[1]Hesham Samir
[2]Bassem Amin
[3]Maged Ghoneima

CD-SEM Image Defect Detection and Classification Using Transformers

Abstract: - In the integrated circuits manufacturing process, a critical dimension scanning electron microscope machine provides high-resolution images of geometries on semiconductor wafers and measurements of the printed polygons after the photolithography process. These images and measurements are crucial for calibrating lithography process models such as the etch model, optical proximity correction model, and resist model, which are necessary for simulating the photolithography process and hence reducing the probability of defects occurrence. However, The test chips used for model calibration often contain imperfections or anomalies that require accurate detection and classification to be removed from the model calibration process. Traditionally, engineers manually filter semiconductor wafer images and their measurements, which is a time-consuming and error-prone task. Therefore, there have been attempts to use machine learning to automate image filtering. In this research, we introduce a novel approach using an image preprocessing stage followed by a transformer-based object detection model to automate detecting and classifying defects in semiconductor wafer images containing an array of lines or a matrix of contacts. The proposed approach shows improved accuracy compared to traditional methods.

Keywords: CD-SEM, Defect detection, Metrology, Semiconductors.

I. INTRODUCTION

Lithography is a fundamental process in semiconductor fabrication, playing a pivotal role in defining the critical dimensions of integrated circuits (ICs). The primary objective of semiconductor lithography is to transfer patterns from a mask to the semiconductor wafer. This process involves projecting the mask's image onto a radiation-sensitive resist material coated on the wafer using optical components of an exposure tool. The patterns transferred include gates, isolation trenches, contacts, metal interconnects, and vias for interconnecting metal layers [1].

A CD-SEM (Critical Dimension Scanning Electron Microscope) machine is a type of scientific instrument used in the semiconductor industry and nanotechnology research to measure and analyze the critical dimensions of extremely small structures and features on a semiconductor wafer or other microelectronic devices [2]. CD-SEM machine exports images of printed circuits and some measurements at some points in the wafer. This data is used to calibrate photo-lithography process models by knowing the variation between the dimensions in the layout and the actual printed dimension [3]. Some of these images might contain failures during the printing process which is called lithographic failure. Traditional wafer defects are usually caused by particles or defects on the mask or in the resist layers. However, as printed dimensions get smaller, random defects can also occur due to reactions within the resist film during pattern formation [4]. Fig.1 shows different types of defects in CD-SEM images.

Traditional approaches to defect detection typically require manual examination, a process that is not only labor-intensive but also prone to errors and variability. The emergence of machine learning and deep learning technologies has led to a significant shift towards automating the detection process using more advanced and reliable techniques. These methods have demonstrated superior capabilities in handling the complexities associated with CD-SEM images, such as high noise levels, varying contrast, and complex defect patterns.

This research aims to automate the process of filtering images produced during the semiconductor fabrication process using transformers. The paper is structured as follows: Section 2 provides a review of relevant studies. Section 3 outlines the proposed method for defect detection and classification. Section 4 presents an analysis of the dataset preparation and the model training process. Section 5 details the results and comparisons, and Section 6 concludes the paper.

[1] Master's student, Computer and Systems Engineering, Ain Shams University, Egypt. 2000513@eng.asu.edu.eg

[2] Associate Professor, Computer and Systems Engineering, Ain Shams University, Egypt. babdullah@eng.asu.edu.eg

[3] Associate Professor, Mechatronics, Ain Shams University, Egypt. maged_ghoneima@eng.asu.edu.eg

Copyright © JES 2024 on-line : journal.esrgroups.org

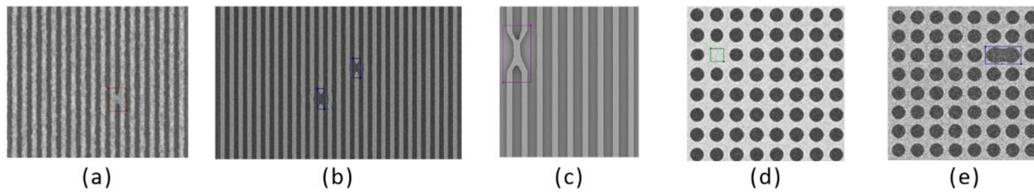

Figure 1: Different defect categories: (a) bridge, (b) gap/break,
(c) line-collapse, (d) missing contact, (e) merging contacts

II. RELATED WORK

This section provides a concise overview of the existing literature on CD-SEM image defect detection in semiconductor manufacturing.

