Characterizing Continual Learning Scenarios for Tumor Classification in Histopathology Images

Veena Kaustaban[1]   VEENA.KAUSTABAN@ROCHE.COM
Qinle Ba[1]   QINLE.BA@ROCHE.COM
Ipshita Bhattacharya[1*]   IPSHITA.SB.BHATTACHARYA@GMAIL.COM
Nahil Sobh[1]   NAHIL.SOBH@CONTRACTORS.ROCHE.COM
Satarupa Mukherjee[1]   SATARUPA.MUKHERJEE@ROCHE.COM
Jim Martin[1]   JIM.MARTIN@CONTRACTORS.ROCHE.COM
Mohammad Saleh Miri[1]   SALEH.MIRI@ROCHE.COM
Christoph Guetter[1]   CHRISTOPH.GUETTER@ROCHE.COM
Amal Chaturvedi[1]   AMAL.CHATURVEDI@ROCHE.COM
[1]Roche Sequencing Solutions, Santa Clara, CA

Editors: Under Review for MIDL 2023

Abstract

Deep-learning models have achieved unprecedented performance in fundamental computational tasks in digital pathology (DP) based analysis, such as image classification, cell detection and tissue segmentation. However, such models are known to suffer from catastrophic forgetting when adapted to unseen data distribution with transfer learning. With an increasing need for deep-learning models to handle ever-changing data distributions, including evolving patient population and new diagnosis assays, it is crucial to introduce methods for alleviating the such model forgetting. To this end, continual learning (CL) models are promising candidates. However, to our best knowledge, there's no systematic study of CL models in DP-specific applications. Here, we propose various CL scenarios in DP settings, where histopathology image data from different sources/distributions arrive sequentially, the knowledge of which is integrated into a single model without training all the data from scratch. To benchmark the performance of recently proposed continual learning algorithms in the proposed CL scenarios, We augmented a dataset for colorectal cancer H&E classification to simulate shifts of image appearance and evaluated CL methods on this dataset. Furthermore, we leveraged a breast cancer H&E dataset along with the colorectal cancer dataset to assess continual learning methods for learning from multiple tumor types. We revealed promising results of CL in DP applications, potentially paving the way for application of these methods in clinical practice.

Keywords: Digital Pathology, Continual Learning, Tumor Tissue Classification

1. Introduction

A general strategy to learn from multiple datasets considers training all the existing and newly arrived data from scratch, leading to increasingly high demand of computing resource, time and data storage. On the other hand, updating a model with transfer learning is known to be ineffective due to catastrophic forgetting (Kirkpatrick et al., 2017). On the contrary, continual learning algorithms (Kirkpatrick et al., 2017; Chaudhry et al., 2018; Lopez-Paz and Ranzato, 2017; Rebuffi et al., 2017) aim at efficiently and effectively adapting a model

---

* Work done at Roche

to new data streams without forgetting of learned knowledge. These algorithms have largely been tested on less complex datasets from non-biomedical domains. Here, we report the first study to assess the feasibility of CL in DP-based applications (Kaustaban et al., 2022).

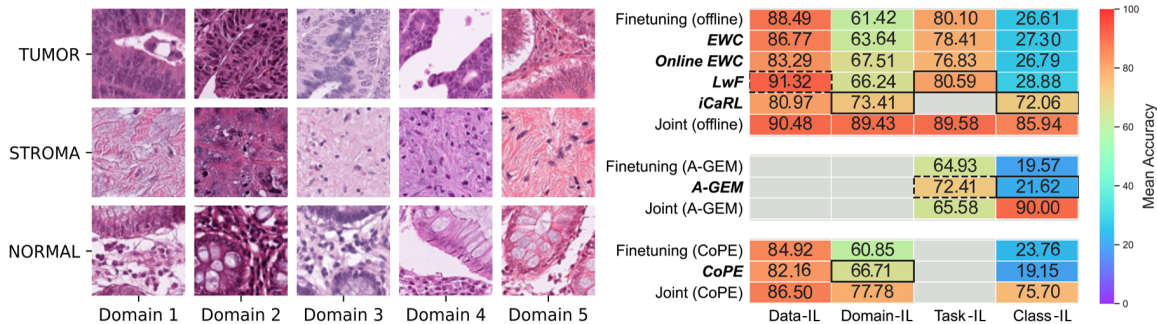|	Data-IL	Domain-IL	Task-IL	Class-IL
Finetuning (offline)	88.49	61.42	80.10	26.61
EWC	86.77	63.64	78.41	27.30
Online EWC	83.29	67.51	76.83	26.79
LwF	91.32	66.24	80.59	28.88
iCaRL	80.97	73.41		72.06
Joint (offline)	90.48	89.43	89.58	85.94
Finetuning (A-GEM)			64.93	19.57
A-GEM			72.41	21.62
Joint (A-GEM)			65.58	90.00
Finetuning (CoPE)	84.92	60.85		23.76
CoPE	82.16	66.71		19.15
Joint (CoPE)	86.50	77.78		75.70

Figure 1: Left: Example augmented CRC images from 5 domains. Right: Test accuracy of offline and online methods on augmented CRC against their corresponding baselines.

