DEEP LEARNING LAB SERIES SCHEDULE

- 7/22 Class #1 - Introduction to Deep Learning
- 7/29 Office Hours for Class #1

- 8/5 Class #2 - Getting Started with DIGITS interactive training system for image classification
- 8/12 Office Hours for Class #2

- 8/19 Class #3 - Getting Started with the Caffe Framework
- 8/26 Office Hours for Class #3

- 9/2 Class #4 - Getting Started with the Theano Framework
- 9/9 Office Hours for Class #4

- 9/16 Class #5 - Getting Started with the Torch Framework
- 9/23 Office Hours for Class #5

USEFUL LINKS

- Deep Learning Lab Course information & recordings:
  developer.nvidia.com/deep-learning-courses
- Recorded presentations from past conferences:
  www.gputechconf.com/gtcnew/on-demand-gtc.php
- Parallel Forall (GPU Computing Technical blog):
  devblogs.nvidia.com/parallelforall
- Become a Registered Developer:
  developer.nvidia.com/programs/cuda/register

INTRODUCTION TO DEEP LEARNING WITH GPUS

July 2015

AGENDA

1 What is Deep Learning?
2 Deep Learning software
3 Deep Learning deployment

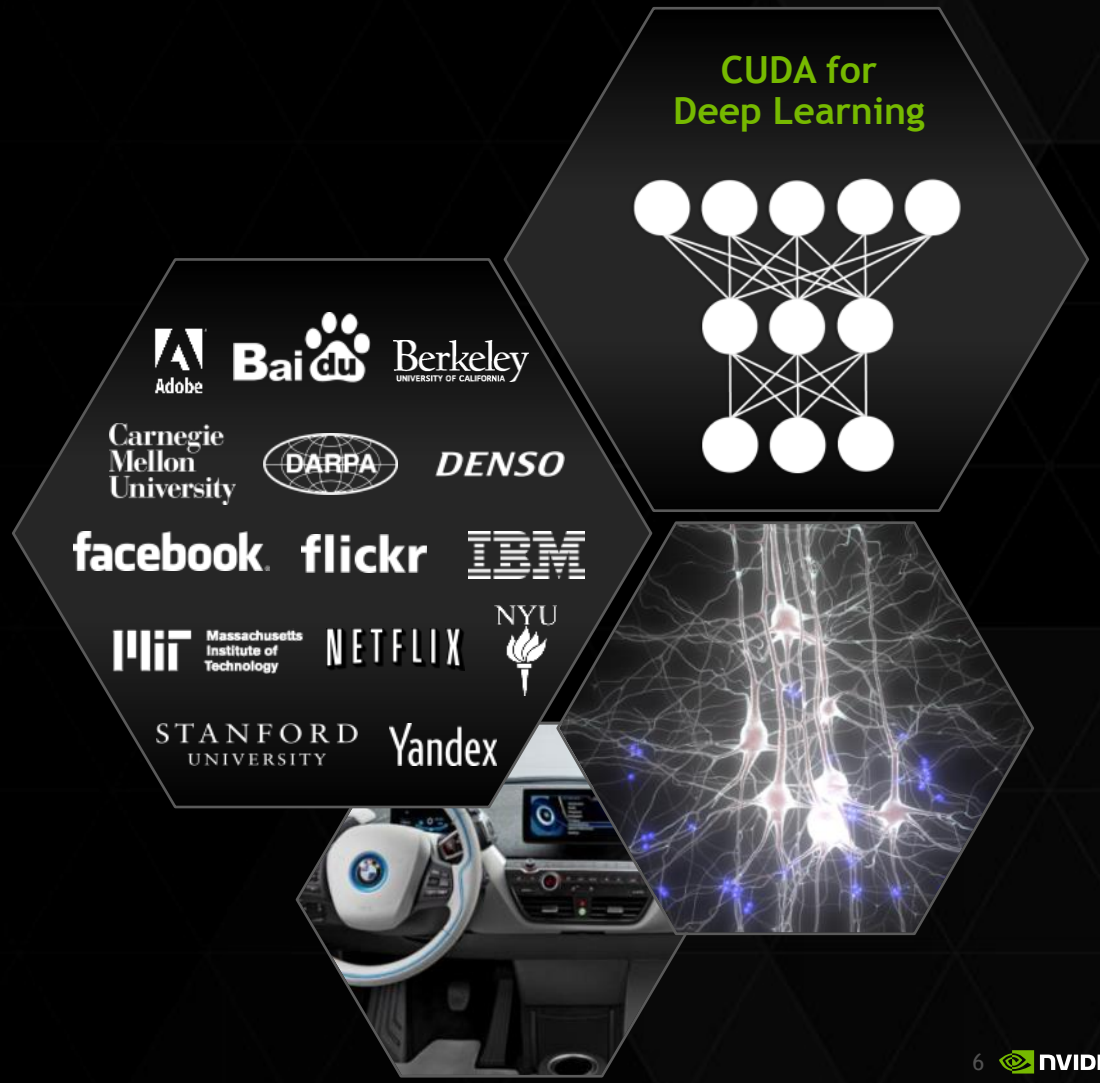
What is Deep Learning?

DEEP LEARNING & AI

Deep Learning has become the most popular approach to developing Artificial Intelligence (AI) – machines that perceive and understand the world

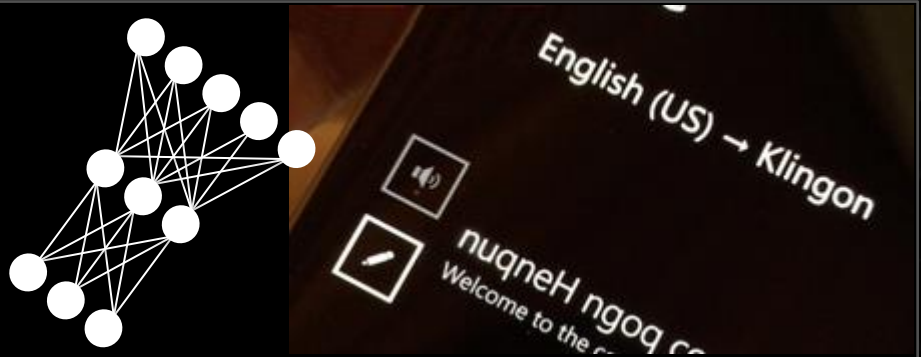
The focus is currently on specific perceptual tasks, and there are many successes.

