An Evaluation of Continual Learning for Advanced Node Semiconductor Defect Inspection

Amit Prasad[1,*] Bappaditya Dey[1,*] (✉) Victor Blanco[1] and Sandip Halder[2,′]

Interuniversity Microelectronics Centre, Kapeldreef 75, 3001, Belgium[1]
SCREEN SPE Germany GmbH, Germany[2]
Equal Contribution[*]
This research was conducted during Sandip Halder's tenure at imec[′]
amit.prasad.ext@imec.be, Bappaditya.Dey@imec.be

Abstract. Deep learning-based semiconductor defect inspection has gained traction in recent years, offering a powerful and versatile approach that provides high accuracy, adaptability, and efficiency in detecting and classifying nano-scale defects. However, semiconductor manufacturing processes are continually evolving, leading to the emergence of new types of defects over time. This presents a significant challenge for conventional supervised defect detectors, as they may suffer from catastrophic forgetting when trained on new defect datasets, potentially compromising performance on previously learned tasks. An alternative approach involves the constant storage of previously trained datasets alongside pre-trained model versions, which can be utilized for (re-)training from scratch or fine-tuning whenever encountering a new defect dataset. However, adhering to such a storage template is impractical in terms of size, particularly when considering High-Volume Manufacturing (HVM). Additionally, semiconductor defect datasets, especially those encompassing stochastic defects, are often limited and expensive to obtain, thus lacking sufficient representation of the entire universal set of defectivity. This work introduces a task-agnostic, meta-learning approach aimed at addressing this challenge, which enables the incremental addition of new defect classes and scales to create a more robust and generalized model for semiconductor defect inspection. We have benchmarked our approach using real resist-wafer SEM (Scanning Electron Microscopy) datasets for two process steps, ADI and AEI, demonstrating its superior performance compared to conventional supervised training methods.

Keywords: Continual learning · Catastrophic forgetting · Semiconductor manufacturing · Defect classification · Lithography · Metrology

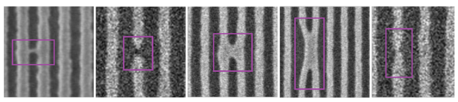
1 Related Work

In the semiconductor process (mainly, Litho-Etch) domain, numerous approaches have been suggested for defect classification and localisation [2], [3], [1]. To the best of the authors' knowledge, the concept of incremental learning [5] for multi-class, multi-instance defect detection on SEM images has previously not been explored.

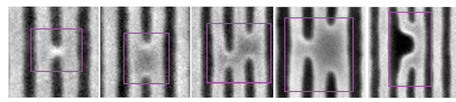
2 Methodology

2.1 Dataset

Original (resist) wafer SEM (Scanning Electron Microscopy) images were obtained during ADI (After Development Inspection) and AEI (After Etch Inspection) stages. Figure 1 illustrates exemplary defect types in both process steps. The instance distribution per defect class is captured in Table 1.



(a) ADI defects. Left to right: Microbridge, Gap, Bridge, Line collapse, Probable-gap



(b) AEI defects. Left to right: Thin bridge, Single bridge, Multi bridge non-horizontal, Multi bridge horizontal, Line collapse

Fig. 1: SEM images with a) ADI defects and b) AEI defects

Table 1: Instance distribution per class

		ADI Instances					AEI Instances		
Label	**Defect type**	**Training**	**Validation**	**Test**	**Label**	**Defect type**	**Training**	**Validation**	**Test**
0	**Microbridge**	380	47	78	5	**Thin bridge**	241	29	29
1	**Gap**	1046	156	174	6	**Single bridge**	240	29	31
2	**Bridge**	238	19	17	7	**Multi bridge non-horizontal**	160	21	19
3	**Line collapse**	550	66	76	8	**Multi bridge Horizontal**	80	10	10
4	**Probable-gap**	315	49	54	9	**Line collapse**	202	40	34
	Total	2529	337	399		Total	923	129	123

2.2 Notations and Preliminaries

The following notations have been used in this work.