2. Methods

Dataset 1 - CRC. A colorectal cancer (CRC) dataset (Kather et al., 2019) was used to generate simulated data streams, which comprise whole-slide H&E images (136 patients) with tile-wise class labels for 9 tissue types. We randomly selected 8700 images (7000 train; 2700 validation) from each class for training and 7150 images for testing. **Dataset 2 - PatchCam.** To characterize the effectiveness of CL methods on learning from multiple tissue types, PatchCam from (Bejnordi et al., 2017) was used for breast cancer classification after stain-normalization (Macenko et al., 2009), which has class labels of "normal" and "tumor" from 400 breast cancer H&E whole-slide images. **Augmented CRC - Simulating**
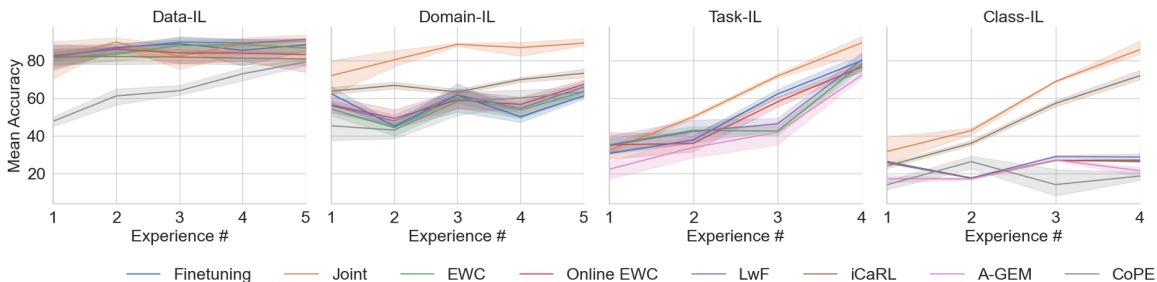

Figure 2: Online and offline CL performance at each experience against offline baselines.

Domain Shifted Data Streams. We generated an augmented CRC dataset to simulate the commonly observed variations of stain appearance from multiple data sources (Fig. 1 Left). Specifically, we first performed stain unmixing of CRC images and then applied augmentation in the optical density space before remixing the stain-unmixed intensity images. In addition to the original images (Domain 1), we generated images simulating different

dye concentrations for H&E stain (Domain 2: increased stain intensity), faded stains due to slide aging (Domain 3: decreased eosin intensity) and changes of reagent manufacturer, scanners or stainers (Domain 4: change in hue; Domain 5: change in hue and saturation). **Continual Learning with Augmented CRC.**: We identified four CL scenarios suitable for DP applications based on how new streams of data differ from the previous ones. Specifically, data streams contain (1) an equal number of images from all domains (Data-IL), (2) an equal number of images from one domain (Domain-IL), (3) images from new classes from all domains (Class-IL), or (4) images from new classes from all domains for each of the multi-task heads (Task-IL). Please see (Kaustaban et al., 2022) for more details. **Continual Learning Algorithms.**: Recently proposed CL methods can be largely classified into three categories: replay, regularization and parameter isolation (Vokinger et al., 2021).

- *Replay methods* select original images (Rebuffi et al., 2017), deep representations (Van de Ven and Tolias, 2019) or model-generated pseudo samples (Shin et al., 2017) via various heuristics, which are stored in memory and replayed at later learning stages to overcome forgetting. *iCaRL* stores a subset of most representative examples per class in memory. *A-GEM* constraints model updates by projecting the estimated gradient on the direction determined by randomly selected samples from a replay buffer. *CoPE* enables rapidly evolving prototypes with balanced memory and a loss function that updates the representations or pseudo prototypes.

- *Regularization methods* include a regularization term in the loss function to penalize model updates that could lead to large deviation from an existing model to avoid forgetting of learned knowledge. *EWC/Online EWC* includes a regularization term in the loss function to penalize large changes to network weights that are important for previous tasks. *LwF* distills knowledge from a previous model to an updated model trained on new data with an additional distillation loss term for replayed data.

- *Parameter isolation methods* assign different model parameters to each task head. Note that parameter isolation only applies to Task-IL.

3. Results

To our surprise, continual learning was most effective for the challenging Class-IL scenario (Figure 1 Right). Comparing offline CL methods, iCaRL had the best overall performance, even comparable to upper-bound joint training in the Class-IL scenario (Figure 1 Right: top block, Figure 2). For online CL (Figure 1 Right: lower blocks), which learned from hundreds of small data batches in a near continuous stream, we observed that A-GEM outperformed upper-bound baseline in Task-IL scenario (Figure 2). However, online methods did not outperform offline ones. We further assessed sequential learning from multiple tumor types in domain-IL for CRC and breast cancer datasets and found that (a) adding more examples and (b) "hard to easy curriculum" generated better results.

In summary, we found that continual learning for DP, despite being challenging, is feasible and promising. CL methods were computationally efficient, taking only about 28% of the runtime as joint training. Though patient data evolve quickly nowadays, FDA has not approved algorithms based on CL (Vokinger et al., 2021) and extensive research is needed to establish related regulations. Our evaluation approaches and proposed method to generate domain-shifted datasets can potentially serve as the first step towards this goal.

References

Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22): 2199–2210, 2017.

Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with A-GEM. *arXiv preprint arXiv:1812.00420*, 2018.

Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1):e1002730, 2019.

Veena Kaustaban, Qinle Ba, Ipshita Bhattacharya, Nahil Sobh, Satarupa Mukherjee, Jim Martin, Mohammad Saleh Miri, Christoph Guetter, and Amal Chaturvedi. Characterizing continual learning scenarios for tumor classification in histopathology images. In *Medical Optical Imaging and Virtual Microscopy Image Analysis: First International Workshop, MOVI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings*, pages 177–187. Springer, 2022.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE international symposium on biomedical imaging: from nano to macro*, pages 1107–1110. IEEE, 2009.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. iCaRL: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *Advances in neural information processing systems*, 30, 2017.

Gido M Van de Ven and Andreas S Tolias. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734*, 2019.

Kerstin N Vokinger, Stefan Feuerriegel, and Aaron S Kesselheim. Continual learning in medical devices: FDA's action plan and beyond. *The Lancet Digital Health*, 3(6):e337–e338, 2021.