

# Sources of Irreproducibility / Testing Pipeline using OSG and Cloud Resources

AI & Reproducibility, Repeatability, and Replicability  
International Data Week 2022

# Sources of Irreproducibility

## Implementation Factors

- Initialization seeds
- Ancillary software
- Ancillary software version
- Bugs in software
- Non-deterministic ordering of floating-point operations
- Parallel execution
- Compiler settings
- Processing unit

# What Can Be Done?

Not all sources of irreproducibility can be controlled for.

However, many of the sources of irreproducibility can be examined by performing multiple runs in heterogeneous computing environments consisting of different hardware and software environments.

The biggest challenge for researchers can be getting access to multiple heterogeneous computing environments.

# Sources of Irreproducibility

## Implementation Factors

- Initialization seeds
- Ancillary software
- Ancillary software version
- Bugs in software
- Non-deterministic ordering of floating-point operations
- Parallel execution
- Compiler settings
- Processing unit

# Open Science Grid

- Distributed High-Throughput Computing
  - For problems that can be run as numerous and self-contained jobs
- Supports Machine Learning and AI executed with Multiple Independent Training Tasks, Different Parameters, and/or Data Subsets
- More than 20 Institutions Participating
- Uses HTCondor Batch Scheduler
  - Run Non-Interactively
  - Uses Bash Scripts to Setup Environment and Copy Output
- Free to qualifying researchers:  
<https://www.osgconnect.net>



# Cloud Resources

- AWS offers:
  - Multiple types of Intel, AMD and ARM CPUs
  - Multiple NVIDIA and AMD GPUs
  - Other processors:
    - TPUs
    - FPGAs
- However, each researcher must setup the instances and schedule the resources on run on their own.
- Must pay to use

# Keras Examples Used

## Computer Vision

- Simple MNIST convnet:

[https://github.com/keras-team/keras-io/blob/master/examples/vision/mnist\\_convnet.py](https://github.com/keras-team/keras-io/blob/master/examples/vision/mnist_convnet.py)

## Natural Language Processing

- Bidirectional LSTM on IMDB:

[https://github.com/keras-team/keras-io/blob/master/examples/nlp/bidirectional\\_lstm\\_imdb.py](https://github.com/keras-team/keras-io/blob/master/examples/nlp/bidirectional_lstm_imdb.py)

## Structured Data

- Imbalanced classification: credit card fraud detection:

[https://github.com/keras-team/keras-io/blob/master/examples/structured\\_data/imbalanced\\_classification.py](https://github.com/keras-team/keras-io/blob/master/examples/structured_data/imbalanced_classification.py)

**Examples were modified to run deterministically**

# Varying Hardware Test

Using Open Science Grid Servers

The job was configured with:

- Various CPUs (4 threads)
- Various GPUs (1 GPU)
- Singularity container based on Docker container:  
tensorflow/tensorflow:2.8.0-gpu

The placement of what hardware the jobs ran on was not controlled.



# Varying Hardware Test

25 tests ran on 4 different sites and 8 different servers.

Hardware included Intel Xeon and AMD EPYC CPU and Tesla V100, Quatro RTX 6000 and NVidia A100 GPUs.

The same 3 Karas examples were used and each example was ran 5 times to ensure repeatability. The accuracy from the imbalance example and the mnist example were repeatable on the same hardware. The following heat map shows the accuracy of these two examples:

