Building Autonomous Machines with JETSON
Machine Learning, Vision, and GPUs

Francois Courteille
Senior Solution Architect, Accelerated Computing
# JETSON™ EMBEDDED PLATFORM

Deep learning for next-generation intelligent, autonomous machines

## AGENDA

<table>
<thead>
<tr>
<th>Background on NVIDIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Learning &amp; GPUs</td>
</tr>
<tr>
<td>NVIDIA Embedded Platform</td>
</tr>
<tr>
<td>Deep Learning &amp; Autonomy</td>
</tr>
</tbody>
</table>
THE WORLD LEADER IN VISUAL COMPUTING
CREDENTIALS BUILT OVER TIME

Majority of HPC Applications are GPU-Accelerated, 410 and Growing

300K CUDA Developers, 4x Growth in 4 years

100% of Deep Learning Frameworks are Accelerated
END-TO-END TESLA PRODUCT FAMILY

**HYPERSONE HPC**

*Tesla M4, M40*

Hyperscale deployment for DL training, inference, video & image processing

---

**MIXED-APPS HPC**

*Tesla K80*

HPC data centers running mix of CPU and GPU workloads

---

**STRONG-SCALING HPC**

*Tesla P100*

Hyperscale & HPC data centers running apps that scale to multiple GPUs

---

**FULLY INTEGRATED DL SUPERCOMPUTER**

*DGX-1*

For customers who need to get going now with fully integrated solution
“Google’s AI engine also reflects how the world of computer hardware is changing. (It) depends on machines equipped with GPUs... And it depends on these chips more than the larger tech universe realizes.”

WIRED
BLISTERING PACE OF INNOVATION FOR AI

Deep Learning Training Performance
*Caffe AlexNet*

- **K40**
- **K80 + cuDNN1**
- **M40 + cuDNN4**
- **P100 + cuDNN5**

**AlexNet training throughput based on 20 iterations,**

**CPU:** 1x E5-2680v3 12 Core 2.5GHz, 128GB System Memory, Ubuntu 14.04

**M40 bar:** 8x M40 GPUs in a node

**P100:** 8x P100 NVLink-enabled
HOW FAR HAS AUTONOMY COME?

ImageNet classification accuracy

<table>
<thead>
<tr>
<th>Year</th>
<th>ImageNet Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>72%</td>
</tr>
<tr>
<td>2011</td>
<td>74%</td>
</tr>
<tr>
<td>2012</td>
<td>84%</td>
</tr>
<tr>
<td>2013</td>
<td>88%</td>
</tr>
<tr>
<td>2014</td>
<td>93%</td>
</tr>
<tr>
<td>2015</td>
<td>95.1%</td>
</tr>
</tbody>
</table>

Human: 94.9%

Deep Learning on GPUs
DEEP LEARNING FOR AUTONOMOUS VEHICLES
## Jetson TX1

<table>
<thead>
<tr>
<th>Specification</th>
<th>Jetson TX1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GPU</strong></td>
<td>1024 GFLOPS 256-core Maxwell</td>
</tr>
<tr>
<td><strong>CPU</strong></td>
<td>4x 64-bit ARM A57 CPUs</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>4 GB LPDDR4</td>
</tr>
<tr>
<td><strong>Video decode</strong></td>
<td>4K 60Hz H.264 / H.265</td>
</tr>
<tr>
<td><strong>Video encode</strong></td>
<td>4K 30Hz H.264 / H.265</td>
</tr>
<tr>
<td><strong>CSI</strong></td>
<td>Up to 6 cameras</td>
</tr>
<tr>
<td><strong>Display</strong></td>
<td>2x DSI, 1x eDP 1.4, 1x DP 1.2/HDMI</td>
</tr>
<tr>
<td><strong>Wi-Fi</strong></td>
<td>802.11 2x2 ac</td>
</tr>
<tr>
<td><strong>Networking</strong></td>
<td>1 Gigabit Ethernet</td>
</tr>
<tr>
<td><strong>PCI-E</strong></td>
<td>Gen 2 1x1 + 1x4</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td>16 GB eMMC, SDIO, SATA</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>3x UART, 3x SPI, 4x I2C, 4x I2S, GPIOs</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>10-15W, 6.6V-19.5VDC</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>50mm x 87mm</td>
</tr>
</tbody>
</table>
JETSON AT THE LEADING EDGE

