Learning Symbolic Representations Through Joint GEnerative and DIscriminative Training (Extended Abstract)

Emanuele Sansone[1], **Robin Manhaeve**[1]
KU Leuven
{emanuele.sansone, robin.manhaeve}@kuleuven.be

Abstract

We introduce **GEDI**, a Bayesian framework that combines existing self-supervised learning objectives with likelihood-based generative models. This framework leverages the benefits of both **GE**nerative and **DI**scriminative approaches, resulting in improved symbolic representations over standalone solutions. Additionally, GEDI can be easily integrated and trained jointly with existing neuro-symbolic frameworks without the need for additional supervision or costly pre-training steps. We demonstrate through experiments on real-world data, including SVHN, CIFAR10, and CIFAR100, that GEDI outperforms existing self-supervised learning strategies in terms of clustering performance by a significant margin. The symbolic component further allows it to leverage knowledge in the form of logical constraints to improve performance in the small data regime and to overcome the problem of representational collapse.

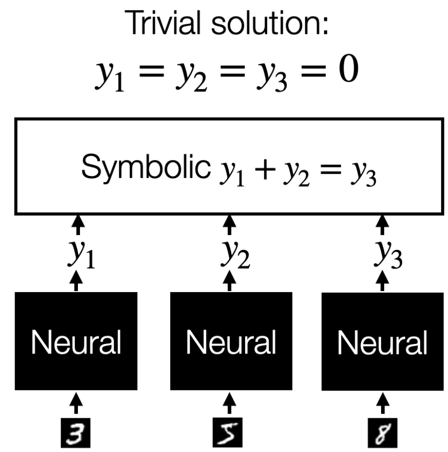

Figure 1: The problem of representational collapse.

1 Introduction

Imagine that we want to create a system capable of discovering symbols and relationships from both data and existing knowledge, whenever it is available. Neuro-symbolic learning represents a promising solution for achieving this goal [Garcez *et al.*, 2022]. However, current neuro-symbolic solutions rely on either expensive pre-training strategies or additional supervision to effectively make use of the training feedback provided by the symbolic component [Manhaeve *et al.*, 2021]. In this work, we highlight that these limitations are caused by the existence of trivial minima in the training objective landscape. One example of such a problem is what we call ”representational collapse”.

To illustrate this problem, let us consider a simple yet non-trivial example. Suppose that we have only raw data, in the form of a tuple of three images, each containing a single handwritten digit (e.g. $<$ , ,  $>$), along with information about the logical relationships between these digits (e.g. the third digit is the sum of the first two). Notably, this example is more fundamental and different than the traditional digit addition task (i.e. $+$  $= 8$) used to evaluate common neuro-symbolic systems, because no information about the value of the sum is provided. Importantly, existing neuro-symbolic solutions that do not leverage pre-training or additional supervision fail to represent the images correctly. As we show in the experiments and illustrated in Figure 1, existing neuro-symbolic frameworks trained from scratch will learn a ”collapsed” representation that satisfies the constraints imposed by the symbolic component, but that does not carry any semantic information about the input images (e.g. all three images are recognized as the digit 0).

The problem of representational collapse currently impedes the possibility of jointly learning representations and their relations [Evans *et al.*, 2021a; Evans *et al.*, 2021b], and therefore solving it represents an important advancement for research in neuro-symbolic learning.

To address the problem of representational collapse, we propose a Bayesian framework that unifies bottom-up approaches based on self-supervised representation learning with top-down approaches based on neuro-symbolic learning. Firstly, we demonstrate that several existing self-supervised learning techniques, such as negative-free [Ozsoy *et al.*, 2022; Liu *et al.*, 2022; Zbontar *et al.*, 2021; Bardes *et al.*, 2022a; Bardes *et al.*, 2022b; Ermolov *et al.*, 2021], contrastive [O. Henaff, 2020; Chen *et al.*, 2020; Lee, 2022; Xu *et al.*, 2022] and cluster-based approaches [Caron *et al.*, 2018; Caron *et al.*, 2020], and likelihood-based generative models, such as energy-based models [Kim and Ye, 2022], can be unified within a coherent Bayesian framework called GEDI. GEDI leverages the complementary properties of discriminative approaches, which are suitable for representation learning, and of generative approaches, which capture information about the underlying density function generating the data, to improve its representation learning capabilities. Secondly, we demonstrate that GEDI can be easily extended to the neuro-symbolic setting thanks to its probabilistic nature. Importantly, GEDI has two main advantages: it can overcome the problem of representational collapse common of existing neuro-symbolic approaches and it can also allow for learning symbolic representations in the small data regime, currently out of reach for existing self-supervised learning techniques.