Today, some of the world's largest internet companies, as well as the foremost research institutions, are using GPUs for deep learning in research and production



PRACTICAL DEEP LEARNING EXAMPLES

Image Classification, Object Detection, Localization,
Action Recognition, Scene Understanding

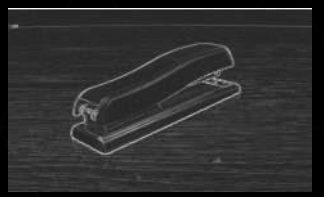
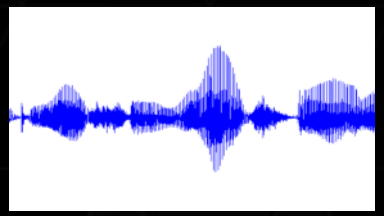
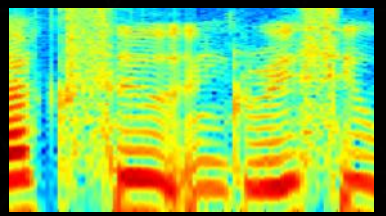
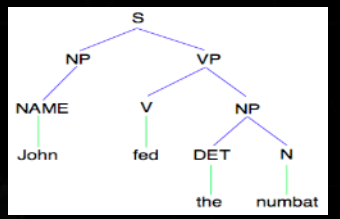


Speech Recognition, Speech Translation,
Natural Language Processing

Pedestrian Detection, Traffic Sign Recognition

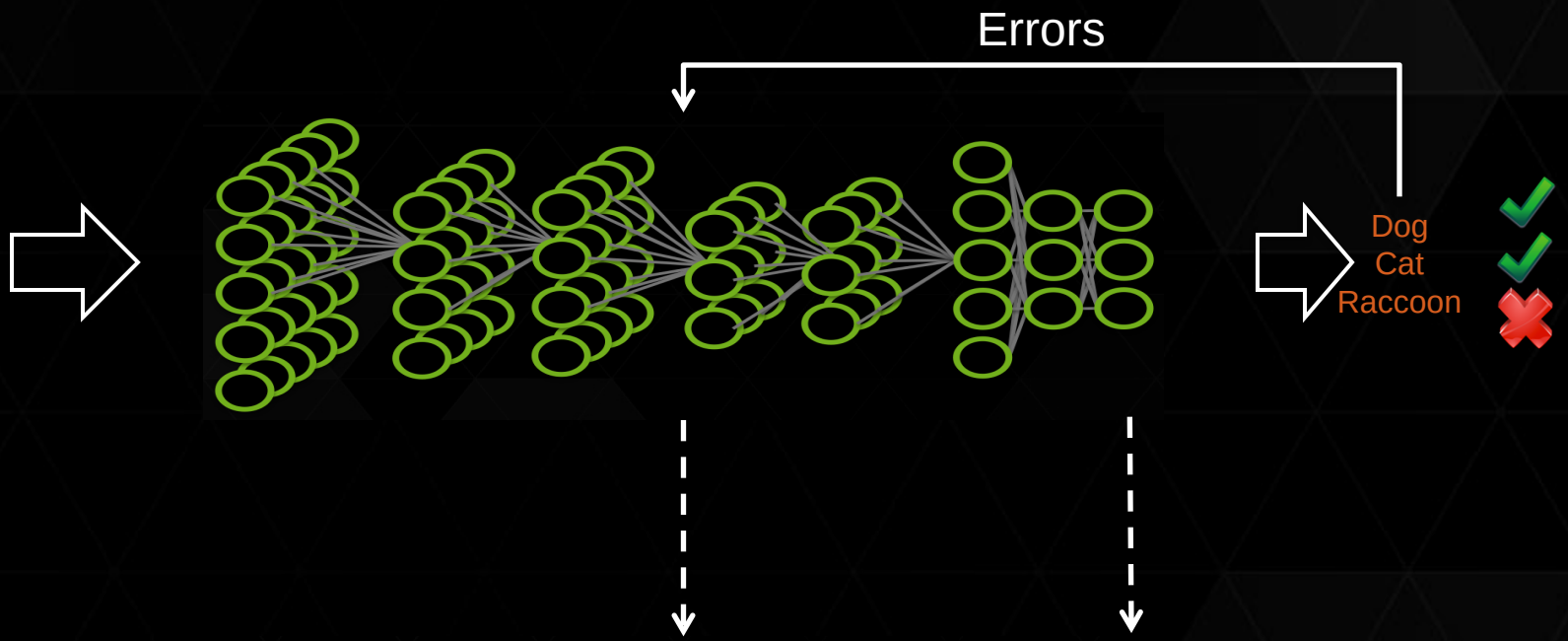
Breast Cancer Cell Mitosis Detection,
Volumetric Brain Image Segmentation