Definition 1. ***Task*** ***(T_p)***: *This is defined as supervised training of a defect detection framework for p classes (0 to $p-1$) in the dataset of the form $(x_i, y_i)_{i=1}^{m}$ (m instances with defect feature x_i and corresponding label y_i). This is denoted by T_p.*

Definition 2. ***Finetuned task*** ***(F_p^q)***: *This is defined as supervised training of a defect detection framework for next q classes (p to $q-1$) in the dataset of the form $(x_i, y_i)_{i=1}^{m}$, which has previously been trained on the initial p classes (0 to $p-1$). However, it's important to note that identifying these initial p classes is not guaranteed. This is denoted by F_p^q.*

Definition 3. ***Incremental task*** ***($\mathcal{T}_p^q$)***: *This is defined as* ***incremental*** *supervised training of a defect detection framework for next q classes (p to $q-1$) in the dataset of the form $(x_i, y_i)_{i=1}^{m}$, which has previously been trained on the initial p classes (0 to $p-1$), enabling it to identify all $(p+q)$ classes. This is denoted by $\mathcal{T}_p^q$.*

2.3 Structure of study

In this work we present the following case studies.

1. **Case study 1** (see Section 3) examines effectiveness of the framework in **incrementally** learning new defect classes and minimizing **forgetting** of previously trained defect classes on the ADI dataset.
2. **Case study 2** (see Section 4) assesses framework for **incrementally** learning new defect classes in AEI images and minimizing **forgetting** of previously trained defect classes across the entire ADI dataset.
3. **Case study 3** (see Section 5) compares three training strategies: (i) conventional supervised training strategy with all defect classes at once, (ii) conventional supervised training with first p defect classes and then fine-tune on new q defect classes, (iii) proposed **incremental** supervised training strategy with first p defect classes and then fine-tune on new q defect classes.

We use the Faster-RCNN [6] model for all studies. Moreover, for incremental tasks, the approach utilized is presented in [4] which also uses FRCNN.

3 Case study 1

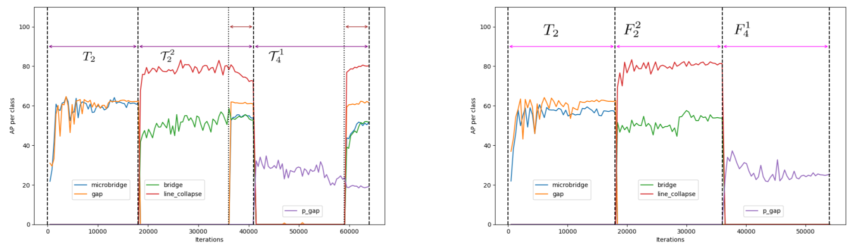
The model starts training with the task T_2 (initially trained for 2 defect classes, microbridge and gap), followed by two consecutive incremental training tasks: $\mathcal{T}_2^2$ (adding 2 more defect classes, bridge and line-collapse), and finally $\mathcal{T}_4^1$ (adding the last defect class as probable gap), using the ADI dataset. For an evaluation of performance, average precision (AP) per defect class vs iterations is plotted, marking checkpoints where new defect classes were introduced and where continual learning takes place. The results are compared to the conventional fine-tuning approach, where the model has been trained on tasks F_2^2 and F_4^1, while keeping all experimental conditions constant. In Figure 2 a), it is evident how effective incremental learning is for progressively learning defect classes and minimizing catastrophic forgetting. Conversely, in Figure 2 b), it is apparent how swiftly catastrophic forgetting occurs in the case of fine-tuning.

4 Case study 2

Defect classes from the AEI dataset are incrementally added following training on the ADI dataset. The model, following task $\mathcal{T}_4^1$, undergoes training on tasks $\mathcal{T}_5^2$ and $\mathcal{T}_7^3$. Similarly, following task F_4^1, the model undergoes fine-tuning for tasks F_5^2 and F_7^3. The Figure 3 illustrates the comparison between proposed incremental learning and conventional fine-tuning (using AP vs iteration plot).