Peter De Bisschop et al. developed the (pix)NOK metric, standing for "Not OK pixels," a specific measure of CD-SEM images to quantify the extent of defects or failures. This metric counts the pixels outside specified tolerance limits for a given feature's dimensions. Using this metric allows a quantifiable and standardized assessment of feature quality, and it is an effective way to determine the pass/fail status of specific areas on the wafer [4].

Hiroshi Fukuda et al. introduced the use of autoencoders to detect defects in CD-SEM images. First, the encoder and decoder are trained to compress and decompress CD-SEM images with no defects. Then, these trained encoder and decoder networks are used to generate images from input SEM images. If there is a difference between the input image and the output image, the input image is considered defective since the output image is always without any defects [5].

Bappaditya Dey et al. proposed an ensemble strategy to combine the output predictions from different models, achieving better performance in the classification and detection of defects. This approach is complemented by an unsupervised machine learning model for denoising CD-SEM images, which enhances the inspection process by reducing false positives and optimizing the effect of stochastic noise [6]. Additionally, they presented another method for defect classification and detection in CD-SEM images using the deep learning-based Mask R-CNN model, leveraging its powerful feature extraction and instance segmentation capabilities to accurately identify and classify defects [7].

To the best of our knowledge, this is the first instance where transformers have been utilized to detect and classify defects in CD-SEM images of semiconductors.

III. PROPOSED APPROACH

The paper "Attention Is All You Need" authored by Vaswani et al., introduced the transformer model that depends entirely on an attention mechanism instead of complex recurrent or convolutional neural networks which are commonly used in previous models [8]. The Transformer architecture consists of an encoder-decoder structure, where both the encoder and decoder are built using layers of self-attention and feed-forward neural networks [9].

Transformers have transformed the field of natural language processing (NLP) and have become foundational to many widely adopted models like BERT [10], GPT [11], and T5 [12]. Their ability to handle long-range dependencies and parallelize training processes has significantly improved the efficiency and effectiveness of NLP tasks, for example, text translation and summarization. Moreover, transformers have extended their influence to other domains, including computer vision and speech recognition, by enabling models to learn complex patterns and relationships within data. We considered the problem of detecting defects in CD-SEM images as a sequence problem as there will be missing or added shapes in a certain pattern. Fig.2 shows an example of a CD-SEM image defect in a matrix of contacts layout. The defect here is a missing contact at the center of the image. As we can see, the defective object is just an empty background, but we can identify this defect from the surrounding contacts.

The proposed method includes detection transformers (DETR) [13] as the base model. DETR Combines transformers with the common Convolutional Neural Network (CNN) [14] to benefit from the power of identifying sequences in transformers in an object detection model. This mechanism allowed the model to understand the overall context of images and the relationships between objects. This capability is particularly valuable for our problem, where the shape of the defect is undefined; instead, defects can be identified based on discrepancies from expected patterns in their surroundings.

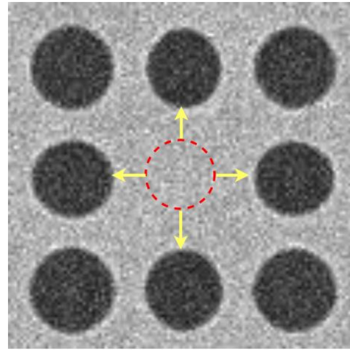

Figure 2: Missing contact defect with an attention mechanism.

The DETR model offers several additional advantages [15] for object detection as follows:

- Simplify the traditional object detection pipeline by treating it as a direct set prediction problem.
- Process images in parallel during inference, which benefits real-time applications.
- Remove the necessity for post-processing steps like Non-Maximal Suppression (NMS).

Fig.3 illustrates the flow diagram of our proposed approach. After the wafer fabrication process, the wafer is inputted into a CD-SEM machine to take a picture of the wafer. The output image undergoes a preprocessing stage to improve its quality. The preprocessing step is an automated brightness adjustment technique to brighten dark images since DETR performs better on bright images. This approach provides consistency and effectiveness for the usage of this model. Practically, this method is better than training the model on dark images.

