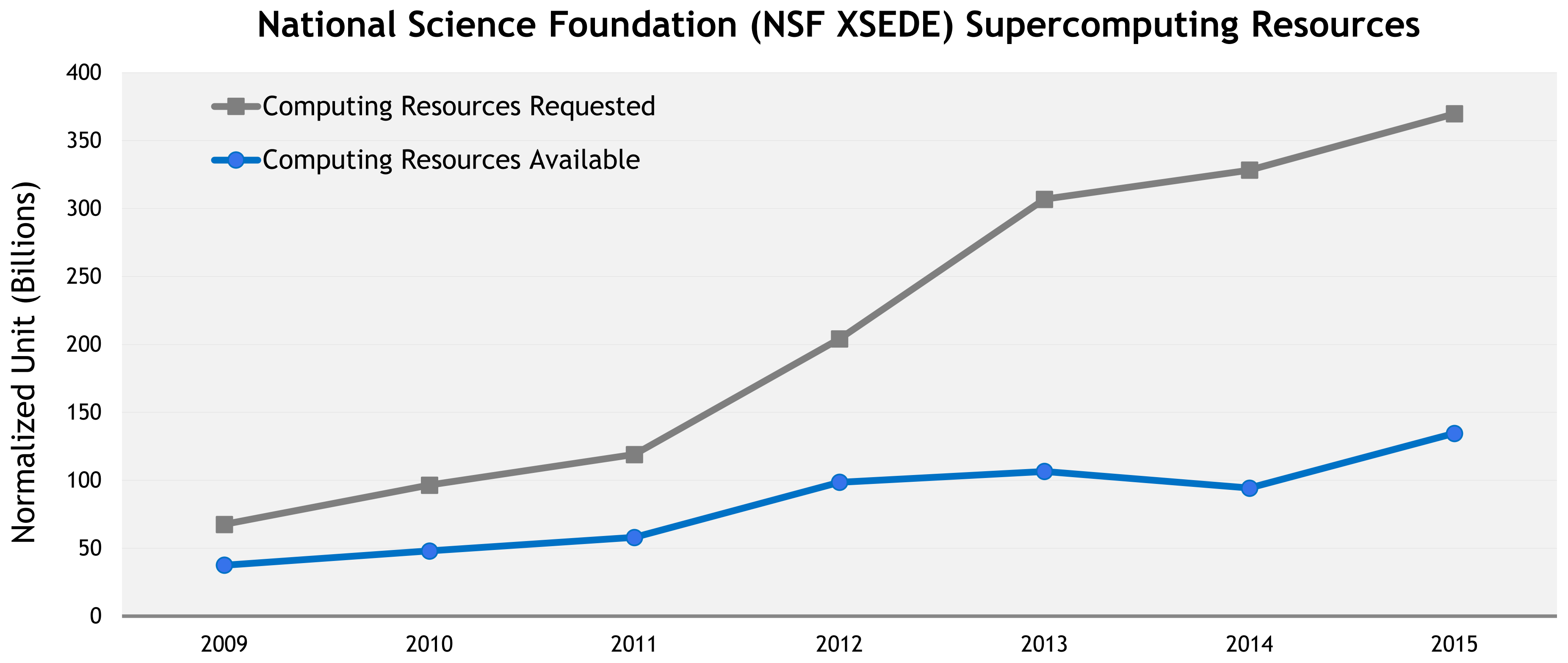
SCALING DEEP LEARNING TO EXASCALE
ACM GORDON BELL PRIZE 2018

Pedro Mario Cruz e Silva
Solutions Architect Manager, Latin América | Global Energy Team

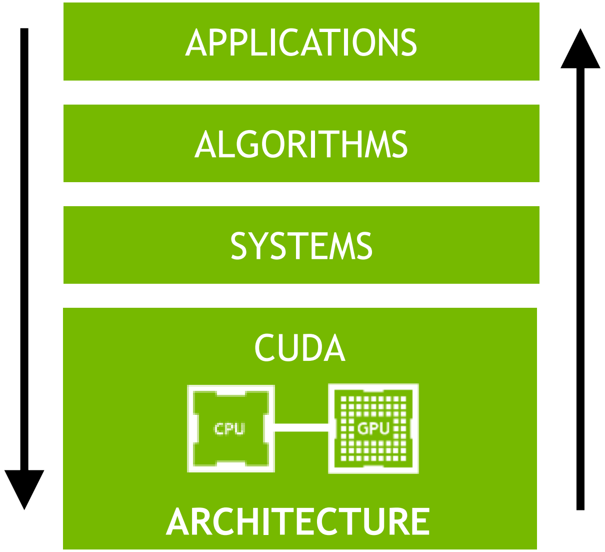
200B CORE HOURS OF LOST SCIENCE

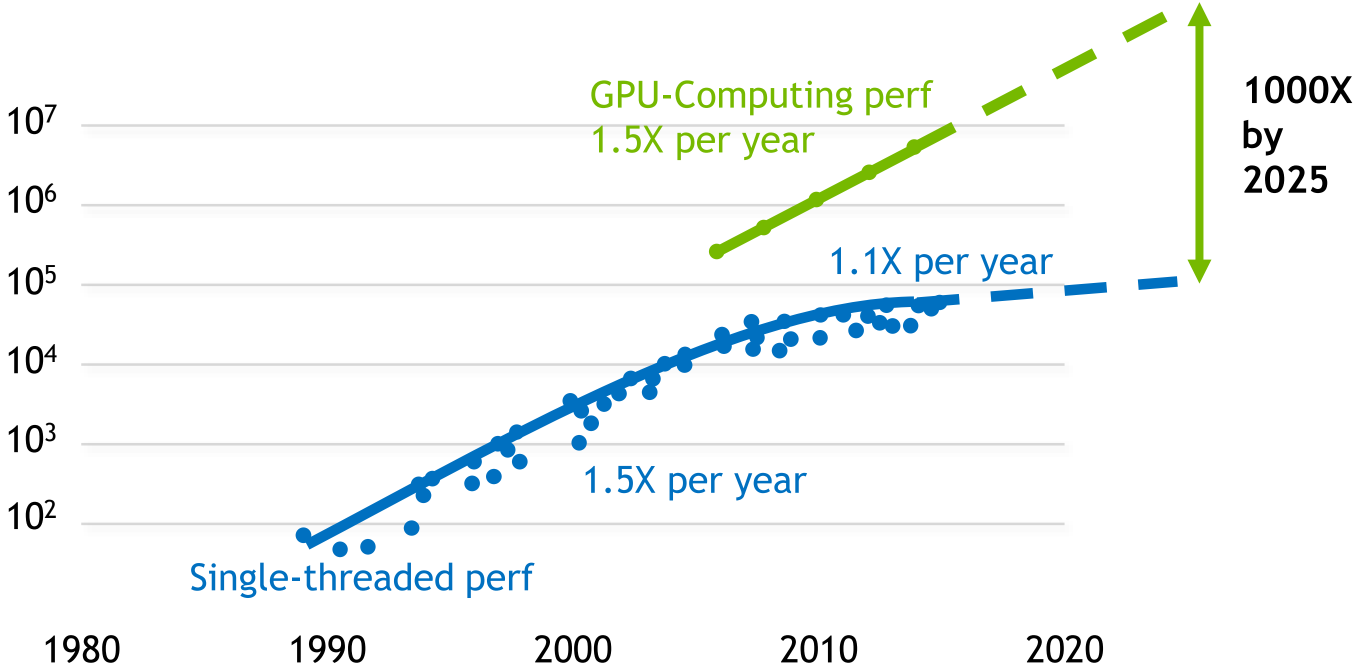
Data Center Throughput is the Most Important Thing for HPC



Source: NSF XSEDE Data: https://portal.xsede.org/#/gallery
NU = Normalized Computing Units are used to compare compute resources across supercomputers and are based on the result of the High Performance LINPACK benchmark run on each system

RISE OF GPU COMPUTING





Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten New plot and data collected for 2010-2015 by K. Rupp

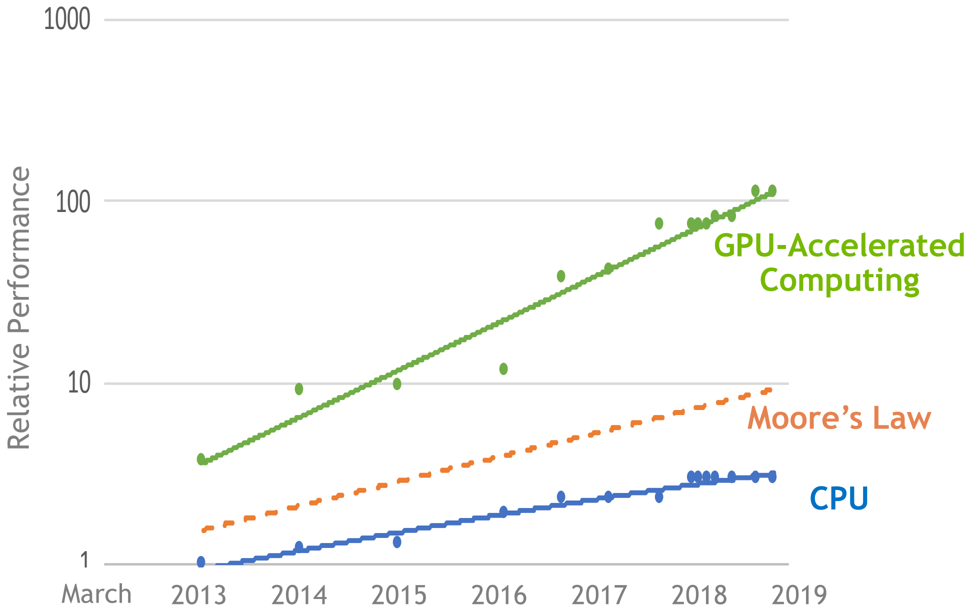
BEYOND MOORE'S LAW

Progress Of Stack In 6 Years

2013

cuBLAS: 5.0
cuFFT: 5.0
cuRAND: 5.0
cuSPARSE: 5.0
NPP: 5.0
Thrust: 1.5.3
CUDA: 5.0
Resource Mgr: r304
Base OS: CentOS 6.2

Accelerated Server With Fermi



Measured performance of Amber, CHROMA, GTC, LAMMPS, MILC, NAMD, Quantum Espresso, SPECFEM3D

