Efficient Meta Continual Learning on the Edge

Young D. Kwon
Computer Laboratory, University of Cambridge
United Kingdom
ydk21@cam.ac.uk

ABSTRACT

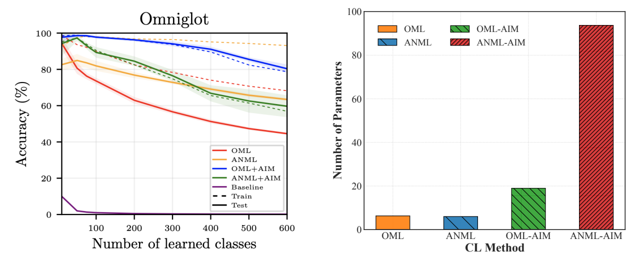
Continual Learning (CL) methods are designed to help deep neural networks adapt and learn new tasks/knowledge without forgetting previously learned tasks. In recent years, researchers have proposed many CL methods: (1) regularization-based CL, (2) rehearsal-based CL, and (3) dynamic architecture-based CL. However, those CL methods typically require a moderate or large number of training samples to learn new tasks/classes since they are essentially a supervised learning approach. This limits its applicability to real-world applications that run on the edge, where the labeled user data is not abundant. Hence, Meta CL methods are proposed to solve the limitation by reducing the amount of required training data to a few samples (e.g., 10-20 samples). However, Meta CL methods also have limitations. Thus, in this work, I first identify the limitations of the Meta CL methods, i.e., they require larger model sizes. Then I propose potential directions to tackle this challenge by ensuring high performance while minimizing the storage overhead.

CCS CONCEPTS

• **Computer systems organization → Embedded and cyber-physical systems**.

KEYWORDS

Continual Learning, Quantization, Edge Computing.



(a) Performance [12] **(b) Storage Overhead**

Figure 1: Preliminary analysis of the Meta CL methods.

1 INTRODUCTION

With the rise of mobile, wearable devices, and the Internet of Things (IoT), the proliferation of sensory type data has fostered the adoption of deep neural networks (DNN) in the modeling of a variety of mobile sensing applications [11]; Then, a crucial characteristic common to the mobile applications on the edge is the need for a trained model to adapt to accommodate new classes/tasks and to a dynamically changing environment. In these settings, the ability to *continually* learn [7, 9, 13], that is, to learn consecutive tasks without forgetting how to perform previously learned tasks, becomes essential.

2 MOTIVATION & METHOD

To enable Continual Learning (CL), many approaches are proposed in the literature. The first group of CL methods is a *regularization-based CL* [8, 16, 18] where regularization terms are added to the loss function to minimize changes to important weights of a model for previous tasks to prevent forgetting. Another group of approaches is a *rehearsal-based CL* [2, 3, 10, 14] where updating the model requires training data from the new class and also a few training samples from earlier classes. Lastly, *dynamic architecture-based CL* [4, 15, 17] modifies the architecture of a model to make it learn new knowledge without interfering with old ones. Although the proposed CL methods largely improve the forgetting issue of a learned model, they are limited since labeled training data is required to learn new tasks/classes continuously. Also, the number of required samples for new tasks is large. Hence, the applicability of the CL is limited to real-world mobile applications where labeled user data is scarce and the computing resources for training are constrained. Also, For rehearsal-based CL, which generally ensures high accuracy, requires exemplars to be saved. This incurs additional memory and storage.

Hence, Meta CL [1, 6, 12] is proposed to resolve the challenges of the CL methods, alleviating the issues mentioned above by relying on only a few samples of new classes. However, as shown in Figure 1a, Meta CL's performance degrades when a large number of classes are added. New Meta CL method (OML-AIM and ANML-AIM) [12] is proposed to ensure high accuracy, however, it uses 3-15x more parameters compared to OML [6] and ANML [1]. For example, the models of OML and ANML have 6M parameters, whereas OML-AIM's model has 19M parameters, and ANML-AIM's model contains 93M parameters as shown in Figure 1b.