TRADITIONAL MACHINE PERCEPTION – HAND TUNED FEATURES

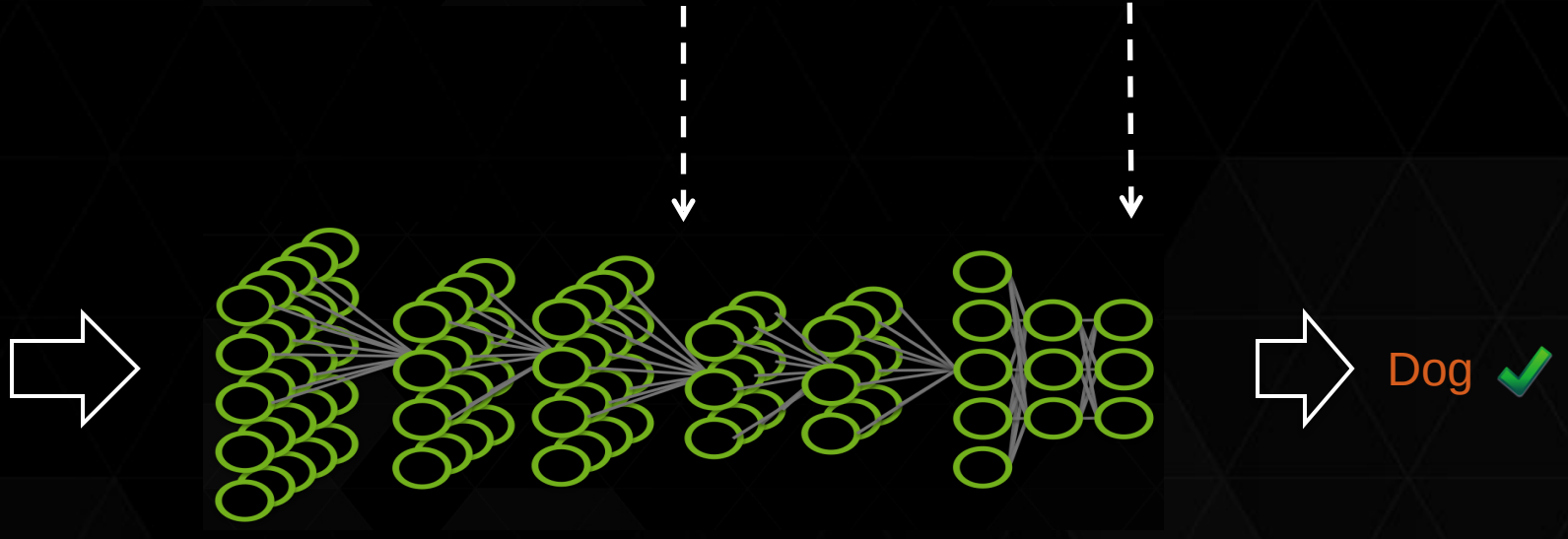
Raw data	Feature extraction	Classifier/ detector	Result
		SVM, shallow neural net, …	
		HMM, shallow neural net, …	Speaker ID, speech transcription, …
		Clustering, HMM, LDA, LSA …	Topic classification, machine translation, sentiment analysis…

DEEP LEARNING APPROACH

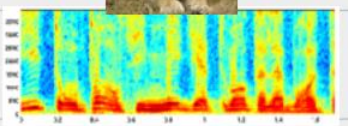
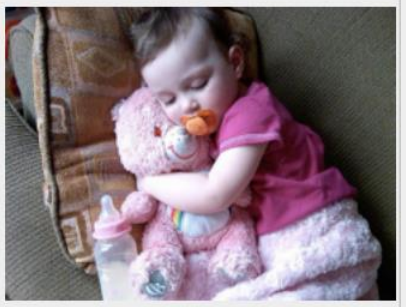
Train:



Deploy:



SOME DEEP LEARNING USE CASES

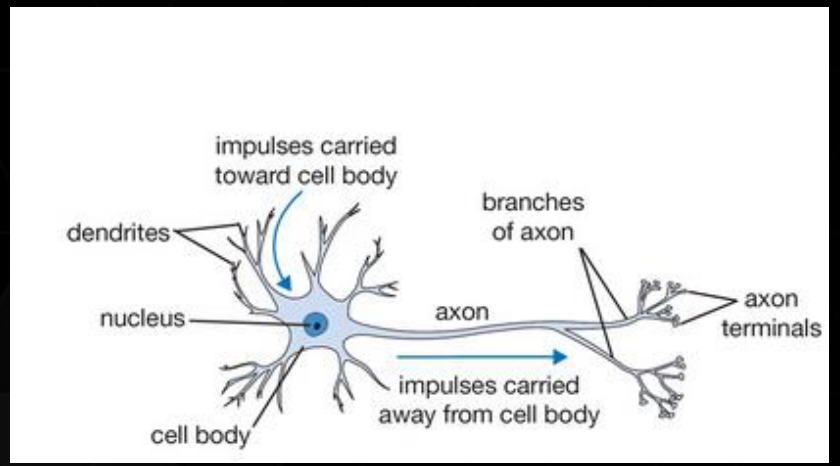
Input	Output
Pixels: 	“lion”
Audio: 	“see at tuhl res taur aun ts”
<query, doc>	P(click on doc)
“Hello, how are you?”	“Bonjour, comment allez-vous?”
Pixels: 	“A close up of a small child holding a stuffed animal”

Jeff Dean, Google, GTC 2015

ARTIFICIAL NEURAL NETWORK (ANN)

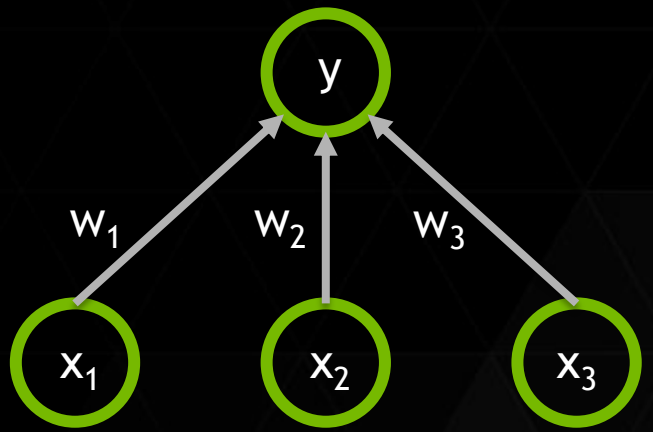
A collection of simple, trainable mathematical units that collectively learn complex functions

