

Title: Frugal Learning with Deep Generative Networks for Visual Scene Recognition

CNRS LIP6, UPMC Sorbonne University

Deep learning is currently witnessing a major success in different computer vision tasks including image and video classification [1]. The purpose of deep learning is to train convolutional or transformer networks that map raw data into suitable representations prior to their classification [2]. However, the success of these networks is highly dependent on the availability of large collections of hand-labeled training data that capture the distribution of the learned categories. In many practical scenarios, mainly those involving streams of data (videos, etc.), large collections covering the inherent variability of the learned categories are neither available nor can be holistically processed. Hence, training deep networks should be achieved as a part of a frugal lifelong process, a.k.a. continual or incremental learning [3, 4].

The traditional mainstream design of deep networks is based on back-propagation and stochastic gradient descent. The latter collects gradients through mini-batches and updates network parameters in order to learn different categories. However, in lifelong learning, tasks involve only parts of data/categories, and this potentially leads to catastrophic forgetting (CF) defined as the inability of a learning model to “memorize” previous tasks when handling new ones. In cognitive science, CF is considered as an extreme case of the stability-plasticity dilemma [5, 6] where excessive plasticity causes an easy fit to new knowledge and less on previous ones. This is also related to concept (or distribution) drift [7] that may happen when a learning model keeps ingesting data [8, 9].

Whereas in most of the learning models (especially shallow ones [10, 11, 12]), CF could be overcome, its handling in deep neural networks is a major challenge and existing solutions can only mitigate its effect. Indeed, CF results from the high non-linearity and entanglement of gradients when achieving back-propagation in deep networks (in contrast to shallow ones). Existing straightforward solutions bypass this effect by storing huge collections of data and replaying the learning process using all these collections; whereas replay is highly effective, it is known to be time and memory demanding and may result into resource saturation even on sophisticated hardware devices. Other solutions, with less time and memory footprint (e.g., regularization) can only mitigate the effect of CF. Another category of methods, based on dynamic networks provides a suitable balance between resource consumption and task memorization.

The goal of this thesis subject is to design novel deep continual learning models for visual recognition [13]. One of the main challenges is to design discriminative as well as generative networks that learn visual categories effectively while also attenuating catastrophic forgetting. The proposed solutions will be built upon deep variational auto-encoders (VAE) and generative adversarial networks (GANs) that allow mitigating CF with a reasonable growth in the number of training parameters, and memory footprint. The objectives also include (but not limited to)

- The design of new GANs/VAEs for image generation, augmentation and replay.
- The disentanglement and interpretation of different factors in these generative models including semantics, appearances and dynamics of the visual contents.
- The inclusion of these generative models in a whole framework that achieves continual learning while also handling the challenging issue of catastrophic forgetting.
- These networks should be designed in order to run, not only on standard GPUs, but also on edge devices including mobile phones and connected objects endowed with low computational and energy resources.
- Applications include continual object detection, image classification and segmentation in still and video sequences.

Keywords. Deep machine learning, frugal and continual learning, generative adversarial networks, variational auto-encoders, visual scene recognition, image and video classification and segmentation, etc.

Thesis Director: Hichem SAHBI, CNRS Researcher, HDR, LIP6 Lab, Sorbonne University

Contact: hichem.sahbi@lip6.fr

PhD Student Background. We are seeking a highly motivated PhD candidate, with a preferred background in applied mathematics or computer science with more emphasis on statistics, machine learning and/or image processing, computer vision, and familiar with existing machine learning tools and programming platforms.

Related bibliography

- [1] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 831–839, 2019.
- [2] Huaiyu Li, Weiming Dong, and Bao-Gang Hu. Incremental concept learning via online generative memory recall. ArXiv preprint arXiv:1907.02788, 2019.
- [3] 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.
- [4] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. arXiv preprint arXiv:1606.04671, 2016.
- [5] Gail A Carpenter and Stephen Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer vision, graphics, and image processing, 37(1):54–115, 1987.
- [6] Xilai Li, Yingbo Zhou, Tianfu Wu, Richard Socher, and Caiming Xiong. Learn to grow: A continual structure learning framework for overcoming catastrophic forgetting. arXiv preprint arXiv:1904.00310, 2019.
- [7] Chrisantha Fernando, Dylan Banarse, Charles Blundell, Yori Zwols, David Ha, Andrei A Rusu, Alexander Pritzel, and Daan Wierstra. Pathnet: Evolution channels gradient descent in super neural networks. arXiv preprint arXiv:1701.08734, 2017.
- [8] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
- [9] Jathushan Rajasegaran, Munawar Hayat, Salman H Khan, Fahad Shahbaz Khan, and Ling Shao. Random path selection for continual learning. In Advances in Neural Information Processing Systems, pages 12669–12679, 2019.
- [10] Hichem Sahbi and Donald Geman. "A Hierarchy of Support Vector Machines for Pattern Detection." Journal of Machine Learning Research 7.10 (2006).
- [11] Hichem Sahbi, Jean-Yves Audibert, and Renaud Keriven. "Context-dependent kernels for object classification." IEEE transactions on pattern analysis and machine intelligence 33.4 (2010): 699-708.
- [12] Hichem Sahbi. Learning laplacians in chebyshev graph convolutional networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2064–2075, 2021.
- [13] Hichem Sahbi and Haoming Zhan. FFNB: Forgetting-free Neural Blocks for Deep Continual Learning. In the proceedings of the British Machine Vision Conference (BMVC), 2021.