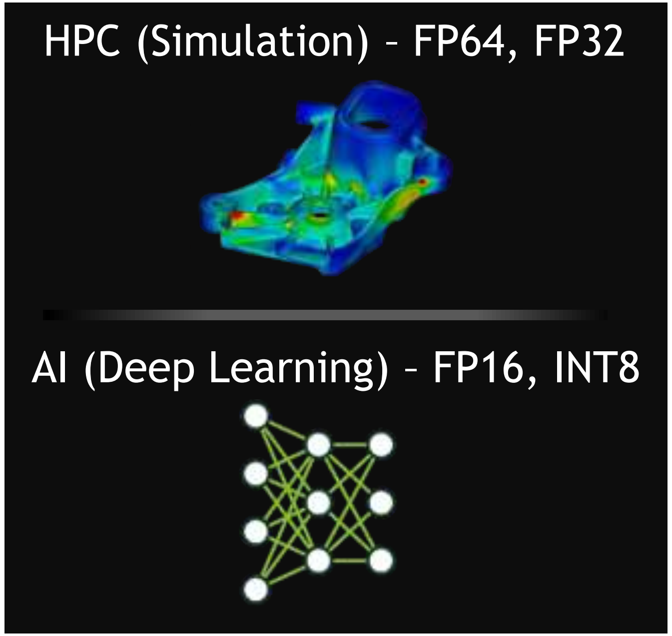
2019

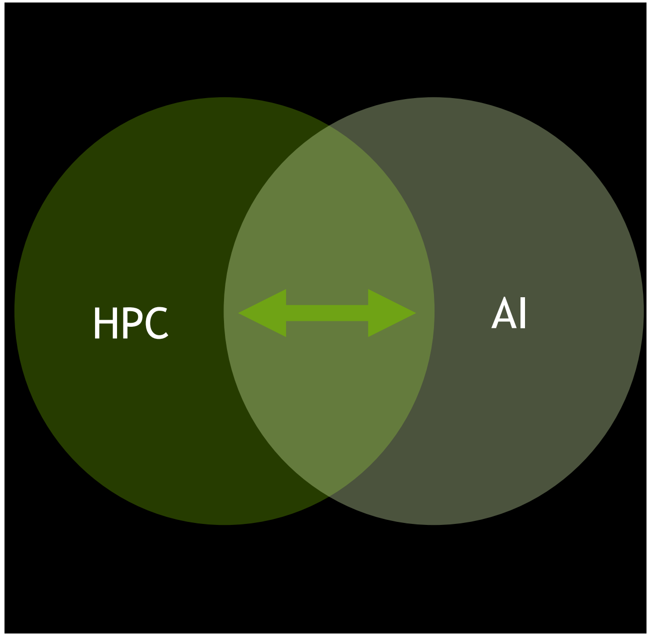
cuBLAS: 10.0
cuFFT: 10.0
cuRAND: 10.0
cuSOLVER: 10.0
cuSPARSE: 10.0
NPP: 10.0
Thrust: 1.9.0
CUDA: 10.0
Resource Mgr: r384
Base OS: Ubuntu 16.04

Accelerated Server with Volta

DIGITAL SCIENCE
HPC + AI + DATA

GPU FUSES HPC & AI COMPUTING





AI – A NEW INSTRUMENT FOR SCIENCE

HPC

> Algorithms based on first principles theory.

> Proven models for accurate results

AI

> Neural Networks that learn patterns from large data sets

> Improve predictive accuracy and faster response time.

Dramatically Improves Accuracy and Time-to-Solution

Commercially viable fusion energy

Understanding cosmological dark energy and matter

Clinically viable precision medicine

Improvement and validation of the Standard Model of Physics

Climate/weather forecasts with ultra-high fidelity

"ACCELERATING EULERIAN FLUID SIMULATION WITH CONVOLUTIONAL NETWORKS"

Tompson, J., Schlachter, K., Sprechmann, P., & Perlin, K. (2016). Accelerating Eulerian Fluid Simulation With Convolutional Networks. *arXiv preprint arXiv:1607.03597*.

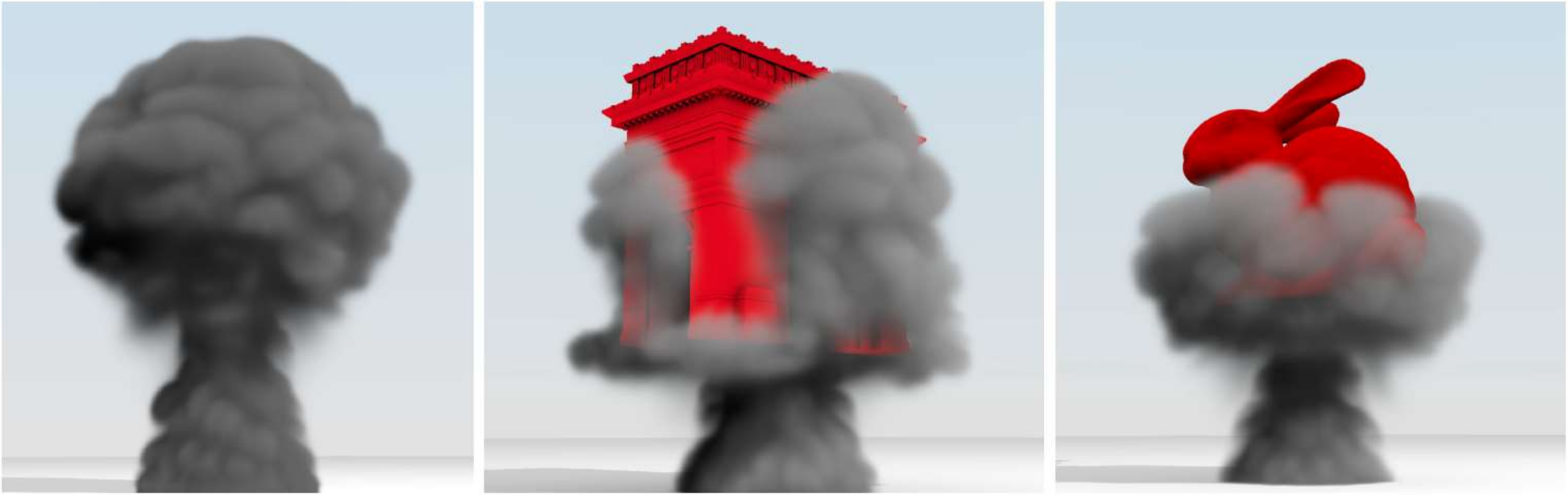


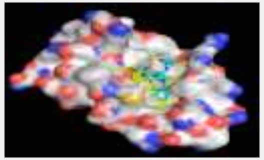
Fig. 1: Smoke simulation using our system - our method is capable of fast and accurate simulation of the Euler Equations for incompressible fluid flow at interactive frame-rates. Videos can be found at: `http://cims.nyu.edu/~schlacht/CNNFluids.htm`.

AI FOR SCIENCE

Transformative Tool To Accelerate The Pace of Scientific Innovation

90% accuracy
Fusion Sustainment
Clean Energy

33% Faster
Track Neutrinos
Particle Physics



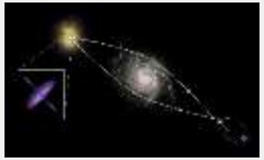
70% accuracy
Score Protein Ligand
Drug Discovery

11% higher accuracy
Monitor Earth's Vital
Climate

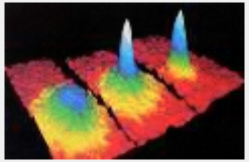
Improves Accuracy
Enabling realization of full scientific potential

5,000X Faster
Process LIGO Signal
Understanding Universe

300,000X Faster
Predict Molecular Energetics
Drug Discovery



Weeks to 10 milliseconds
Analyze Gravitational Lensing
Astrophysics



14X Faster
Generate Bose-Einstein
Condensate (Physics)

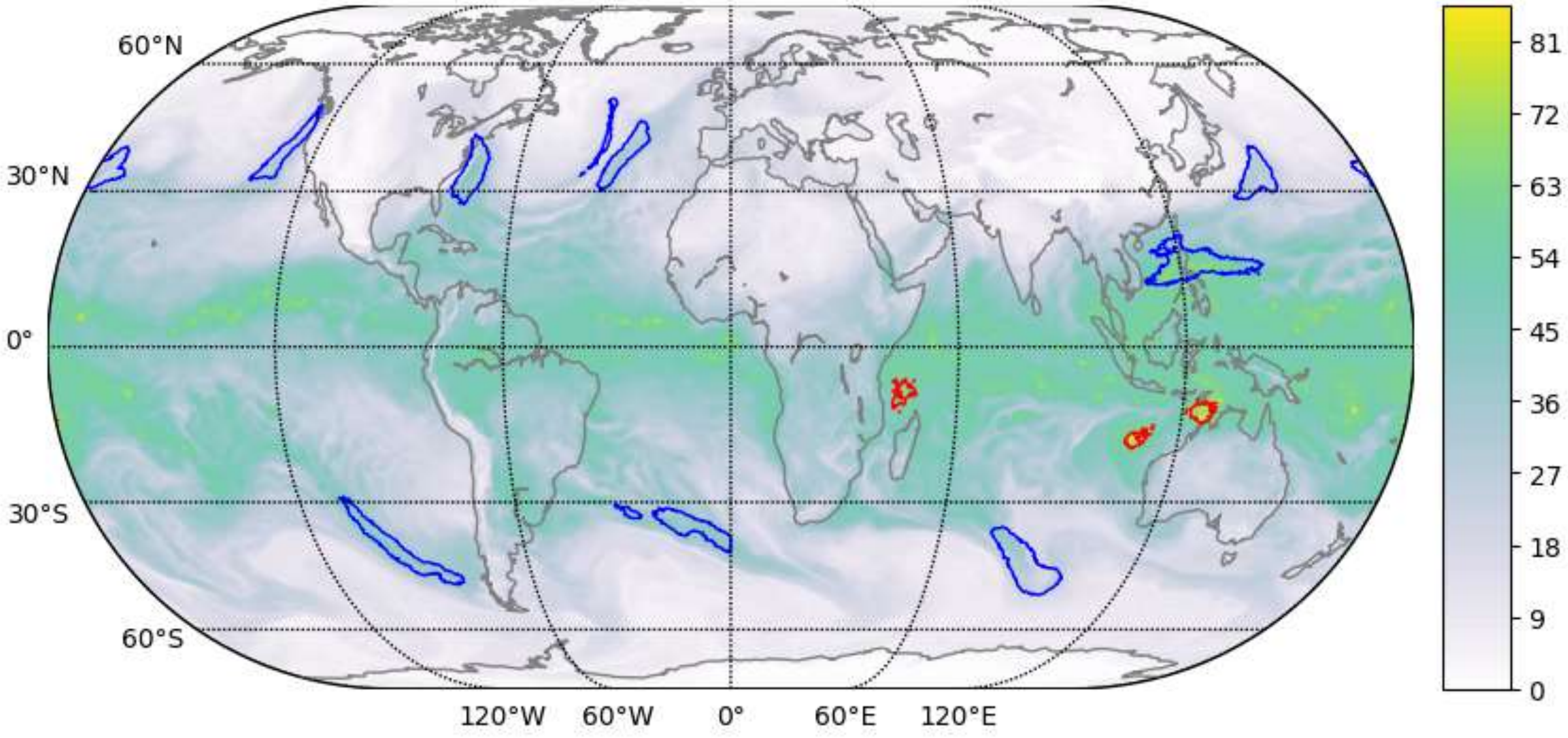
Accelerates Time to Solution
Unlocking the use of science in exciting new ways

THE PROBLEM

IMAGE SEGMENTATION

Pattern Detection for Characterizing Extreme Weather

Atmospheric rivers (ARs) are labeled in blue, while tropical cyclones (TCs) are labeled in red



CLIMATE DATASET AND GROUND TRUTH LABELS

Climate data used is currently 3.5 TB

- 0.25-degree Community Atmosphere Model (CAM5) output
- Climate variables are stored on an 1152x768 spatial grid, with a temporal resolution of 3 hours
- All available 16 variables (water vapor, wind, precipitation, temperature, pressure, etc).
- Process climate model output with the Toolkit for Extreme Climate Analysis to identify TCs.
- A floodfill algorithm is used to create spatial masks of ARs

- There are about 63K high-resolution samples in total
- Split into 80% training, 10% test and 10% validation sets
- The pixel mask labels correspond to 3 classes:
  1) Tropical Cyclone (TC)
  2) Atmospheric River (AR)
  3) Background (BG)

