Continual Learning with Out-of-Distribution Data Detection for Defect Classification

Cheng-Hsueh Lin*, Chia-Yu Lin†, Li-Jen Wang* and Ted T. Kuo*
*College of Artificial Intelligence, National Yang Ming Chiao Tung University, Tainan, Taiwan
†Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan
Corresponding Author: Chia-Yu Lin (sallylin0121@ncu.edu.tw)

2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan) | 979-8-3503-2417-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCE-Taiwan58799.2023.10226969

***Abstract*—We propose a framework for defect detection in production lines that leverages deep learning models, out-of-distribution (OOD) detection, and continual learning to address the challenges of unknown defects and catastrophic forgetting. The proposed method divides classifier training into chronicle tasks, each introducing new defect classes and leveraging OOD detection to classify unknown defects. We evaluate the framework on a highly unbalanced product defect dataset and demonstrated that it outperformed existing approaches, improving the average F-score by 10%. Our method also improve the performance of the PODNet and DER models, but not the WA model due to its poor performance on our dataset. These results suggest that the proposed method has the potential to improve defect detection in production lines, especially for small-quantity-wide-variety production scenarios.**

***Index Terms*—Continual learning, out-of-distribution,**

I. Introduction

Defect detection is essential in production lines. AOI has been traditionally used to detect defects. However, it can only determine whether products have defects but not categorize them. Deep learning models have been adopted to classify defects, but two challenges remain when applying them in production lines. The first challenge is the presence of outliers or unknown defects due to increased demand for small-quantity-wide-variety production. Most models are trained on closed sets of defects, which can lead to misclassification if an unknown defect is encountered. The second challenge is the catastrophic forgetting problem, where the model forgets accumulated knowledge when learning new ones. We propose a continual learning with an out-of-distribution (OOD) detection framework to address these challenges. The framework leverages OOD detection and continual learning to classify known and unknown defects. Classifier training is divided into multiple chronicle tasks, each triggering as enough new defect classes accumulate. In each task, we classify new defects or outliers as a designated class, the unknown class, and apply continual learning to classify them further.

Our framework is verified using a production dataset, showing promising results to outperform existing approaches. This framework enables manufacturers to improve defect detection in production lines, especially for the small-quantity-wide-variety production paradigm shift of demands by staging the classification of known and unknown defects.

II. Related Work

A. Out-Of-Distribution Detection

Out-Of-Distribution (OOD) detection is a technique used to identify defects that were not encountered during training. Various OOD methods employed scoring mechanisms, such as Maximum softmax probability (MSP) [1], to estimate scores based on the output of classifiers and determine if the defect is OOD data. Other methods, like Mahalanobis detector [2], projected the features onto an embedding space and used the Mahalanobis distance to identify OOD defects. Auxiliary datasets were also used to train the models for detecting OOD data, such as Outlier exposure (OE) [3] and Deep abstaining classifier (DAC) [4], which trained classifiers with auxiliary datasets to detect OOD data. DAC had several benefits, including detecting a wide range of OOD data, even dissimilar to the in-distribution data, without requiring excessive computational resources and complex implementation. Therefore, we have adopted a similar approach as the DAC for our OOD detection.

B. Continual Learning

Continual Learning-based classifiers aim to learn new classes without forgetting previously learned knowledge. Existing approaches could be categorized into three types: *replay-based*, *regularization-based*, and *parameter isolation-based*. Replay-based methods stored a subset of previously learned data to prevent forgetting. However, these methods could be computationally expensive and required careful data selection. Regularization-based approaches used additional terms to limit model growth and prevent catastrophic forgetting. Parameter isolation-based methods isolated task-specific parameters, preventing forgetting by dividing the network into modules.

Recent works have combined these approaches to improve the model, e.g., Weight alignment (WA) [5] and Pooled outputs distillation network (PODNet) [6] integrated replay-based and regularization-based methods by adding a distillation loss and training the model with memory. Dynamically expandable representation (DER) [7] combined replay-based and parameter isolation-based methods by freezing past feature extractors, copying the last extractors for learning new classes, and combining all features to train the classifier.

III. Method

The proposed framework divide the continual learning process into sequential training tasks chronically. In the first task,

the model is trained to classify b classes of defects. However, during the training, an OOD detector is used to create an additional class, $Other$, for OOD defects unlikely to belong to any of the b classes. In the following tasks, n new defect classes are introduced along with the previously b known classes to train the new classifier for $(b'+1)$ classes of defects, where $b' = b + n$ and the $+1$ represents the $Other$ class.

The problem is set up as follows: a base model $\mathcal{M}^0$ trained by the dataset $\mathcal{D}^0$ with the class set $\mathcal{C}^0$, which contains b classes of defects initially. After enough new classes of defects are collected, a training task, t, starts. Let T represent the task set of all tasks and D^t be the dataset used in task t. In each task, the dataset D^t contains n new classes of samples than D^{t-1}. The objective of each task t is to help the classifier to recognize the additional n new classes in $\mathcal{C}^t$ while maintaining the previously learned knowledge to classify the $b'^{(t-1)}$ classes in $\mathcal{C}^{old}$, where $\mathcal{C}^{old} = \bigcup_{i=1}^{t-1} \mathcal{C}^i$.

