DEEP LEARNING DEMYSTIFIED

Will Ramey
Director, Developer Programs
NVIDIA Corporation
DEFINITIONS

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.

MACHINE LEARNING
Machine learning begins to flourish.

DEEP LEARNING
Deep learning breakthroughs drive AI boom.
DEEP LEARNING IS SWEEPING ACROSS INDUSTRIES

- Internet Services:
  - Image/Video classification
  - Speech recognition
  - Natural language processing

- Medicine:
  - Cancer cell detection
  - Diabetic grading
  - Drug discovery

- Media & Entertainment:
  - Video captioning
  - Content based search
  - Real time translation

- Security & Defense:
  - Face recognition
  - Video surveillance
  - Cyber security

- Autonomous Machines:
  - Pedestrian detection
  - Lane tracking
  - Recognize traffic signs
DEEP LEARNING IS TRANSFORMING HPC

“Seeing” Gravity In Real Time

Accelerating Drug Discovery

insideHPC.com Survey
November 2016

92% believe AI will impact their work

93% using deep learning seeing positive results
AI IS CRITICAL FOR INTERNET APPLICATIONS

Users Expect Intelligence In Services

Growing Use of Deep Learning at Google

# of directories containing model description files

Across many products/areas:
- Android
- Apps
- drug discovery
- Gmail
- Image understanding
- Maps
- Natural language understanding
- Photos
- Robotics research
- Speech
- Translation
- YouTube
- ... many others ...

Unique project directories

Time
THE EXPANDING UNIVERSE OF MODERN AI

"THE BIG BANG"
Big Data
GPU
Algorithms

RESEARCH
Berkeley
Georgia Institute of Technology
NYU
University of Toronto

CORE TECHNOLOGY / FRAMEWORKS
Facebook
Google
IBM
Microsoft
NVIDIA
OpenAI
University of Montreal

AI-as-a-PLATFORM
Amazon Web Services
Microsoft Azure

START-UPS
api.ai
Personal Assistant
nervana
Automotive
riverbed
Enforcement & Medical

INDUSTRY LEADERS

© 2016 NVIDIA Corporation. All rights reserved. All other trademarks are property of their respective owners.
A NEW COMPUTING MODEL
Algorithms that Learn from Examples

Expert Written Computer Program

car
vehicle
coupe

Traditional Approach
➢ Requires domain experts
➢ Time consuming
➢ Error prone
➢ Not scalable to new problems
A NEW COMPUTING MODEL
Algorithms that Learn from Examples

Traditional Approach
➢ Requires domain experts
➢ Time consuming
➢ Error prone
➢ Not scalable to new problems

Deep Learning Approach
✓ Learn from data
✓ Easily to extend
✓ Speedup with GPUs
DEEP LEARNING

Untrained Neural Network Model
DEEP LEARNING

TRAINING
Learning a new capability from existing data

Untrained Neural Network Model

Deep Learning Framework

TRAINING DATASET

"cat"

"dog"

"cat"
DEEP LEARNING

TRAINING
Learning a new capability from existing data

- Untrained Neural Network Model
- Deep Learning Framework
- Training Dataset
- Trained Model New Capability

- "dog"
- "cat"
- X
- ✔️
DEEP LEARNING

TRAINING
Learning a new capability from existing data

Untrained Neural Network Model

Deep Learning Framework

TRAINING DATASET

Trained Model New Capability

INFEERENCE
Applying this capability to new data

NEW DATA

App or Service Featuring Capability

Trained Model Optimized for Performance
# CHALLENGES

<table>
<thead>
<tr>
<th>Deep Learning Needs</th>
<th>Why</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Scientists</td>
<td>New computing model</td>
</tr>
<tr>
<td>Latest Algorithms</td>
<td>Rapidly evolving</td>
</tr>
<tr>
<td>Fast Training</td>
<td>Impossible -&gt; Practical</td>
</tr>
<tr>
<td>Deployment Platforms</td>
<td>Must be available everywhere</td>
</tr>
</tbody>
</table>
NVIDIA DEEP LEARNING INSTITUTE

Hands-on Training for Data Scientists and Software Engineers

Helping the world to solve challenging problems using AI and deep learning

On-site workshops and online courses presented by certified instructors

Covering complete workflows for proven application use cases
Self-Driving Cars, Healthcare, Intelligent Video Analytics, IoT/Robotics, Finance and more

www.nvidia.com/dli
ADVANCE YOUR DEEP LEARNING TRAINING AT GTC
Don’t miss the world’s most important event for GPU developers

Silicon Valley, May 8-11
Beijing, September 26-27
Munich, October 10-11

Israel, October 18
Washington DC, November 1-2
Tokyo, December 12-13
DEEP LEARNING SOFTWARE

NVIDIA DIGITS™
Interactively manage data and train deep learning models for image classification without the need to write code.