THE TEAM

NERSC & NVIDIA

Exascale Deep Learning for Climate Analytics

Thorsten Kurth*
tkurth@lbl.gov

Sean Treichler†
sean@nvidia.com

Joshua Romero†
joshr@nvidia.com

Mayur Mudigonda*
mudigonda@berkeley.edu

Nathan Luehr†
nluehr@nvidia.com

Everett Phillips†
ephillips@nvidia.com

Ankur Mahesh*
amahesh@lbl.gov

Michael Matheson‡
mathesonma@ornl.gov

Jack Deslippe*
jrdeslippe@lbl.gov

Massimiliano Fatica†
mfatica@nvidia.com

Prabhat*
prabhat@lbl.gov

Michael Houston†
mhouston@nvidia.com

***Abstract*—We extract pixel-level masks of extreme weather patterns using variants of Tiramisu and DeepLabv3+ neural networks. We describe improvements to the software frameworks, input pipeline, and the network training algorithms necessary to efficiently scale deep learning on the Piz Daint and Summit systems. The Tiramisu network scales to 5300 P100 GPUs with a sustained throughput of 21.0 PF/s and parallel efficiency of 79.0%. DeepLabv3+ scales up to 27360 V100 GPUs with a sustained throughput of 325.8 PF/s and a parallel efficiency of 90.7% in single precision. By taking advantage of the FP16 Tensor Cores, a half-precision version of the DeepLabv3+ network achieves a peak and sustained throughput of 1.13 EF/s and 999.0 PF/s**

communities. For instance, the state of California receives over 50% of its rainfall through Atmospheric Rivers (ARs), and Water Resource Management planners are interested in understanding if AR tracks will shift in the future, potentially resulting in a dramatic shortfall in fresh water supply. In the state of Florida, homeowners are interested in understanding if Tropical Cyclones (TCs) or hurricanes will become more intense and start making landfall more often. This has a direct impact on home prices and the insurance industry. TCs have caused the US economy over $200B worth of damage in 2017,

HARDWARE

NVIDIA POWERS TODAY'S FASTEST SUPERCOMPUTERS

22 of Top 25 Greenest

ORNL Summit
World's Fastest
27,648 GPUs| 149 PF

LLNL Sierra
World's 2nd Fastest
17,280 GPUs| 95 PF

Piz Daint
Europe's Fastest
5,704 GPUs| 21 PF

Total Pangea 3
Fastest Industrial
3,348 GPUs| 18 PF

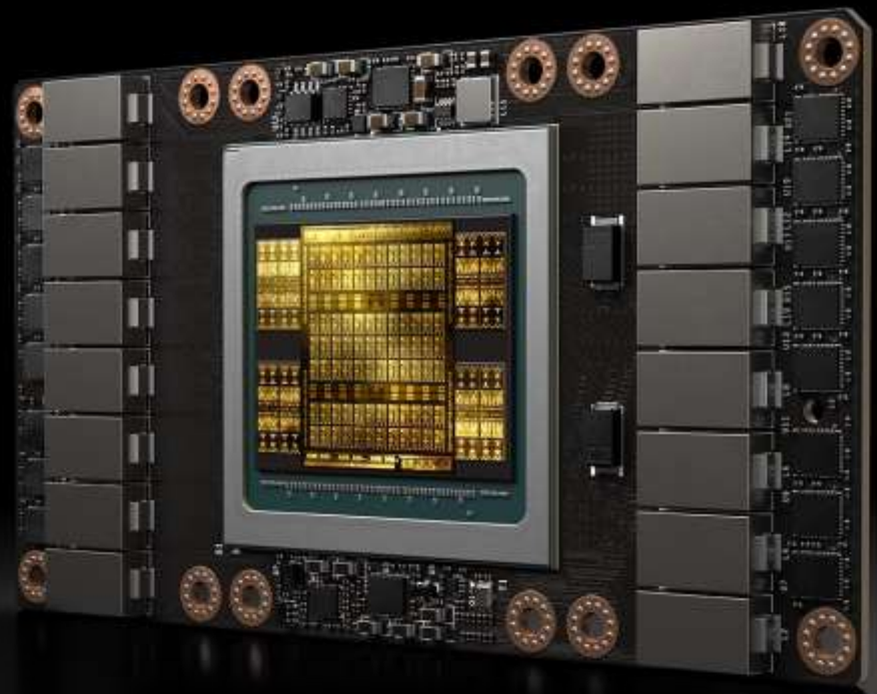
ABCI
Japan's Fastest
4,352 GPUs| 20 PF

NVIDIA POWERS WORLD'S FASTEST SUPERCOMPUTER

Summit Becomes First System To Scale The 100 Petaflops Milestone

122 PF
HPC

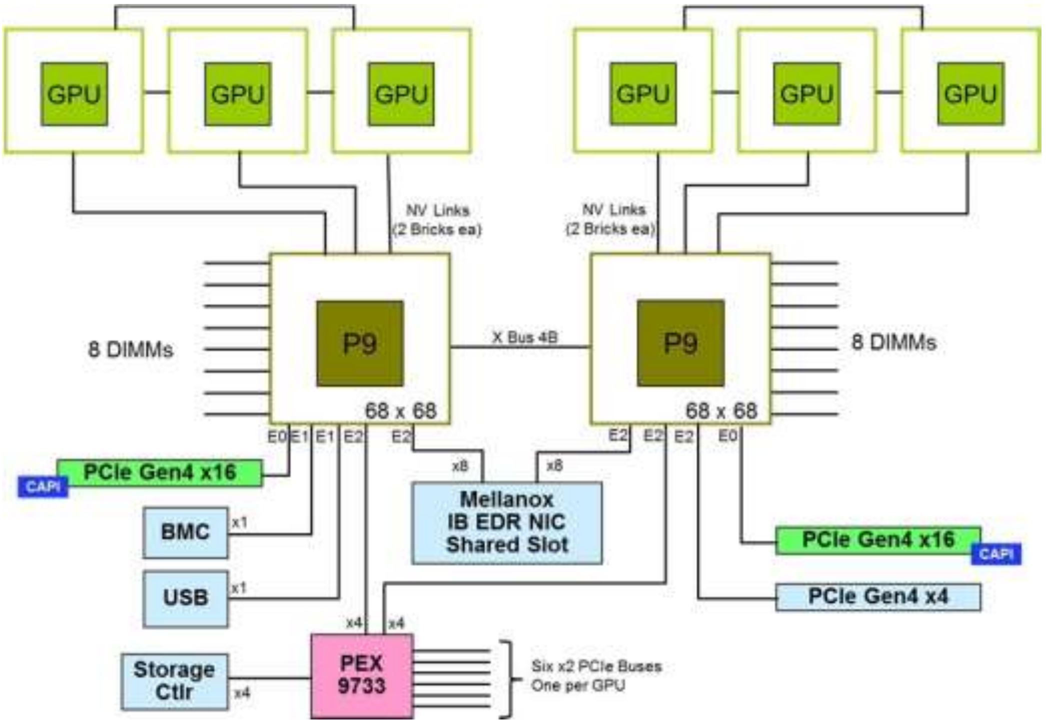
3 EF
AI



27,648
Volta Tensor Core GPUs

IBM AC922

6xV100 + 2xP9 (Water Cooled)



TESLA V100
TENSOR CORE GPU

World's Most Powerful
Data Center GPU

5,120 CUDA cores
640 NEW Tensor cores
7.8 FP64 TFLOPS | 15.7 FP32 TFLOPS
| 125 Tensor TFLOPS
20MB SM RF | 16MB Cache
32 GB HBM2 @ 900GB/s |
300GB/s NVLink

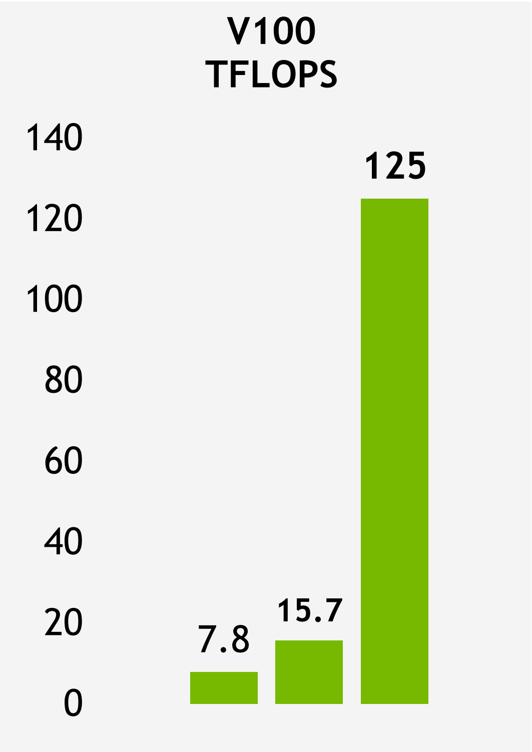
TENSOR CORE

4x4x4 matrix multiply and accumulate

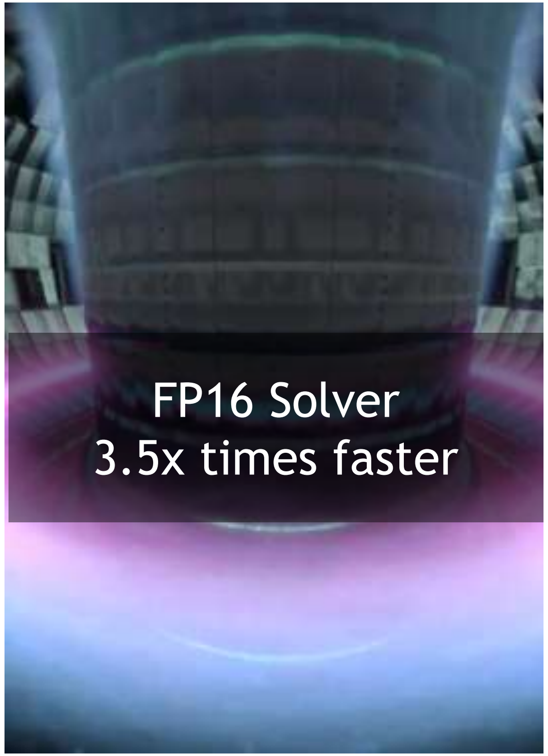
$$\underset{\text{FP16 or FP32}}{D} = \underset{\text{FP16}}{\begin{pmatrix} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} \end{pmatrix}} \underset{\text{FP16}}{\begin{pmatrix} B_{0,0} & B_{0,1} & B_{0,2} & B_{0,3} \\ B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,0} & B_{3,1} & B_{3,2} & B_{3,3} \end{pmatrix}} + \underset{\text{FP16 or FP32}}{\begin{pmatrix} C_{0,0} & C_{0,1} & C_{0,2} & C_{0,3} \\ C_{1,0} & C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,0} & C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \end{pmatrix}}$$

