WEEK 6 Progress

Jaison Lewis

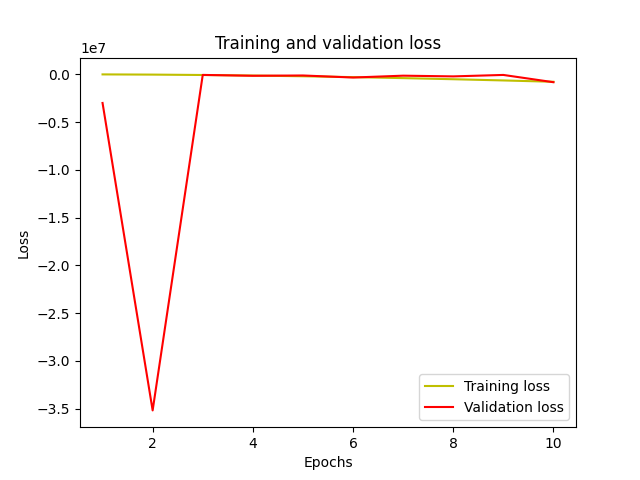
------------------------------------------------------------------------------------------------------------------------

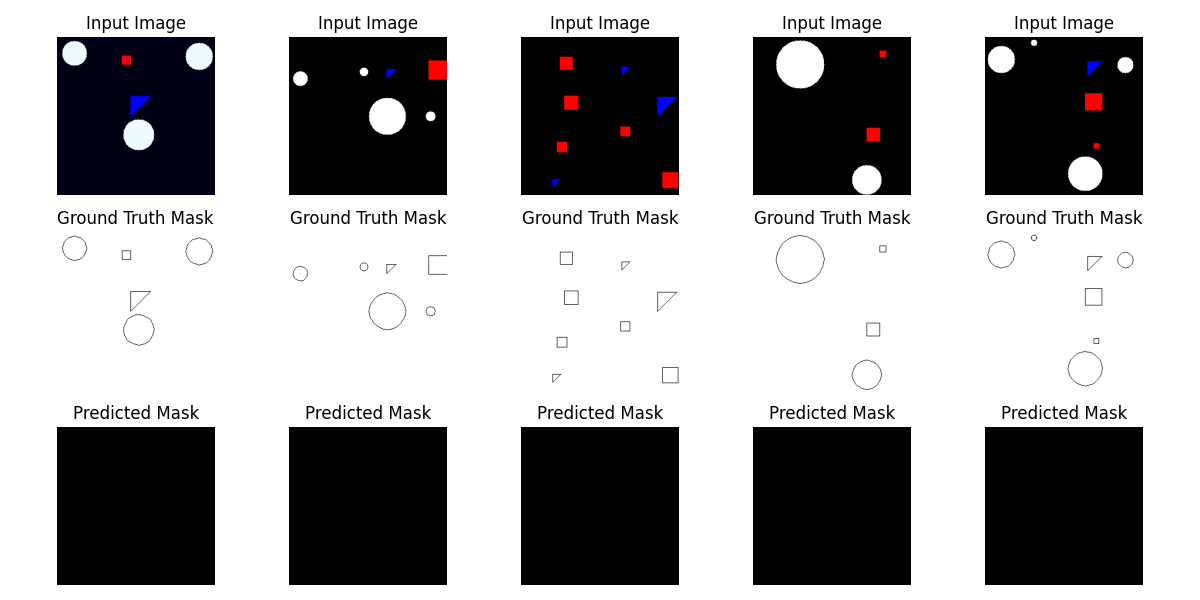
Applied for access to the cluster. Awaiting approval.

Read extensively about AI HPC benchmarking and the various options available to benchmark after attending the benchmarking meeting on Tuesday. I didn’t have enough time to present anything concrete on Deep500, but I have since written a bit about it.

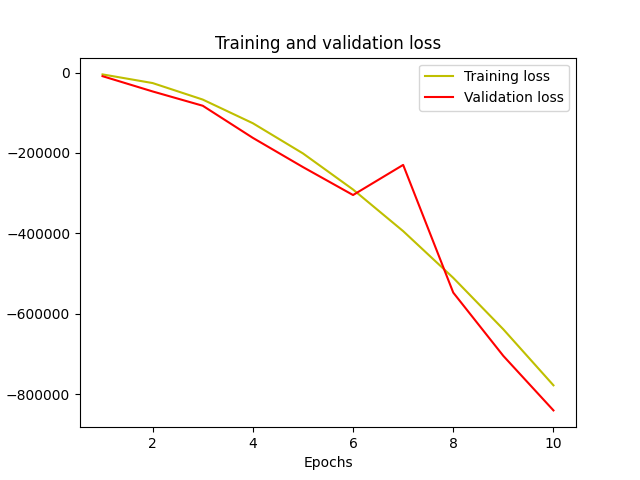
I also spent a lot of time just reading about benchmarking and looking at all the benchmarks Narges has listed.