# Varying Hardware Test

imbalanced\_classification and mnist\_convnet set\_random\_seed GPU Runs

Server	imbalanced_classification	mnist_convnet
WSU-GRID:voh7:21680789	0.993925690650939	0.991999983787536
WSU-GRID:vov9:21680789.1	0.993925690650939	0.991999983787536
WSU-GRID:voh7:21680789.2	0.993925690650939	0.991999983787536
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21680789.3	0.996611714363098	0.991900026798248
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21680789.4	0.996611714363098	0.991900026798248
GPN-GP-ARGO-Backfill:5149ad6dab4d:21680889	0.992205202579498	0.992200016975402
WSU-GRID:voh7:21680889.1	0.993925690650939	0.991999983787536
WSU-GRID:voh7:21680889.2	0.993925690650939	0.991999983787536
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21680889.3	0.996611714363098	0.991900026798248
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21680889.4	0.996611714363098	0.991900026798248
WSU-GRID:vov9:21681043	0.993925690650939	0.991999983787536
WSU-GRID:voh7:21681043.1	0.993925690650939	0.991999983787536
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681043.2	0.996611714363098	0.991900026798248
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681043.3	0.996611714363098	0.991900026798248
GPN-GP-ARGO-Backfill:a1471c8c2ee7:21681043.4	0.992205202579498	0.992200016975402
WSU-GRID:vov9:21681260	0.993925690650939	0.991999983787536
WSU-GRID:voh7:21681260.1	0.993925690650939	0.991999983787536
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21681260.2	0.996611714363098	0.991900026798248
OSG_US_SIUE_CC:cc-condor.hpc.siue.edu:21681260.3	0.992205202579498	0.992200016975402
OSG_US_SIUE_CC:cc-condor.hpc.siue.edu:21681260.4	0.992205202579498	0.992200016975402
GPN-GP-ARGO-Backfill:46bf332f6a0b:21681627	0.992205202579498	0.992200016975402
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681627.1	0.996611714363098	0.991900026798248
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21681627.2	0.996611714363098	0.991900026798248
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681627.3	0.996611714363098	0.991900026798248
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21681627.4	0.996611714363098	0.991900026798248

Run Number

# Varying Hardware Test

The bidirectional Keras test didn't have repeatable results when ran on GPU hardware.

Determinism on GPU hardware can be difficult.

The following is a heat map of all 5 runs of all 25 test runs:

# Varying Hardware Test

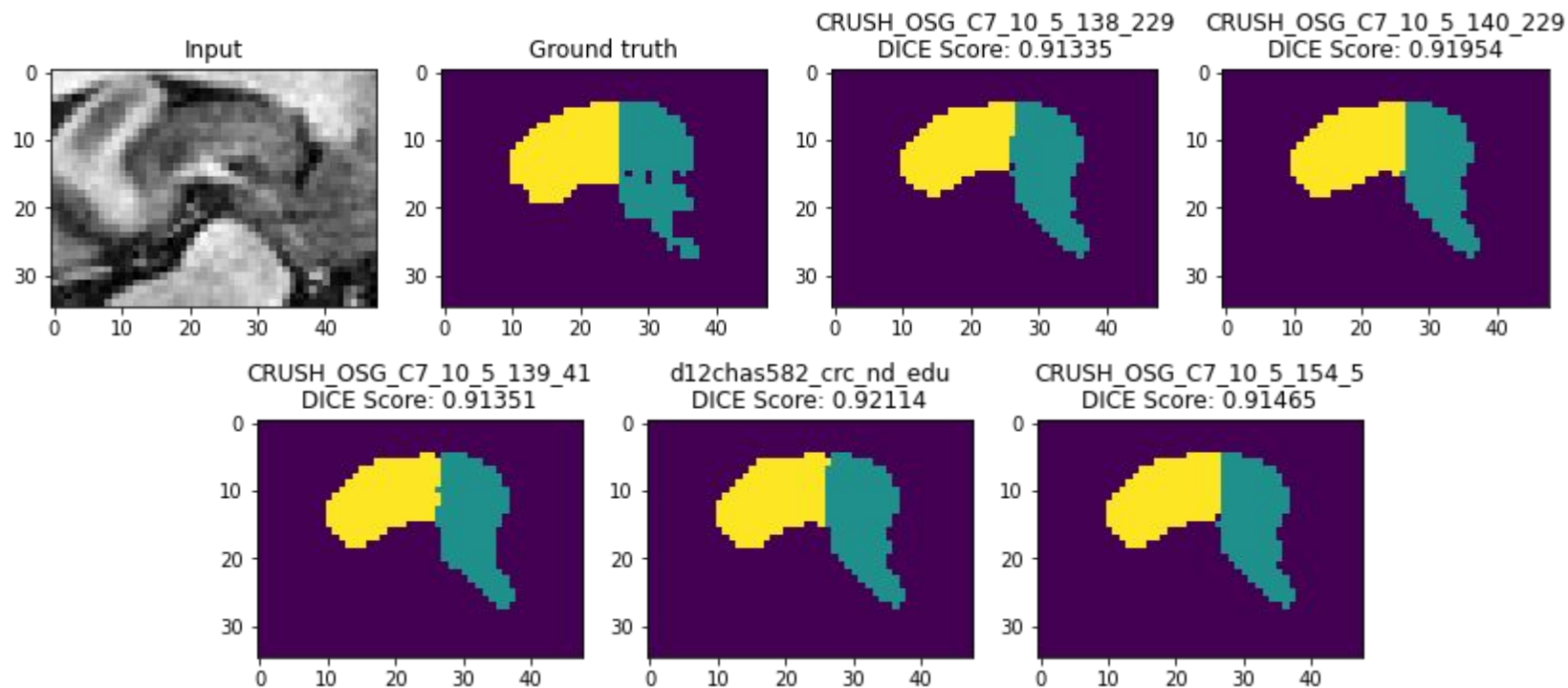
Bidirectional LSTM on IMDB set\_random\_seed on Different Hardware

