Foreword

The current period of progress in artificial intelligence was triggered when Krizhevsky et al. [2012] demonstrated that an artificial neural network designed twenty years earlier [LeCun et al., 1989] could outperform complex state-of-the-art image recognition methods by a huge margin, simply by being a hundred times larger and trained on a dataset similarly scaled up.

This breakthrough was made possible thanks to Graphical Processing Units (GPUs), highly parallel consumer-grade computing devices developed for real-time image synthesis and repurposed for artificial neural networks.

Since then, under the umbrella term of "deep learning," innovations in the structures of these networks, the strategies to train them, and dedicated hardware have allowed for an exponential increase in both their size and the quantity of training data they take advantage of [Sevilla

et al., 2022]. This has resulted in a wave of successful applications across technical domains, from computer vision and robotics to speech processing, and since 2020 in the development of Large Language Models with general proto-reasoning capabilities [Chowdhery et al., 2022].

Although the bulk of deep learning is not difficult to understand, it combines diverse components such as linear algebra, calculus, probabilities, optimization, signal processing, programming, algorithmics, and high-performance computing, making it complicated to learn.

Instead of trying to be exhaustive, this little book is limited to the background necessary to understand a few important models. This proved to be a popular approach, resulting in more than 500,000 downloads of the PDF file in the 12 months following its announcement on Twitter.

If you did not get this book from its official URL

https://fleuret.org/public/lbdl.pdf

please do so, to allow the estimation of the number of readers.

François Fleuret,
May 19, 2024