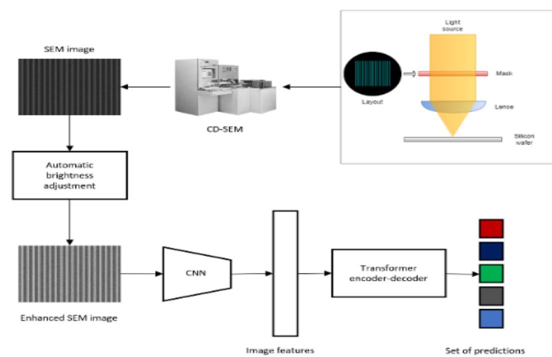

Figure 3: The model flow diagram.

The automatic brightness adjustment process operates in the following manner. First The average brightness $\bar{I}$ of the image is calculated as:

$$\bar{I} = \left(\frac{1}{N}\right)\sum_{i=1}^{n} I_{\mathrm{i}} \tag{1}$$

where N is the total number of pixels in the image, and I_i are the intensities of individual pixels.
Then, the brightness adjustment factor α is determined by:

$$\alpha = \frac{I_{target}}{\bar{I}} \tag{2}$$

where I_{target} is the target brightness level.
Finally, the adjusted RGB values of each pixel are given by:

$$R' = \alpha R, G' = \alpha G, B' = \alpha B \tag{3}$$

where R', G', and B' are the adjusted values of the red, green, and blue channels.

The next stage is the DETR model. First, the CD-SEM image passes through CNN to extract features from the CD-SEM image. Then, the image features pass through the transformer encoder-decoder to generate a constant set of predictions.

IV. EXPERIMENTS

A. *Data Acquisition*

As far as we are aware, no publicly accessible dataset of CD-SEM images with labeled defects is currently available. This lack of data poses a challenge for developing and benchmarking defect detection algorithms in the semiconductor manufacturing industry. So, we prepared CD-SEM images containing multiple defects and labeled them with the defect category. We used the PROLITH [16] tool to simulate the semiconductor fabrication process and generate CD-SEM images. The defect is added to the layout at different locations to simulate random printing failures. This simulation is repeated under different process window conditions such as dose and focus to generate multiple variations of defects. Fig.4 shows the process of generating simulated CD-SEM images with missing contact using PROLITH.

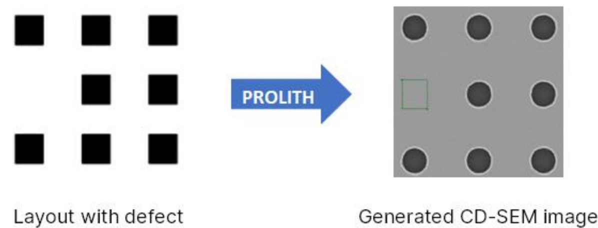

Figure 4 Generating CD-SEM images with defects by adding defects in the layout and simulating the semiconductor manufacturing process using PROLITH.

The generated dataset contains 2,106 images with defects including bridges, gaps, and collapsing lines within line/space patterns. It also contains missing and merging contacts within contact hole patterns. We used the Labelme [17] tool to label defects manually. Table 1 illustrates the number of defects across train, validation, and test sets.

Table 1: Data distribution across different sets.

Defect type	**Train**	**Validation**	**Test**
Bridge	910	115	119
Gap	845	85	81
Missing Contact	403	58	54
Merging Contact	284	41	41
Collapse	135	17	22

B. *Evaluation Criteria*

In the assessment of the proposed method's performance, it is essential to establish clear and robust evaluation criteria. The primary objective is to accurately quantify the model's effectiveness in identifying defects represented by the bounding boxes in images. We developed an approach that calculates the number of correctly identified defects (true positives), incorrectly identified defects (false positives), and missed defects (false negatives) for each defect category, as detailed below:

- For every actual defect represented by a ground truth bounding box, we identify all overlapping predicted bounding boxes of the same type. The first overlapping predicted box is labeled as a true positive, while subsequent overlapping boxes are considered false positives due to duplication.
- Predicted bounding boxes that do not overlap with ground truth boxes are also classified as false positives.
- Any ground truth bounding boxes without any corresponding predicted boxes are classified as false negatives.

