LEAPS IN VISUAL COMPUTING

JEN-HSUN HUANG, CO-FOUNDER & CEO | GTC 2015

FOUR ANNOUNCEMENTS

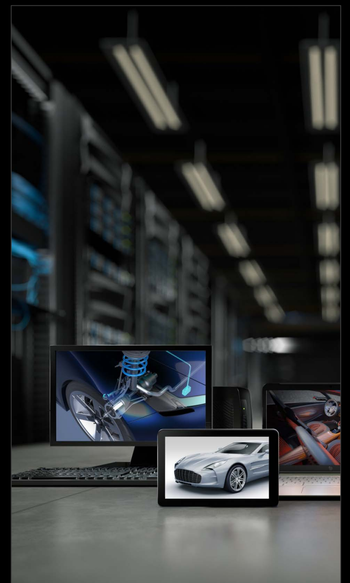
A New GPU
and
Deep Learning

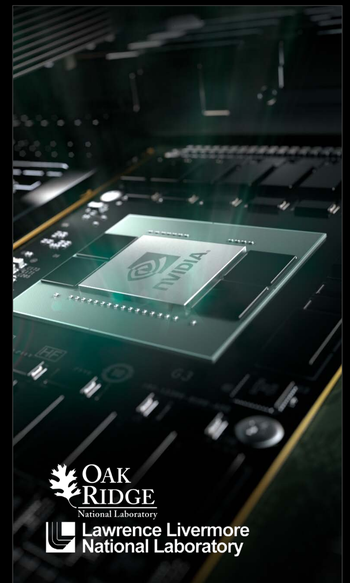
A Very Fast Box
and
Deep Learning

Roadmap Reveal
and
Deep Learning

Self-Driving Cars
and
Deep Learning

AMAZING YEAR IN VISUAL COMPUTING





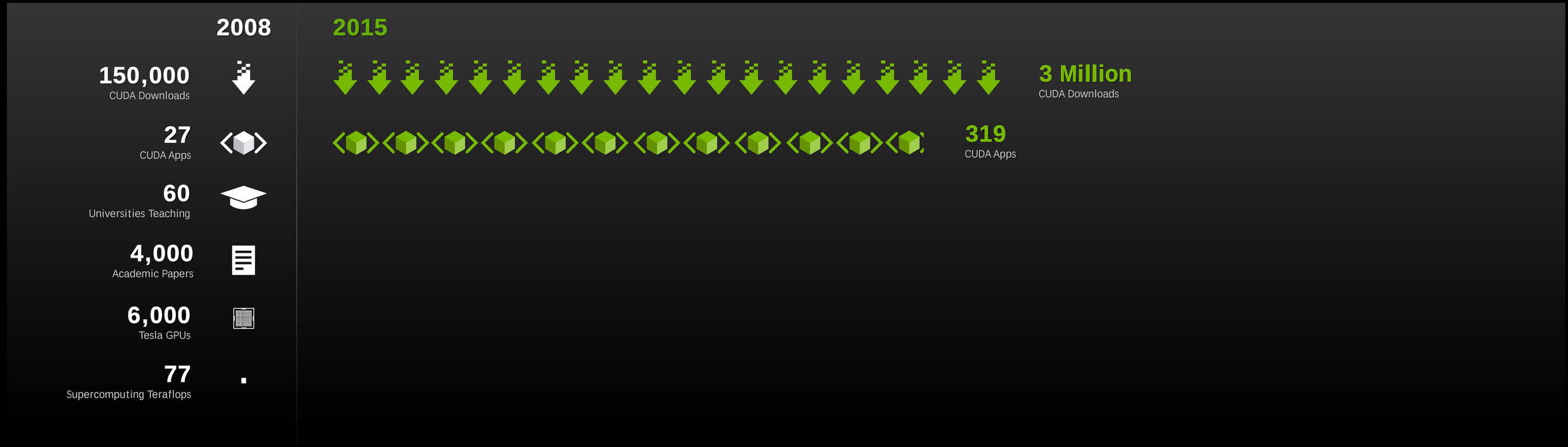
10X GROWTH IN GPU COMPUTING

	2008
150,000 CUDA Downloads	
27 CUDA Apps	
60 Universities Teaching	
4,000 Academic Papers	
6,000 Tesla GPUs	
77 Supercomputing Teraflops	.

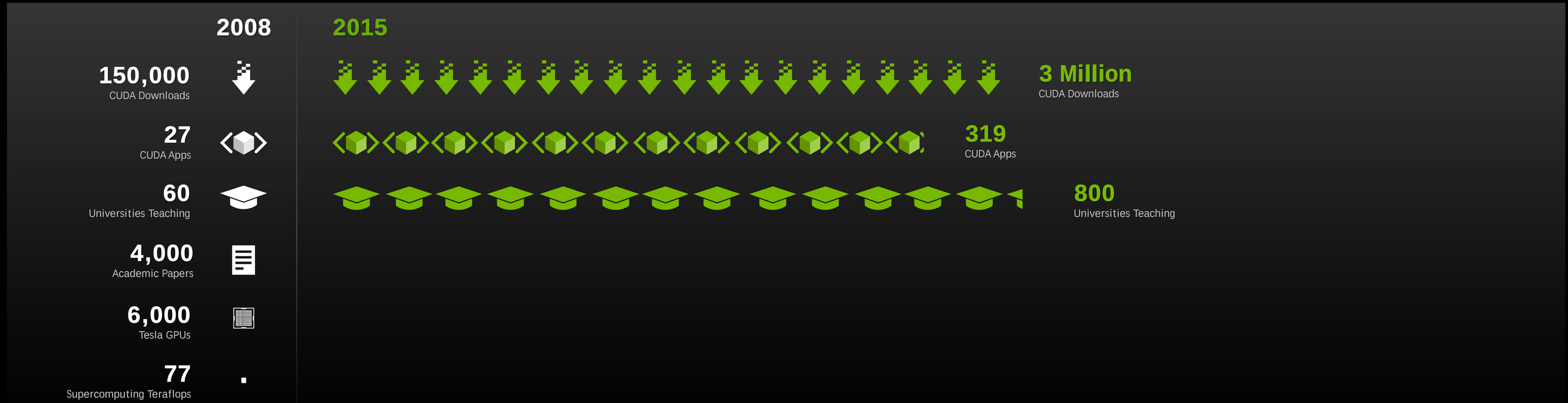
10X GROWTH IN GPU COMPUTING

	2008	2015
150,000 CUDA Downloads		3 Million CUDA Downloads
27 CUDA Apps		
60 Universities Teaching		
4,000 Academic Papers		
6,000 Tesla GPUs		
77 Supercomputing Teraflops		

