

Lab 1: Debugging and Predicting Runtimes on Rosie

Introduction

In this lab, we learn how to use Rosie to edit and run/debug code in VSCode, as well as how to use Rosie to run a large training job in the course Singularity image with different hardware allocated to it. We then compare and discuss the different times it took each allocation of hardware to complete the training.

Predict Relative Runtime from GPU Specs

Since the throughput of the NVIDIA V100 is about double that of the T4, and only one GPU is used, I expect the time the NVIDIA V100 takes to train a neural network to be about half the time the T4 takes.

Experimentally Determine the Runtimes

Hardware Used to Train	Time to Train (s)	Predicted Time to Train 20 Epochs (s)	Time to Train 20 Epochs (s)
T4 with 1 CPU	1591.6	31,832	31,327.1
T4 with 18 CPUs	733.4	14,668	10,128.4
T4 with a GPU	67.5	1350	1615
DGX with a GPU	37.2	744	759.5

Discussion

For the most part, the actual results were very close to my predictions. The only one that appears to be off by a large percentage is the T4 with 18 CPUs. The real time was actually much faster than the predicted, with the average time to train over one epoch being 506.42 seconds, over 200 seconds faster than the one epoch run. Additionally, my prediction that the time it takes for the V100 to train is about half the time the T4 takes appears to be correct. The “T4 with a GPU” time is the time for the T4, and the “DGX with a GPU” time is the time for the V100, so the ratio I got for one epoch is $37.2/67.5 = 0.55$, while the ratio for 20 epochs is $759.5/1615 = 0.47$.