Our OOD detector is trained following the DAC [4] approach, but we choose not to use orthogonal datasets. Instead, we train the detector using newly collected defects that were previously unseen by the model. To $Other$ class, we construct the class sets $\mathcal{C}^{OOD}$ for each task t, where $\mathcal{C}^{OOD} = \bigcup_{i=t+1}^{T} \mathcal{C}^i$. This approach more accurately reflects the small-quantity-wide-variety production lines. Our experiments demonstrate that the OOD detector achieved higher F-scores, even for future scarce defects. Additionally, the framework leverages WA [5], PODNet [6], and DER [7] as a baseline backbone to classify incremental unknown classes, resulting in improved model accuracy.

IV. Experiments

We conduct experiments on a highly unbalanced product defect dataset comprising 35,316 images and 27 classes. To mitigate the effect of class imbalance, we augment the training set images using flip, rotate, and lightness techniques for classes with less than 500 images. We use ResNet18 as the backbone network and train the model using a batch size of 64 images, randomly resized and cropped into 224x224 pixels in size, and update by the Adam optimizer with a learning rate 1e-3 controlled by cosine annealing scheduler for 500 epochs each task. To emulate the small-quantity-wide variety production scenarios, our OOD detector is initially trained with $b = 16$ classes but test with the entire dataset containing full dataset of 27 classes for its capability to classify the b^t classes and $Other$ for those unseen defects in a task, t. Furthermore, we incrementally add three new classes, i.e., $n = 3$, while keeping 20 images per learned class in the exemplar set to train the incremental learning model to classify the n classes while keeping the knowledge of b^{t-1} classes in each task.

Our framework improves the average F-score by 10% for all methods, with DER performing the best. Compared to the original DER, our method increases last accuracy by 9.44%, average accuracy by 15.7%, and average F-score by 17.07%, accurately predicting OOD data. Our method also improves PODNet, but not WA due to its poor performance on our dataset. We attribute DER's superior performance to its replay and parameter isolation approach. Table I summarizes the three methods' performance, while Table II compares OOD with DER F-score with and without our training method, showing an improvement of at least 47% for each task.

TABLE I
Experiment Results

Methods	*Last Acc.*	*Avg Acc.*	*Avg F-score*
WA [5]	55.2	68.36	54.23
PODNet [6]	78.5	81.32	63.07
DER [7]	70.2	79.54	71.54
Ours (WA)	54	65.36	64.06
Ours (PODNet)	76.5	86.36	74.37
Ours (DER)	**79.64**	**95.24**	**88.61**

TABLE II
DER's F-score w/ and w/o the proposed OOD Dection

Methods	Task 1	Task 2	Task 3	Task 4	Task 5
DER [7]	43.4	15.75	12.92	5.78	5.47
Ours (DER)	91.09	83.87	75.28	60	57.46

V. Conclusion

We proposed a framework to detect out-of-distribution (OOD), and address the challenges of unknown defects and catastrophic forgetting for production lines. The experimental results showed that our framework generally improved the performance of existing continual learning methods. In the future, we can enhance the model to handle the defect datasets that contain subtle differences and cannot be classified well.

Acknowledgment

This work is jointly sponsored by AUO Corporation, AUO · NYCU Joint Research and Development Center, National Central University, and National Science and Technology Council (NSTC) under the project NSTC 111-2622-8-A49-023 and NSTC 110-2222-E-008-008-MY3.

References

[1] Dan Hendrycks and Kevin Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *ArXiv*, vol. abs/1610.02136, 2016.

[2] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," *Advances in neural information processing systems*, vol. 31, 2018.

[3] Dan Hendrycks, Mantas Mazeika, and Thomas G. Dietterich, "Deep anomaly detection with outlier exposure," *ArXiv*, vol. abs/1812.04606, 2018.

[4] Sunil Thulasidasan, Sushil Thapa, Sayera Dhaubhadel, Gopinath Chennupati, Tanmoy Bhattacharya, and Jeff A. Bilmes, "An effective baseline for robustness to distributional shift," *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2021.

[5] Bowen Zhao, Xi Xiao, Guojun Gan, Bin Zhang, and Shutao Xia, "Maintaining discrimination and fairness in class incremental learning," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[6] Arthur Douillard, Matthieu Cord, Charles Ollion, Thomas Robert, and Eduardo Valle, "Podnet: Pooled outputs distillation for small-tasks incremental learning," in *European Conference on Computer Vision*.

[7] Shipeng Yan, Jiangwei Xie, and Xuming He, "Der: Dynamically expandable representation for class incremental learning," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.