Biological neuron



From Stanford cs231n lecture notes

Artificial neuron

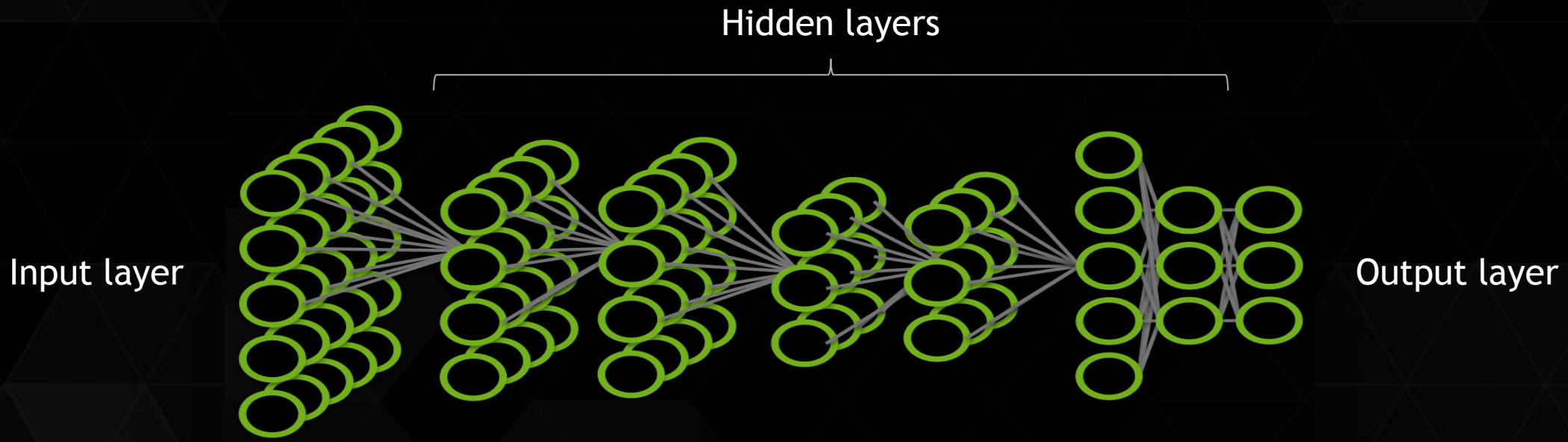


$y=F(w_1x_1+w_2x_2+w_3x_3)$

$F(x)=\max(0,x)$

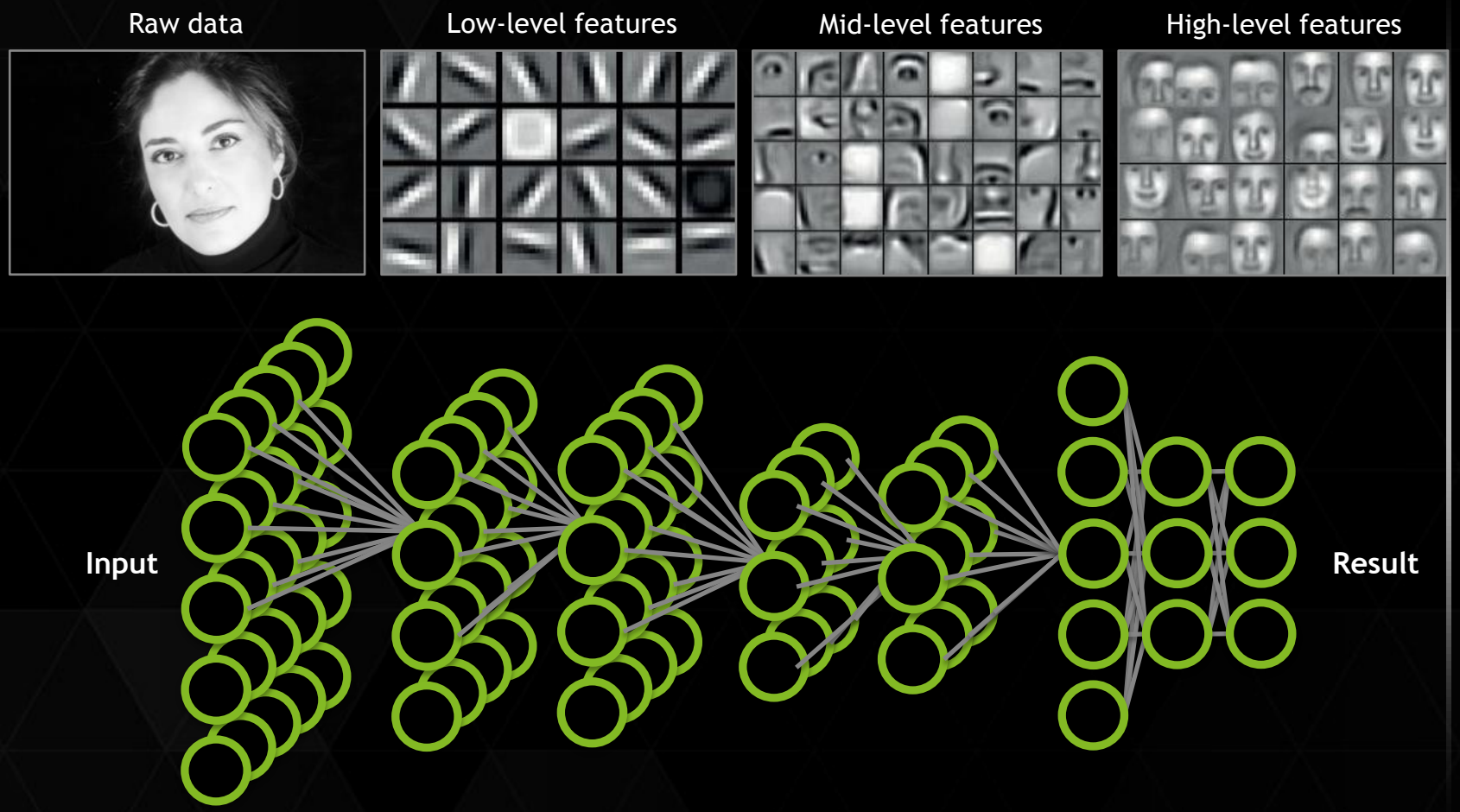
ARTIFICIAL NEURAL NETWORK (ANN)

A collection of simple, trainable mathematical units that collectively learn complex functions



Given sufficient training data an artificial neural network can approximate very complex functions mapping raw data to output decisions

DEEP NEURAL NETWORK (DNN)



Application components:

