

Deep Image Segmentation for Defect Detection in Photo-lithography Fabrication

Omari Paul*, Sakib Abrar, Richard Mu, Manar D. Samad

College of Engineering, Tennessee State University, Nashville, TN 37209 US

Abstract—Surface acoustic wave (SAW) sensors with increasingly unique and refined designed patterns are often developed using the lithographic fabrication processes. Emerging applications of SAW sensors often require the use of novel materials, which may present uncharted fabrication outcomes. The fidelity of the SAW sensor performance is often correlated with the ability to restrict the presence of defects in post-fabrication. Therefore, it is critical to have effective means to detect the presence of defects within the SAW sensor. However, due to the need for precision identification and classification of surface features for increased confidence in model accuracy, labor-intensive manual labeling is often required. One approach to automating defect detection is to leverage effective machine learning techniques to analyze and quantify defects within the SAW sensor. In this paper, we propose a machine learning approach using a deep convolutional neural network to segment surface features semantically. Experimental results demonstrate an average dice score of 0.72 (0.13) across five cross-validation folds in segmenting the defective region for a novel SAW sensor variant.

Index Terms—convolutional neural network, autoencoder, defect detection, image segmentation, photo-lithography, surface acoustic wave sensor

I. INTRODUCTION

The revolution of deep learning has enabled automated and accurate classification and pattern analysis of images in myriad research and applications. Fast and automatic detection of defects in manufacturing and fabrication is crucial to ensure large-scale industrial productivity and quality [1]. The detection of manufacturing defects is primarily posed as an image classification problem where the presence of a defect is confirmed in a given sample or product unit. Therefore, classifying defective images paves the path for sorting or correcting defective units in production. However, such a binary classification task does not identify the region of the defect or quantify the extent of the defect to assess the production quality and efficacy. In photo-lithography sensor fabrication, the manufactured sensors may contain defects at varying scales resulting in varying sensor throughput. A mere presence of a fabrication defect may not always infer defective sensors with unsatisfactory performance [2]. We assume that the region or extent of defects is proportional to the drop in the sensor performance. Therefore, segmenting the region of defects in sensor fabrication is an important step toward quantifying and localizing the defect, which cannot be achieved via traditional image classification tasks. In this paper, we propose a deep autoencoder-based image reconstruction framework to segment and localize defects in surface

acoustic wave (SAW) sensor fabrication. Unlike the defect detection performed on other sensor fabrication, this variant of SAW sensor is unique and challenging because we are developing the metallic interdigital transducing (IDT) fingers on a flexible polymer substrate instead of the conventional rigid crystalline substrate. It is important for substrates to be as flat as possible; hence, the flexibility of the substrate often conflicts with the flatness during the fabrication of the sensor. A lack of flatness is one of the leading causes of defects during lithographic fabrication [3]. This paper shows one of the first studies on localizing and segmenting defective regions on SAW sensor fabrication using a deep learning method.

The remainder of the paper is organized as follows. Section II shares a background reviewing the literature on automatic defect detection in fabrication. Section III presents the sensor fabrication steps, microscopic imaging of sensors, image processing, and deep learning-based experiments and evaluation steps. Section IV provides the key deep learning and segmentation results with some discussions. The paper concludes in Section V.

II. BACKGROUND REVIEW

This section highlights the literature on defects in photolithography and approaches to automated defect detection.