Fig.5 illustrates how our evaluation criteria are applied. Rectangles in this image represent the bounding boxes of defects in the input image.

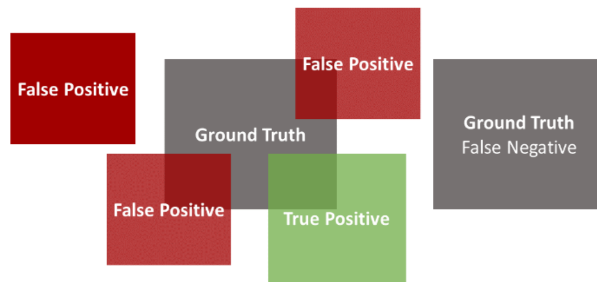

Figure 5: Evaluation criteria diagram

After collecting true positives, false positives, and false negatives, we calculated precision, recall, and the F1 score, which are commonly used metrics for evaluating classification models.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{5}$$

$$F1\ Score = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

To better understand the performance of the proposed flow, we examine its predictions through a confusion matrix. This matrix offers a detailed view of the model's performance by recording the counts of correctly identified defects, misclassified defects, and missed defects for each class. Analyzing this matrix allows us to assess the model's classification accuracy and pinpoint potential areas for enhancement.

C. Training

We used the transfer learning method to fine-tune the DETR model pre-trained on the coco [18] dataset. To enhance the model's generalization capabilities, our training pipeline incorporates a range of randomized data augmentation methods, including horizontal flipping, resizing, and cropping. The training process was executed on NVIDIA V100 GPU hosts. Fig.6 displays the mean average precision (mAP) at an Intersection Over Union (IOU) [19] threshold of 0.5 as it evolves over 50 epochs during the training process. The x-axis represents the epoch number, ranging from 1 to 50, while the y-axis quantifies the mAP. The mean average precision started from 0 and reached 0.9 in the second epoch and then oscillated between 0.8 and 1.

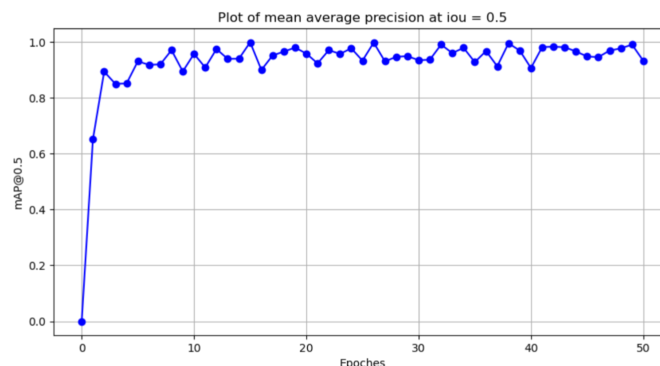

Figure 6: Plot of mAP at IOU threshold 0.5 across different

V. RESULTS AND ANALYSIS

In this section, we present the results obtained from our experiments on the DETR model and provide a comparison with commonly used object detection models.

Fig.7 Displays a sample of detected defects using our method, along with the defect classification and detection score. In most cases, the model succeeded in detecting the defected object with a high score.

