Working Title

Josh Murr

September 10, 2020

Abstract

Research in recent years has continually shown new and unexpected applications for Deep Neural Networks which seem to show no limit to their ability. This is in part down to the ubiquity and capability of modern Graphics or Tensor Processing Units which allow for huge amounts of computation at lightning speed. GPT-3 has shown us that simply making a model larger is one possible way to improve results[2] which could easily set (or continue) the trend that more data and more compute power are the keys to better models. This may in part be true, but this trend widens the gulf of accessibility and also understanding with cutting edge machine learning research. *Learning to See* by Memo Akten et al[1] provides a glimpse inside a Conditional Generative Adverserial Network (*cGAN*) in an immediate and impactful way showing it's ability and also it's limitations; in todays climate of uncertainty and disbelief this is an important thing. Getting work like this on more modest devices presents the challenge of shrinking such Deep Learning models and improving their efficiency so that they do not require vast compute power to be effective. In this paper I (we?) explore different methods of model compression and efficient architectures and analyse their respective efficiency gains with a view to making such systems more accessible and interactive.

1 Introduction

2 Background Context

3 Method

4 Results

5 Discussion

6 Conclusion

References

[1] Memo Akten, Rebecca Fiebrink, and Mick Grierson. "Learning to See: You Are What You See". In: (2020). DOI: 10.1145/3306211.3320143. eprint: arXiv:2003.00902.

[2] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. eprint: `arXiv:2005.14165`.