Task objective
e.g. Identify face

Training data
10-100M images

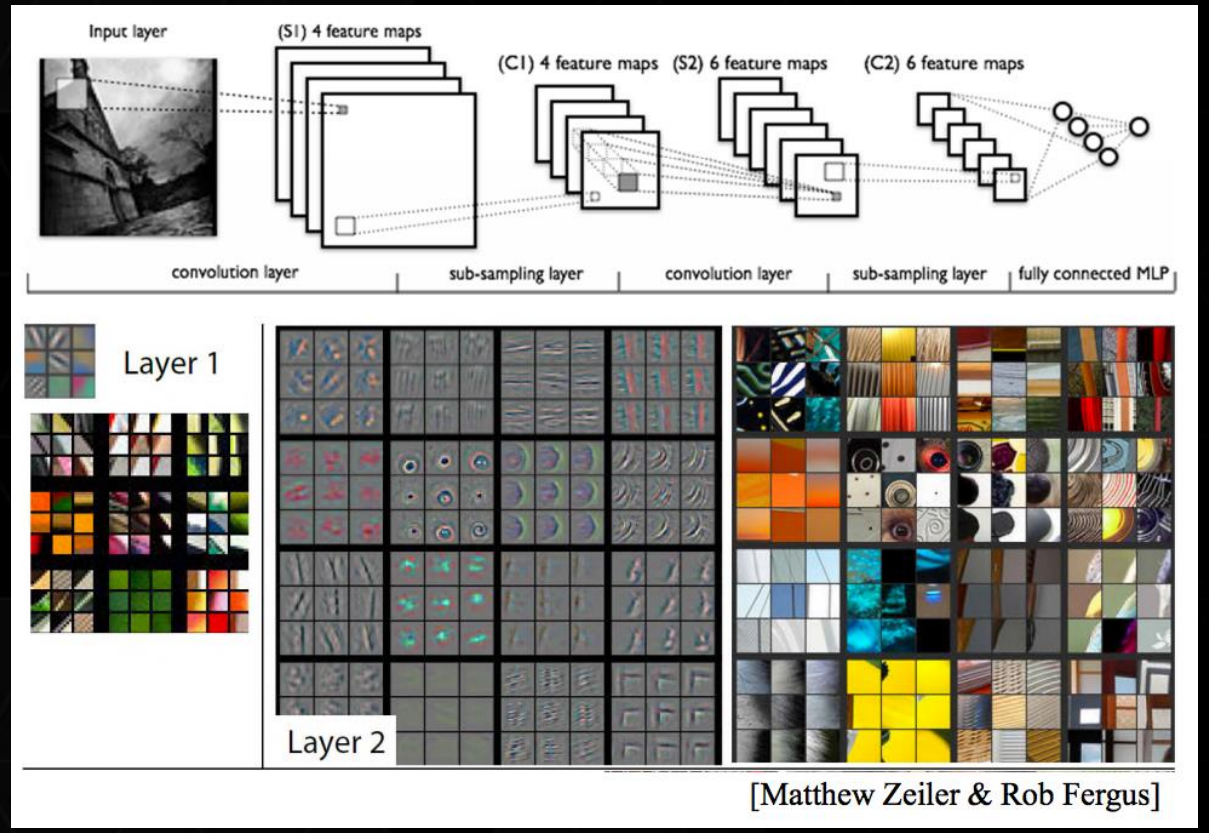
Network architecture
~10 layers
1B parameters

Learning algorithm
~30 Exaflops
~30 GPU days

DEEP LEARNING ADVANTAGES

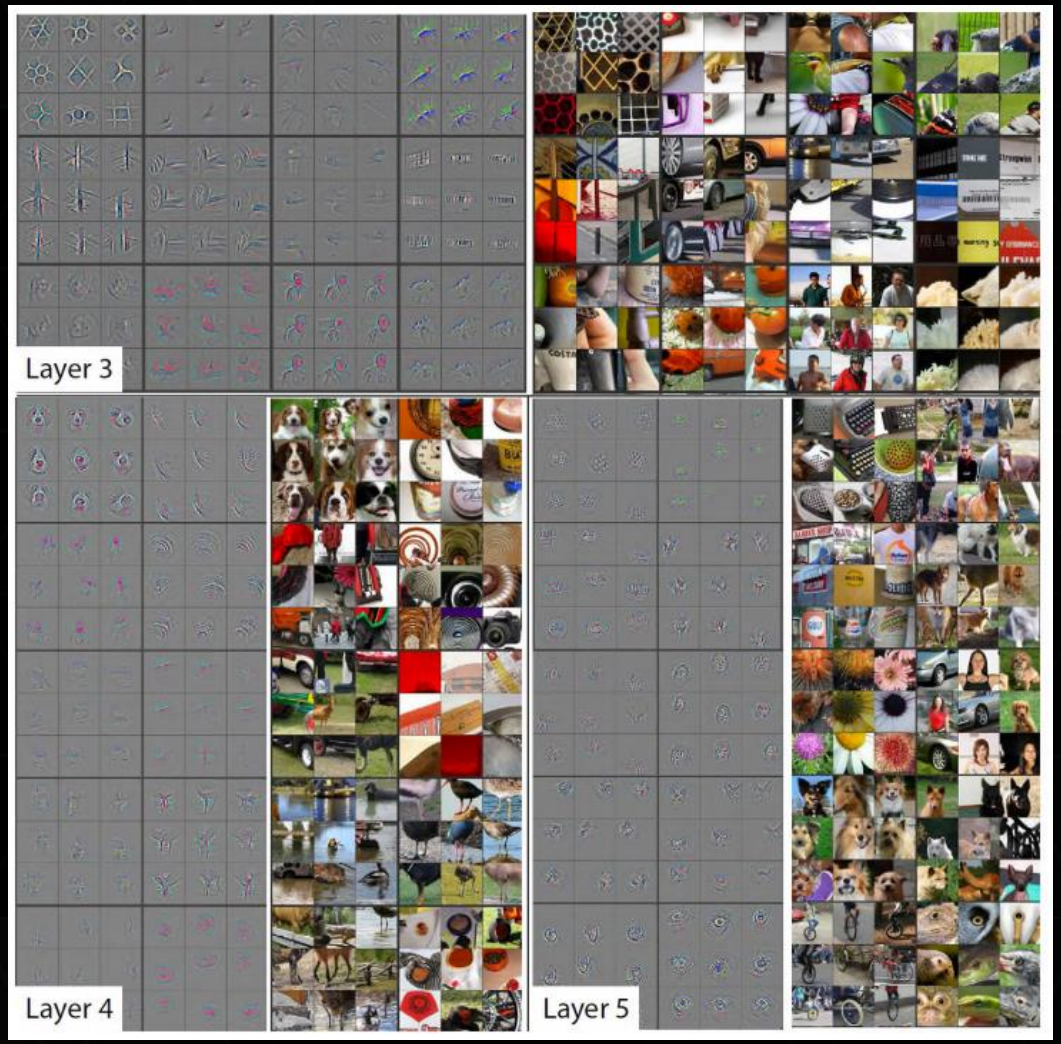
- Robust
  - No need to design the features ahead of time – features are automatically learned to be optimal for the task at hand
  - Robustness to natural variations in the data is automatically learned
- Generalizable
  - The same neural net approach can be used for many different applications and data types
- Scalable
  - Performance improves with more data, method is massively parallelizable