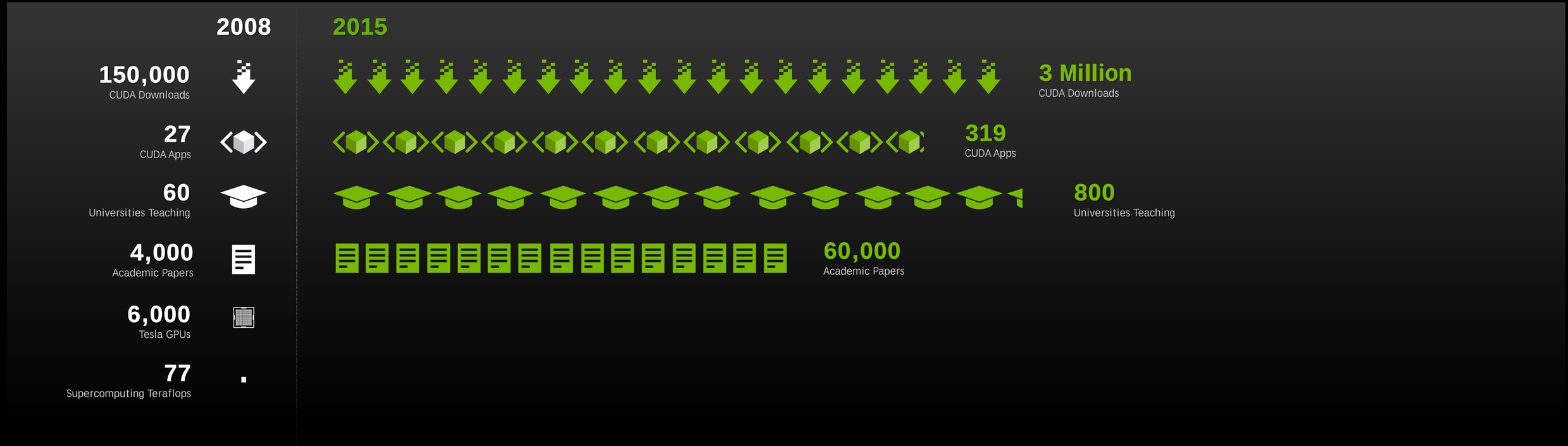
10X GROWTH IN GPU COMPUTING



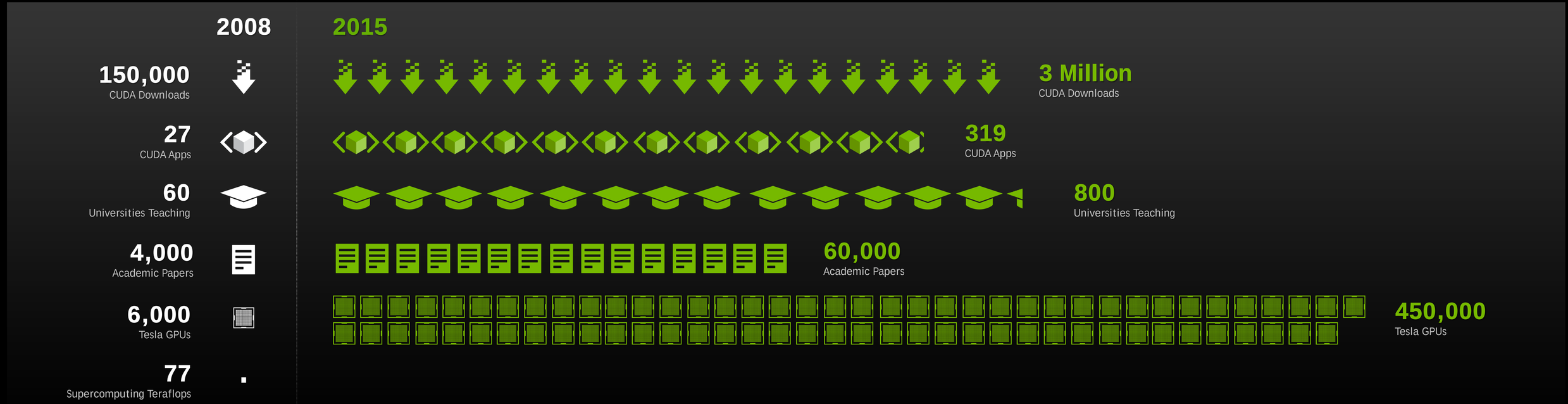
10X GROWTH IN GPU COMPUTING



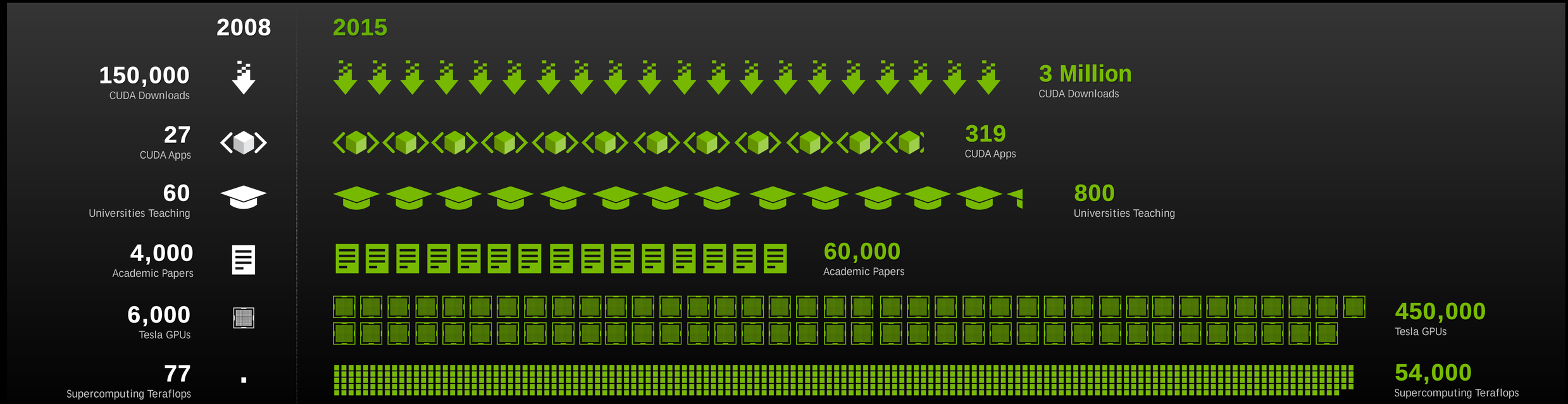
10X GROWTH IN GPU COMPUTING



10X GROWTH IN GPU COMPUTING



10X GROWTH IN GPU COMPUTING

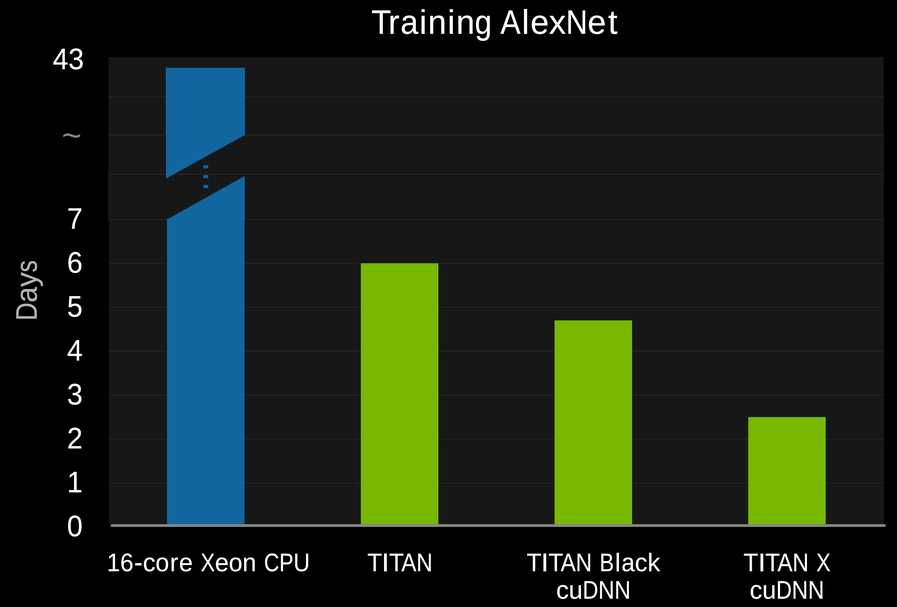


TITAN X

THE WORLD'S FASTEST GPU

8 Billion Transistors
3,072 CUDA Cores
7 TFLOPS SP / 0.2 TFLOPS DP
12GB Memory

TITAN X FOR DEEP LEARNING



TITAN X

THE WORLD'S FASTEST GPU

8 Billion Transistors
3,072 CUDA Cores
7 TFLOPS SP / 0.2 TFLOPS DP
12GB Memory

$999

FOUR ANNOUNCEMENTS

A New GPU
and
Deep Learning

A Very Fast Box
and
Deep Learning

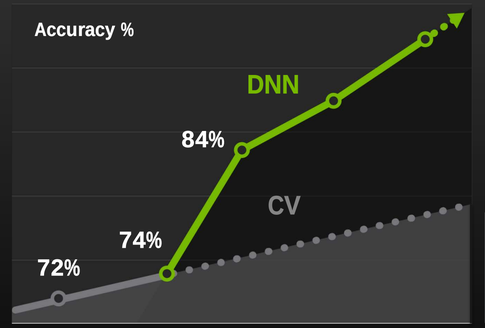
Roadmap Reveal
and
Deep Learning

Self-Driving Cars
and
Deep Learning

A SHORT HISTORY OF DEEP LEARNING

Convolutional Neural Networks for Handwritten Digital Recognition

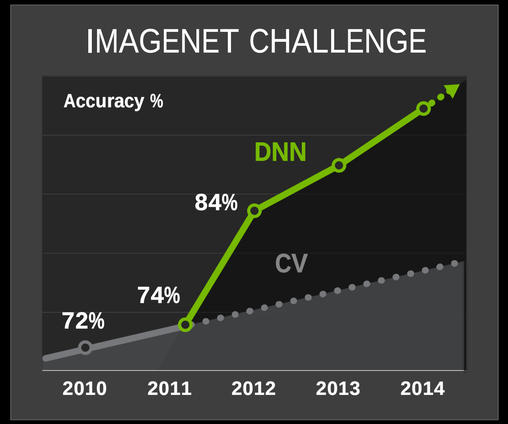
LECUN, BOTTOU, BENGIO, HAFFNER, 1998



ImageNet Classification with NVIDIA GPUs

KRIZHEVSKY, HINTON, ET AL., 2012

1995 2000 2005 2010 2015



“Deep Image: Scaling up Image Recognition”

— Baidu: 5.98%, Jan. 13, 2015

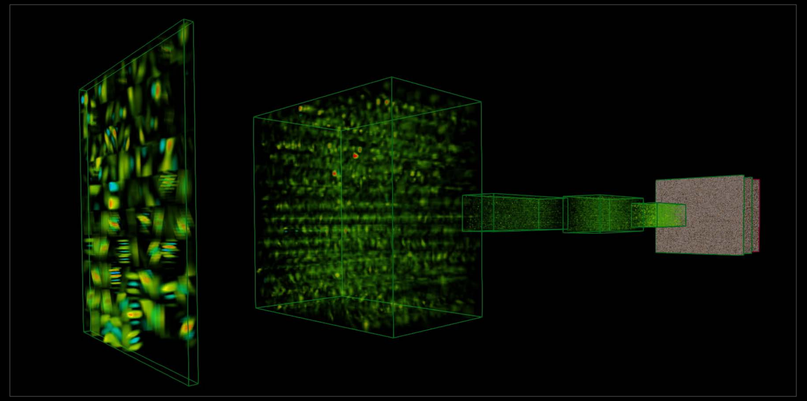
“Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”