5 Case study 3

Inference results are shown in Figure 4 (with corresponding labels, bounding boxs and confidence scores) are from 3 training strategies, first is the model

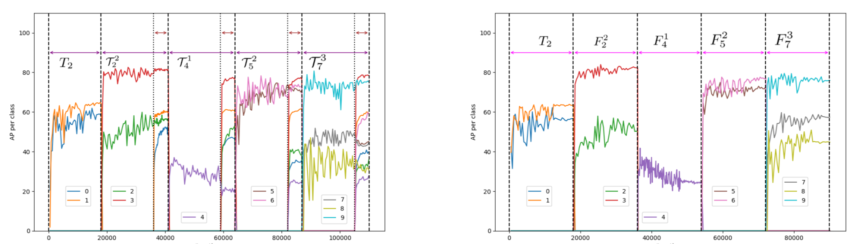


(a) Model trained incrementally for tasks $\mathcal{T}_2^2$ and $\mathcal{T}_4^1$ after training on task T_2

(b) Model finetuned on tasks F_2^2 and F_4^1 after task T_2

Fig. 2: Comparison between (a) proposed incremental learning and (b) conventional fine-tuning method.

trained on task T_{10} (incorporating all defect classes simultaneously) while the other models are derived from tasks $\mathcal{T}_7^3$ and F_7^3. The labels are referenced from Table 1. Notably, it is observed that the model after task T_7^3 performs comparably to the model trained on task T_{10}. However, the model obtained after task F_7^3 demonstrates forgetfulness or mislabeling of defects it encountered earlier, as it has only recently been exposed to labels 7, 8, and 9.



(a) Model trained incrementally for tasks $\mathcal{T}_2^2$, $\mathcal{T}_4^1$, $\mathcal{T}_5^2$, $\mathcal{T}_7^3$ after training on task T_2

(b) Model finetuned on tasks F_2^2 and F_4^1, F_5^2 and F_7^3 after task T_2

Fig. 3: (a) Proposed incremental learning vs (b) conventional fine-tuning method for incremental learning of AEI defects, after training across the ADI dataset.

6 Conclusion

In this study, we demonstrated the effectiveness of a continual learning strategy in progressively learning the classification and localization of semiconductor defect classes in aggressive pitches, while mitigating catastrophic forgetting.

References

1. Dehaerne, E., Dey, B., Esfandiar, H., Verstraete, L., Suh, H.S., Halder, S., Gendt, S.D.: YOLOv8 for defect inspection of hexagonal directed self-assembly

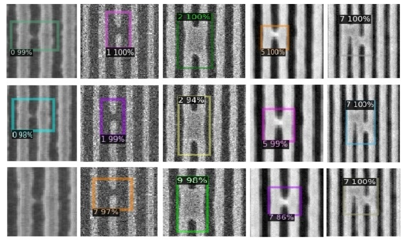


Fig. 4: Upper row: Model trained for defect detection on all classes at once.
Middle row: Model obtained after incremental training on task T_7^3.
Lower row: Model obtained after training on task F_7^3
Defect types (ground truth), left to right: Microbridge, Gap, Bridge, Thin bridge, Multi bridge non-horizontal.

patterns: a data-centric approach. In: Behringer, U.F., Loeschner, H., Finders, J. (eds.) 38th European Mask and Lithography Conference (EMLC 2023). vol. 12802, p. 128020S. International Society for Optics and Photonics, SPIE (2023). https://doi.org/10.1117/12.2675573, https://doi.org/10.1117/12.2675573

2. Dehaerne, E., Dey, B., Halder, S.: A comparative study of deep-learning object detectors for semiconductor defect detection. In: 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS). pp. 1–2 (2022). https://doi.org/10.1109/ICECS202256217.2022.9971022
3. Dey, B., Goswami, D., Halder, S., Khalil, K., Leray, P., Bayoumi, M.A.: Deep learning-based defect classification and detection in SEM images. In: Robinson, J.C., Sendelbach, M.J. (eds.) Metrology, Inspection, and Process Control XXXVI. vol. PC12053, p. PC120530Y. International Society for Optics and Photonics, SPIE (2022). https://doi.org/10.1117/12.2622550, https://doi.org/10.1117/12.2622550
4. Joseph, K.J., Rajasegaran, J., Khan, S., Khan, F.S., Balasubramanian, V.N.: Incremental object detection via meta-learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(12), 9209–9216 (2022). https://doi.org/10.1109/TPAMI.2021.3124133
5. Masana, M., Liu, X., Twardowski, B., Menta, M., Bagdanov, A.D., van de Weijer, J.: Class-incremental learning: survey and performance evaluation on image classification (2022)
6. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. p. 91–99. NIPS'15, MIT Press, Cambridge, MA, USA (2015)