TENSOR CORES FOR SCIENCE

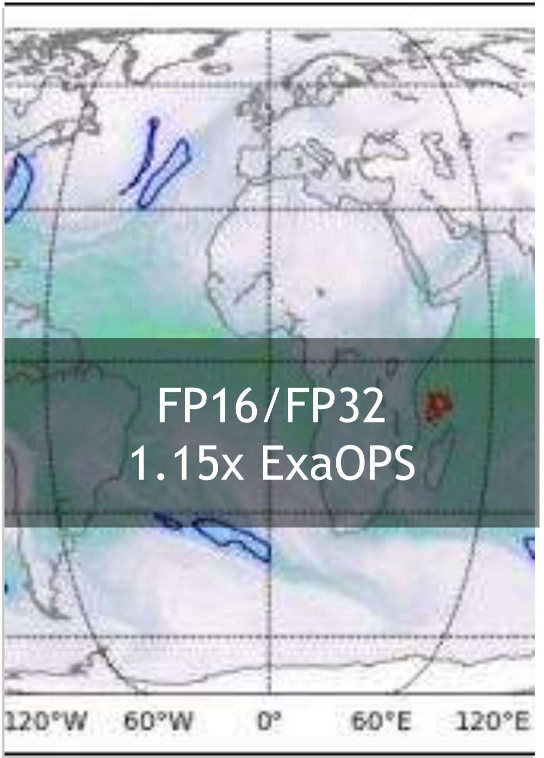
Multi-precision computing



FP64+ MULTI-PRECISION



PLASMA FUSION APPLICATION



AI-POWERED WEATHER PREDICTION

EARTHQUAKE SIMULATION

SOFTWARE: PERFORMANCE & PRDUCTIVITY

POWERING THE DEEP LEARNING ECOSYSTEM

NVIDIA SDK Accelerates Every Major Framework

COMPUTER VISION

OBJECT DETECTION

IMAGE CLASSIFICATION

SPEECH & AUDIO

VOICE RECOGNITION

LANGUAGE TRANSLATION

NATURAL LANGUAGE PROCESSING

RECOMMENDATION ENGINES

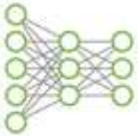
SENTIMENT ANALYSIS

DEEP LEARNING FRAMEWORKS

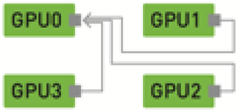
Caffe · Caffe2 · Chainer · Cognitive Toolkit · KALDI · mxnet · PaddlePaddle · PYTORCH · TensorFlow

NVIDIA DEEP LEARNING SDK and CUDA

cuDNN

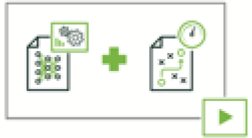


NCCL



cuBLAS

TensorRT



cuSPARSE

DeepStream SDK

developer.nvidia.com/deep-learning-software

NVIDIA DEVELOPER COMPUTEWORKS GAMEWORKS JETPACK DESIGNWORKS DRIVE Join Login

NVIDIA SDK

The Essential Resource for GPU Developers

DEEP LEARNING

Deep Learning SDK

High-performance tools and libraries for deep learning

AUTONOMOUS VEHICLES

NVIDIA DRIVE Platform

Deep learning, HD mapping and supercomputing solutions, from ADAS to fully autonomous

VIRTUAL REALITY

NVIDIA VRWorks™

A comprehensive SDK for VR headsets, games and professional applications

GAME DEVELOPMENT

NVIDIA GameWorks™

Advanced simulation and rendering technology for game development

ACCELERATED COMPUTING

NVIDIA ComputeWorks

Everything scientists and engineers need to build GPU-accelerated applications

DESIGN & VISUALIZATION

NVIDIA DesignWorks™

Tools and technologies to create professional graphics and advanced rendering applications

AUTONOMOUS MACHINES

NVIDIA JetPack™

Powering breakthroughs in autonomous machines, robotics and embedded computing

SMART CITIES

NVIDIA Metropolis

Edge-to-cloud development platform for smart cities

GPU ACCELERATED LIBRARIES

"Drop-in" Acceleration for Your Applications

DEEP LEARNING

- cuDNN
- TensorRT
- DeepStream SDK

SIGNAL, IMAGE & VIDEO

- cuFFT
- NVIDIA NPP
- CODEC SDK

LINEAR ALGEBRA

- cuBLAS
- cuSPARSE
- CUDA Math library
- cuSOLVER
- cuRAND

PARALLEL ALGORITHMS

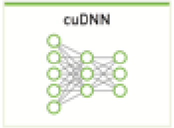
- nvGRAPH
- NCCL
- Thrust

NVIDIA DEEP LEARNING SDK

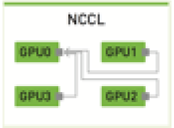
High Performance GPU-acceleration for Deep Learning

Powerful tools and libraries for designing and deploying GPU-accelerated deep learning applications

- High performance building blocks for training and deploying deep neural networks on NVIDIA GPUs
- Industry vetted deep learning algorithms and linear algebra subroutines for developing novel deep neural networks
- Multi-GPU and multi-node scaling that accelerates training on up to eight GPU



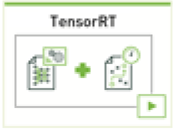
Deep Learning Primitives



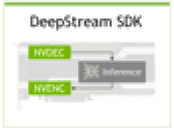
Multi-GPU Communication

Linear Algebra

Sparse Matrix Operations



Programmable Inference Accelerator



Deep Learning for Video Analytics

developer.nvidia.com/deep-learning-software

NVIDIA cuDNN

Deep Learning Primitives

High performance building blocks for deep learning frameworks

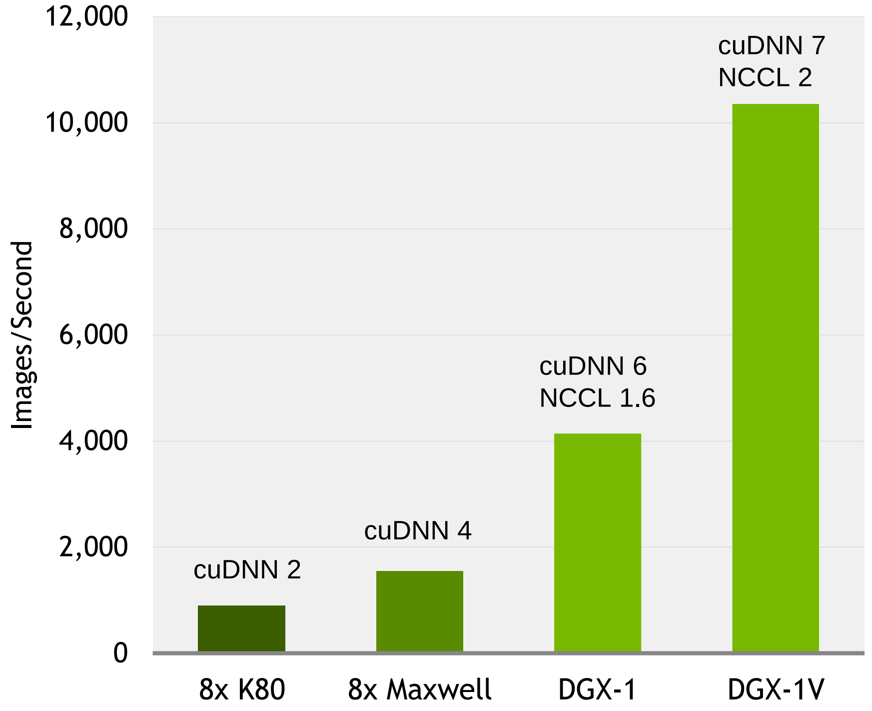
Drop-in acceleration for widely used deep learning frameworks such as Caffe2, Microsoft Cognitive Toolkit, PyTorch, Tensorflow and others

Accelerates industry vetted deep learning algorithms, such as convolutions, LSTM RNNs, fully connected, and pooling layers

Fast deep learning training performance tuned for NVIDIA GPUs

developer.nvidia.com/cudnn

Deep Learning Training Performance



“NVIDIA has improved the speed of cuDNN with each release while extending the interface to more operations and devices at the same time.”

— Evan Shelhamer, Lead Caffe Developer, UC Berkeley

NVIDIA COLLECTIVE COMMUNICATIONS LIBRARY (NCCL)

Multi-GPU and Multi-node Collective Communication Primitives

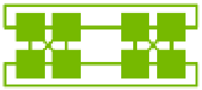
Open-source High-performance multi-GPU and multi-node collective communication primitives optimized for NVIDIA GPUs

Fast routines for multi-GPU multi-node acceleration that maximizes inter-GPU bandwidth utilization

Easy to integrate and MPI compatible. Uses automatic topology detection to scale HPC and deep learning applications over PCIe and NVLink

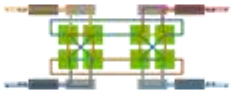
Accelerates leading deep learning frameworks such as Caffe2, Microsoft Cognitive Toolkit, MXNet, PyTorch and more

developer.nvidia.com/nccl



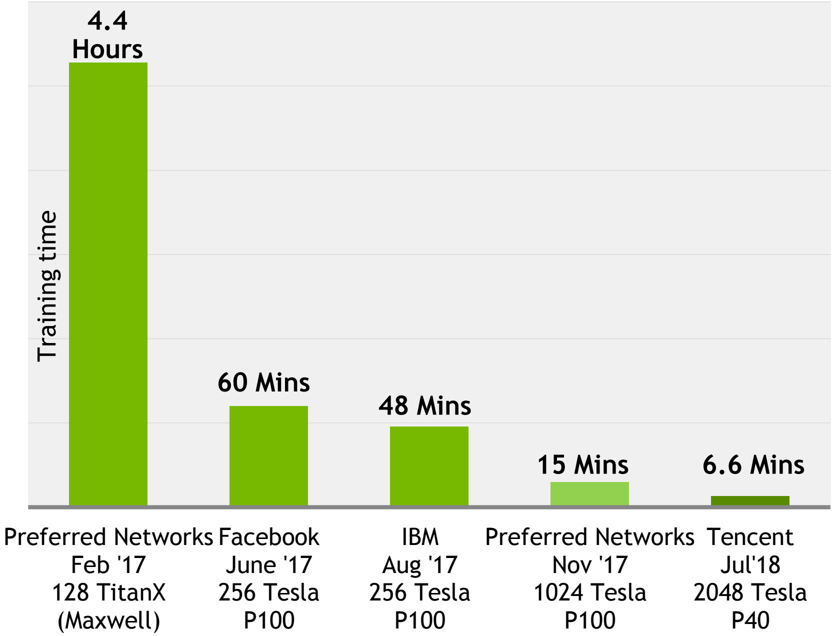
Multi-GPU:
NVLink, PCIe

Multi-Node:
InfiniBand verbs,
IP Sockets



Automatic
Topology
Detection

Scaling training to 2048 GPUs



ResNet-50 | Dataset: Imagenet | Trained for 90 Epochs

HOROVOD (UBER)

Horovod: fast and easy distributed deep learning in TensorFlow

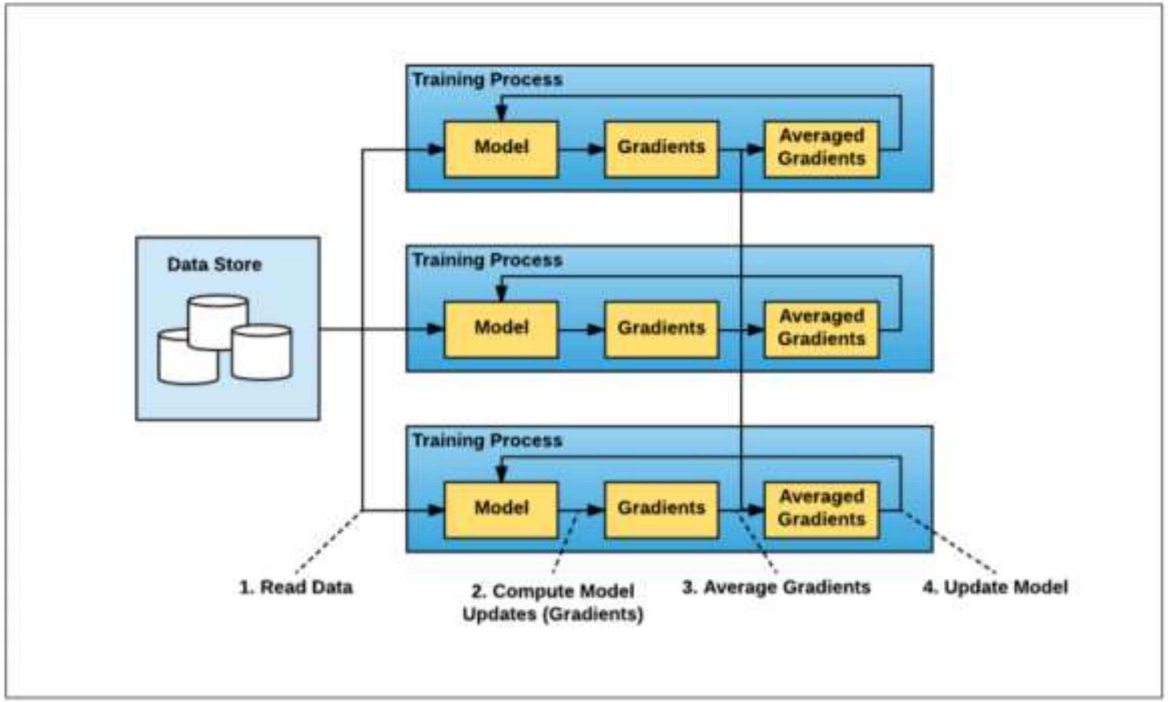


Figure 2: The "data parallel" approach to distributed training involves splitting up the data and training on multiple nodes in parallel. In synchronous cases, the gradients for different batches of data are calculated separately on each node but averaged across nodes to apply consistent updates to the model copy in each node.