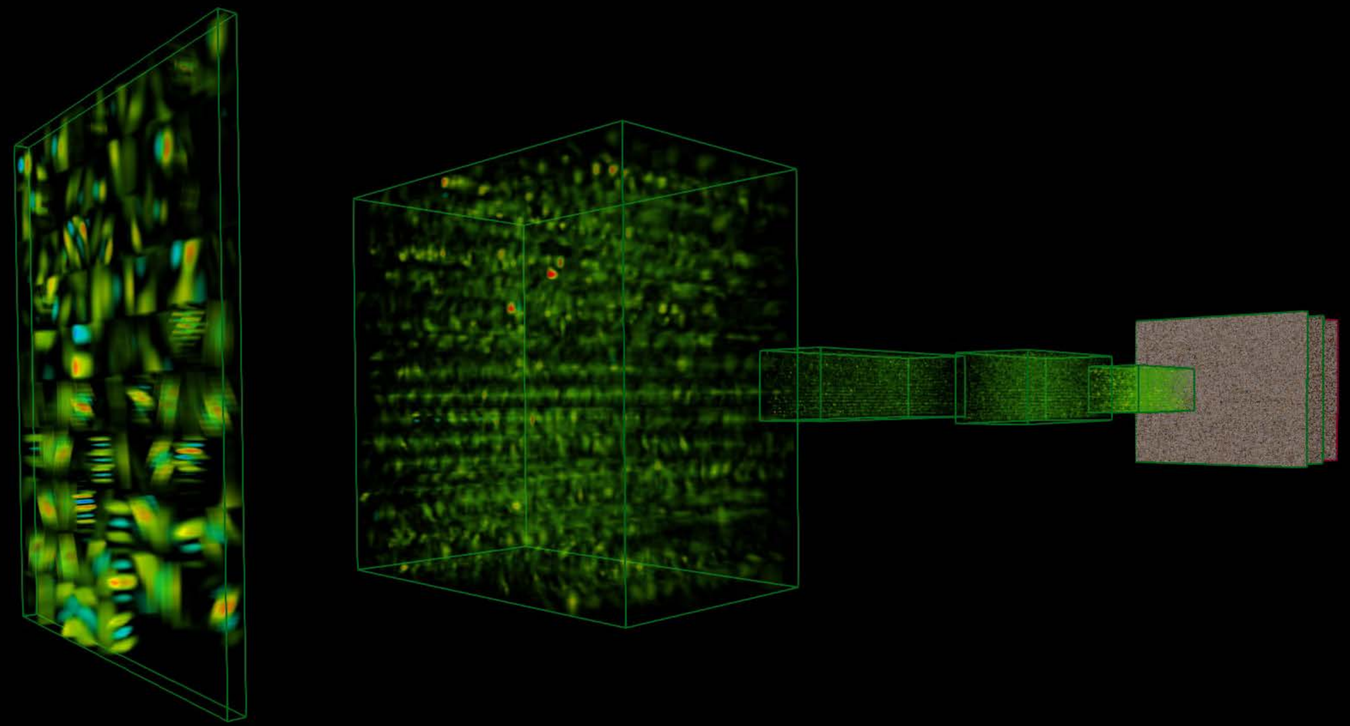
— Microsoft: 4.94%, Feb. 6, 2015

“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariant Shift”

— Google: 4.82%, Feb. 11, 2015

THE BIG BANG





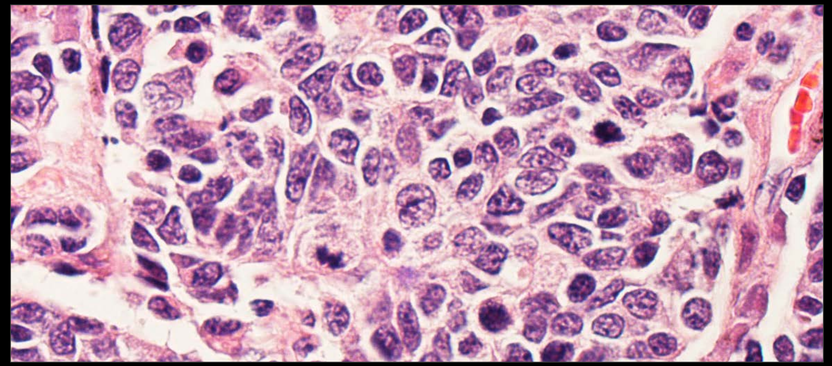
DEEP LEARNING VISUALIZED

GPU-ACCELERATED DEEP LEARNING

DEEP LEARNING REVOLUTIONIZING MEDICAL RESEARCH

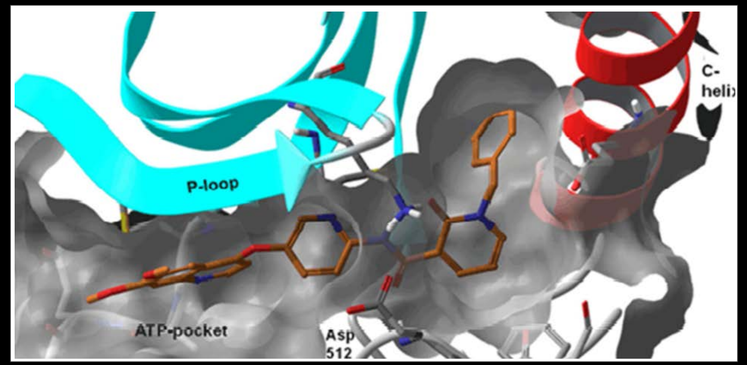
Detecting Mitosis in Breast Cancer Cells

— IDSIA



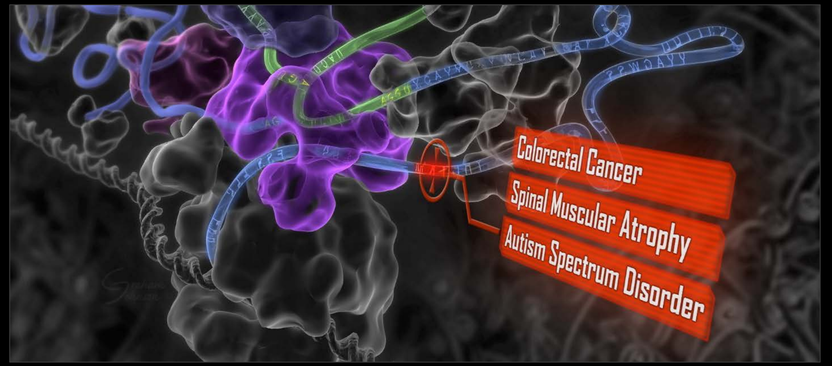
Predicting the Toxicity of New Drugs

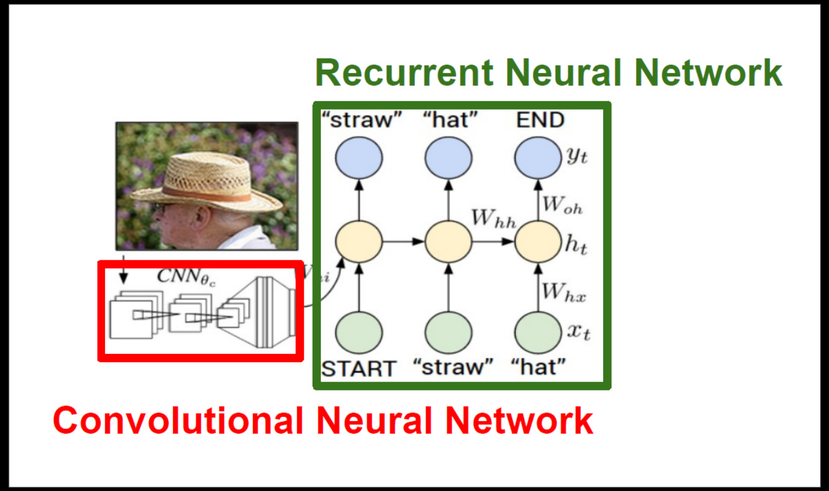
— Johannes Kepler University



Understanding Gene Mutation to Prevent Disease

— University of Toronto

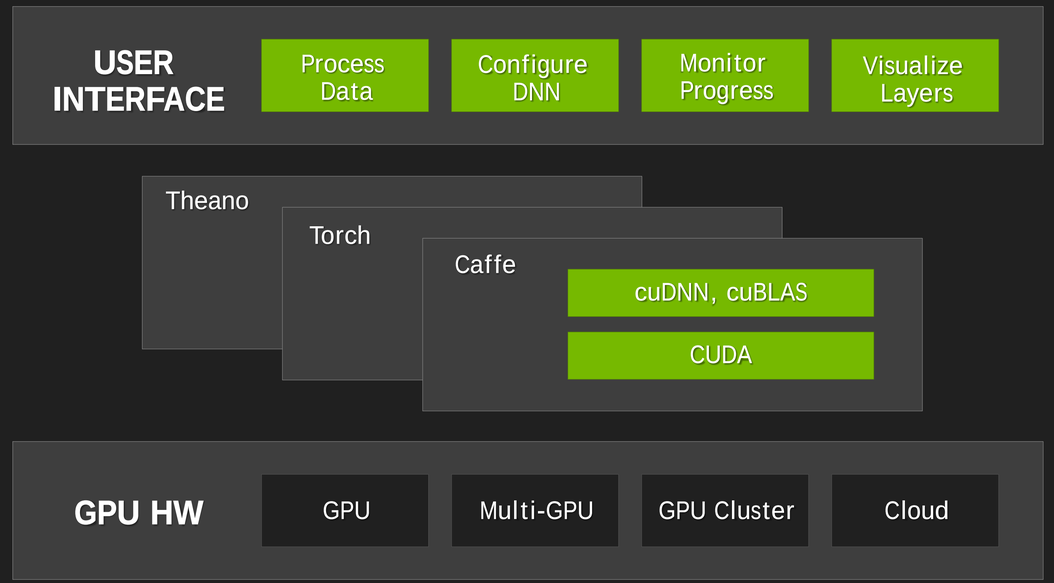




a bird perched on a branch of a tree

"Automated Image Captioning with ConvNets and Recurrent Nets"