CONVOLUTIONAL NEURAL NETWORK (CNN)

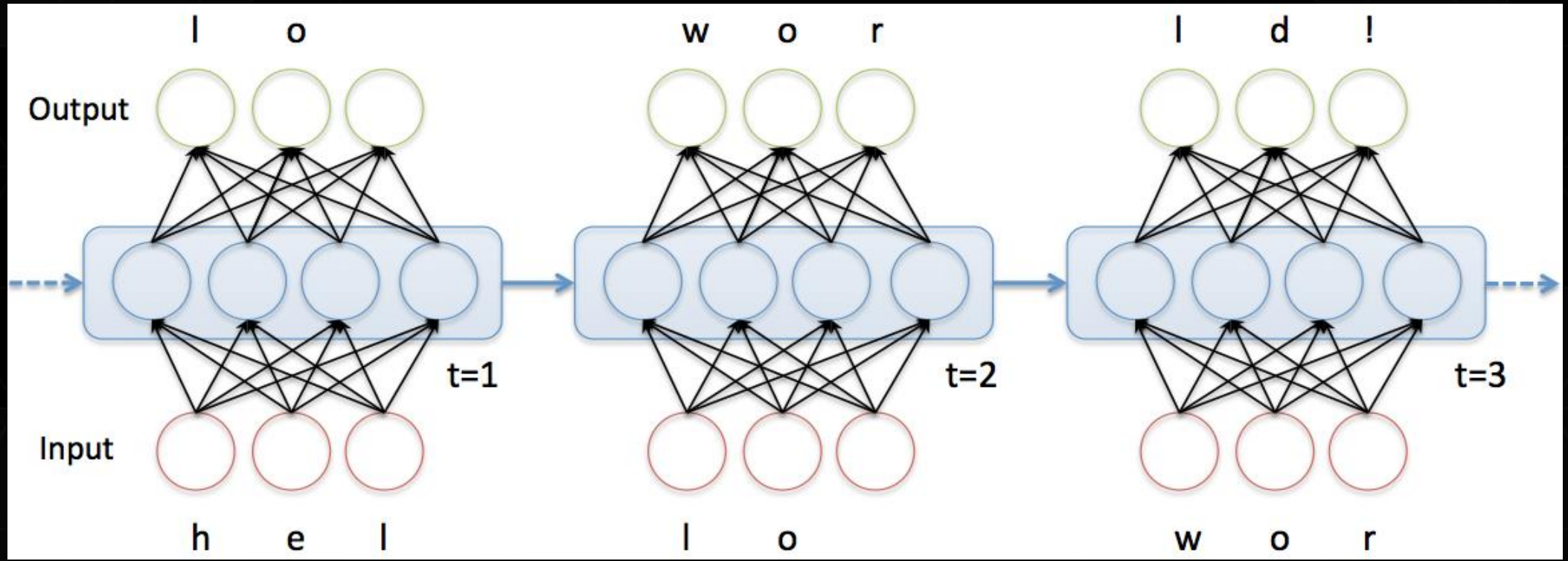


- Inspired by the human visual cortex
- Learns a hierarchy of visual features
- Local pixel level features are scale and translation invariant
- Learns the “essence” of visual objects and generalizes well

CONVOLUTIONAL NEURAL NETWORK (CNN)



RECURRENT NEURAL NETWORK (RNN)



DNNS DOMINATE IN PERCEPTUAL TASKS

- Handwriting recognition MNIST (many), Arabic HWX (IDSIA)
- OCR in the Wild [2011]: StreetView House Numbers (NYU and others)
- Traffic sign recognition [2011] GTSRB competition (IDSIA, NYU)
- Asian handwriting recognition [2013] ICDAR competition (IDSIA)
- Pedestrian Detection [2013]: INRIA datasets and others (NYU)
- Volumetric brain image segmentation [2009] connectomics (IDSIA, MIT)
- Human Action Recognition [2011] Hollywood II dataset (Stanford)
- Object Recognition [2012] ImageNet competition (Toronto)
- Scene Parsing [2012] Stanford bgd, SiftFlow, Barcelona datasets (NYU)
- Scene parsing from depth images [2013] NYU RGB-D dataset (NYU)
- Speech Recognition [2012] Acoustic modeling (IBM and Google)
- Breast cancer cell mitosis detection [2011] MITOS (IDSIA)

Slide credit: Yann Lecun, Facebook & NYU

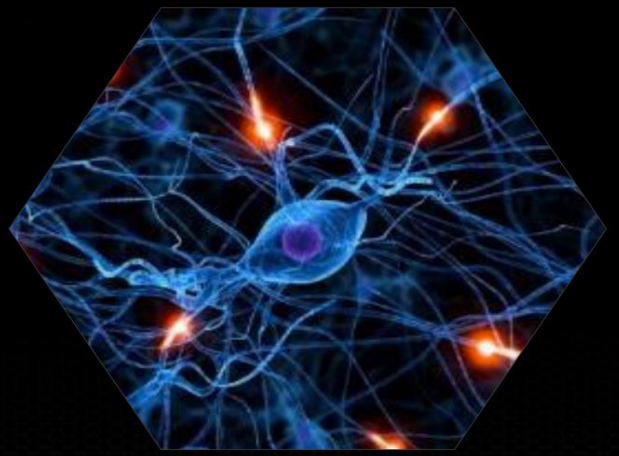
WHY IS DEEP LEARNING HOT *NOW*?