Sergeev, Alexander, and Mike Del Balso. "Horovod: fast and easy distributed deep learning in TensorFlow." arXiv preprint arXiv:1802.05799 (2018).

FULLY CONVOLUTIONAL NETWORKS (FCN)

"Fully Convolutional Networks for Semantic Segmentation", Shellhammer et al, 2015

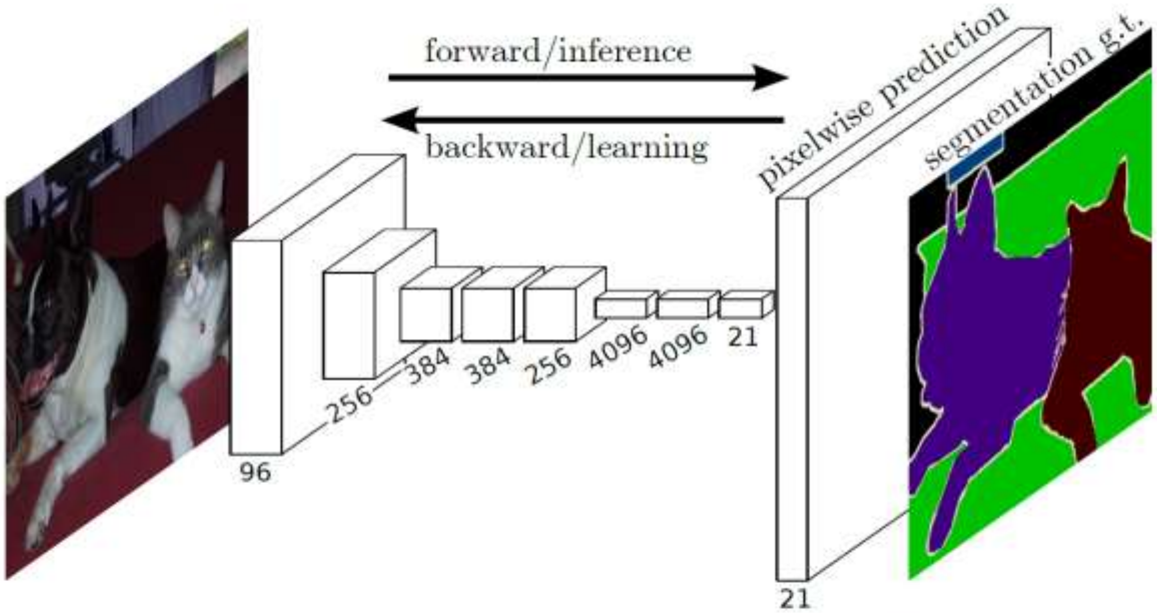
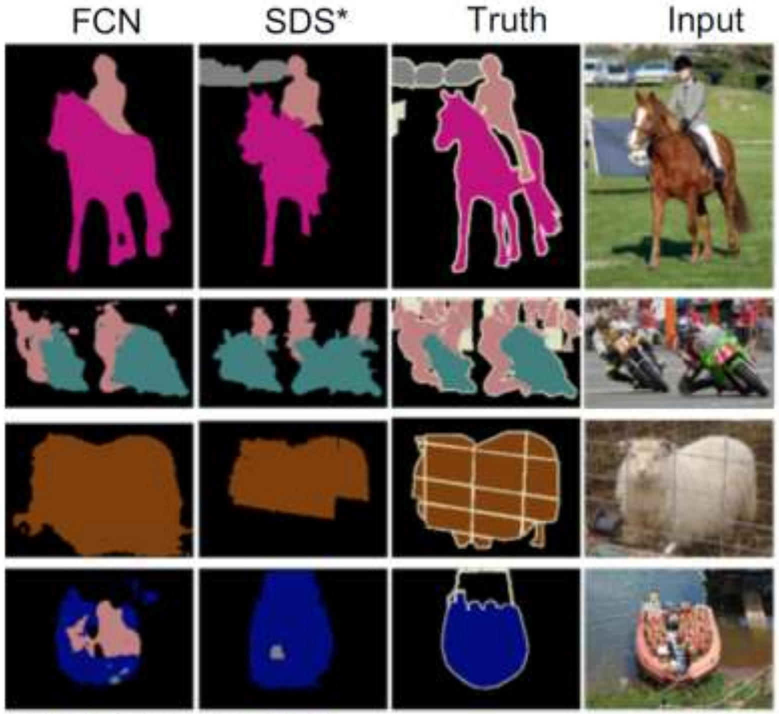


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

SEMANTIC SEGMENTATION

FCN *vs* SDS



Slide credit: Jonathan Long

Relative to prior state-of-the-art SDS:

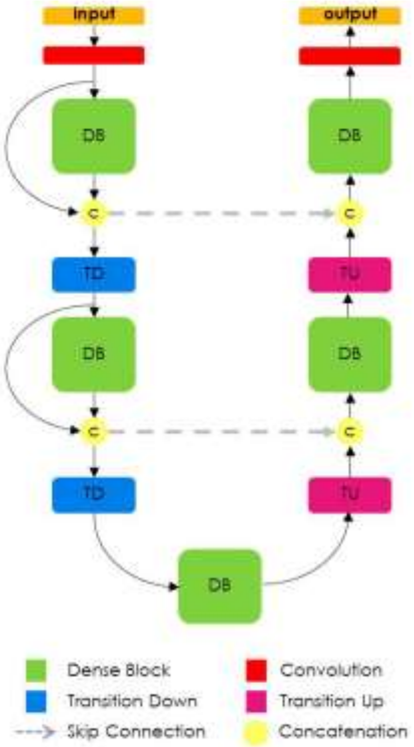
- 20% relative improvement for mean IoU
- 286× faster

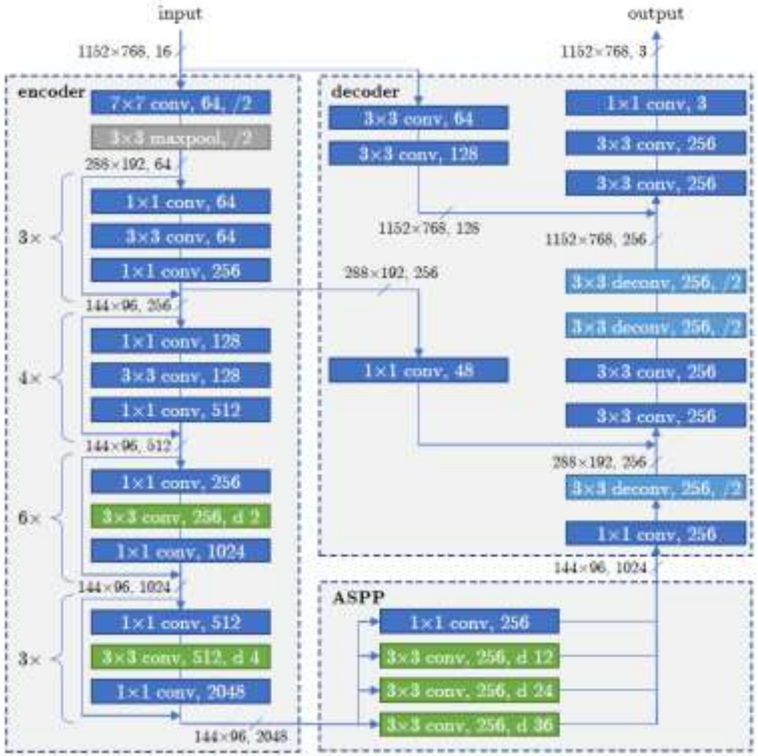
*Simultaneous Detection and Segmentation Hariharan et al. ECCV14

25

DEEP NEURAL NETWORKS

Tiramisu (left) and DeepLabv3+ (right)





INNOVATIONS (1)

Deep Learning Innovations

- Weighted loss
- Layer-wise adaptive rate control (LARC)
- Multi-channel segmentation
- Gradient lag
- Modifications to the neural network architectures

INNOVATIONS (2)

System Innovations

- High speed parallel data staging
- Optimized data ingestion pipeline
- Hierarchical all-reduce

INNOVATIONS (2)

System Innovations

- High speed parallel data staging
  - Read 1500 images per node (250 per GPU)
  - 800GB of high-speed SSD storage on each node
  - Distributed data staging system that first divides the data set into disjoint pieces, read to some nodes, than copy to other nodes

INNOVATIONS (2)

System Innovations

- Optimized data ingestion pipeline
  - TensorFlow input pipeline that reads the input files and converts them into the tensors that are fed through the network (read and convert to TFRecords)
  - Eliminate serialization by enabling the prefetching option of TensorFlow datasets
  - HDF5 Library serializes calls
  - By using the Python multiprocessing module, transform these parallel worker threads into parallel worker processes, each using its own instance of the HDF5 library

INNOVATIONS (2)

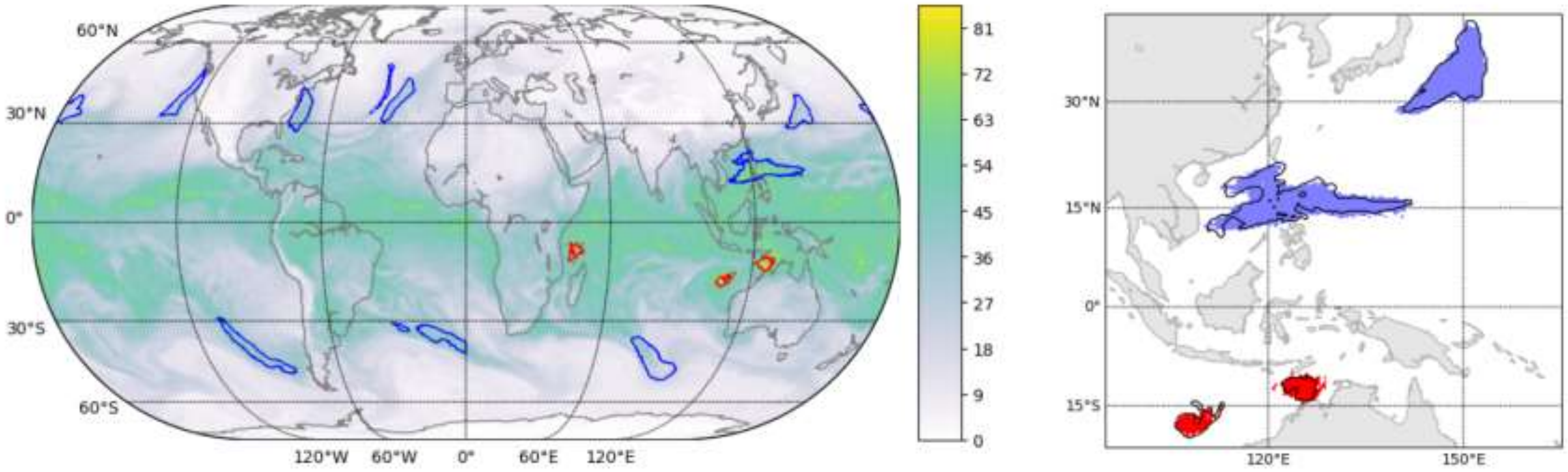
System Innovations

- Hierarchical all-reduce
  - Horovod is a Python module that uses MPI to transform a single-process TensorFlow application into a data-parallel implementation
  - Each MPI rank creates its own identical copy of the TensorFlow operation graph.
  - The first issue was a bottleneck on the first rank, which acts as a centralized scheduler for Horovod operations. Solution: organize in a communication tree.
  - The existing Horovod implementation is able to reduce data residing on GPUs in two different ways, either by a standard MPI_Allreduce or by using the NVIDIA Collective Communications Library (NCCL)
  - NCCL is better for intra-node (exploits NVLINK) and Standard MPI is better for inter-node

