encoder. In the case of A-UNet, the feature map with defect information was given more weight in the mask map obtained by the attention module. The IoU increased from 68.04% to 68.63%. The results showed that the A-UNet model, which only introduced the attention module, performed better than the Res-UNet model, which only introduced the ResNet encoder. This suggests that the attention module plays a more critical role in the network compared to the ResNet encoder.In addition,we compared FRelu with the traditional Relu activation function in Table VII. Using FRelu increased IoU by 0.23%.

By comparing the performance of our models on both the training and test sets, we observed that UNet suffered from severe overfitting. Res-UNet showed the best performance in reducing overfitting, while A-UNet achieved good segmentation results on both training and test sets. This suggests that the ResNet encoder can effectively reduce overfitting, while the attention module can greatly improve segmentation performance. The proposed RA-UNet model, which combines both of them, not only improves segmentation performance but also effectively reduces overfitting.

TABLE VIII
EFFECT OF BATCH SIZE AND THE PROPORTION OF COMBINED LOSS FUNCTIONS ON IOU

Method	Batch size	IoU	$\lambda : \mu$	IoU
RA-UNet(Res34_RAMB4)	4	69.27%	2:3	69.63%
RA-UNet(Res34_RAMB4)	**6**	**71.11%**	**1:2**	**71.11%**
RA-UNet(Res34_RAMB4)	8	69.63%	1:1	69.40%
RA-UNet(Res34_RAMB4)	10	69.22%	3:2	69.37%

F. Model Fine-tuning

In the previous experiments, we confirmed that the encoder used ResNet34 and that the number of RAMBs was 4. In this subsection, we fine-tuned this model by changing the batch size and loss function to achieve higher IoU. Table VIII show that when Batch size=6 and $\lambda : \mu$=1:2, the model has a higher segmentation precision and the IoU increases from 69.63% to 71.11%.

V. CONCLUSION

In this work, new task of defect morphology analysis of wafer SEM images has been introduced,and the proposed SEM image segmentation method for wafer defects based on RA-UNet has been presented. The ResNet encoder has been used for feature extraction, and an attention module has been added between the encoder and the decoder to focus on the extracted features that relate to the defect area. The FRelu activation function has been used in the attention module to adaptively process complex shape layouts. The experimental results demonstrate that our proposed model has achieved a good segmentation effect on SEM images of wafer defects, with an IoU of 71.11%.

This work provides a new application of deep learning in semiconductor defect analysis. Future work includes combining defects with integrated circuit layout to improve the speed of defect root cause analysis and enhance yield improvement.

REFERENCES

[1] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 26, no. 1, pp. 131–137, 2004.

[2] D. Jiang, W. Lin, and N. Raghavan, "A novel framework for semiconductor manufacturing final test yield classification using machine learning techniques," *IEEE Access*, vol. 8, pp. 197 885–197 895, 2020.

[3] F. Chen, S.-C. Lin, K. Y.-Y. Doong, and K. Young, "Logic product yield analysis by wafer bin map pattern recognition supervised neural network," in *IEEE Int. Symp. Semicond. Manuf. Conf. Proc.*, vol. 1, pp. 501–504, 2003.

[4] C.-C. Huang and X.-P. Lin, "Study on machine learning based intelligent defect detection system," in *MATEC Web Conf.*, vol. 201, p. 01010. EDP Sciences, 2018.

[5] J. Yu, X. Zheng, and J. Liu, "Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map," *Comput. Ind.*, vol. 109, pp. 121–133, 2019.

[6] J. Liu, M. Fu, F. Liu, J. Feng, and K. Cui, "Window feature-based two-stage defect identification using magnetic flux leakage measurements," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 12–23, 2017.

[7] G. Wang, Q. Xiao, M. Guo, and J. Yang, "Optimal frequency of ac magnetic flux leakage testing for detecting defect size and orientation in thick steel plates," *IEEE Trans. Magn.*, vol. 57, no. 9, pp. 1–8, 2021.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3431–3440, 2015.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lect. Notes Comput. Sci.*, pp. 234–241. Springer, 2015.

[10] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Lect. Notes Comput. Sci.*, pp. 3–11. Springer, 2018.

[11] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP IEEE Int Conf Acoust Speech Signal Process Proc*, pp. 1055–1059. IEEE, 2020.

[12] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested u-structure for salient object detection," *Pattern Recognit.*, vol. 106, p. 107404, 2020.

[13] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[14] N. Enshaei, S. Ahmad, and F. Naderkhani, "Automated detection of textured-surface defects using unet-based semantic segmentation network," in *Proc. Annu. Conf. Progn. Health Manag. Soc., PHM*, pp. 1–5. IEEE, 2020.

[15] T. Tziolas, T. Theodosiou, K. Papageorgiou, A. Rapti, N. Dimitriou, D. Tzovaras, and E. Papageorgiou, "Wafer map defect pattern recognition using imbalanced datasets," in *Int. Conf. Inf., Intell., Syst. Appl., IISA*, pp. 1–8. IEEE, 2022.

[16] J. Yu and J. Liu, "Two-dimensional principal component analysis-based convolutional autoencoder for wafer map defect detection," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8789–8797, 2020.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, 2016.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 7132–7141, 2018.

[19] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3156–3164, 2017.

[20] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[21] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Adv. neural inf. proces. syst.*, vol. 28, 2015.

[22] Y. Jeon and J. Kim, "Active convolution: Learning the shape of convolution for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4201–4209, 2017.

[23] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 764–773, 2017.

[24] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation," in *Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit*, pp. 3126–3135, 2019.

[25] H. Lu, Y. Dai, C. Shen, and S. Xu, "Index networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 242–255, 2020.

[26] D. Im, D. Han, S. Choi, S. Kang, and H.-J. Yoo, "Dt-cnn: Dilated and transposed convolution neural network accelerator for real-time image segmentation on mobile devices," in *Proc IEEE Int Symp Circuits Syst*, pp. 1–5. IEEE, 2019.

[27] J. Lu and K.-y. Tong, "Visualized insights into the optimization landscape of fully convolutional networks," *arXiv preprint arXiv:1901.08556*, 2019.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[29] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. - Int. Conf. 3D Vis., 3DV*, pp. 565–571. IEEE, 2016.