—Andrej Karpathy, Fei-Fei Li



DIGITS

DEEP GPU TRAINING SYSTEM FOR DATA SCIENTISTS

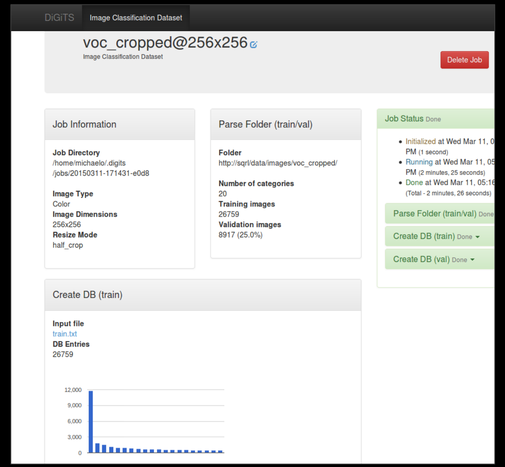
Design DNNs

Visualize activations

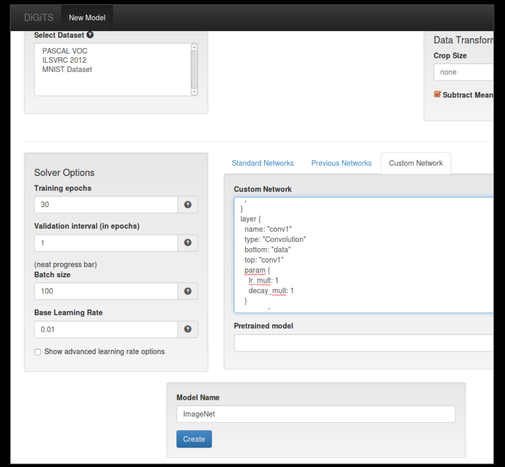
Manage multiple trainings

DIGITS

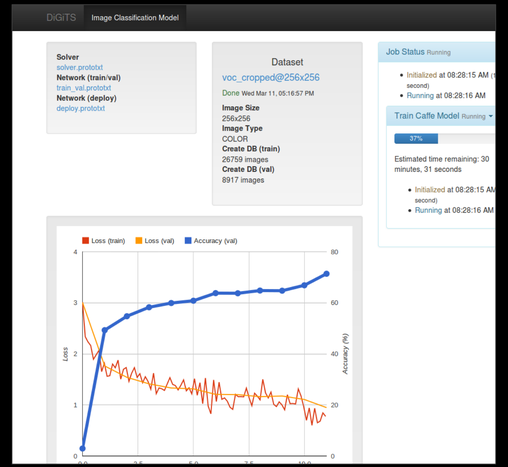
Process Data



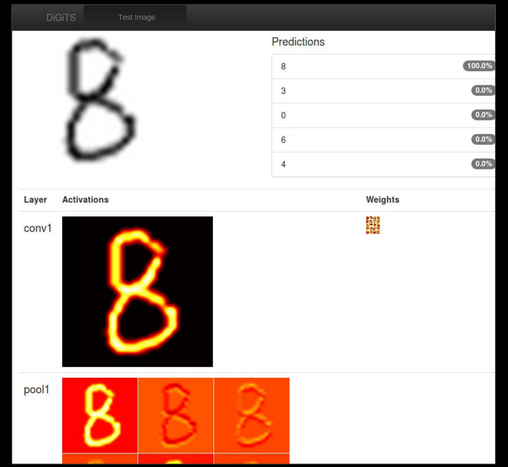
Configure DNN

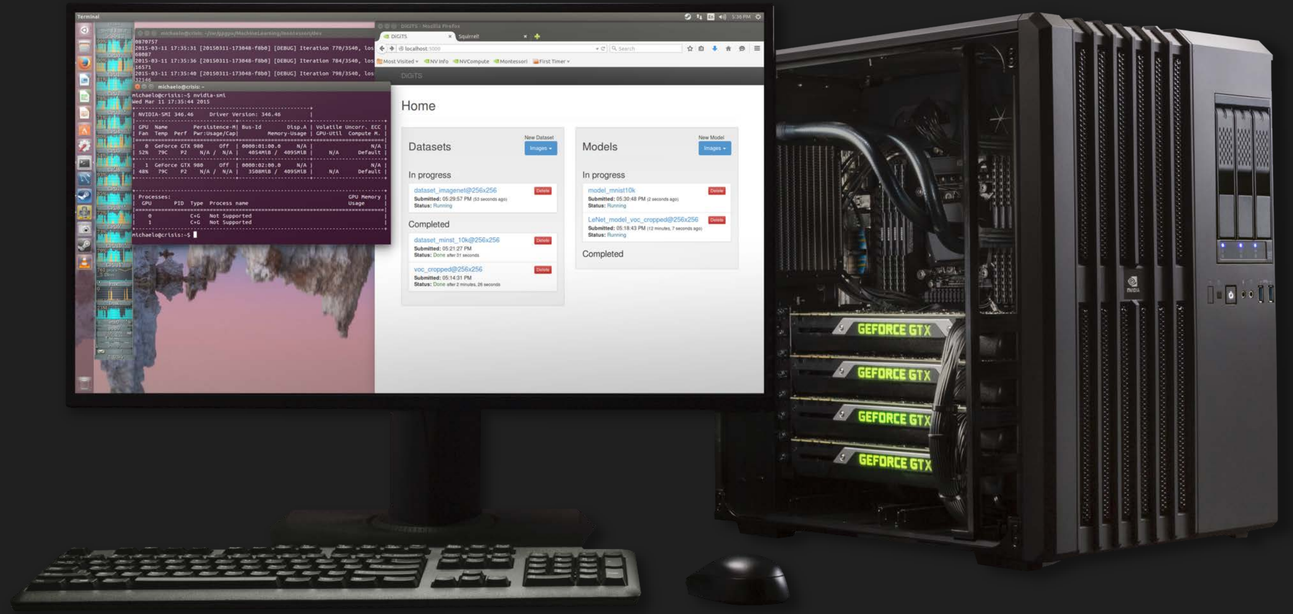


Monitor Progress



Visualize Layers





DIGITS DEVBOX

World’s fastest GPU

Max GPU out of a plug

Multi-GPU training & inference

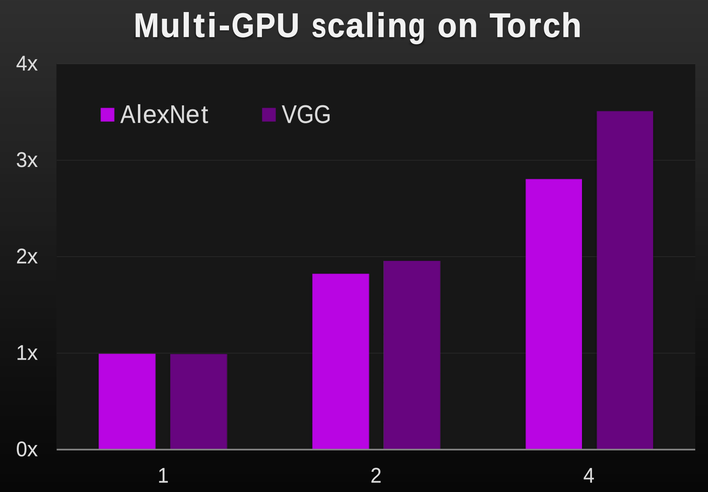
DIGITS DEVBOX — EARLY RESULTS

"DIGITS makes it way easier to design the best network for the job"

— Simon Osindero

A.I. Architech

flickr

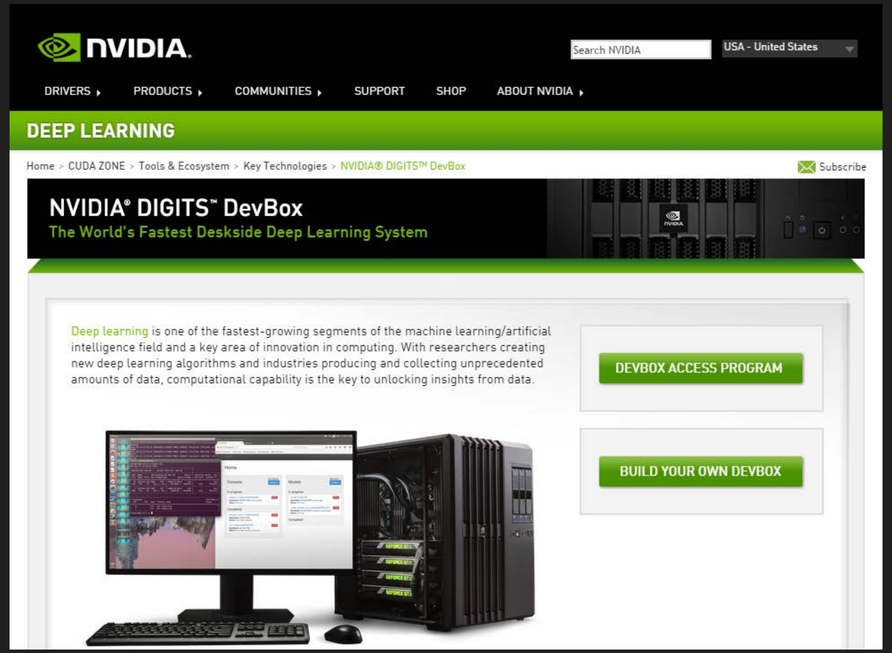


"I've never seen AlexNet run this fast…TitanX is a monster, Crazy Fast"