RESULTS

CLIMATE RESULTS

Atmospheric Rivers (AR) in Blue and Tropical Cyclones (TC) in Red

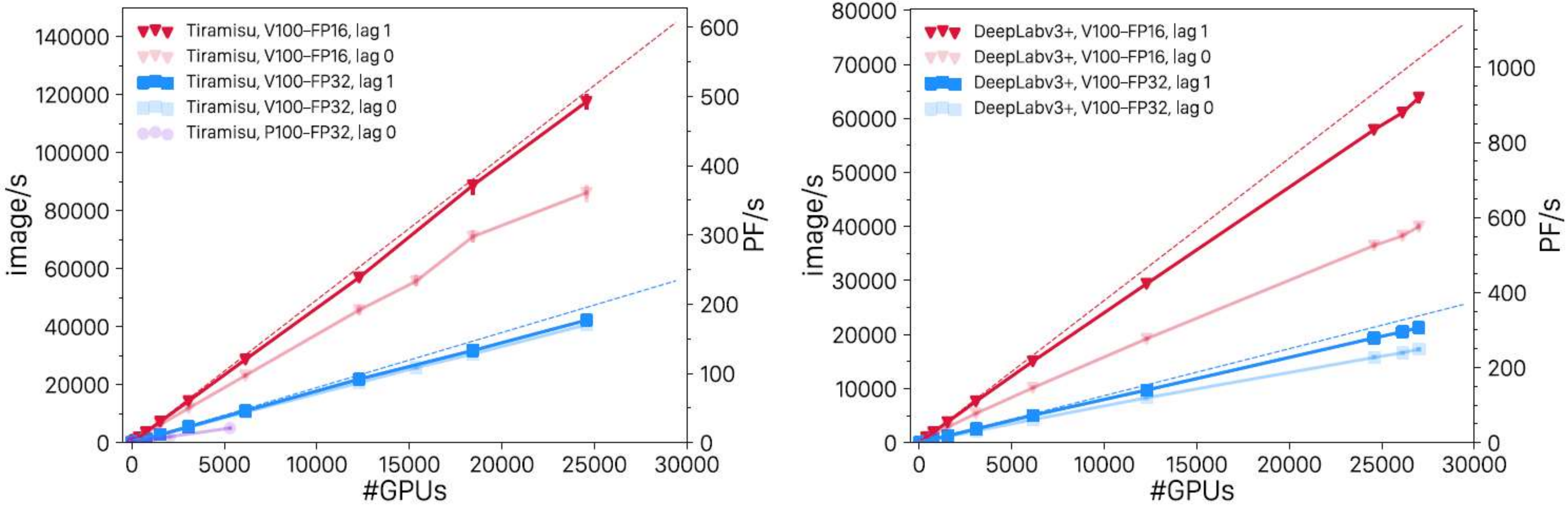


(a) Segmentation masks overlaid on a globe. Colors (white->yellow) indicate IWV (integrated water vapor, kg/m^2), one of the 16 input channels used by the network.

(b) Detailed inset showing predictions (red and blue) vs. labels used in training (black).

Fig. 7: Segmentation results from modified DeepLabv3+ network. Atmospheric rivers (ARs) are labeled in blue, while tropical cyclones (TCs) are labeled in red.

SCALING RESULTS



(a) Tiramisu (b) DeepLabv3+

Fig. 4: Weak scaling results in terms of images/sec and sustained performance in PF/s on Summit (FP16 and FP32, Tiramisu and DeepLabv3+) and Piz Daint (FP32, Tiramisu). The dashed lines represent the ideal scaling lines for the different architectures and precisions.

OVERALL RESULTS

Tiramisu (PizDaint) and DeepLabv3 (Summit)

	#GPUs	GPU Arch	Peak (PFLOPS)	Sustained (PFLOPS)	Efficiency
Tiramisu (FP32)	5300	P100 (CUDA)	26.6	21	79.0%
DeepLabv3+ (FP32)	27360	V100 (CUDA)	359.2	325.8	90.7%
DeepLabv3+ (FP16)	27360	V100 (Tensor Cores)	1130.0	999	88.4%

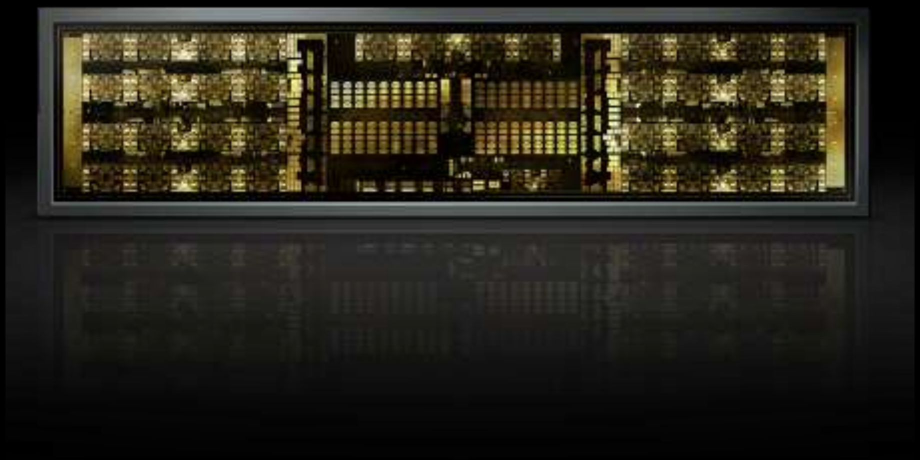
PUSHING AI COMPUTING LIMITS

NVIDIA® DGX-1™

NVSWITCH

World's Highest Bandwidth On-node Switch

7.2 Terabits/sec or 900 GB/sec
18 NVLINK ports | 50GB/s per port bi-directional
Fully-connected crossbar
2 billion transistors | 47.5mm x 47.5mm package



NVIDIA DGX-2

THE LARGEST GPU EVER CREATED

2 PFLOPS | 512GB HBM2 | 10 kW | 350 lbs

MORE SCIENTIFIC EXAMPLES

GALAXY CLASSIFICATION

Merging vs Not-Merging

Astronomy & Astrophysics manuscript no. Paper-Real-Sim-Mergers
May 13, 2019

Identifying Galaxy Mergers in Observations and Simulations with Deep Learning

W. J. Pearson[1,2], L. Wang[1,2], J. W. Trayford[3], C. E. Petrillo[2], F. F. S. van der Tak[1,2]

[1] SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD, Groningen, The Netherlands
e-mail: `w.j.pearson@sron.nl`
[2] Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands
[3] Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

Received DD Month YYYY / Accepted DD Month YYYY

GALAXY CLASSIFICATION

Merging vs Not-Merging

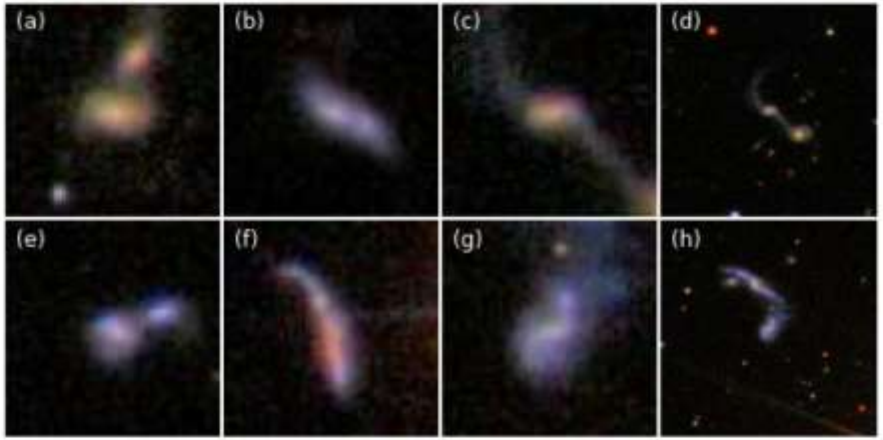


Fig. 20. Examples of SDSS FN galaxies from the simulation network for (a) a galaxy with a clear merging counterpart, (b) a clearly disturbed system, (c) a galaxy whose merger companion is outside of the 64×64 pixel image and (d) the larger 256×256 pixel image showing the merger companion outside of panel (c). Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d).

Table 3. Confusion matrix for SDSS images classified by the observation network.

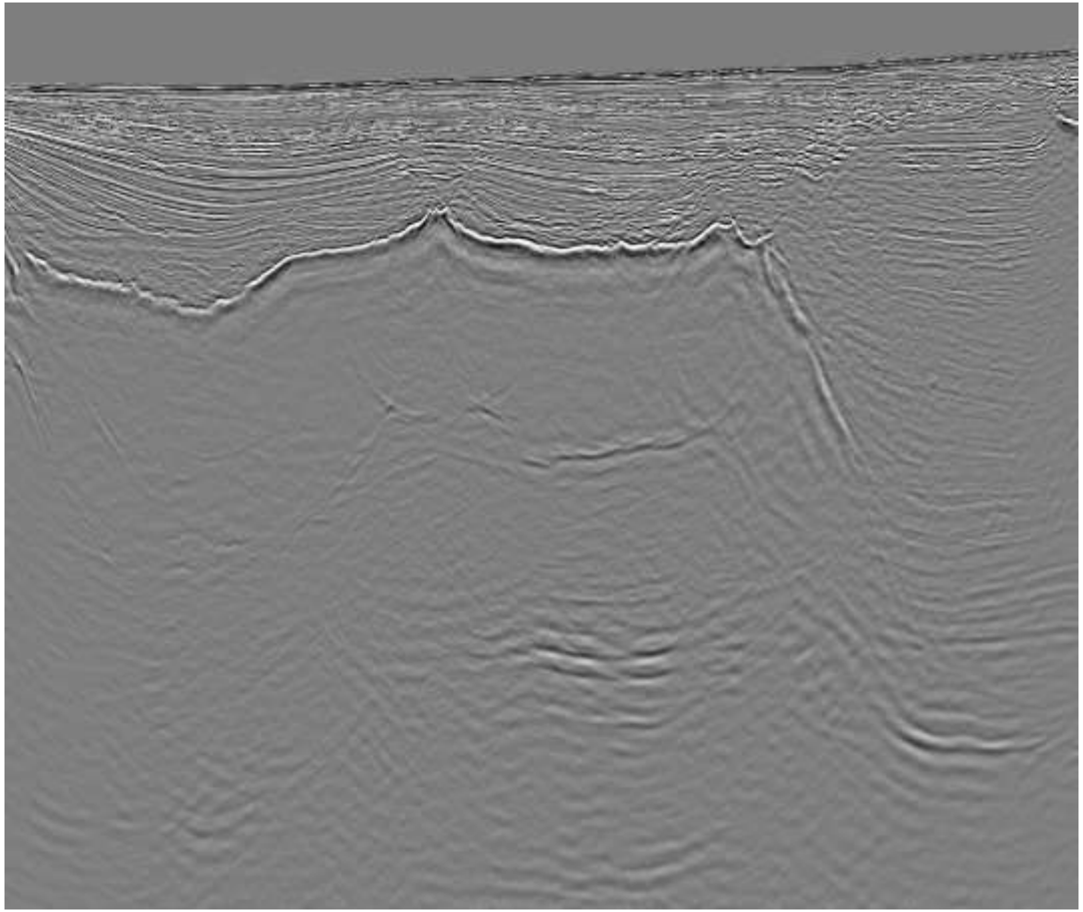
		Network Classification		
		Merger	Non-merger	Total
Catalogue Classification	Merger	276 TP	24 FN	300
	Non-merger	27 FP	273 TN	300
	Total	303	297	

