Mitigating Catastrophic Forgetting in Advanced Node Semiconductor Defect Inspection Using YOLOv5 with Elastic Weight Consolidation

Ester Devlieghere, Bappaditya Dey, Victor Blanco, Sandip Halder
imec, Kapeldreef 75, 3001 Leuven, Belgium
bappaditya.dey@imec.be

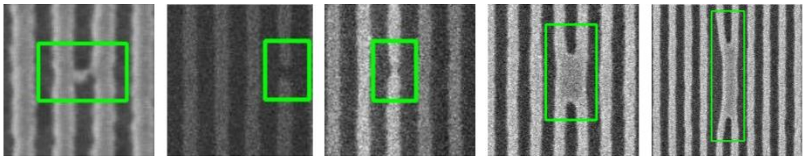
Ke Han, Jiangping Wang, Lingling Pu, Roger Lahaye
ASML, San Jose, California, USA

Abstract-As the demand for increasingly miniaturized and high-performance semiconductor devices intensifies, the patterns printed on wafers continue to shrink, making defect detection significantly more challenging[1]. In recent years, deep learning (DL)-based computer vision algorithms[1,2,3] have shown promising results, often outperforming conventional approaches in detecting nanometer-scale defects in scanning electron microscope (SEM) images at advanced technology nodes developed under the extreme ultraviolet lithography (EUVL) paradigm. However, conventional DL models are prone to catastrophic forgetting when trained sequentially on new tasks, such as the appearance of new datasets or new defect types. This issue poses a major challenge, as the common workaround involves storing previous datasets and multiple historical model versions to enable retraining from scratch or fine-tuning whenever new defect data appears. Such a storage-heavy strategy is impractical in high-volume manufacturing (HVM) settings due to substantial memory and data-management overhead. Continual Learning (CL) techniques[4] have recently emerged as a promising solution, yet many still depend on storing a small subset of samples per defect category or modifying model architectures during incremental updates, both of which remain impractical for deployment in real HVM settings. This work investigates elastic weight consolidation[5] (EWC), a regularization-based CL method that modifies only the loss function, enabling seamless integration into existing defect detectors. Given the limited and costly nature of semiconductor defect datasets, this study leverages two real-world SEM datasets from after-development inspection (ADI) and after-etch inspection (AEI) processing steps, both featuring line-space (LS) patterns. EWC is implemented within the one-stage object detector You Only Look Once (YOLO)v5, selected for its demonstrated effectiveness[6] in this application domain. The results show that EWC significantly improves retention of prior knowledge compared to non-continual learning baselines.

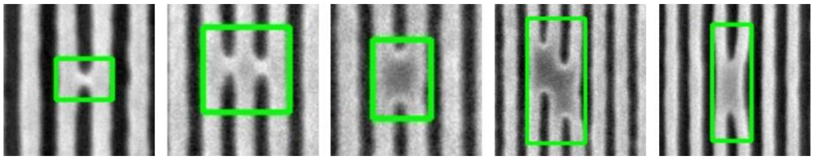
---

[1] B. Dey, D. Goswami, S. Halder, K. Khalil, P. Leray, and M. A. Bayoumi, Proc. SPIE 12053, 120530Y (2022).
[2] B. Dey, V. De Ridder, V. Blanco, S. Halder, and B. Van Waeyenberge, Proc. Int. Symp. ELMAR, 141 (2024).
[3] M. Hwang, B. Dey, E. Dehaerne, S. Halder, and Y. Shin, Proc. SPIE 12496, 1249608 (2023).
[4] B. Dey, A. Prasad, A. Belgharat, V. Blanco, and S. Halder, Proc. SPIE 13426, 134263J (2025).
[5] A. Aich, arXiv:2105.04093 (2021).
[6] B. Dey, E. Dehaerne, and S. Halder, Proc. SPIE 12293, 1229305 (2022).

Preliminary Results and Analysis:



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



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

Figure 1: Examples of defects in the ADI and AEI datasets.

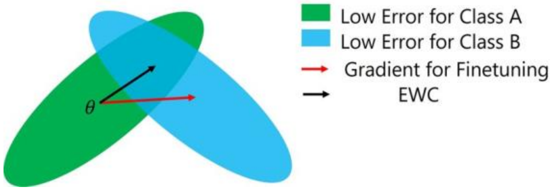
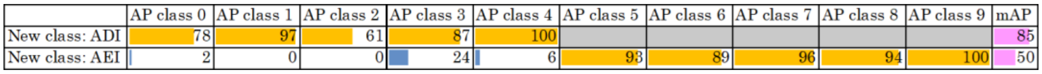
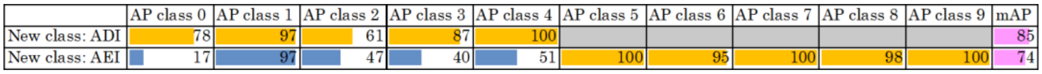


Figure 2: Conceptual illustration of EWC. The green and blue regions represent low-loss areas for tasks A and B, respectively. Standard training (red arrow) optimizes task B at the cost of forgetting A, while EWC (black arrow) constrains updates to preserve performance across both tasks[5].



	AP class 0	AP class 1	AP class 2	AP class 3	AP class 4	AP class 5	AP class 6	AP class 7	AP class 8	AP class 9	mAP
New class: ADI	78	97	61	87	100						85
New class: AEI	2	0	0	24	6	93	89	96	94	100	50

(a) Without EWC



	AP class 0	AP class 1	AP class 2	AP class 3	AP class 4	AP class 5	AP class 6	AP class 7	AP class 8	AP class 9	mAP
New class: ADI	78	97	61	87	100						85
New class: AEI	17	97	47	40	51	100	95	100	98	100	74

(b) With EWC

Figure 3: Comparison of the defect detector's performance with and without Elastic Weight Consolidation (EWC). The model was first trained on the ADI dataset and then sequentially on the AEI dataset. Performance metrics shown include detection accuracy and retention of previously learned defect classes.