We conduct experiments on toy and real-world data (viz. SVHN, CIFAR10 and CIFAR100) comparing against state-of-the-art self-supervised approaches, i.e. Barlow Twins [Zbontar *et al.*, 2021] and SwAV [Caron *et al.*, 2020], thus showing the superiority of our unified framework. Additionally, we show that GEDI outperforms DeepProbLog [Manhaeve *et al.*, 2018], a state-of-the-art neuro-symbolic framework, on the above mentioned problematic example.

To summarize, the key contributions of the paper are:

- We introduce the problem of representational collapse in neuro-symbolic learning.
- We provide a Bayesian perspective of self-supervised learning objectives based on GEDI.
- GEDI can be easily extended to the neuro-symbolic setting, thus showing for the first time the capability to overcome the problem of representational collapse.
- GEDI effectively exploits the reasoning component enabling to learn self-supervised learning models in the small data regime.

References

[Bardes *et al.*, 2022a] A. Bardes, J. Ponce, and Y. Lecun. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. In *ICLR*, 2022.

[Bardes *et al.*, 2022b] A. Bardes, J. Ponce, and Y. LeCun. VICREGL: Self-Supervised Learning of Local Visual Features. In *NeurIPS*, 2022.

[Caron *et al.*, 2018] M. Caron, P. Bojanowski, A. Joulin, and M. Douze. Deep Clustering for Unsupervised Learning of Visual Features. In *ECCV*, 2018.

[Caron *et al.*, 2020] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. In *NeurIPS*, 2020.

[Chen *et al.*, 2020] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*, 2020.

[Ermolov *et al.*, 2021] A. Ermolov, A. Siarohin, E. Sangineto, and N. Sebe. Whitening for Self-Supervised Representation Learning. In *ICML*, 2021.

[Evans *et al.*, 2021a] Richard Evans, Matko Bošnjak, Lars Buesing, Kevin Ellis, David Pfau, Pushmeet Kohli, and Marek Sergot. Making Sense of Raw Input. *Artificial Intelligence*, 299:103521, 2021.

[Evans *et al.*, 2021b] Richard Evans, José Hernández-Orallo, Johannes Welbl, Pushmeet Kohli, and Marek Sergot. Making Sense of Sensory Input. *Artificial Intelligence*, 293:103438, 2021.

[Garcez *et al.*, 2022] Artur d'Avila Garcez, Sebastian Bader, Howard Bowman, Luis C Lamb, Leo de Penning, BV Illuminoo, Hoifung Poon, and COPPE Gerson Zaverucha. Neural-symbolic learning and reasoning: a survey and interpretation. *Neuro-Symbolic Artificial Intelligence: The State of the Art*, 342(1), 2022.

[Kim and Ye, 2022] B. Kim and J. C. Ye. Energy-Based Contrastive Learning of Visual Representations. In *NeurIPS*, 2022.

[Lee, 2022] K. Lee. Prototypical Contrastive Predictive Coding. In *ICLR*, 2022.

[Liu *et al.*, 2022] X. Liu, Z. Wang, Y. Li, and S. Wang. Self-Supervised Learning via Maximum Entropy Coding. In *NeurIPS*, 2022.

[Manhaeve *et al.*, 2018] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt. DeepProbLog: Neural Probabilistic Logic Programming. In *NeurIPS*, 2018.

[Manhaeve *et al.*, 2021] Robin Manhaeve, Sebastijan Dumančić, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Neural probabilistic logic programming in deepproblog. *Artificial Intelligence*, 298:103504, 2021.

[O. Henaff, 2020] Olivier O. Henaff. Data-Efficient Image Recognition with Contrastive Predictive Coding. In *ICML*, 2020.

[Ozsoy *et al.*, 2022] S. Ozsoy, S. Hamdan, S. Arik, D. Yuret, and A. T. Erdogan. Self-Supervised Learning with an Information Maximization Criterion. In *NeurIPS*, 2022.

[Xu *et al.*, 2022] H. Xu, H. Xiong, and G. J. Qi. K-Shot Contrastive Learning of Visual Features with Multiple Instance Augmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[Zbontar *et al.*, 2021] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. In *ICML*, 2021.