1) *Defects in sensor fabrication*: With the burgeoning demand for micro-electromechanical systems (MEMS) class devices, the surface acoustic wave (SAW) sensor has shown promise in meeting the requirements of a wide variety of applications. The applications of SAW sensors range from biosensors in the detection of harmful biological agents [4] to resonant systems that are capable of quantum mode coupling for qubit storage [5]. Many of these micro/nanoscale sensor technologies are developed using advanced lithographic techniques such as photolithography, electron-beam lithography, and ion lithography. The post-lithographic fabrication processes are often laden with subtle defects leading to degradation in sensor performance. According to Aubert et al., defects located on the surface of SAW devices of size 1–2-micron depth have been reported to produce errors in temperature measurement [6]. Floer et al. reported that SAW sensors with defective lithium niobate substrates could degrade the quality of the sensor-acquired signals [2]. The post-fabrication defects from lithography also affect integrated circuits (IC), resulting in circuit shorting hot spots [7]. These hot spots are a direct result of scratches and particles on the backside of wafers that

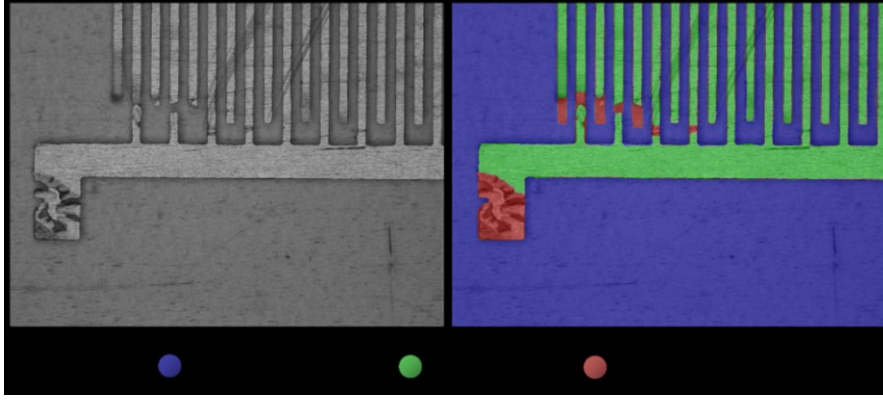


Fig. 1: Original surface acoustic wave sensor IDTs (left). Corresponding RGB mask with defective regions coded in red (right).

affect the photolithographic process by distorting the wafer's flatness during UV exposure [3].

2) *Detection of sensor fabrication defects:* The defect inspection process has been classically conducted by human surveillance, which is subject to human error. This human-level surveillance is not an efficient means for large-scale manufacturing and production due to the subtlety of defects and variability in human ratings. Specifically, as the market size, demand, and intricacies of lithographically produced technology continue to increase, the need for more effective means of inspection will inevitably follow suit. Zhang et al. report that manufacturing is becoming more challenging as transistor size continues to shrink, leading to the need for more precise and large-scale defect analysis [8]. It has also been reported that the strain on the semiconductor manufacturing sites due to a global chip shortage has led engineers to investigate methods for improving the wafer yields at the site. This improvement may be achieved by optimizing final test processes via early and accurate defect detection during manufacturing [9]. Therefore, the need for a more automated approach to detecting sensor manufacturing defects is in obvious high demand.

3) *Automating fabrication defects detection:* Advanced machine learning techniques can play an important role in automating and objectively assessing the defects in post-fabrication processes. The semiconductor industry has progressively adopted machine learning techniques to analyze integrated circuits. Torres et al. have proposed a combination of artificial neural networks (ANN) and support vector machines (SVM) to detect hot spots following high-performance lithographic processes [10]. Ye et al. have leveraged machine learning for hotspot detection in ICs and reported an improved area under the curve (AUC) in defect classification considering data imbalances [11].

One major challenge for traditional machine learning is that it requires the “hand-crafting” of image features prior to training a classifier model. In contrast, deep learning techniques are capable of learning optimal image features concurrently with the training of neural network models. Therefore, deep learning has automated the feature learning and classification

processes to replace conventional machine learning approaches in image representation learning. Cheon et al. use a convolutional neural network (CNN) architecture to replace “naked eyeballing” and “machine learning of handcrafted features” for wafer surface defect classification. They achieve this by identifying mislabeled images classified by the trained CNN to extract effective feature vectors. A k-nearest neighbor (k-NN) classifier algorithm is then used to group the defect classes by measuring the total squared distance between each image feature vector and its k-nearest neighbor in the same class. If the squared distance between the image feature vector and cluster exceeds a certain threshold, then it is labeled as an “unknown class”. Otherwise, it receives the label given by the unmodified CNN classifier [12].

