Sustainability of neural network applications in training and inference - Some approaches and practices – Extended Abstract

Angela Ciocan Voinea¹,² and Vincent Courboulay³

¹ La Rochelle University – L3i, Avenue Michel Crépeau, La Rochelle, 17042, France
² Pôle Emploi – DSI Nantes, Rue Konrad Adenauer, Nantes, 44000, France
³ La Rochelle University – L3i, Avenue Michel Crépeau, La Rochelle, 17042, France

Keywords
sustainable AI, reducing energy consumption, power consumption, DNN, Green AI

A Deep Neural Networks (DNNs) have gained tremendous importance for companies in recent times. This has become possible with the emergence of big data and the great possibility of storage of these data, on the one hand, and the computing resources which allowed the development of deeper deep learning algorithms with a considerable number of parameters for the other hand [1],[2],[3],[4]. The development of more efficient hardware, such as GPUs and TPUs, has made it possible to train DNNs faster and more efficiently, which has made them more accessible to companies of all sizes [5],[6],[7],[8],[9]. The ability of DNNs to learn from vast amounts of data and extract useful insights has made them a powerful tool for companies across various industries, and as a result, DNNs have become a key tool for companies looking to gain a competitive edge in their respective industries [10],[11],[12],[13],[14]. As DNNs continue to advance in their ability to perform complex tasks, concerns have arisen about their sustainability [15],[16],[17],[18].

Overall, considering the environmental impact of DNNs development requires an exhaustive approach that considers the entire lifecycle of a model, from data collection to model deployment [19],[20],[21]. Adopting sustainable practices and technologies can help reduce the environmental impact of DNN development while still achieving high performance on various tasks. Two key phases in the lifecycle of DNNs applications that can significantly impact its development and deployment are the training and inference. Both require significant computational resources, resulting in high energy consumption and carbon emissions [22],[23],[24],[25].

Considering the environmental impact of both training and inference, and adopting sustainable practices and technologies, can help reduce the environmental footprint of DNN models throughout their entire lifecycle. This holistic approach to DNN development can contribute to a more sustainable future for artificial intelligence and technology in general.

In this paper, we discuss techniques such as Model Compression, Knowledge Distillation, Transfer Learning and Data Management practices that can contribute to a more sustainable approach to DNN development and exploring strategies for reducing energy consumption without compromising performance [26],[27],[28],[29],[30]. This paper aims to highlight the importance of sustainability in DNNs development and to inspire further research into more sustainable approaches to training and inference.
References