Learn more

developer.nvidia.com/deep-learning

Deep Learning Frameworks
Design and train deep learning models using a high-level interface. Choose a deep learning framework that best suits your needs based on your choice of programming language, platform, and target application.

Learn more

NVIDIA Deep Learning SDK
This SDK delivers high-performance multi-GPU acceleration and industry-vetted deep learning algorithms, and is designed for easy drop-in acceleration for deep learning frameworks.

cuDNN
TensorRT
DeepStream SDK
NCCL

cuSPARSE
cuBLAS
END-TO-END PRODUCT FAMILY

TRAINING

FULLY INTEGRATED DL SUPERCOMPUTER

DGX-1

DATA CENTER

Tesla P100

Tesla P4

Tesla P100

Titan X Pascal

Tesla P100

DATA CENTER

AUTOMOTIVE

Drive PX

INFEERENCE

EMBEDDED

Jetson TX
## CHALLENGES

<table>
<thead>
<tr>
<th>Deep Learning Needs</th>
<th>Why</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Scientists</td>
<td>New computing model</td>
</tr>
<tr>
<td>Latest Algorithms</td>
<td>Rapidly evolving</td>
</tr>
<tr>
<td>Fast Training</td>
<td>Impossible -&gt; Practical</td>
</tr>
<tr>
<td>Deployment Platforms</td>
<td>Must be available everywhere</td>
</tr>
</tbody>
</table>
## CHALLENGES

<table>
<thead>
<tr>
<th>Deep Learning Needs</th>
<th>NVIDIA Delivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Scientists</td>
<td>Deep Learning Institute, GTC, DIGITS</td>
</tr>
<tr>
<td>Latest Algorithms</td>
<td>DL SDK, GPU-Accelerated Frameworks</td>
</tr>
<tr>
<td>Fast Training</td>
<td>DGX, P100, TITAN X</td>
</tr>
<tr>
<td>Deployment Platforms</td>
<td>TensorRT, P100, P4, Drive PX, Jetson</td>
</tr>
</tbody>
</table>
1. What problem are you solving, what are the DL tasks?
2. On what platform(s) will you train and deploy?
3. What data do you have/need, and how is it labeled?
4. Which deep learning framework & tools will you use?
## WHAT PROBLEM ARE SOLVING?
### Defining the AI/DL Tasks

<table>
<thead>
<tr>
<th>INPUTS</th>
<th>QUESTION</th>
<th>AI/DL TASK</th>
<th>EXAMPLE OUTPUTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text Data</td>
<td>Is “it” present or not?</td>
<td>Detection</td>
<td>Cancer Detection</td>
</tr>
<tr>
<td>Images</td>
<td>What type of thing is “it”?</td>
<td>Classification</td>
<td>Tumor Identification</td>
</tr>
<tr>
<td>Video</td>
<td>To what extent is “it” present?</td>
<td>Segmentation</td>
<td>Tumor Size/Shape Analysis</td>
</tr>
<tr>
<td>Audio</td>
<td>What is the likely outcome?</td>
<td>Prediction</td>
<td>Survivability Prediction</td>
</tr>
<tr>
<td></td>
<td>What will likely satisfy the objective?</td>
<td>Recommendation</td>
<td>Therapy Recommendation</td>
</tr>
</tbody>
</table>
SELECTING A DEEP LEARNING FRAMEWORK

Considerations

1. Type of problem
2. Training & deployment platforms
3. DNN models available, layer types supported
4. Latest algos & GPU acceleration: cuDNN, NCCL, etc.
5. Usage model/interfaces: GUI, command line, programming language, etc.
6. Easy to install and get started: containers, docs, code samples, tutorials, ...
7. Enterprise integration, vendors, ecosystem
START SIMPLE, LEARN FAST

How One NVIDIAN Uses Deep Learning to Keep Cats from Pooping on His Lawn