Kyeong et al. use CNNs for classifying the presence of multiple defect patterns in wafer bin maps (WBMs) using defect-detecting dies that reveal patterns correlated to defects. The increase in the size of WBMs has increased the probability of mixed localized noisy defect patterns. CNNs are able to perform classification robustly against the noisy mixed defect data. [13]. Borisov et al. have demonstrated, through experimental results, that a deep CNN produces increased accuracy in detecting hotspots in ICs [14]. Cha et al. identify that defect detection is one of the biggest challenges in the semiconductor manufacturing industry [15]. They propose a CNN autoencoder with skip connections in the network architecture to produce high-fidelity images for training a classifier. Their architecture consists of three encoder and three decoder blocks that are trained to reproduce the input image at the output with minimal loss. Within their work, the primary purpose of the autoencoder was to generate image samples for training a separate classifier model. After training on the synthetic autoencoder data, their classification model is found to produce 98% accuracy when tested on the real data set.

A. Contributions

In this paper, we propose a design and application of a deep CNN autoencoder architecture to localize and segment the region of defects in grayscale microscopic images of novel polymer-based SAW sensors. The contributions of this paper

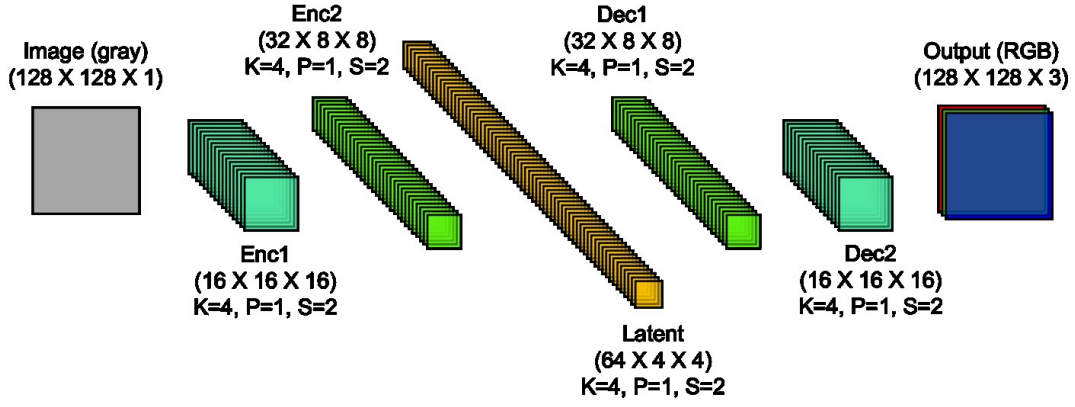


Fig. 2: Proposed deep convolutional autoencoder architecture. Enc and dec represent encoder and decoder parts, respectively. In encoder 2, for example, 32 feature maps, each with 8x8 dimensions, are convolved with 32 4x4 filters, stride of 2, and feature map padding of 1 to produce 64 feature maps of size 4x4 in the latent layer. K = filter size (KxK), P = padding, S = stride.

are as follows. First, our image segmentation and localization objectives are different from CNN-based defect classification performed in the literature. Second, the proposed variant of the SAW sensor presents a set of challenges in polymer-based photo-lithographically produced micro-sensors, which are relatively novel, making the results of defects unpredictable and challenging to identify. Third, in a novel methodological design, our research leverages the autoencoder’s ability to map a one-channel grayscale image input to a three-channel RGB output to segment and localize the defective regions in the SAW sensor electrodes. The color mapping is performed to automate the process of segmenting and quantifying the presence of defective pixels in our autoencoder-generated images. This automated approach is a more efficient route to post-fabrication defect analysis than conducting human vision-based defect analysis of SAW microsensors. Fourth, our proposed self-supervised autoencoder learning approach alleviates the need for a large image sample size as required in supervised classification tasks.