— Soumith Chintala

Research Engineer

facebook



DIGITS DEVBOX

Available May 2015
$15,000

FOUR ANNOUNCEMENTS

A New GPU
and
Deep Learning

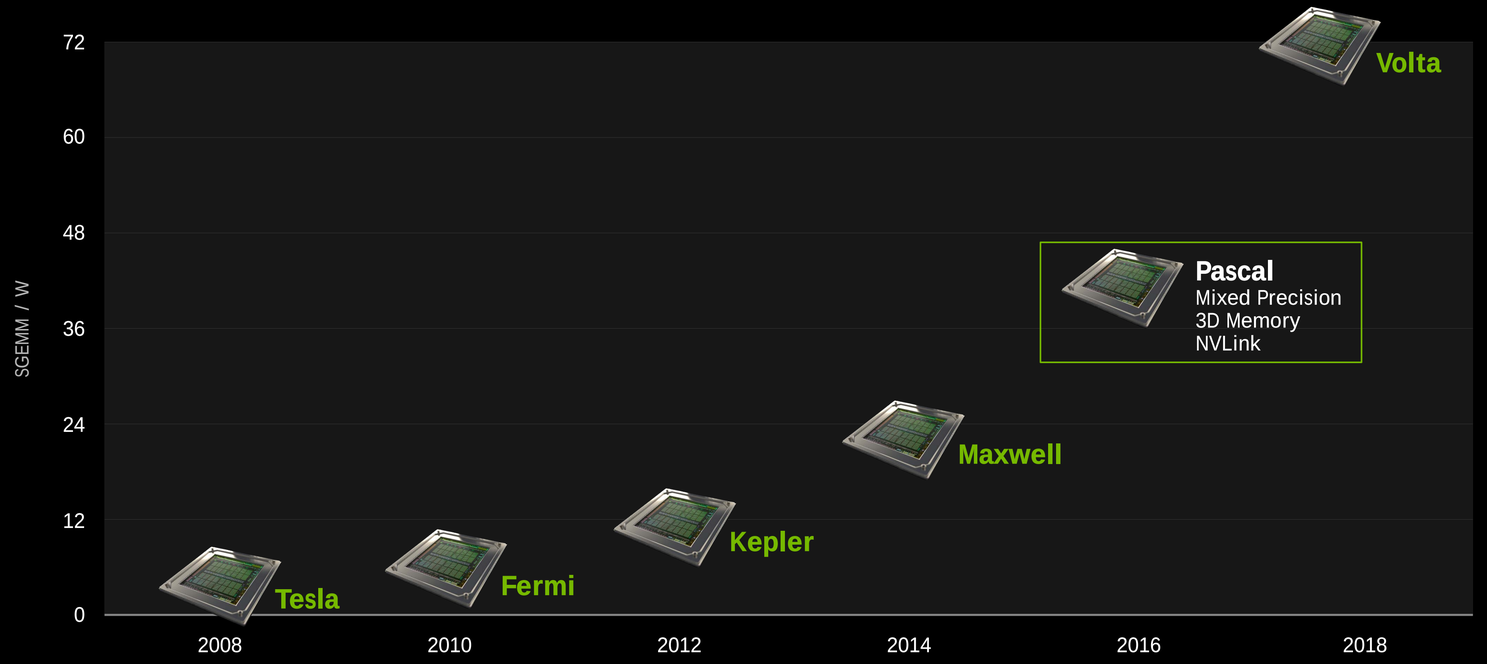
A Very Fast Box
and
Deep Learning

Roadmap Reveal
and
Deep Learning

Self-Driving Cars
and
Deep Learning

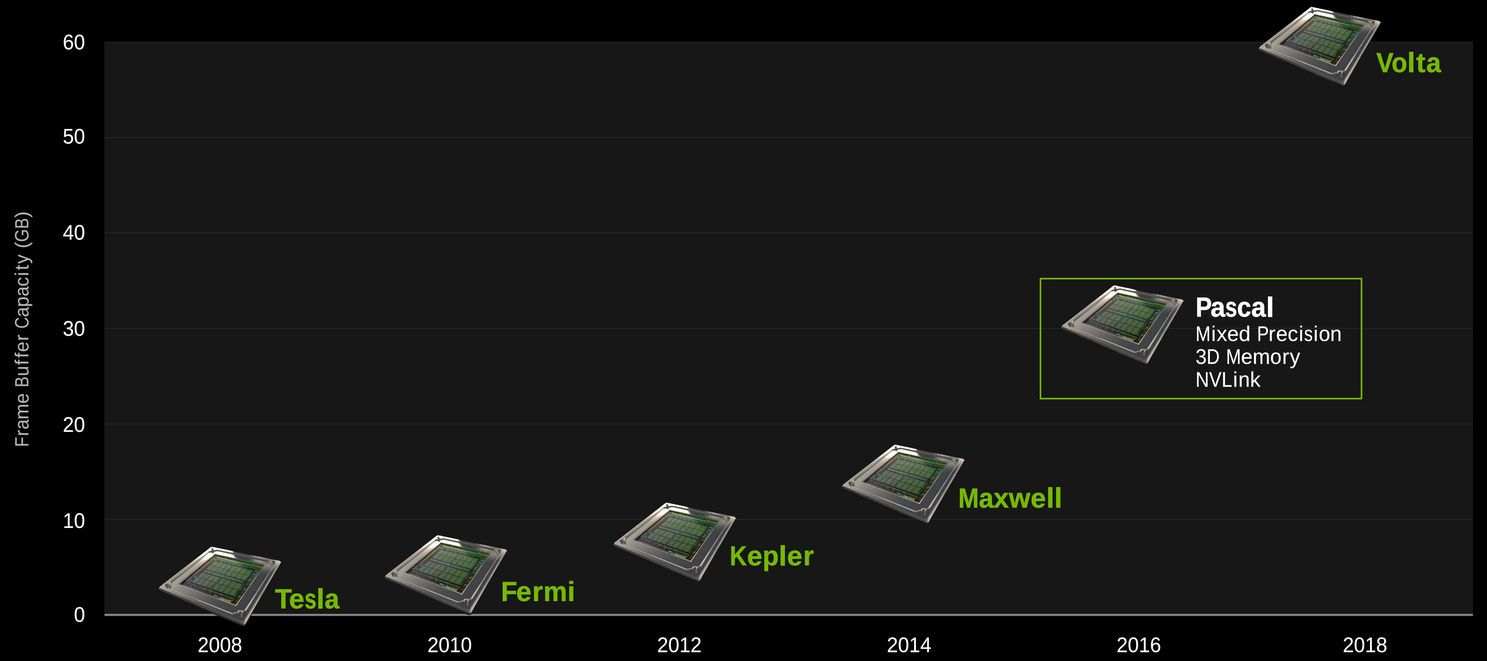
GPU ROADMAP

Pascal 2x SGEMM/W



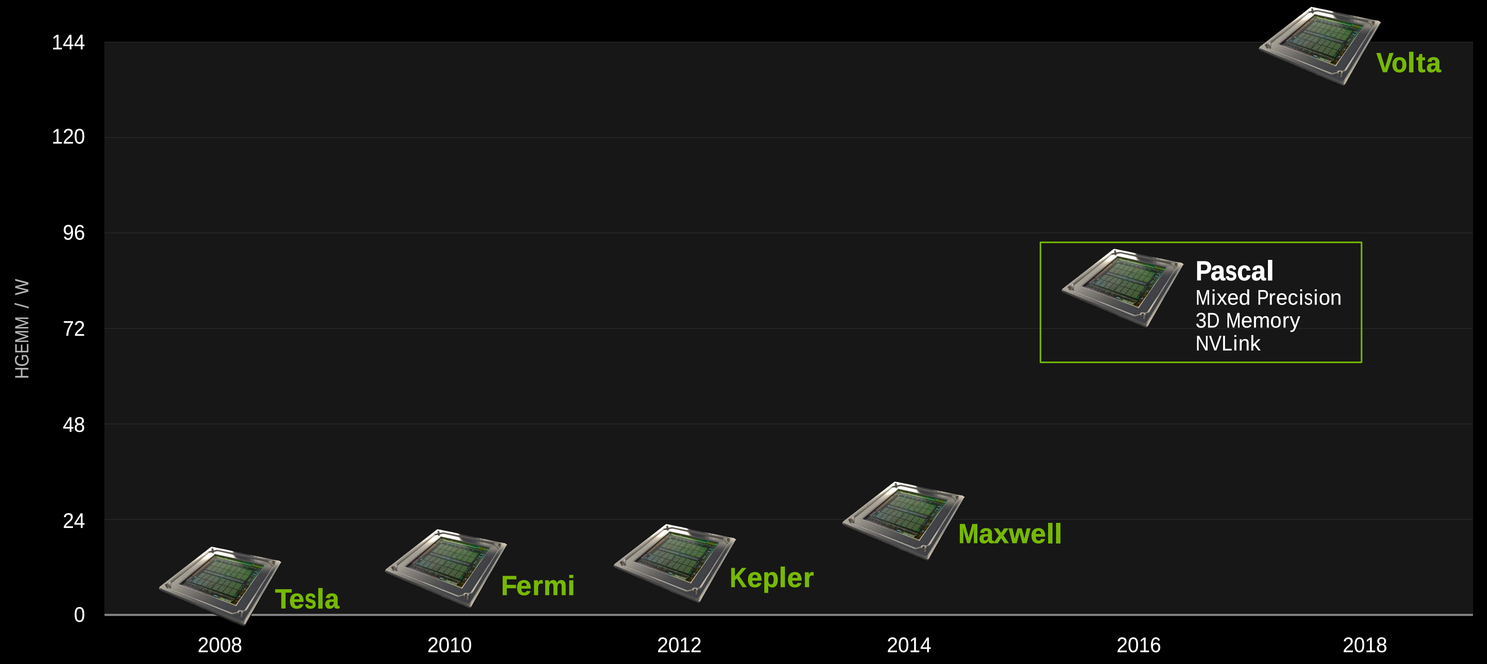
GPU ROADMAP

Pascal 2.7x Memory Capacity



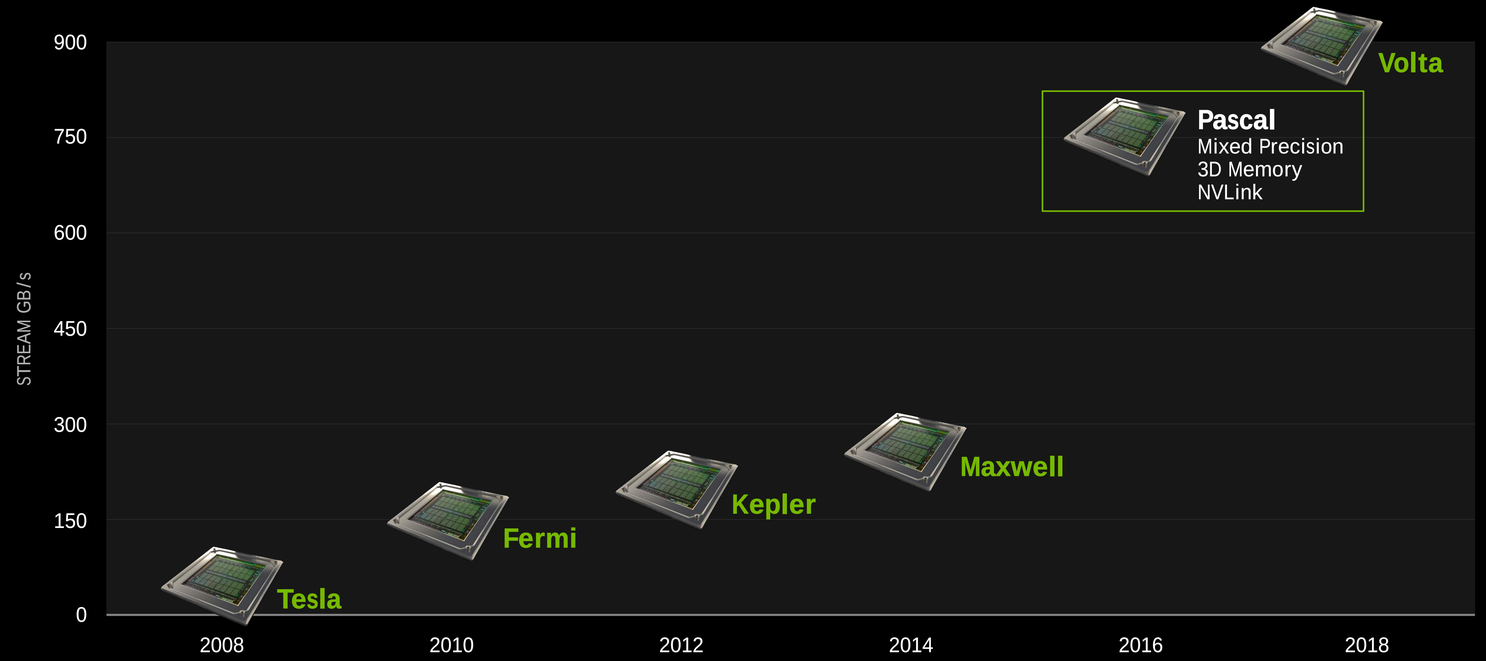
GPU ROADMAP

Pascal 4x Mixed Precision

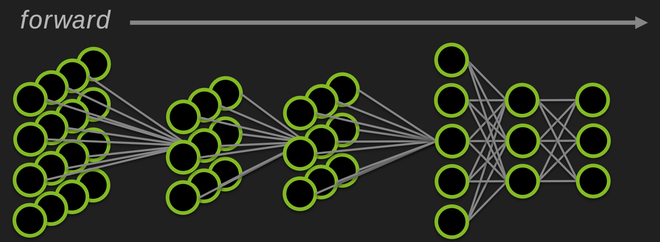


GPU ROADMAP

Pascal 3x Bandwidth



PASCAL 10X MAXWELL



CONVOLUTION
(compute)

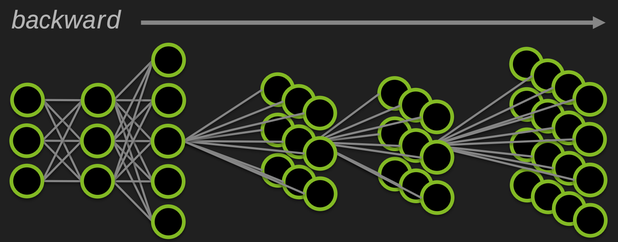
4x (FP16)

Mixed Precision

FULLY CONNECTED
(bandwidth)

6x

3D Memory



FULLY CONNECTED