1º LATIN AMERICA SUPER-COMPUTER: PETROBRAS FENIX

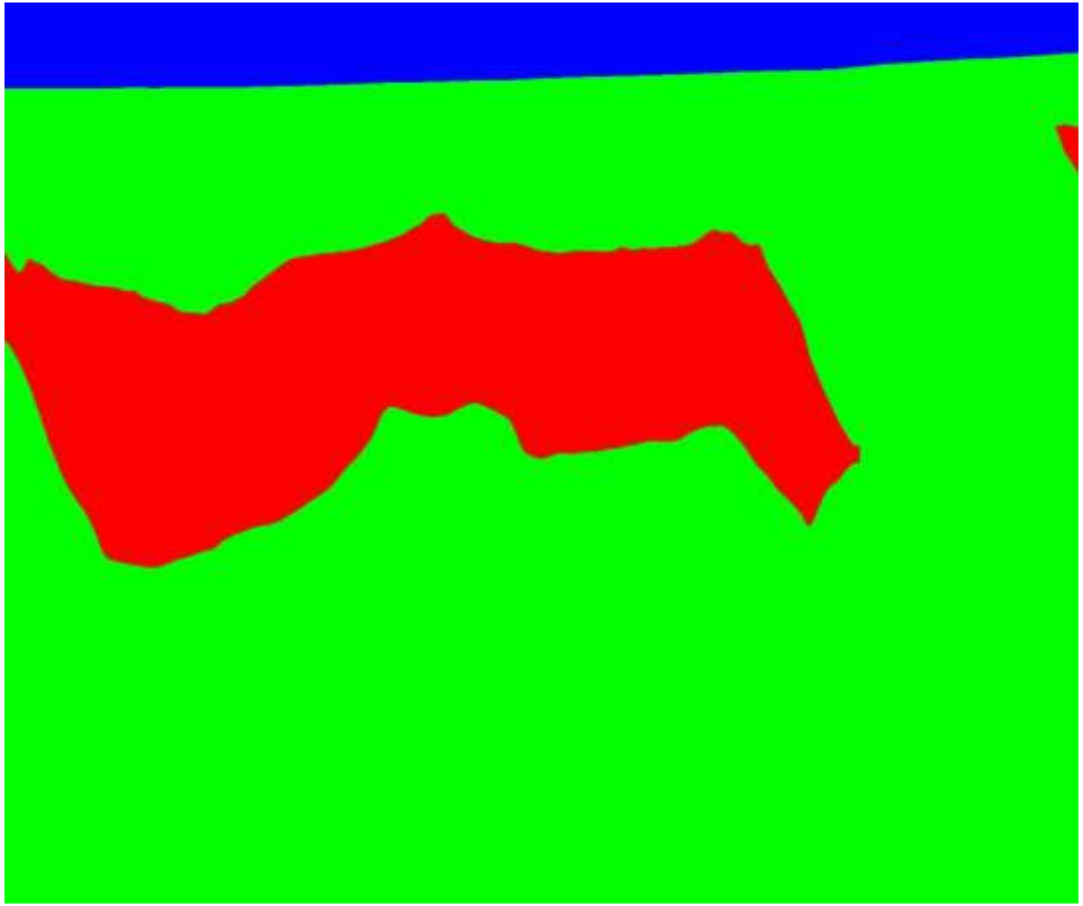
Fenix is at the 142nd position in the Top500 List.
576x V100 = 288x Nodes w/ 2xGPU/node

Source: Top500.org.

TRAINING SET



Features (Seismic)

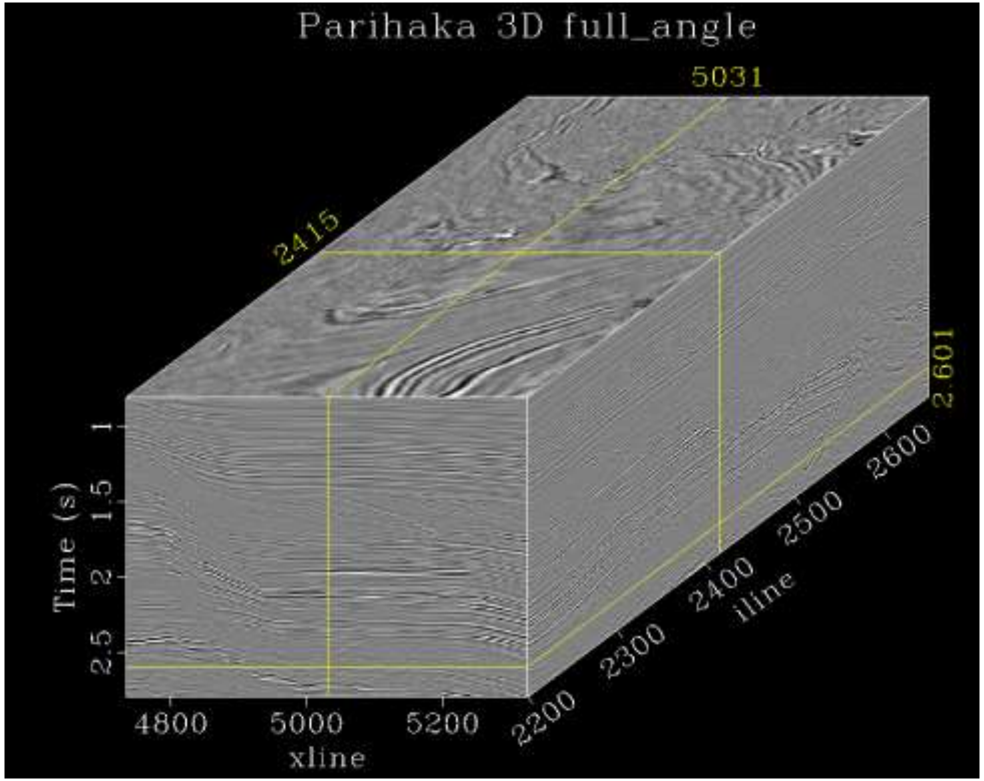


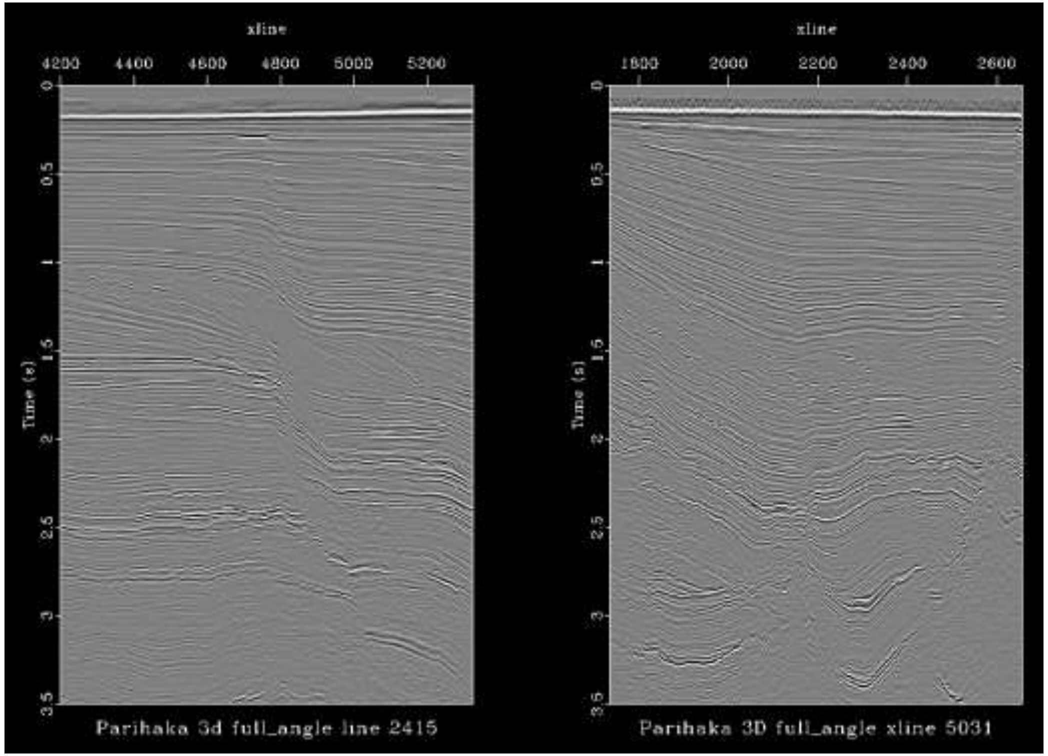
Labels

TRAINING IMAGES

Parihaka dataset (SEGY)

https://wiki.seg.org/wiki/Parihaka-3D





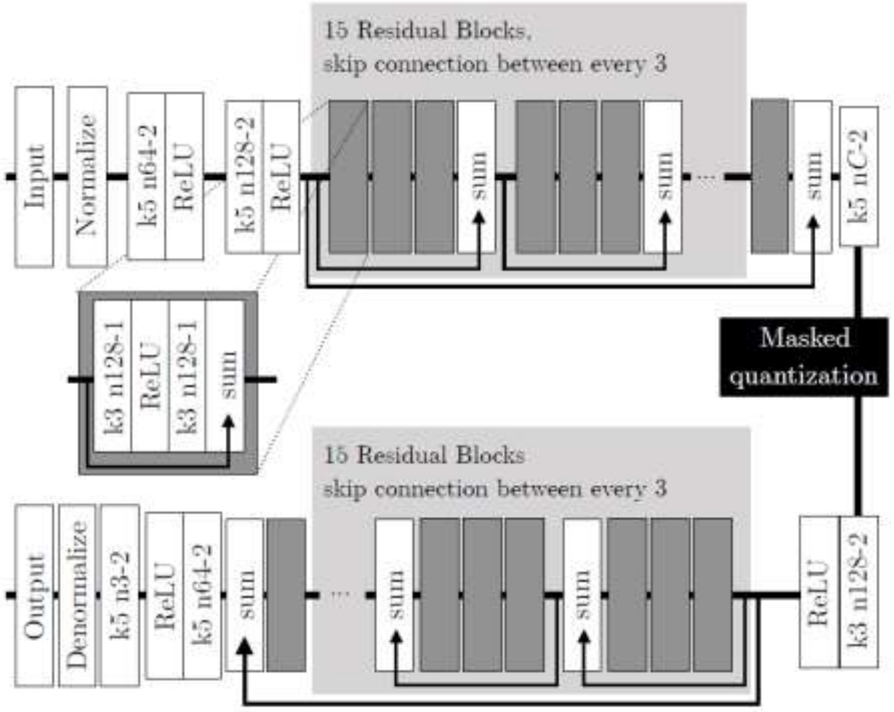
METHOD

Conditional Problabilistc Deep Auto-Encoder

Based on the state-of-art on image compression work (CVPR-18):