III. METHODS

This section discusses the experimental steps and procedures of the proposed work.

A. Sensor fabrication and imaging

In sensor fabrication, we have used the photolithographic process to develop novel flexible PVDF-based microsensors. The process involves several sequential stages: metallization, photoresist deposition, exposure, development, and etching. After the final etching step in sensor fabrication, microscopic imaging is used to capture IDTs. Figure 1 shows a gray scale image with defects present in the IDT’s post-photolithography. The goal of this paper is to reconstruct a color image (shown in Figure 1) with the segmented defect region in red color taking the grayscale microscopic image as the input.

B. Image processing

The objective of training a deep CNN autoencoder is to reconstruct color images with segmented defect regions from grayscale image inputs. To facilitate this training, ground truth color images with segmented defect regions are generated by manually color coding three observed regions: 1) substrate (blue), 2) IDT (green), and 3) defect (red), as shown in Figure 1. The color coding of ground truth images is performed using the paintbrush tool within the adobe Photoshop application. Each original SAW optical image has a corresponding RGB color-coded mask as depicted in Figure 3. To obtain a decent amount of training image samples, the original image is subsampled, and each subsample is subjected to different image transformations to augment the image sample size.

C. Deep CNN autoencoder

The proposed deep autoencoder architecture with its filter and parameter settings is shown in Figure 2. The latent layer yields the representation from which the color images are reconstructed via sequential deconvolution in multiple layers. The CNN architecture is excellent at extracting features from complex data, which may contain multiple foreground features, such as multiple objects contained in image data. The convolutional filters in CNNs can extract edges, corners, and location-invariant geometric patterns in images for pattern recognition tasks [16]. Autoencoders are designed to compress and encode input data into a lower dimensional latent space vector. From the latent space, the vector can be reconstructed back (decoder) into the original or any target image representation (image labeled with defects).

D. Model training and evaluation

Following subsampling and augmentation, the grayscale (input) and color (target) image pairs are split into five folds. In a five-fold cross-validation scheme, four folds of image sample pairs are used to train the deep CNN autoencoder to

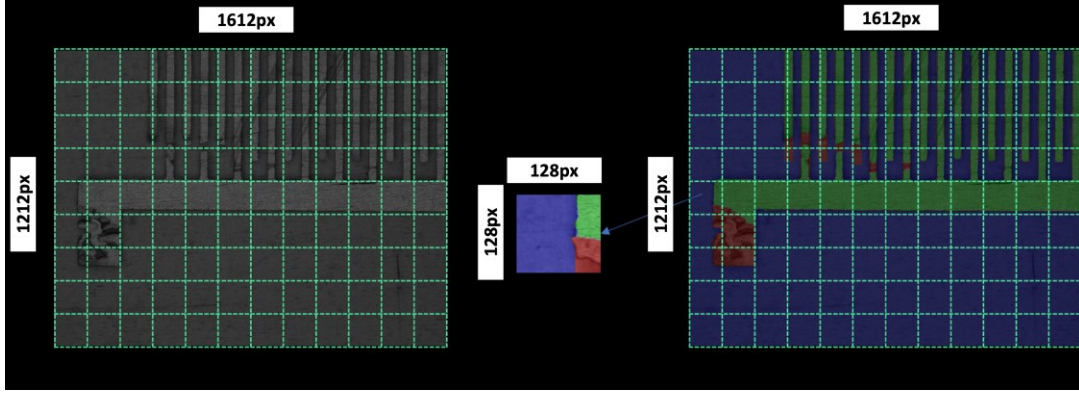


Fig. 3: Illustration of square grids to generate 108 image subsamples of the original SAW microscopic image.