(bandwidth)

6x

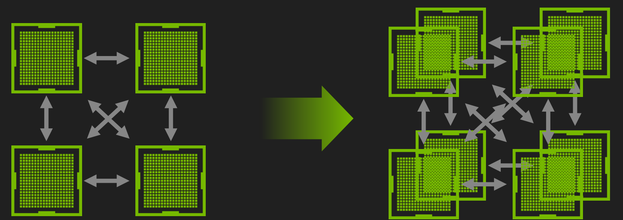
3D Memory

CONVOLUTION
(compute)

4x

Mixed Precision

5x



WEIGHT UPDATE
(interconnect)

10x

NVLINK

2x

** Very rough estimates*

FOUR ANNOUNCEMENTS

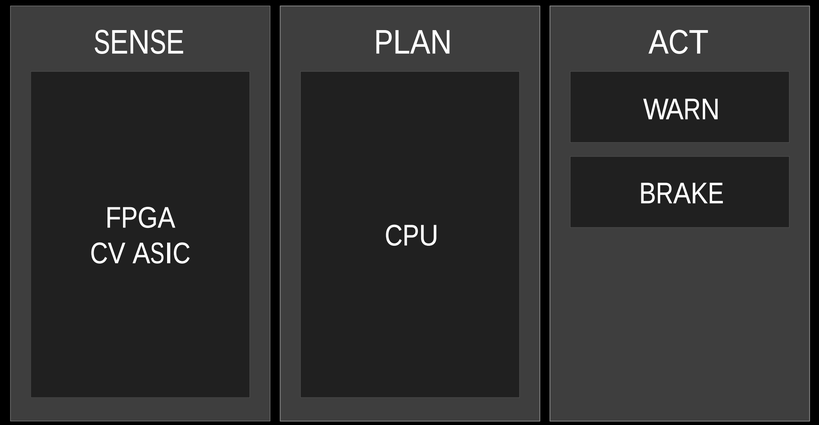
A New GPU
and
Deep Learning

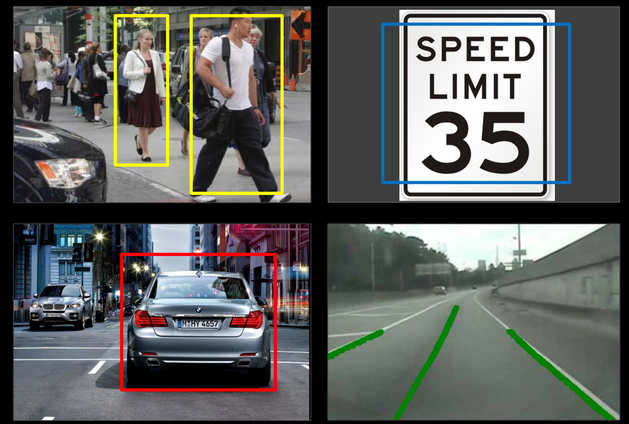
A Very Fast Box
and
Deep Learning

Roadmap Reveal
and
Deep Learning

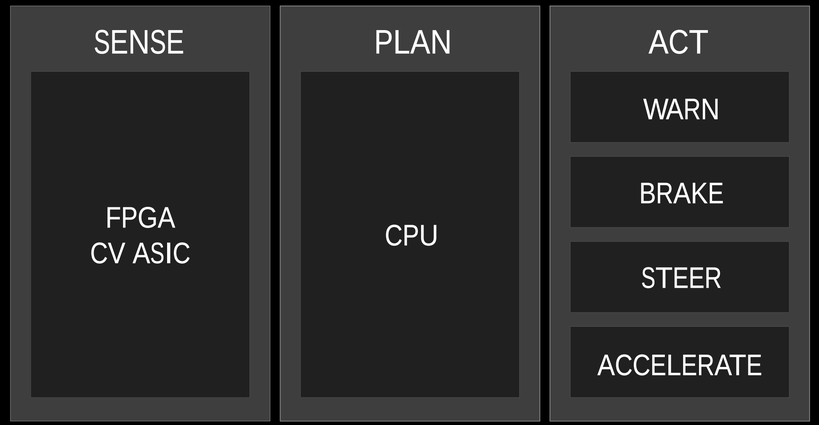
Self-Driving Cars
and
Deep Learning

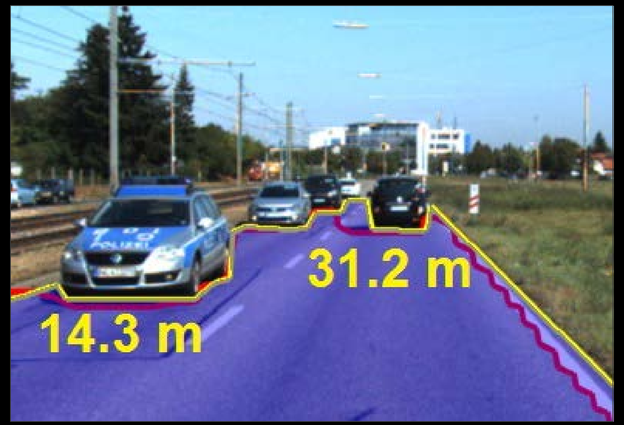
TODAY’S ADAS



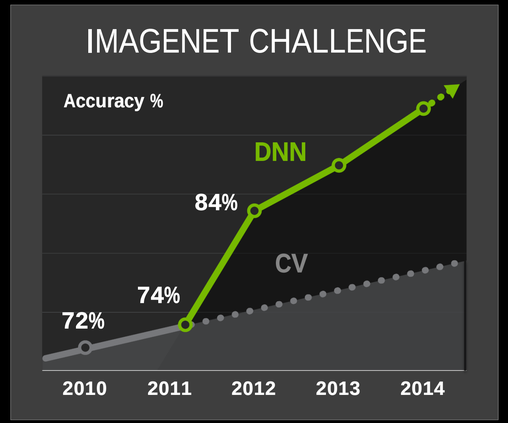


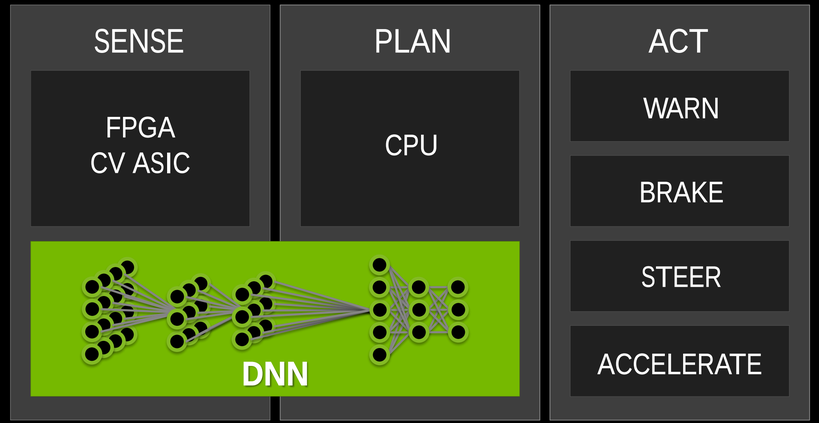
NEXT-GENERATION ADAS



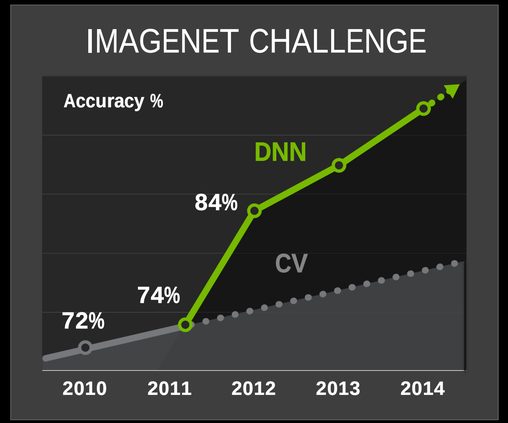


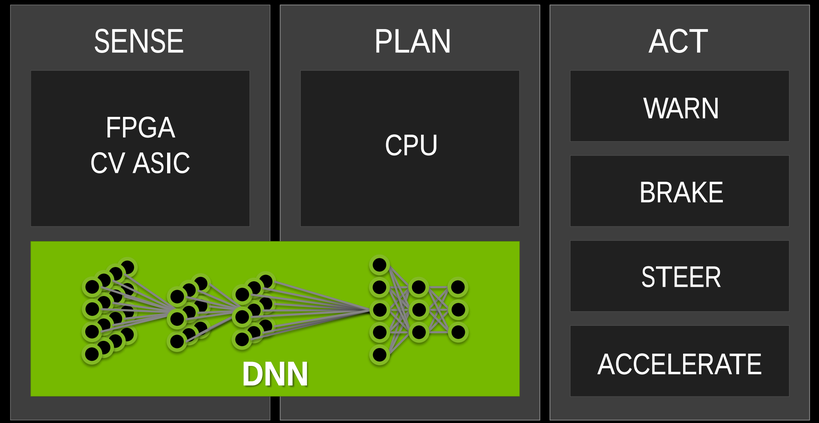
NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