Three Driving Factors…

Big Data Availability

- facebook: 350 millions images uploaded per day
- Walmart: 2.5 Petabytes of customer data hourly
- YouTube: 100 hours of video uploaded every minute

New DL Techniques



GPU acceleration

GPUs and Deep Learning

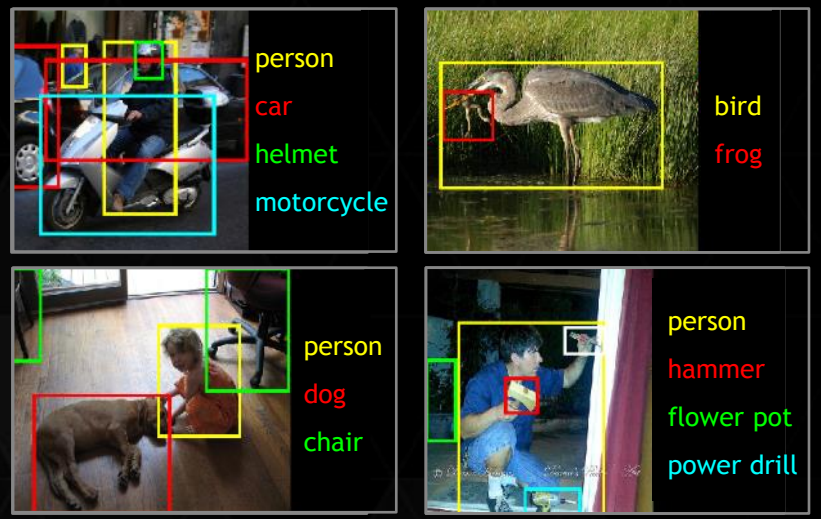
GPUs — THE PLATFORM FOR DEEP LEARNING

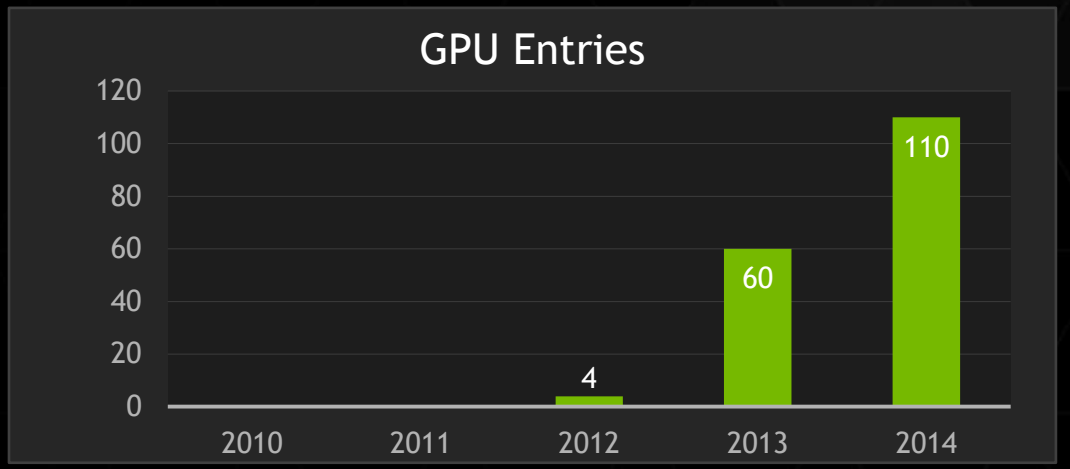
Image Recognition Challenge

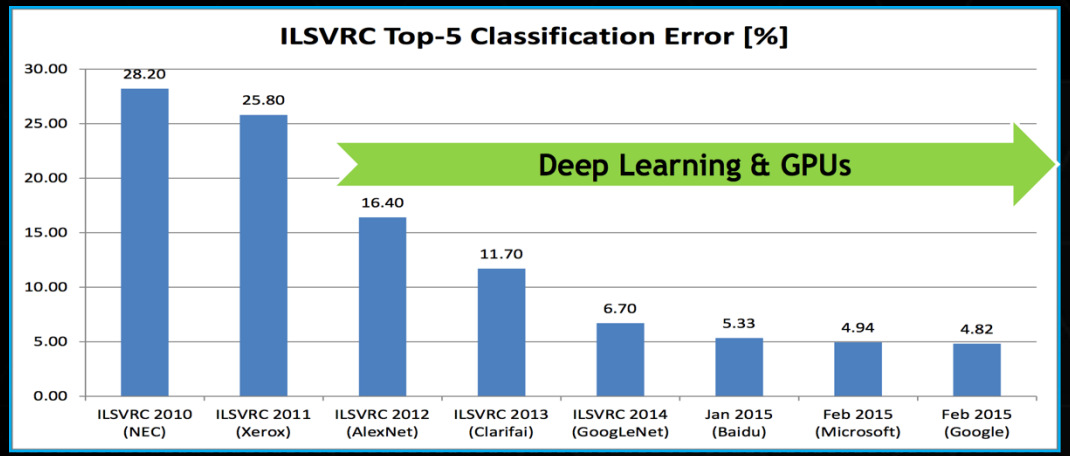
1.2M training images • 1000 object categories