reconstruct the color-coded defect regions from input grayscale images. The left-out test fold images are used to detect the defect regions passing through the trained autoencoder. We use the two most commonly used metrics in image segmentation: the Jaccard index and Dice coefficients [17], [18]. These metrics measure the overlap between hand-annotated color images and the model-annotated color images from the test grayscale images. The Jaccard index is also known as the intersection-over-union (IoU), which is the area of overlap between the predicted segment and the ground truth. The dice coefficient and IoU are positively correlated. The IoU and dice scores are defined using Equations shown in (1) and (2).

$$IoU = \frac{\text{Area of overlap}}{\text{Area of union}} \quad (1)$$

$$Dice = \frac{2 * \text{Area of overlap}}{\text{Total pixels combined}} \quad (2)$$

IV. RESULTS AND DISCUSSION

This section discusses the results we obtained after conducting the proposed experiments on a Dell Precision 5820 workstation running Ubuntu 20.04 with 64GB RAM and an NVIDIA GeForce RTX 3080 GPU with 10GB memory.

A. Image generation

The original image of 1612x1212 pixels is subsampled into 108 images, each of 128x128 dimensions, as shown in Figure 3. Each of the 108 subsamples is subject to eight different image transformations to produce an additional eight samples per subsample. These transformations increase the sample size to a total of 108x8 or 864 subsamples. The eight image transformations include rotations of an image by 0°, 90°, 180°, and 270° and horizontally flipped versions of those four rotated images.

B. Deep model architecture

The CNN autoencoder starts by taking an unlabeled grayscale input image of size 128x128 pixels and is convolved with 16 filters of size 4x4 with a padding of 1 and stride of 2. The padding of 1 contributes to maintaining the size of the original image after convolution. The stride feature

TABLE I: Layer-wise parameter details of the convolutional autoencoder.

Layer	Output Shape	Parameters #
1 Conv2d	[16, 64x64]	784
2 Conv2d	[32, 32x32]	8,224
3 Conv2d	[64, 16x16]	32,832
4 ConvTranspose2d	[32, 32x32]	32,800
5 ConvTranspose2d	[16, 64x64]	8,208
6 ConvTranspose2d	[3, 128x128]	771
Total Trainable Parameters:		83,619

is responsible for downsampling the convolved image into a compressed representation with a higher receptive field. The model hyperparameters (filter size, padding, and stride) are consistent throughout the encoder and decoder architecture. The first encoder layer yields 16 feature maps of size 16x16 stored in a 16x16x16 tensor. The second encoder layer is a 32x8x8 tensor after convolution. The third encoder layer is the latent feature layer with a 64x4x4 tensor size. Starting from this latent layer, deconvolution operations are sequentially performed in the decoder part of the autoencoder. The goal of the decoder part is to reconstruct the target masked image from the latent feature. Accordingly, the first decoded layer is a tensor of size 32x8x8 which is the same size as the second encoded layer. As expected, the second decoded layer is the same size as the first encoded layer, and the output will be of size 128x128 pixels except in color. To calculate the loss at each epoch, the mean squared error between the input grayscale and the annotated color mask is leveraged. The optimization method used for backpropagation and filter updates is adaptive momentum (ADAM) optimization. A learning rate of 0.001 and a weight decay (L2 penalty) of 0.00001 are also used in the optimization process. Table I shows the number of trainable parameters at each layer of the proposed deep architecture.

TABLE II: Intersection-Over-Union (IoU) and Dice scores for color image masks reconstructed by the trained deep convolutional in five-fold cross-validation.