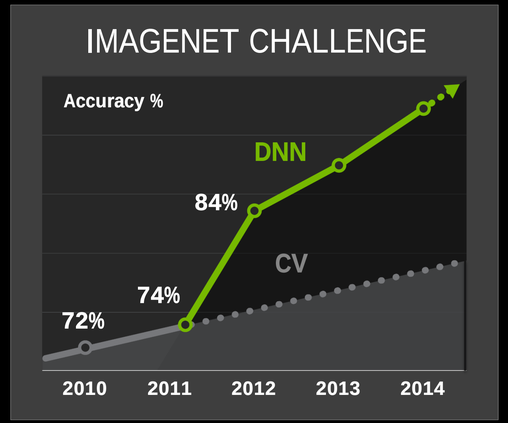


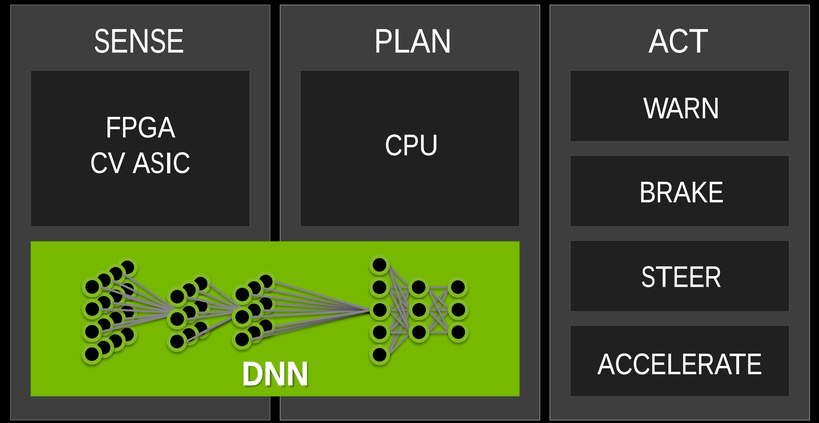
NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



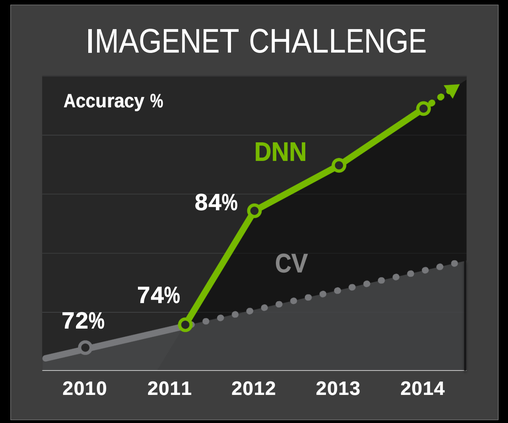


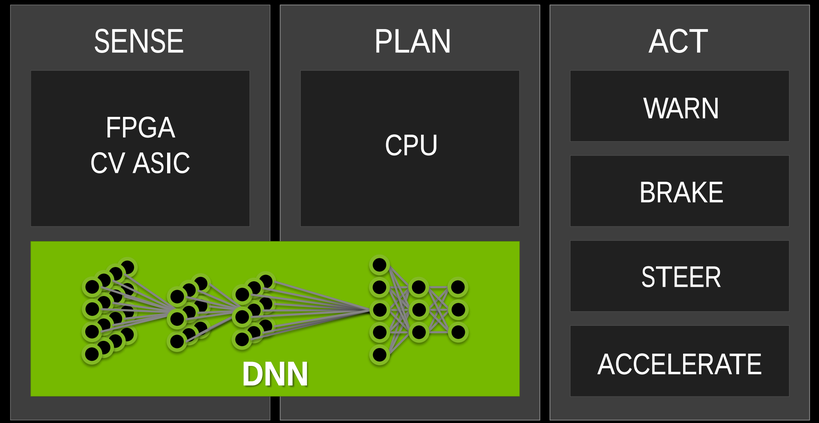
NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



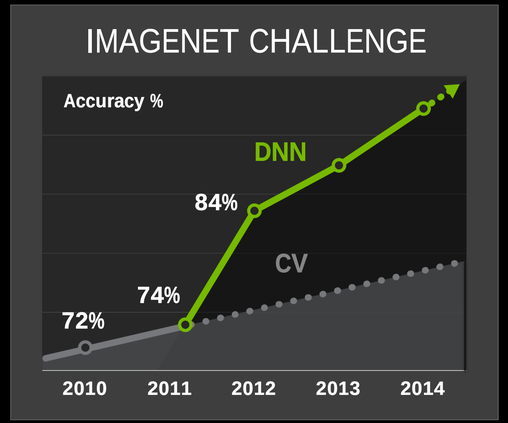


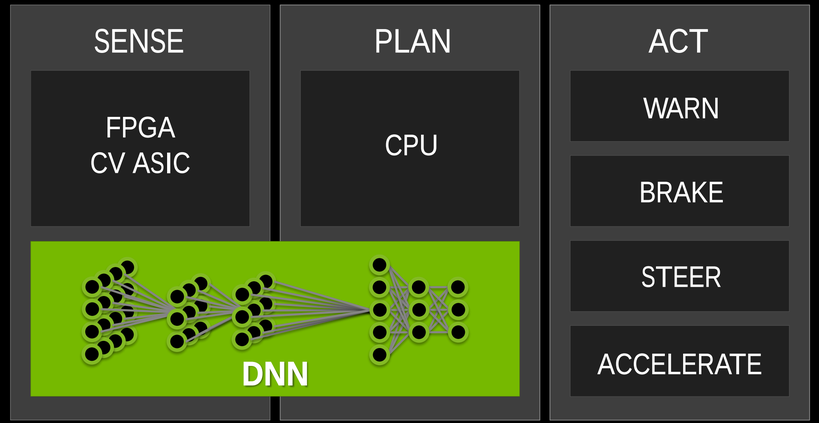
NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



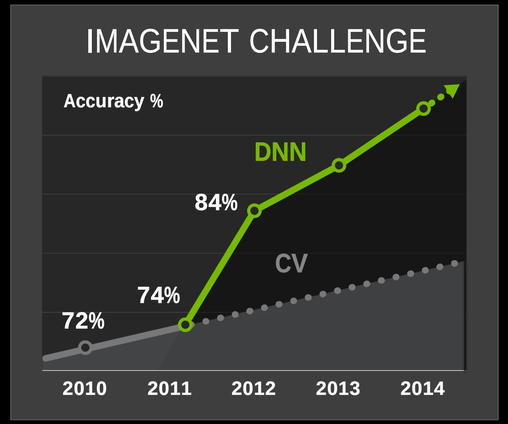


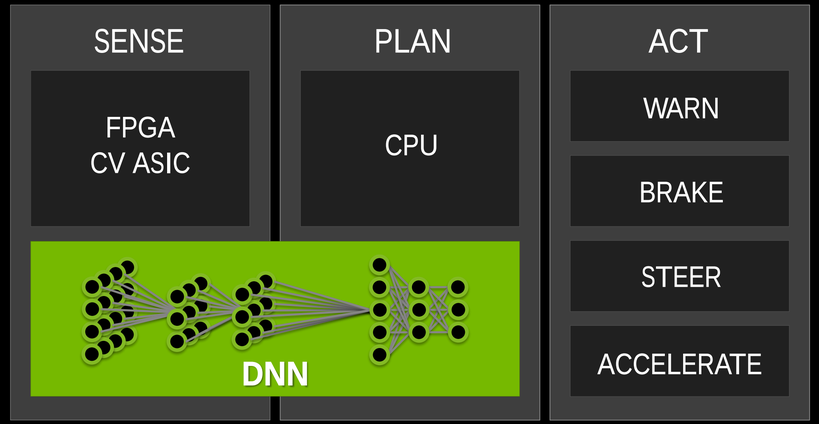
NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