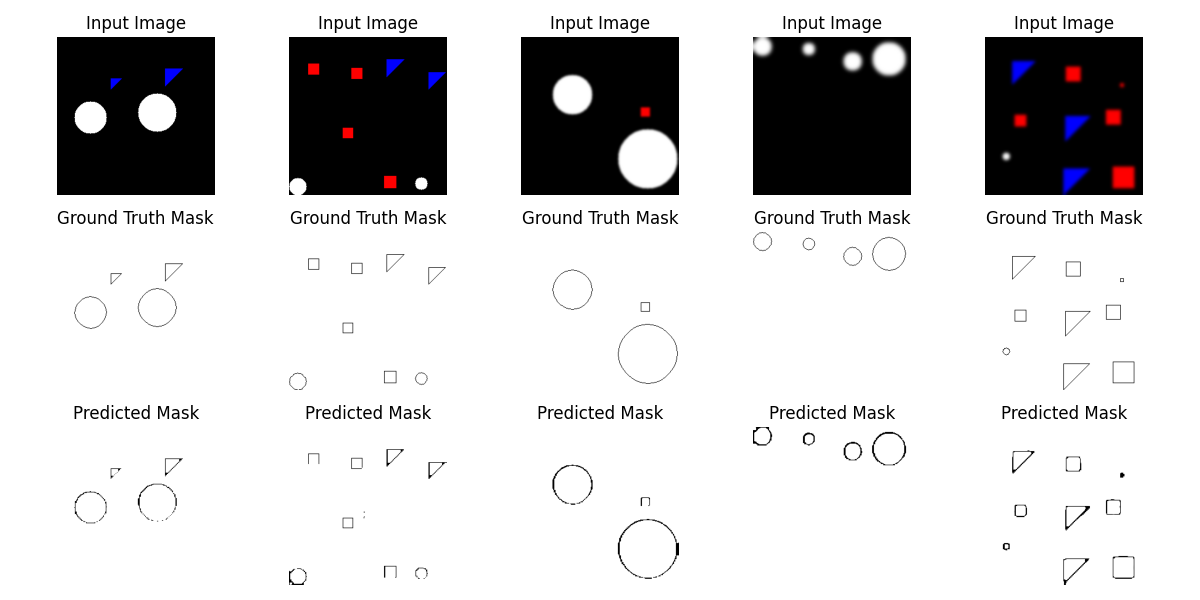
I also had some time to run two resnet50 on the synthetic blemished data with impressive results. For the Synthetic blemished data, it seems Resnet 50 with normalised images is the best option.

**Synthetic Blemished data Resnet50 Not Normalised 10Epochs 256x256:**

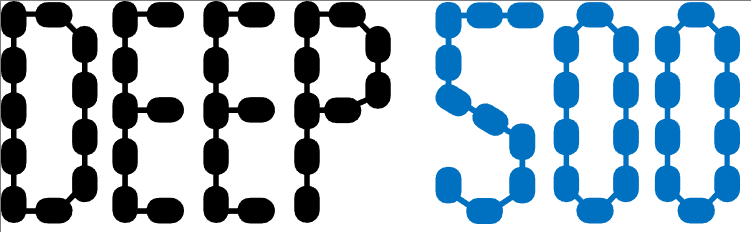


**Synthetic Blemished data Resnet50 Normalised 10Epochs 256x256:**

****

****

**Deep500**



What is it?

Deep500 by SPCL (Scalable Parallel Computing Lab) part of ETH Zurich. It is essentially a library of tools that can be customised and used to measure any level of a deep neural network, including (L0) Operators, (L1) Network Evaluations, (L2) Training and (L3) Distributed training. Its supported frameworks are Tensorflow, Pytorch and Caffe2. It aims to benchmark by covering all the use cases of an AI HPC rather than a narrow focus and rigid set of rules.

The idea of Deep500 is to split deep learning into the four levels mentioned and have separate interfaces, metrics and reference implementations to benchmark each. The design offers the ability to mix, match and tweak the benchmark to your specific use case. Deep500 also uses ONNX, making it hardware-agnostic, more standardised and interoperable

Why Deep500?  
Compared to industry standard MLPerf, Deep500 considers more factors that affect performance and is lot more adaptable. MLperf, on the other hand, is a bit more rigid with set models and rules. Deep500 also provides more performance metrics like time-to-accuracy, framework overhead and communication volume. Deep500 also improves reproducibility by using the ONNX (Open Neural Network Exchange) format.  
ONNX is an open format for representing deep learning models and it allows models to be shared between frameworks and tools. It lets users concentrate on models rather than the framework.

Since these two benchmarking systems are so different in how they work, it might also be a good idea to run both in conjunction.

Papers and links:  
Papers and presentations:   
<https://ieeexplore.ieee.org/document/8821020>

<https://deep500.org/slides/hoefler.pdf>

<https://deep500.org/slides/dubey.pdf>

<https://deep500.org/slides/kurth.pdf>

<https://deep500.org/slides/zhai.pdf>

Github:   
<https://github.com/deep500>

Video:  
<https://www.youtube.com/watch?v=URTI6y8tJ4g>

Check out this interesting chart from the Deep500 paper, which gives an overview of all the Deep Learning (DL) frameworks that can be integrated.

