Computer Vision Center
(Universitat Autonoma de Barcelona)
Edifici O, Campus UAB
08193 - Bellaterra - Cerdanyola - Spain

29th July 2025

Thesis Review

Title: Data representations in non-stationary optimization
Author: Wojciech Masarczyk
Summary: The quality of the thesis is excellent and I recommend its acceptance. I also support awarding the degree Cum Laude.

The introduction clearly proposes the main research question of the thesis, and briefly outlines the main contributions for each of the chapters. In Chapter 2, the thesis explores whether synthetic data can be generated in a way that prevents forgetting when learned in sequence. They introduce a two-step optimization method using meta-gradients to create such data. The chapter is more exploratory in nature, and the preliminary results suggest that synthetic data can be used to mitigate forgetting. However, experiments on larger datasets and a deeper analysis are required for this to be a solid contribution.

In Chapter 3, the thesis compares the quality of generative and discriminative representations for the task of continual learning, considering both catastrophic forgetting (negative backward transfer) and generalization (positive forward transfer). Discriminative models forget considerably more than generative ones, highlighting the latter's resilience. The analysis includes both per task performance analysis and Centered Kernel Alignment analysis. The conclusions of this chapter are important because they also generalize to other non-discriminative representation learning strategies, such as self-supervised learning. This work has contributed to a better understanding of the strength of these representations within a continual learning context. The work has been published in the International Conference on Neural Information Processing.

Chapter 4 focuses on an interesting experiment, where a learner is first learning a task, and subsequently is provided with data with artificially simple features to solve the same classification problem. The results of this experiment provide intriguing insights into deep neural networks. This way of training leads to feature erosion, a dramatic drop in feature quality. The network focuses on the simple feature and forgets the previously learned features used for classification. In addition, experiments show that this hurts the plasticity of the network, permanently damaging the network.

Chapter 5 makes several fundamental observations on network training. Neural networks divide into two stages: early layers craft linearly separable features, and then a tunnel of later layers compresses them. This tunnel appears early in training and contributes little to final accuracy despite its depth. Its length grows when the network's capacity greatly exceeds the task's intrinsic complexity (the extractor remains constant, only becomes smaller for wider networks). Interestingly, the tunnel hurts out-of-distribution generalization and can aggravate continual learning forgetting. The conclusions of this chapter may have consequences for rethinking network depth, especially in the context of continual learning. The task-agnostic nature of the compression is particularly fascinating. This chapter has been published at NeurIPS 2023.

Chapter 6 studies the impact of the softmax function on representation collapse, learning dynamics, compression, and generalization to OOD data. The paper shows the surprising effect of the logit norm at initialization on the learning dynamics of deep neural networks. This can have profound effects on the generalization of the representation. The chapter then shows that adjusting the softmax temperature can steer the learning dynamics and lead to more generalizable representations that do not suffer from rank collapse. The chapter also establishes a fundamental trade-off between OOD generalization and OOD detection, highlighting the role of softmax temperature in balancing this trade-off.

Conclusion: This thesis is of very high quality. The thesis addresses several fundamental aspects of deep neural network training and identifies previously unobserved phenomena that deepen our understanding and can have a considerable impact on AI in general. The thesis is well written, and the graphics are of high quality. The empirical evaluation of the posed scientific hypotheses is extensive. The analysis is of high quality and scientific depth and will certainly lead to follow-up research within the community. The thesis has led to several publications, including one at NeurIPS, the most important and impactful AI conference in the field. **I consider the quality of the thesis to be excellent and recommend its acceptance. Furthermore, given the exceptional quality of the work, I would also support awarding the degree Cum Laude.**

JOOST VAN DE
WEIJER - DNI
X9417760L

Digitally signed by
JOOST VAN DE WEIJER
- DNI X9417760L
Date: 2025.07.29
08:42:10 +02'00'

Joost van de Weijer
Senior Scientist at the Computer Vision Center, Universitat Autònoma de Barcelona
Group Leader of the Learning and Machine Perception team
http://www.cvc.uab.es/LAMP/joost/