In this work, I propose several directions to solve the limitations and challenges of the CL methods. First, to solve the performance degradation problem of OML and ANML, I want to extend the current Meta CL methods as rehearsal-based Meta CL so as to ensure high performance without making the model too large. Besides, I plan to apply model compression techniques (e.g., quantization [5]) to minimize the memory and storage overheads that are incurred due to the stored exemplars.

3 CONCLUSION

Through the extensive literature review, I first identified the limits of the three CL methods and recently developed Meta CL methods. Then, I listed several promising directions to start investigating to improve the limitations of the CL methods.

REFERENCES

[1] Shawn Beaulieu, Lapo Frati, Thomas Miconi, Joel Lehman, Kenneth O. Stanley, Jeff Clune, and Nick Cheney. 2020. Learning to Continually Learn. *arXiv:2002.09571 [cs, stat]* (March 2020).

[2] Jagmohan Chauhan, Young D. Kwon, Pan Hui, and Cecilia Mascolo. 2020. ContAuth: Continual Learning Framework for Behavioral-based User Authentication. *Proc. IMWUT* 4, 4 (Dec. 2020), 122:1–122:23.

[3] Tyler L. Hayes, Kushal Kafle, Robik Shrestha, Manoj Acharya, and Christopher Kanan. 2020. REMIND Your Neural Network to Prevent Catastrophic Forgetting. In *Computer Vision – ECCV 2020 (Lecture Notes in Computer Science)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 466–483.

[4] Ching-Yi Hung, Cheng-Hao Tu, Cheng-En Wu, Chien-Hung Chen, Yi-Ming Chan, and Chu-Song Chen. 2019. Compacting, Picking and Growing for Unforgetting Continual Learning. *Advances in Neural Information Processing Systems* 32 (2019). https://proceedings.neurips.cc/paper/2019/hash/3b220b436e5f3d917a1e649a0dc0281c-Abstract.html

[5] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. 2017. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. *CoRR* abs/1712.05877. arXiv:1712.05877

[6] Khurram Javed and Martha White. 2019. Meta-Learning Representations for Continual Learning. In *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 1820–1830.

[7] Ronald Kemker, Marc McClure, Angelina Abitino, Tyler L Hayes, and Christopher Kanan. 2018. Measuring catastrophic forgetting in neural networks. In *Thirty-second AAAI conference on artificial intelligence*.

[8] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. 2017. Overcoming catastrophic forgetting in neural networks. *Proc. National Academy of Sciences* 114, 13 (March 2017), 3521–3526.

[9] Young D Kwon, Jagmohan Chauhan, Abhishek Kumar, Pan Hui, and Cecilia Mascolo. 2021. Exploring System Performance of Continual Learning for Mobile and Embedded Sensing Applications. In *ACM/IEEE Symposium on Edge Computing*. Association for Computing Machinery (ACM).

[10] Young D. Kwon, Jagmohan Chauhan, and Cecilia Mascolo. 2021. FastICARL: Fast Incremental Classifier and Representation Learning with Efficient Budget Allocation in Audio Sensing Applications. In *Proc. Interspeech 2021*. 356–360.

[11] Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications Magazine* 48, 9 (Sept. 2010), 140–150.

[12] Eugene Lee, Cheng-Han Huang, and Chen-Yi Lee. 2021. Few-Shot and Continual Learning with Attentive Independent Mechanisms. *arXiv:2107.14053 [cs]* (July 2021). http://arxiv.org/abs/2107.14053

[13] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. 2019. Continual lifelong learning with neural networks: A review. *Neural Networks* 113 (May 2019), 54–71.

[14] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. 2017. iCaRL: Incremental classifier and representation learning. In *Proc. CVPR*. 2001–2010.

[15] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. 2016. Progressive neural networks. *arXiv preprint arXiv:1606.04671* (2016).

[16] Sandra Servia-Rodriguez, Cecilia Mascolo, and Young D. Kwon. 2021. Knowing when we do not know: Bayesian continual learning for sensing-based analysis tasks. *arXiv:2106.05872 [cs]* (June 2021).

[17] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. 2018. Lifelong Learning with Dynamically Expandable Networks. (Feb. 2018).

[18] Friedemann Zenke, Ben Poole, and Surya Ganguli. 2017. Continual Learning Through Synaptic Intelligence. In *Proc. ICML*. 3987–3995.