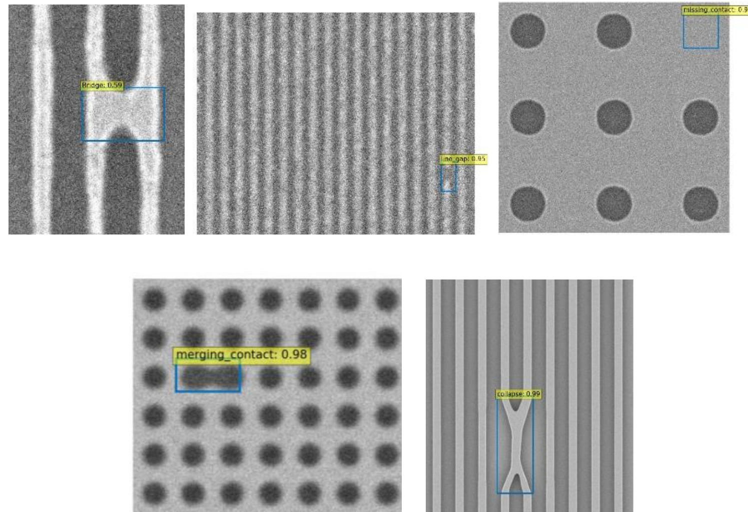

Figure 7: Defect detection results

Table 4 presents the outcomes of deploying the DETR model on the test set, demonstrating its success in identifying defects within the dataset. Specifically, the model achieved an impressive overall F1 score of 96%. Fig.8 displays the confusion matrix resulting from applying our proposed flow to the test dataset.

Table 2: Defect detection evaluation

Defect	Precision	Recall	F1-Score	Support
Bridge	1.00	0.90	0.95	81
Collapse	0.92	1.00	0.96	22
Line Gap	1.00	0.97	0.99	119
Merging Contact	0.98	1.00	0.99	41
Missing Contact	0.96	1.00	0.98	54
accuracy			96%	319
macro average	81%	81%	81%	319
weighted average	98%	96%	97%	319

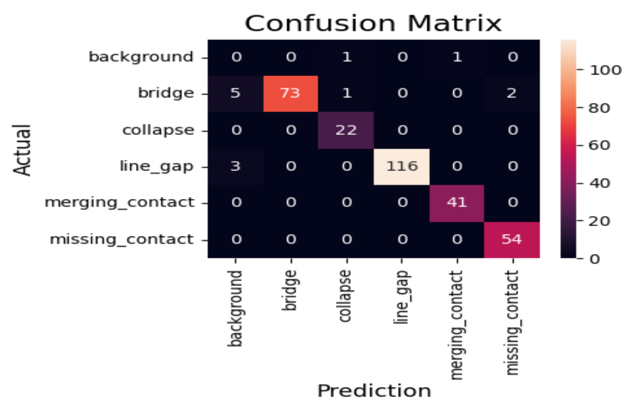

Figure 8: Confusion matrix

Fig.9 Visualized how the DETR model succeeded in detecting defects in a line/space pattern in a CD-SEM image. The first image in these figures represents the CD-SEM image which is the input image to the model. The second image visualizes the attention weights of the last decoder layer which represents what was the model looking at while detecting this object. The third image visualizes the encoder self-attention weights which represent what objects the model attended to. Fig.10 shows a similar analysis for the contact hole pattern. As we can see in the self-attention images, the DETR model effectively identified the nearest objects of interest, thereby enhancing the defect detection process

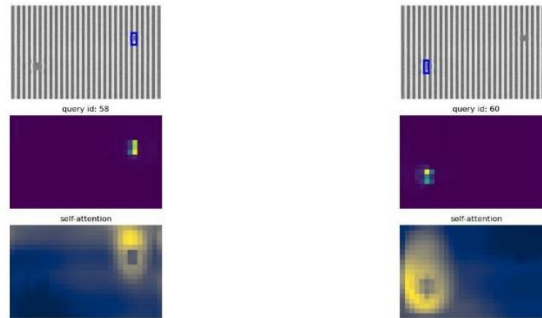

Figure 9: The decoder's query number and encoder's self-attention for detected defects in line/space pattern

Figure 10: Decoder's query number and encoder's self-attention for detected defects in contact hole pattern

Table 3 shows a comparison between the F1 score of our proposed flow and the F1 scores of YOLO v8 [20] and Faster RCNN [21]. All models are trained using the same dataset for 50 epochs. The table shows that our proposed flow produced the highest F1 score in CD-SEM image defect detection.

Table 3: Comparison between different models

Model	**F1 Score**
DETR	0.96
YOLO v8	0.93
Faster RCNN	0.92

VI. CONCLUSION AND FUTURE WORK

Automatic CD-SEM defect detection has become very important with the continuous shrinkage of the critical dimension. This task is challenging because images are often noisy, and defects are typically small. The DETR model demonstrates excellent performance in detecting defects in CD-SEM images, owing to its capability to analyze surrounding objects effectively.