Scores	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
IoU (RGB)	0.97	0.94	0.97	0.99	0.94	0.96 (0.02)
Dice (RGB)	0.98	0.97	0.99	0.99	0.97	0.98 (0.01)
IoU (Defects in red)	0.54	0.46	0.58	0.87	0.43	0.58 (0.17)
Dice (Defects in red)	0.70	0.63	0.73	0.93	0.61	0.72 (0.13)
IoU (IDTs in green)	0.98	0.98	0.97	0.98	0.95	0.97 (0.01)
Dice (IDTs in green)	0.99	0.99	0.98	0.99	0.97	0.98 (0.01)
IoU (Substrate in blue)	0.95	0.96	0.99	0.99	0.95	0.97 (0.02)
Dice (Substrate in blue)	0.98	0.98	0.99	0.99	0.98	0.98 (0.01)

C. Deep model training and testing

The 864 subsamples are randomly grouped into five data folds. In a standard five-fold cross-validation scheme, the deep autoencoder is trained with four-fold grayscale images to reconstruct their corresponding color images with defects. The trained model is then used to reconstruct color images with segmented defects from the test data folds of grayscale images.

The results from the five-fold experiment in Table II measure the overlapping regions between the ground truth color and the autoencoder-predicted color images. The overlaps between color images (including IDTs, substrate, and defect) are very high. These scores are obtained on the test image folds following 100 epochs of training. Because the defective region (in red color) accounts for a small and subtle portion of the image, the corresponding average dice score is relatively low 0.72 (0.13) with an IoU score of 0.58 (0.17). Furthermore, these scores measure pixel-level overlap and accuracy where a near-perfect score may be difficult to achieve given the sample size. The high defect detection scores on fold 4 image samples suggest that the proposed model has some variability in terms of segmentation accuracy. In other words, some defect samples are better segmented than others.

D. Visualizing the predicted defects

The results can be visually analyzed by comparing the ground truth labeled masks to the respective autoencoder-generated masks, as shown in Figure 4. Figure 4 (a) shows a representative example of a correctly segmented defect. Although the shape is not perfectly reconstructed, the location of the defect with an approximate shape is correctly identified. We also identify instances in Figure 4 (b) where the ground truth defect label disagrees with its autoencoder-generated counterpart. This disagreement may be due to human errors during the labeling of defect regions. Figure 4 (c) shows that defects can be too subtle to be detected by the autoencoder, resulting in a false negative scenario.

It can be seen in Figure 4 (a) that the autoencoder is able to reproduce the three main regions of defects by comparing the O_Mask and AE_Mask RGB color images. In Figure 4(b), the autoencoder (AE_Mask) generates a region of defect that

does not appear in the ground truth (O_Mask). Upon careful observation in the grayscale optical image, it can be noted that the autoencoder correctly reproduced a defect that the human labeling process overlooked.

E. Discussion of the results

The accuracy scores indicate a high fidelity in segmenting the two classes (IDT and substrate) features in the image. The defect areas are with subtler and finer features with more complexity than the IDT and substrate features. The false positive and false negative scenarios suggest that more representative image samples are required to improve the segmentation performance of the proposed deep model. It is important to note that obtaining such fabrication images at a large scale is expensive and time-consuming, which is one of the requirements for developing robust deep-learning models. The defect features were more difficult for the autoencoder to capture indicated by the accuracy scores. False positive and false negative predictions for regions of defects can also be observed. These two scenarios have to be further investigated because the false predictions often correct the misidentified regions of defects during labeling caused by human oversight (error). This could suggest that the model is observing details that a human may easily overlook. The second analysis of the false predictions is that the model is identifying some of the noise or minor/granular details in the image as defects where these features may not contribute to sensor performance. The model's ability to potentially capture features that are not present in the ground truth label but should have been identified and labeled serves as a good second opinion in discovering defective regions in the images.

