Super Convergence: Very Fast Training of Residual Networks Using Large Learning Rates
ICLR Reproducibility Challenge

Keivaun Waugh
University of Texas at Austin
keivaunwaugh@gmail.com

Paul Choi
University of Texas at Austin
choipaul96@gmail.com

Abstract

In this paper, we aim to evaluate the reproducibility of the experiments detailed in the paper "Super Convergence: Very Fast Training of Residual Networks Using Large Learning Rates" [1]

1. Introduction

Paul

1.1. Target Questions

2. Method

Due to time and hardware constraints (lack of multiple GPUs), we were unable to replicate all of the experiments that were listed in the paper. Therefore, we had to choose which experiments from which we could glean the most. It was our goal to show the existence or not of the following claims made in the paper. SC:

- allows the network to achieve higher test accuracy than traditional piece-wise constant learning rates (PCLR)
- requires an order of magnitude fewer iterations to converge than PCLR
- is more noticeable with fewer training examples
- works with various ResNet sizes
- works best with a certain step size

When available, we used the hyperparameters that the authors listed in their paper. When these were omitted, we attempted to use standard values that are used on Cifar-10 and ResNets.

2.1. Implementation Details

In the ICLR submitted paper, the authors mentioned that they would release their code upon acceptance to the conference. However, upon Googling the paper name, we found an arXiv version of the paper that links to the authors' source code on GitHub. We chose to not look at this code and reimplement their solution so that we could test the reproducibility under the information (especially hyperparameters) that were listed in the ICLR paper.

As a starting point, we used the open source ResNet training on Cifar-10 code available in the TensorFlow [**?**] repository. We modified the code accordingly for each of the experiments.

A major limitation in our testing was our restriction to a single GPU for all of the experiments. The authors stated that they used 8 Nvidia Titan Black GPUs, which allowed them to test with batch sizes as large as 1536. We were restricted to a batch size of 256. The authors found that increasing the batch size significantly improved the performance of their CLR tests, so the discrepancy between our results and the authors could be due to this.

3. Experiments

Keivaun

3.1. Methodology

4. Conclusion

Paul

4.1. Cost of Reproduction

What cost in terms of resources (computation, time, people, developemnt effort, communication with the authors).

References

[1] Anonymous. Super-convergence: Very fast training of residual networks using large learning rates. *International Conference on Learning Representations*, 2018.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.