Future work involves extending support to various layout shapes to cover all possible defects. This can be achieved by developing a multiple-input model that can process both the target layout and the CD-SEM image of the fabricated wafer. The layout input will provide insights into the intended shapes and potential defects.

VII. Limitations

We limited this study to simple layouts containing an array of lines or a matrix of contacts. The generalizability of the findings to other pattern types may be limited.

Acknowledgment

The authors would like to extend their heartfelt thanks to Germain Fenger, Kevin Ahi, Stewart Wu, and Mohammad Kamel from Siemens EDA for their crucial contributions to the development and implementation of this research. Their valuable insights and expertise, drawn from their positions at Siemens EDA, were essential to the successful completion of this work.

References

[1] U. Okoroanyanwu, The Semiconductor Litho-graphic Process. SPIE, 2010, pp. 463–550, https://www.spiedigitallibrary.org/ebooks/PM/Chemistry-and-Lithography/11/The-SemiconductorLithographic-Process/10.1117/3.821384.ch11?SSO=1. [Online].

[2] C. F. Bevis and D. E. Clapper, “Critical dimension scanning electron microscope,” Aug. 3 2004, uS Patent 6,770,868.

[3] C. Mack, “Improved methods for lithography model calibration,” in Photomask and Next-Generation Lithography Mask Technology XIV, H. Watanabe, Ed., vol. 6607, International Society for Optics and Photonics. SPIE, 2007, p. 66071D. [Online]. Available: https://doi.org/10.1117/12.728961

[4] P. D. Bisschop, “Stochastic printing failures in extreme ultraviolet lithography,” Journal of Micro/Nanolithography, MEMS, and MOEMS, vol. 17, no. 4, p. 041011, 2018. [Online]. Available: https://doi.org/10.1117/1.JMM.17.4.04101

[5] H. Fukuda and T. Kondo, “Anomaly detection in random circuit patterns using autoencoder,” Journal of Micro/Nanopatterning, Materials, and Metrology, vol. 20, no. 4, pp. 044 001–044 001, 2021

[6] B. Dey, D. Goswami, S. Halder, K. Khalil, P. Leray, and M. A.Bayoumi, “Deep learning-based defect classification and detection in sem images,” in Metrology, Inspection, and Process Control XXXVI, J. C. Robinson and M. J. Sendelbach, Eds. SPIE, Jun. 2022. [Online].

[7] B. Dey, E. Dehaerne, K. Khalil, S. Halder, P. Leray, and M. A.Bayoumi, “Deep learning based defect classification and detection in sem images: A mask r-cnn approach,” 2022. [Online]. Available: https://arxiv.org/abs/2211.02185

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N.Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023. [Online]. Available: https://arxiv.org/abs/1706.03762

[9] E. Y. Zhang, A. D. Cheok, Z. Pan, J. Cai, and Y. Yan, “From turing to transformers: A comprehensive review and tutorial on the evolution and applications of generative transformer models,” Sci, vol. 5, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2413-4155/5/4/46

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

[11] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever et al., “Improving language understanding by generative pre-training,” 2018.

[12] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” 2023. [Online]. Available: https://arxiv.org/abs/1910.10683

[13] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” in European conference on computer vision. Springer, 2020, pp. 213–229.

[14] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” 2015. [Online]. Available: https://arxiv.org/abs/1511.08458

[15] T. Shehzadi, K. A. Hashmi, D. Stricker, and M. Z. Afzal, “Object detection with transformers: A review,” 2023. [Online]. Available: https://arxiv.org/abs/2306.04670

[16] KLA-Tencor, “Prolith,” https://www.kla-tencor.com, software.

[17] K. Wada, “labelme: Image polygonal annotation with python,” https://github.com/wkentaro/labelme, 2018.

[18] ar, “Microsoft coco: Common objects in context,” 2015. [Online]. Available: https://arxiv.org/abs/1405.0312

[19] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” 2019. [Online]. Available: https://arxiv.org/abs/1902.09630

[20] D. Reis, J. Kupec, J. Hong, and A. Daoudi, “Real-time flying object detection with yolov8,” 2024. [Online]. Available: https://arxiv.org/abs/2305.09972

[21] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” 2016. [Online]. Available: https://arxiv.org/abs/1506.01497