Mini Project Proposal: Investigating the Effect of Latent Representations on Continual Learning Performance

By Henry Bourne, Supervised by Rihuan Ke

1 Introduction to Continual Learning

Deep learning [5] over the past decade has been the foremost technology used in the creation of "intelligent" systems capable of solving previously unsolvable tasks such as protein folding [13] and superhuman performance at games such as chess and go.

Although a very powerful method, deep learning isn't without its weaknesses. One of these weaknesses is its poor performance in sequential learning, this is where we train the network on data pertaining to one "task" and then go on to train it on data from other tasks. It has been found that when we test the network on data from a previous task the performance it obtains is very low (in comparison to the performance it has straight after being trained on the task). This weakness is known as the problem of **Catastrophic Forgetting (CF)** [3, 7, 9], and the reason it comes about is because when trained on a new task the neural network will change the networks weights such that the network performs optimally on the new task, however, in the process it changes weights such that the network will no longer perform well for the previous task.

If we could create a technique that solved the problem of catastrophic forgetting then we would have a technique that could continually learn (ie. train on tasks sequentially with good performance across all the tasks its ever been trained on) hence we aptly name techniques that aim to solve the problem of catastrophic forgetting as **Continual Learning (CL)** techniques.

CL techniques become invaluable in a multitude of scenarios: such as when you don't have access to all the data that you want to train on at training time, when you can't store data due to data privacy reasons, when datasets are so large you can't store it all in one place, when the tasks you want to be able to perform aren't all known at initial training time, etc.

2 The Literature

The literature in the area of CL is growing quickly as CF becomes a more pressing issue. A good survey of techniques pre 2022 is provided in [2]. Per the taxonomy in the survey most CL techniques fall into one of three categories:

1. **Replay based methods:** which mainly focus on keeping old data and retraining the network on this old data (from previous tasks) so that the network doesn't loose performance [10, 12].
2. **Regularization-based methods:** which focus on adding an extra regularization term to the loss function such that the network consolidates on previous knowledge when learning on new data [4, 14].
3. **Parameter isolation methods:** which focus on dedicating model parameters to different tasks to alleviate possible overwriting [6, 11] .

There also exist many sub-approaches to each of these approaches and approaches that incorporate elements of multiple of these approaches. These approaches are also applied to a whole host of problems, however, in this brief and in the mini-project we will focus on the specific problem of image classification.

3 Our Proposal

Often CL techniques will make use of an encoder [1] during empirical testing to simplify the downstream task of continual learning and has been shown in the replay setting for example to dramatically reduce compute with on par performance to networks trained end-to-end in certain scenarios [8]. It is often argued that adding an encoder to the architecture simplifies the downstream task (of image classification and CL). The encoder can be implemented in many ways but a classical setup is to feed images to the encoder and then to train the architecture on the output of the encoder, the output of the encoder is referred to as the latent space of the input (the image).

What we propose to investigate is to what degree does learning in latent space (eg. training the architecture on the output of an encoder) make simpler the downstream task of CL. We propose conducting an empirical investigation in order to say something rigorous about the effects of "learning in latent space" on CL performance (both in terms of accuracy and compute). If there is benefit to "learning in latent space", ie. there is evidence that it does in fact make the downstream task of CL more simple, we will also investigate *why* it might make it easier to perform CL.

References

[1] Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *arXiv preprint arXiv:2003.05991*, 2020.

[2] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021.

[3] Robert M French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135, 1999.

[4] 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.

[5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[6] Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7765–7773, 2018.

[7] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, pages 109–165. Elsevier, 1989.

[8] Oleksiy Ostapenko, Timothee Lesort, Pau Rodríguez, Md Rifat Arefin, Arthur Douillard, Irina Rish, and Laurent Charlin. Continual learning with foundation models: An empirical study of latent replay. In *Conference on Lifelong Learning Agents*, pages 60–91. PMLR, 2022.

[9] Roger Ratcliff. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological review*, 97(2):285, 1990.

[10] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[11] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *International Conference on Machine Learning*, pages 4548–4557. PMLR, 2018.

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

[13] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1):D439–D444, 2022.

[14] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International conference on machine learning*, pages 3987–3995. PMLR, 2017.