Powering the Next Generation of Autonomous Machines
Jetson TX1 Developer Kit

Jetson TX1 Module
Developer Board
5MP Camera
€630 / £450 retail
€350 / £250 academic discount
Comprehensive developer platform

Jetson SDK
Libraries
Developer tools
Design collateral
Developer Forum
Training and Tutorials
Ecosystem

JETSON SDK: THE DETAILS

Vertically Integrated Packages

- **Machine Learning**
  - cuDNN
  - CuPy
  - Caffe
  - Theano
  - Torch

- **Compute**
  - cuFFT
  - CUDA Math Library
  - cuBLAS
  - NPP
  - cuSPARSE
  - cuRAND
  - Thrust
  - cuSolver

- **Vision**
  - NVIDIA VisionWorks
  - OpenVX
  - OpenCV

- **Graphics**
  - Vulkan
  - OpenGL
  - OpenGL ES

- **Linux for Tegra**
  - Ubuntu
  - Fedora
  - gstreamer
  - ALSA
  - PulseAudio
  - Qt
  - V4L2
  - libjpeg

**Tools**
- Source code editor
- Debugger
- Profiler
- System Trace
- NVTX (NVIDIA Tools eXtension)

Jetson TX1
TRAIN, THEN DEPLOY YOUR NEURAL NETWORKS

Training: Server, GRID, DIGITS

Inference: GIE, Embedded (Jetson)

Solver
Network
Data Scientist

Camera Inputs
Trained Deep Neural Net Model
CUDNN
CUDA-accelerated Deep Learning library

Your Application
(dog breed detector, for example)

Supports industry-standard frameworks:

Out-of-the-box speedups of neural networks:

For both Inference and Training:

DIGITS
Frameworks

Caffe
torch
theano

cuDNN

CUDA

Jetson TX1 // Tesla/Titan/GRID
NVIDIA DIGITS
Interactive Deep Learning GPU Training System
OBJECT DETECTION
New in DIGITS 4

REMOTE SENSING

INTELLIGENT VIDEO ANALYTICS

ADVANCED DRIVER ASSISTANCE SYSTEMS (ADAS)

MEDICAL DIAGNOSTICS

developer.nvidia.com/digits
GPU INFERENCE ENGINE (GIE)

High-performance deep learning inference for production deployment

**DATA CENTER**
- Tesla M4

**AUTOMOTIVE**
- Drive PX

**EMBEDDED**
- Jetson TX1

40x more inference perf/watt with GIE

GoogLenet Images Per Second Per Watt
CPU-Only vs Tesla M4 + GIE on
dual-socket Haswell E5-2698 v3@2.3GHz HT on

developer.nvidia.com/gie
END-TO-END DEEP LEARNING FOR SELF-CONTROL

- Sensory Inputs
- Inference
- Recognition
- Perceptron
- RNN
- Motor PWM

Goal/reward function

Short-term nav

Long-term nav

user application

Motor PWM
ONE ARCHITECTURE — END-TO-END AI

CUDA + Linux throughout the stack.
NVIDIA DGX-1
WORLD’S FIRST DEEP LEARNING SUPERCOMPUTER

170 TFLOPS FP16
8x Tesla P100 16GB
NVLink Hybrid Cube Mesh
Accelerates Major AI Frameworks
Dual Xeon
7 TB SSD Deep Learning Cache
Dual 10GbE, Quad IB 100Gb
3RU - 3200W
THANK YOU!

Q&A: WHAT CAN I HELP YOU BUILD?