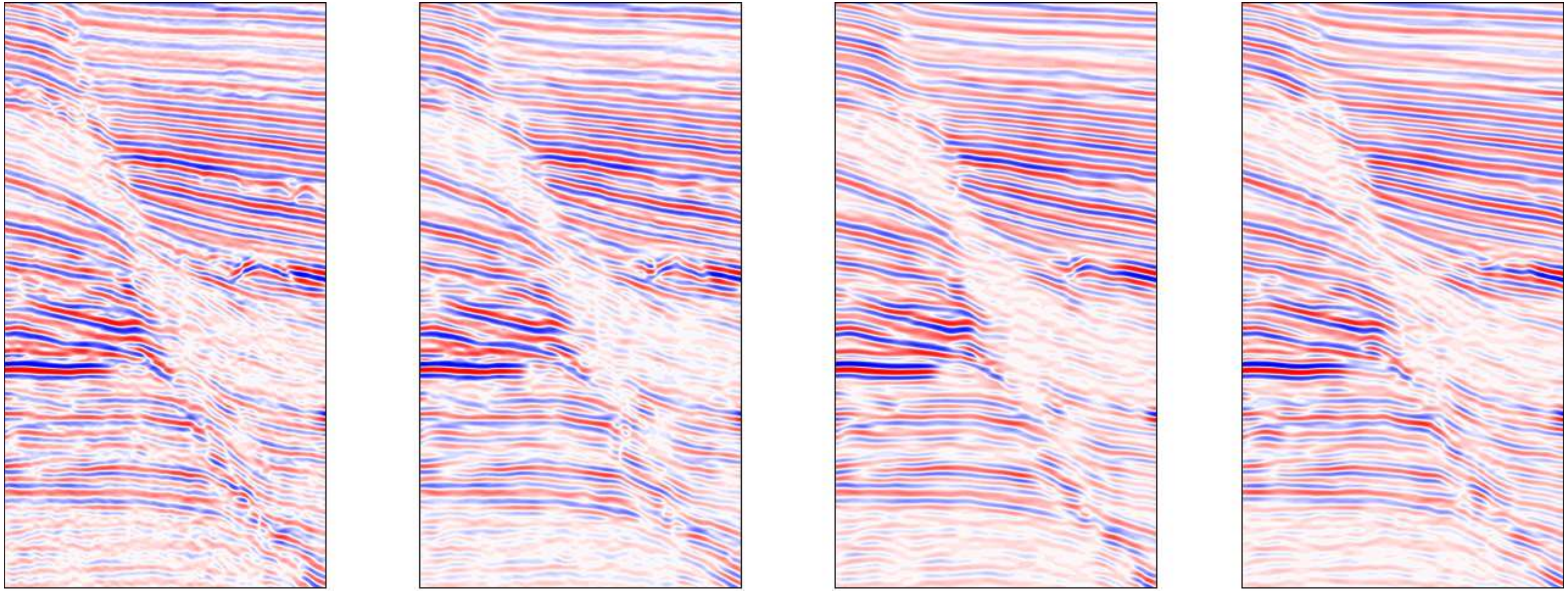
“Conditional Probability Models for Deep Image Compression” – Mentzer et. al.

Original work operates on 8-bits depth images. Changes for 32-bits and specific training protocols were performed for 3D post-stacked seismic data.



EXPERIMENTS

Visual Comparison

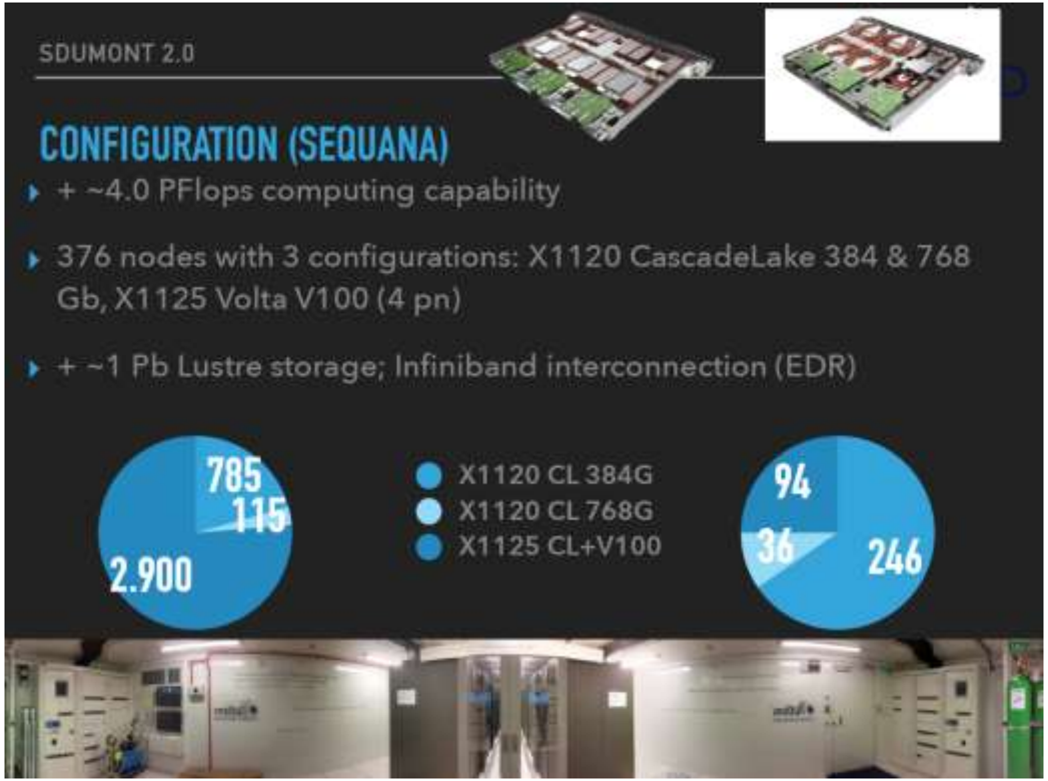


Original | 40.08db | 3.58bpv | 39.9db | 1.56bpv | 37.87db | 0.74bpv

Figure 2 *Visual comparison of different trained models for Parihaka-3D dataset. The results (crossline 4791) refer to training models targeting different compression rates. PSNR and bpv are below the decompressed slices. As we move to lower compression rates (rightmost panel), the reconstruction error increases but most representative seismic features are still preserved.*

NEW SUPERCOMPUTERS IN BRAZIL

LNCC (Rio de Janeiro)
376x V100 = 96x Nodes w/ 4xGPU/node



SENAI-SIMATEC (Salvador)
312x V100 = 78x Nodes w/ 4xGPU/node

LEARN & SHARE MORE

NVIDIA

GTC GPU TECHNOLOGY CONFERENCE

Don’t miss the premier AI conference.

www.nvidia.com/gtc

March 22—26, 2020 | Silicon Valley

CONNECT

Connect with hundreds of experts from top industry, academic, startup, and government organizations

LEARN

Gain insight and valuable hands-on training through over 500+ sessions

DISCOVER

See how GPU technology is creating breakthroughs in deep learning, cybersecurity, data science, healthcare and more

INNOVATE

Explore disruptive innovations that can transform your work

JOIN US AT GTC 2020 | USE VIP CODE XXXXX FOR 25% OFF

NVIDIA

GTC GPU TECHNOLOGY CONFERENCE

THE LATEST DEEP LEARNING DEVELOPER TOOLS

March 22 | Full-Day Workshops
March 23 - 26 | Conference & Training

Get the hands-on experience you need to transform the future of AI, high-performance computing and more with NVIDIA's Deep Learning Institute (DLI).

Register for GTC 2020 to earn certification in full-day workshops, join instructor-led sessions, and start self-paced training.

www.nvidia.com/en-us/gtc/sessions/training/

Join the NVIDIA Developer Program

Access everything you need to develop with NVIDIA products.

Register Now

developer.nvidia.com

DEEP LEARNING

Deep Learning SDK

High-performance tools and libraries for deep learning

ACCELERATED COMPUTING

NVIDIA ComputeWorks

Everything scientists and engineers need to build GPU-accelerated applications

AUTONOMOUS VEHICLES

NVIDIA DRIVE Platform

Deep learning, HD mapping and supercomputing solutions, from ADAS to fully autonomous

SMART CITIES

NVIDIA Metropolis

Edge-to-cloud development platform for smart cities

NVIDIA DEEP LEARNING INSTITUTE

Hands-on self-paced and instructor-led training in deep learning and accelerated computing for developers

Request onsite instructor-led workshops at your organization: **www.nvidia.com/requestdli**

Take self-paced labs online: **www.nvidia.com/dlilabs**

Download the course catalog, view upcoming workshops, and learn about the University Ambassador Program: **www.nvidia.com/dli**

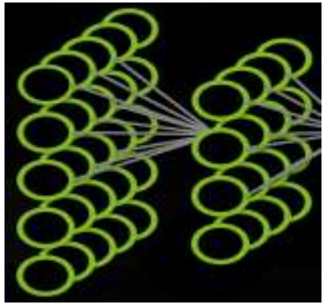
Caffe2

Microsoft Cognitive Toolkit

mxnet

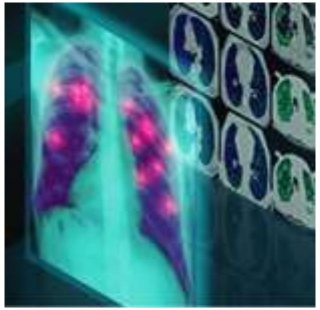
TensorFlow

PYTORCH

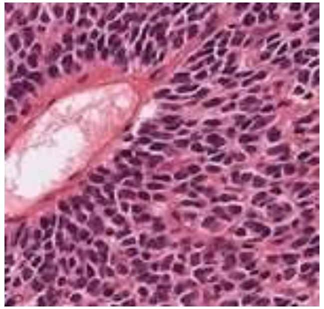


Deep Learning Fundamentals

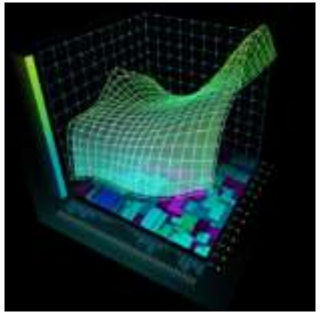
Autonomous Vehicles



Medical Image Analysis



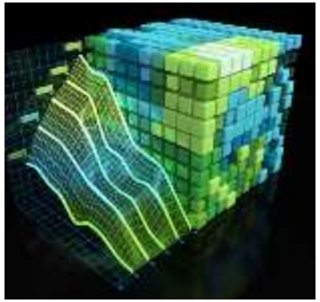
Genomics



Finance

Intelligent Video Analytics

Game Development & Digital Content



Accelerated Computing Fundamentals

More industry-specific training coming soon…

NVIDIA HW GRANT PROGRAM

Titan V Volta

- Scientific Computing
- HPC
- Deep Learning

Quadro P6000

- Scientific Visualization
- Virtual Reality

Jetson TX2 (Dev Kit)

- Robotics
- Autonomous Machines

https://developer.nvidia.com/academic_gpu_seeding

Obrigado
Gracias
Thank you

pcruzesilva@nvidia.com

NVIDIA.