Server	0	1	2	3	4
WSU-GRID:voh7:21680789	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
WSU-GRID:vov9:21680789.1	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
WSU-GRID:voh7:21680789.2	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21680789.3	0.862160027027130	0.862160027027130	0.862160027027130	0.840759992599487	0.862160027027130
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21680789.4	0.866760015487670	0.868680000305175	0.838599979877471	0.862160027027130	0.865440011024475
GPN-GP-ARGO-Backfill:5149ad6dab4d:21680889	0.835439980030059	0.861559987068176	0.865960001945495	0.851760029792785	0.857039988040924
WSU-GRID:voh7:21680889.1	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
WSU-GRID:voh7:21680889.2	0.864440023899078	0.864440023899078	0.865800023078918	0.864440023899078	0.864440023899078
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21680889.3	0.862160027027130	0.862160027027130	0.862160027027130	0.840759992599487	0.862160027027130
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21680889.4	0.865840017795562	0.862160027027130	0.862160027027130	0.862160027027130	0.862160027027130
WSU-GRID:vov9:21681043	0.864440023899070	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
WSU-GRID:voh7:21681043.1	0.865320026874542	0.864440023899078	0.864440023899078	0.864440023899078	0.864960014820098
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681043.2	0.862160027027130	0.862160027027130	0.862160027027130	0.862160027027130	0.862160027027130
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681043.3	0.856480002403259	0.862879991531372	0.863479971885681	0.860880017280578	0.845000028610229
GPN-GP-ARGO-Backfill:a1471c8c2ee7:21681043.4	0.856480002403259	0.862879991531372	0.863479971885681	0.860880017280578	0.845000028610229
WSU-GRID:vov9:21681260	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
WSU-GRID:voh7:21681260.1	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078	0.864440023899078
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21681260.2	0.866760015487670	0.862160027027130	0.862160027027130	0.862160027027130	0.862160027027130
OSG_US_SIUE_CC:cc-condor.hpc.siue.edu:21681260.3	0.860480010509491	0.772119998931884	0.865520000457763	0.846319973468780	0.865520000457763
OSG_US_SIUE_CC:cc-condor.hpc.siue.edu:21681260.4	0.864120006561279	0.8636800005073547	0.862839996814727	0.858759999275207	0.859279990196228
GPN-GP-ARGO-Backfill:46bf332f6a0b:21681627	0.852959990501403	0.869239985942840	0.861280024051666	0.853760004043579	0.856719970703125
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681627.1	0.866760015487670	0.862160027027130	0.862160027027130	0.862160027027130	0.862160027027130
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21681627.2	0.866760015487670	0.862160027027130	0.862160027027130	0.862160027027130	0.862160027027130
ND-CAML_gpu:qa-rtx6k-034.crc.nd.edu:21681627.3	0.866760015487670	0.862160027027130	0.862160027027130	0.862160027027130	0.869440019130706
ND-CAML_gpu:qa-rtx6k-031.crc.nd.edu:21681627.4	0.864359974861145	0.862160027027130	0.866760015487670	0.862160027027130	0.862160027027130

Run Number

125 Runs - Max: .86944 & Min: .77212

# Running nnUNet with Medical Imaging



# Conclusion

- There are sources of irreproducibility that cannot be controlled for.
- Running the experiment in multiple heterogeneous environments during the analysis stage will help validate the conclusions and also help validate the conclusions will be reproducible.
- Having resources like the Open Science Grid and access to Cloud resources can be used to improve reproducibility.