V. CONCLUSIONS

In this study, we proposed a custom application of CNN autoencoder for the segmentation of defect regions, IDTs, and substrate areas in microscopic images of SAW sensors. The segmentation results show a promising approach to automating the defect detection process, especially when human visual inspections can be error-prone. Our proposed deep model is successful in segmenting three regions in grayscale images by transforming the image regions into color codes, which aided image segmentation and defect visualization. The defect

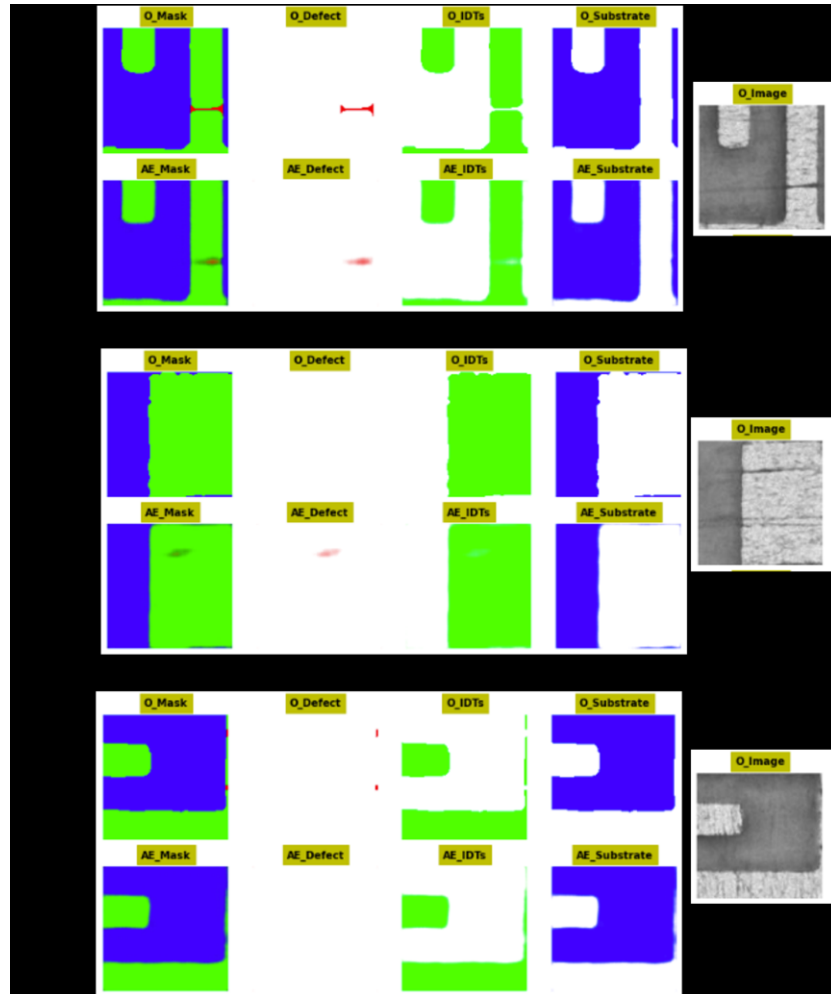


Fig. 4: The ground truth color image labeled with the defective region (O_Mask) and the predicted color image (AE_Mask). The ground truth segments of defect, IDTs, and substrate are shown using O prefix. The predicted segments are shown using the AE prefix. (a) An example of correct segmentation, (b) misidentified (false positive) segments, and (c) missed detection (false negative) where the defect was too subtle to detect.

regions are segmented with less accuracy compared to other regions due to the subtlety of the region. This technology can contribute to the automation of defect analysis in the microsensor manufacturing field to reduce the sole reliance on "eye-balling" defects as well as provide a second opinion for identified defects. In the future, more image samples will be required to improve the robustness of the proposed deep model in similar image segmentation tasks.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the following grants that contributed to funding the research:

REFERENCES

- [1] D. Pan, B. Yu, and J.-R. Gao, "Design for manufacturing with emerging nanolithography," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 32, pp. 1453–1472, 10 2013.
- [2] C. Floer, S. Hage-Ali, P. Nicolay, H. Chambon, S. Zhgoon, A. Shvetsov, J. Streque, H. M'Jahed, and O. Elmazria, "Saw rfid devices using connected idts as an alternative to conventional reflectors for harsh environments," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. PP, pp. 1–1, 09 2019.
- [3] A. Carlson and T. Le, "Correlation of wafer backside defects to photolithography hot spots using advanced macro inspection - art. no. 61523e," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6152, 03 2006.
- [4] M. Zakaria, S. Rahman, S. Leng, M. MAT, and U. Hashim, "Modification of surface acoustic wave (saw) sensor using zinc oxide nanowires."
- [5] R. Manenti, A. Frisk Kockum, A. Patterson, T. Behrle, J. Rahamim, G. Tancredi, F. Nori, and P. Leek, "Circuit quantum acoustodynamics with surface acoustic waves," *Nature Communications*, vol. 8, 10 2017.
- [6] N. K. Thierry Aubert and O. Elmazria, "Langasite as piezoelectric substrate for sensors in harsh environments: Investigation of surface degradation under high-temperature air atmosphere," *Sensors*.
- [7] H. Yang, Y. Lin, B. Yu, and E. F. Y. Young, "Lithography hotspot detection: From shallow to deep learning," vol. 2017-Sept. IEEE, 9 2017, pp. 233–238. [Online]. Available: <http://ieeexplore.ieee.org/document/8226047/>
- [8] H. Zhang, B. Yu, and E. F. Y. Young, "Enabling online learning in lithography hotspot detection with information-theoretic feature optimization," vol. 07-10-Nove. ACM, 11 2016, pp. 1–8. [Online].

Available: <https://dl.acm.org/doi/10.1145/2966986.2967032>

- [9] M. P. McLaughlin, P. Mennell, A. Stamper, G. Barber, J. Paduano, E. Benn, M. Linnane, J. Zwick, C. Khatumria, R. L. Isaacson, N. Hoffman, and C. Menser, "Improved color defect detection with machine learning for after develop inspections in lithography," *IEEE Transactions on Semiconductor Manufacturing*, vol. 35, pp. 418–424, 8 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9807429/>
- [10] D. Ding, A. J. Torres, F. G. Pikus, and D. Z. Pan, "High performance lithographic hotspot detection using hierarchically refined machine learning," *IEEE*, 1 2011, pp. 775–780. [Online]. Available: <http://ieeexplore.ieee.org/document/5722294/>
- [11] W. Ye, Y. Lin, M. Li, Q. Liu, and D. Z. Pan, "Lithoroc: Lithography hotspot detection with explicit rocoptimization," *ACM*, 1 2019, pp. 292–298. [Online]. Available: <https://dl.acm.org/doi/10.1145/3287624.3288746>
- [12] S. Cheon, H. Lee, C. O. Kim, and S. H. Lee, "Convolutional neural network for wafer surface defect classification and the detection of unknown defect class," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, pp. 163–170, 5 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8657760/>
- [13] K. Kyeong and H. Kim, "Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks," *IEEE Transactions on Semiconductor Manufacturing*, vol. 31, no. 3, pp. 395–402, 2018.
- [14] V. Borisov and J. Scheible, "Lithography hotspots detection using deep learning," *IEEE*, 7 2018, pp. 145–148. [Online]. Available: <https://ieeexplore.ieee.org/document/8434561/>
- [15] J. Cha, J. Park, and J. Jeong, "A novel defect classification scheme based on convolutional autoencoder with skip connection in semiconductor manufacturing," vol. 2022-Febru. *IEEE*, 2 2022, pp. 347–352. [Online]. Available: <https://ieeexplore.ieee.org/document/9728861/>
- [16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [17] M. J. Awan, M. S. M. Rahim, N. Salim, A. Rehman, and B. Garcia-Zapirain, "Automated knee mr images segmentation of anterior cruciate ligament tears," *Sensors*, vol. 22, no. 4, p. 1552, 2022.
- [18] A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, "Automatic instrument segmentation in robot-assisted surgery using deep learning," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. *IEEE*, 2018, pp. 624–628.