## ALISON B LOWNDES

AI DevRel | EMEA

@alisonblowndes

JULY 2018





## NASA FRONTIER DEVELOPMENT LAB





















KICK OFF WORKSHOP ESA ESRIN, ROME **25th - 29th June '18** 

RESEARCH SPRINT 30th June - 17th August





## TITAN V THE MOST POWERFUL PC GPU EVER

5,120 CUDA cores 640 NEW Tensor cores

12GB HBM2 Memory 1.7Gbps









# A massive Deep Learning challenge





"Our in-depth technical assessment confirms the Xavier SoC architecture is suitable for use in autonomous driving applications and highlights NVIDIA's commitment to enable safe autonomous driving"; TÜV SÜD, Germany

## Xavier (ASIL-D rated)

# "The Building Blocks of Interpretability"

# https://distill.pub/2018/building-blocks/

Chris Olah, Google et al, March 2018

# **Dynamic Memory Networks**

MetaMind now Salesforce (2yrs ago) https://arxiv.org/pdf/1603.01417v1.pdf



What is this sculpture made out of ?

Answer: metal





What color are the bananas ?







Who is on both photos ?

Answer: girl



Answer: stripes What is the pattern on the cat's fur on its tail?



Did the player hit the ball?





### Designing and building state-of-art Automated Machine Learning software and services for real world problems

### Background

Spin out from the University of Oxford

Combining leading edge machine learning research with advanced software engineering

Working with leading companies to deliver insight & quantifiable benefits

Setting benchmarks for a new wave of machine learning

### Vision

Build services and software that Automate the Data Science lifecycle

Services which understand the entanglement, uncertainty, complexity and power of machine learning

Services that are evergreen and quantum ready





"After finishing this book, you will have a deep understanding of how to set technical direction for a machine learning project."

https://bit.ly/2wNFYIY

# MACHINE LEARNING YEARNING

Technical Strategy for AI Engineers, In the Era of Deep Learning



## NVIDIA DEEP LEARNING INSTITUTE

Online self-paced labs and instructor-led workshops on deep learning and accelerated computing

# Take self-paced labs at www.nvidia.co.uk/dlilabs

View upcoming workshops and request a workshop onsite at www.nvidia.co.uk/dli

Educators can join the University Ambassador Program to teach DLI courses on campus and access resources. Learn more at www.nvidia.com/dli









Fundamentals

Autonomous Vehicles

Healthcare







Intelligent Video Analytics

Robotics







**Digital Content** 

Finance

Accelerated Computing



# **CONVOLUTIONAL NEURAL NETWORKS**



Image

# What ML Models Do We Leverage?



# AI INFERENCE IS THE NEXT GREAT CHALLENGE













kernel

**Brain Computer Interfaces** Focused on treatment for disease and dysfunction eg epilepsy, depression, Parkinsons but ultimately to advance human intelligence by restoring and extending cognitive vibrancy.

"We're either going to have to merge with AI or be left behind"; Elon Musk









Particle Count: 0 Diffuse Count: 0 Shape Match Count: 0 Rigid Body Count: 6500 Rigid Shape Count: 9500 Rigid Joint Count: 12000 Spring Count: 0 Tetra Count: 0 Num Substeps: 4 Num Iterations: 30

#### Device: TITAN X (Pascal)







## OPENMINED.ORG HTTPS://GITHUB.COM/OPENMINED



## **"SEEING" GRAVITY FOR THE FIRST TIME**

In September 2015, 100 years after Einstein predicted them, gravitational waves were observed for the first time. Astronomers at the Laser Interferometer Gravitational-wave Observatory have since used GPU-powered deep learning to process gravitational wave data 100 times faster than previous methods, making real-time analysis possible and putting us one step closer to understanding the universe's oldest secrets.



## **A PLETHORA OF HEALTHCARE STORIES**



**Molecular Energetics** For Drug Discovery



Al for Drug Discovery



**Predicting Growth** Problems





**Image Processing** 





**Gene Mutations** 



**Treatment Outcomes** 



Diagnosis Errors by 85%



Predicting Toxicology



**Detect Colon Polyps** 

Type



Predicting Disease from Medical Records



Enabling Detection of Fatty Acid Liver Disease



## **Radiation Therapy with Al** ViewRay MRIdian (imaging & radiation dose)







# **GPU OPEN ANALYTICS INITIATIVE**

## github.com/gpuopenanalytics



# AI SUPERCOMPUTING WILL TRANSFORM HPC

Extending Reach of HPC By Combining Computational & Data Science





Molecular Dynamics



Structural Analysis



N-body Simulation





"What's happening?"



"Is there cancer?"



"Next move?"



"What does she mean?"