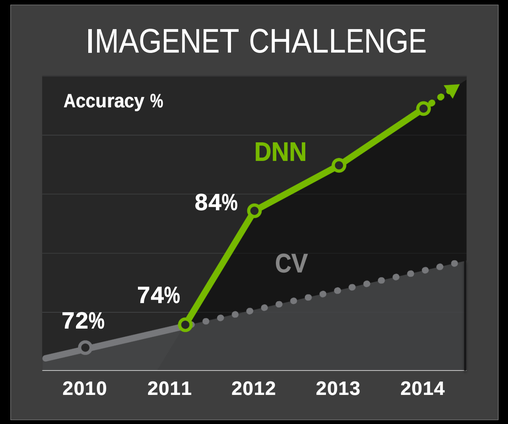


NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER





PROJECT DAVE — DARPA AUTONOMOUS VEHICLE



DNN-based self-driving robot

Training data by human driver

No hand-coded CV algorithms

PROJECT LEADS

Urs Muller: Chief Architect,
Autonomous Driving, NVIDIA

Yann LeCun: Director,
AI Research, Facebook

DAVE IN ACTION

TRAINING DATA

225K Images

TEST DRIVE

No Training

TEST DRIVE

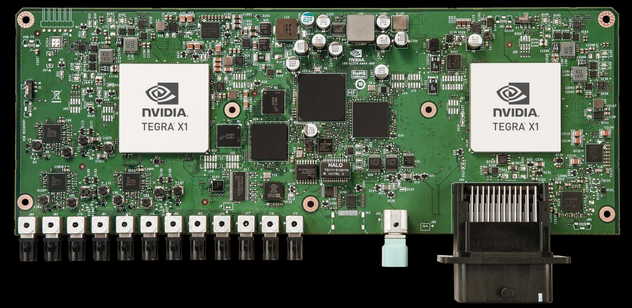
Partially Trained (52K images)

TEST DRIVE

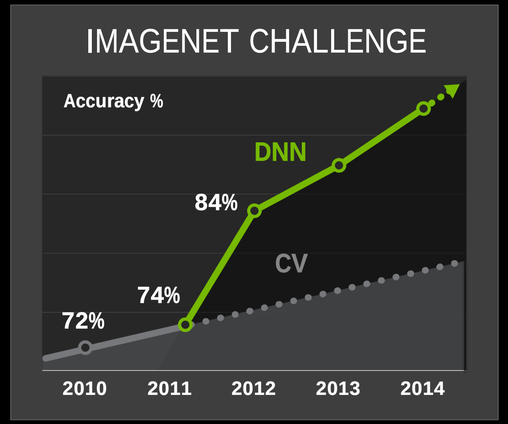
Fully Trained (225K images)

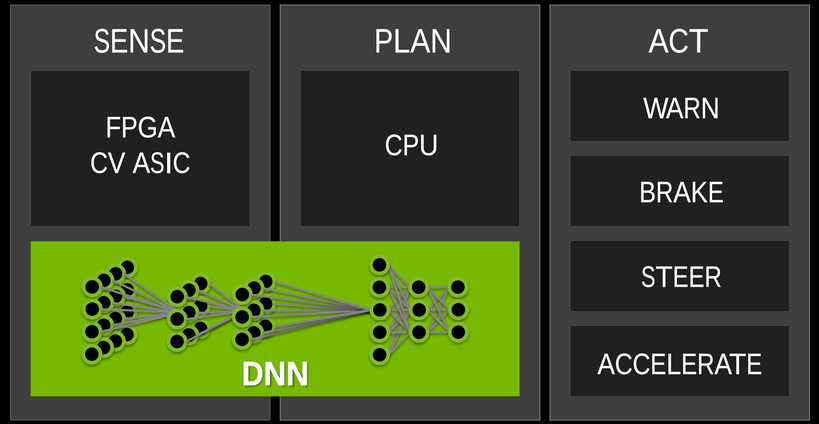
	DAVE	AlexNet on DRIVE PX
Number of Connections	3.1 Million	630 Million
Frames / Second	12	184
Connections / Second	38 Million	116 Billion

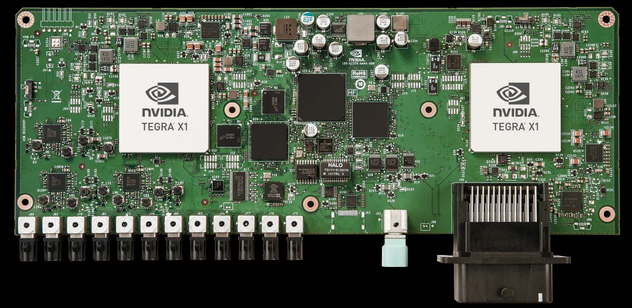
3,000x Faster

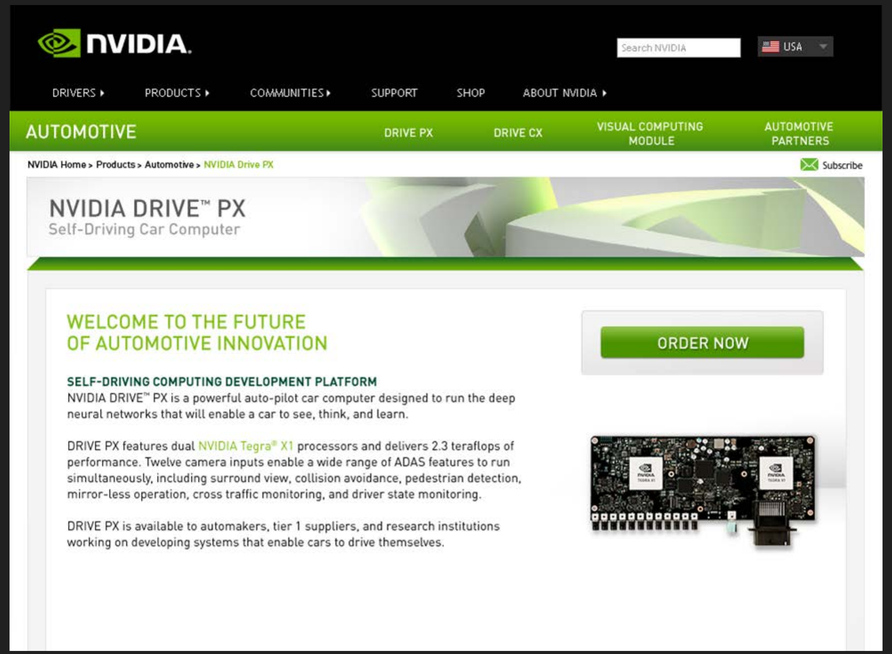


NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER









NVIDIA DRIVE™ PX

SELF-DRIVING CAR COMPUTER

Available May 2015
$10,000

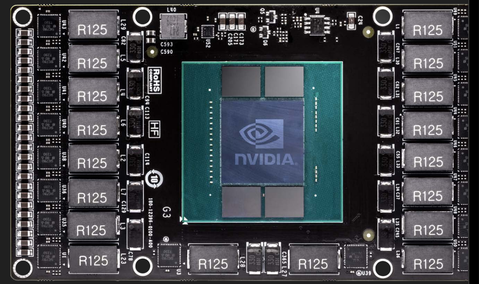
ELON MUSK

LEAPS IN VISUAL COMPUTING

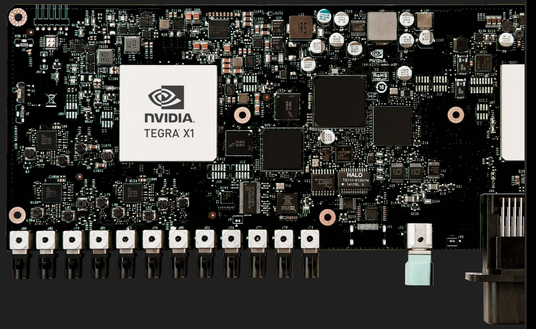
TITAN X
The World's Fastest GPU

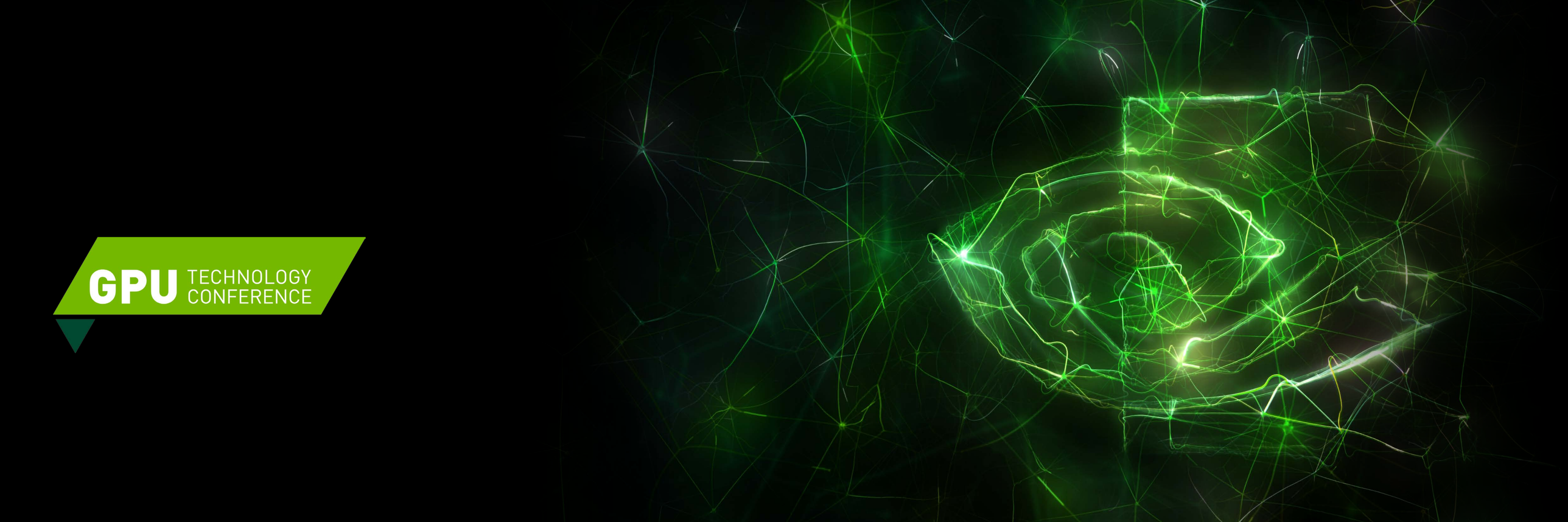
DIGITS DevBox
GPU Deep Learning Platform

Pascal — 10x Maxwell
For Deep Learning



NVIDIA DRIVE PX
Deep Learning Platform for Self-Driving Cars




GPU TECHNOLOGY CONFERENCE