## DATA SCIENCE





Drug Discovery

Clean Energy





Monitoring Climate Change

## **COMPUTATIONAL & DATA SCIENCE**

# **GPU READY APPS** AWARENESS : ADOPTION : UTILIZATION

Customers	Guide	Benefits	
Target	Quick Start	llsor	
DevelopersDeep Learning	step-by-step instructions	✓ Faster discoveries	
	Simple, ISV-approved,	✓ Higher productivity	
IT Managers HPC/Data Centers	<u>4.</u> <u>5.</u>	✓ Optimal results	
	<u>2.</u> <u>3.</u>	✓ Best app practices	
App Users Life Science	<u>1.</u>	✓ Easy setup	

# POWERING THE DEEP LEARNING ECOSYSTEM

NVIDIA SDK accelerates every major framework





developer.nvidia.com/deep-learning-software

### 🧼 nvidia developer

### **NVIDIA SDK**

The Essential Resource for GPU Developers

### **NVIDIA SDK**

## 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

### **SELF-DRIVING CARS**

NVIDIA DriveWorks™

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

### ADDITIONAL RESOURCES

More resources for GPU Developers

### 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

## **NVIDIA DEEP LEARNING SDK UPDATE**

**GPU-accelerated DL Primitives** cuDNN 7 Faster training **Optimizations for RNNs** Leading frameworks support



# **NEW NVIDIA TENSORRT 4**

**Programmable Inference Accelerator** 



Compile and Optimize Neural Networks | Support for Every Framework Optimize for Each Target Platform

# **NVIDIA TensorRT 4**

## RNN and MLP Layers • ONNX Import • NVIDIA DRIVE Support



Accelerate inference of recommenders, speech and machine translation apps with new layers and optimizations 50x Faster Inference for ONNX Models

Easily import and accelerate inference for ONNX frameworks with native ONNX parser in TensorRT

# Support for NVIDIA DRIVE Xavier



Deploy optimized deep learning inference models NVIDIA DRIVE Xavier

Free download to members of NVIDIA Developer Program soon at developer.nvidia.com/tensorrt



## **TENSORCORES** D = A \* B + C (4x4)



# **ANNOUNCING NVIDIA DGX-2**

## THE LARGEST GPU EVER CREATED



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

# MULTI-NODE DGX-1 "A-HA" MOMENTS IN DL CLUSTER DESIGN

## Additional design insights to get you started



# NVIDIA GPU CLOUD

## **Optimized Stacks for Every Cloud**



GPU CONTAINERS				
DEEP LEARNING Caffe Caffe2 Chainer PaddlePaddle CNTK CUDA Digits H2O.ai MXNet PyTorch TensorFlow TensorRT Theano Torch	HPC GAMESS Gromacs LAMMPS NAMD RELION CHROMA MILC CANDLE Lattice Microbes	HPC Viz ParaView Holodeck ParaView IndeX ParaView OptiX IndeX VMD ANALYTICS MapD Kinetica		

20,000+ Registered Organizations | 30 Containers NOW on AWS, GCP, AliCloud, Oracle Cloud, DGX

# **ALWAYS UP-TO-DATE**

## Monthly Updates from NVIDIA to Frameworks and Containers

### **Containerized Applications**



Linux Kernel + CUDA Driver

# JETSON TX2 SUPERCOMPUTER FOR AI AT THE EDGE

2 Core i7 PCs in <10W 256 CUDA cores >1 TFLOPS





# JETSON XAVIER DEVELOPER KIT

\$1299 (US) Available from distributors WW Early access August 2018



# JETSON XAVIER DEVELOPER KIT

## \$1299 (US) Available from distributors WW Early access August 2018







NEW YORK UNIVERSITY DEEP LEARNING

# Robotics Teaching Kit with 'Jet' - ServoCity

Available to Instructors Now! developer.nvidia.com/teaching-kits

# NVIDIA INCEPTION PROGRAM

Accelerates AI startups with a boost of GPU tools, tech and deep learning expertise



## www.nvidia.com/inception

## **Startup Qualifications**

Driving advances in the field of AI Business plan Incorporated Web presence

## Technology

DL startup kit<sup>\*</sup> Pascal Titan X Deep Learning Institute (DLI) credit Connect with a DL tech expert

DGX-1 ISV discount<sup>\*</sup> Software release notification Live webinar and office hours \*By application

## Marketing

Inclusion in NVIDIA marketing efforts GPU Technology Conference (GTC) discount

Emerging Company Summit (ECS) participation<sup>+</sup>

Marketing kit

One-page story template eBook template Inception web badge and banners Social promotion request form Event opportunities list

Promotion at industry events GPU ventures<sup>+</sup> +By invitation





NVIDIA .

Additional consulting by

## Contributors

Written by James McClung & Beverly Clarke

Adam Grzywaczewski Alison B Lowndes Sam Jones Bea Longworth Charlotte Han Tom Bradley

Twin Karmakharm Paul Richmond

Prady Dash Felix Manoharan

Peter Balshaw

Gemma Bond Andrew Csizmadia Antony Morse

With special thanks to Elan Raja III of Scan Computers International.

## **COMPUTING AT SCHOOL** EDUCATE · ENGAGE · ENCOURAGE In collaboration with BCS, The Chartered Institute for IT





The University Of Sheffield.