Hosted by

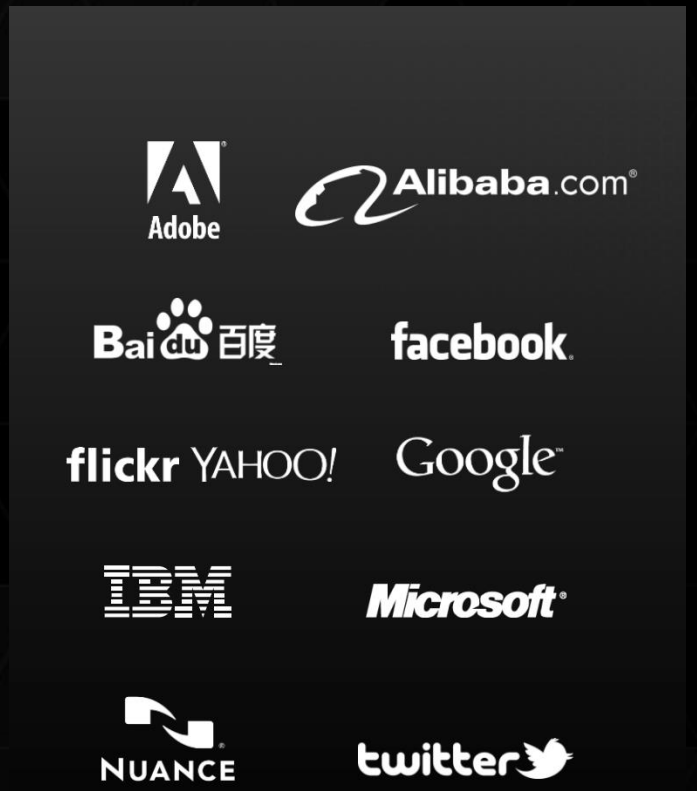
IMAGENET







GPU-ACCELERATED DEEP LEARNING



GPUS MAKE DEEP LEARNING ACCESSIBLE

Deep learning with COTS HPC systems

A. Coates, B. Huval, T. Wang, D. Wu, A. Ng, B. Catanzaro

ICML 2013

"Now You Can Build Google's $1M Artificial Brain on the Cheap"

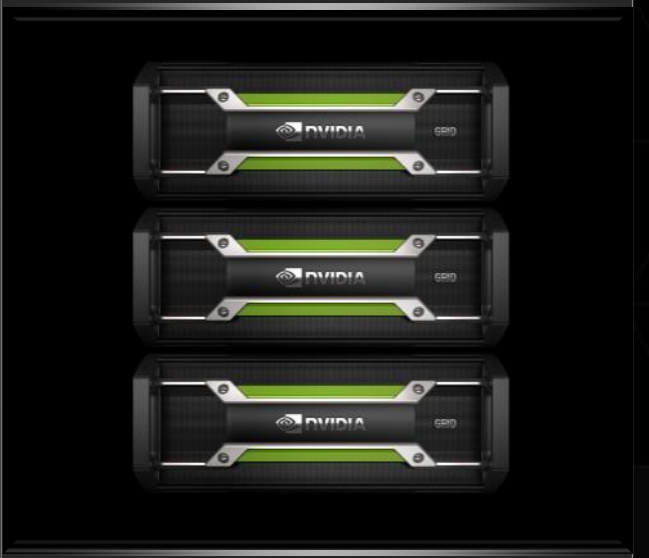
WIRED

GOOGLE DATACENTER

1,000 CPU Servers
2,000 CPUs • 16,000 cores

600 kWatts
$5,000,000

STANFORD AI LAB



3 GPU-Accelerated Servers
12 GPUs • 18,432 cores

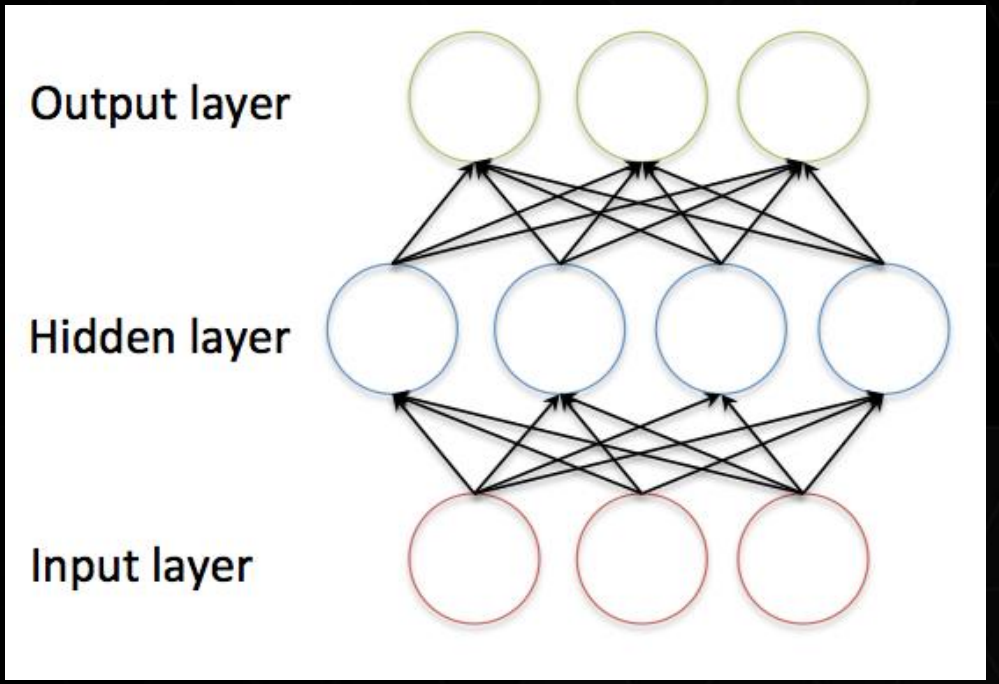
4 kWatts
$33,000

WHY ARE GPUs GOOD FOR DEEP LEARNING?

	Neural Networks	GPUs
Inherently Parallel	✓	✓
Matrix Operations	✓	✓
FLOPS	✓	✓
Bandwidth	✓	✓

GPUs deliver --

- *same or better prediction accuracy*
- *faster results*
- *smaller footprint*
- *lower power*
- *lower cost*



GPU ACCELERATION

Training A Deep, Convolutional Neural Network

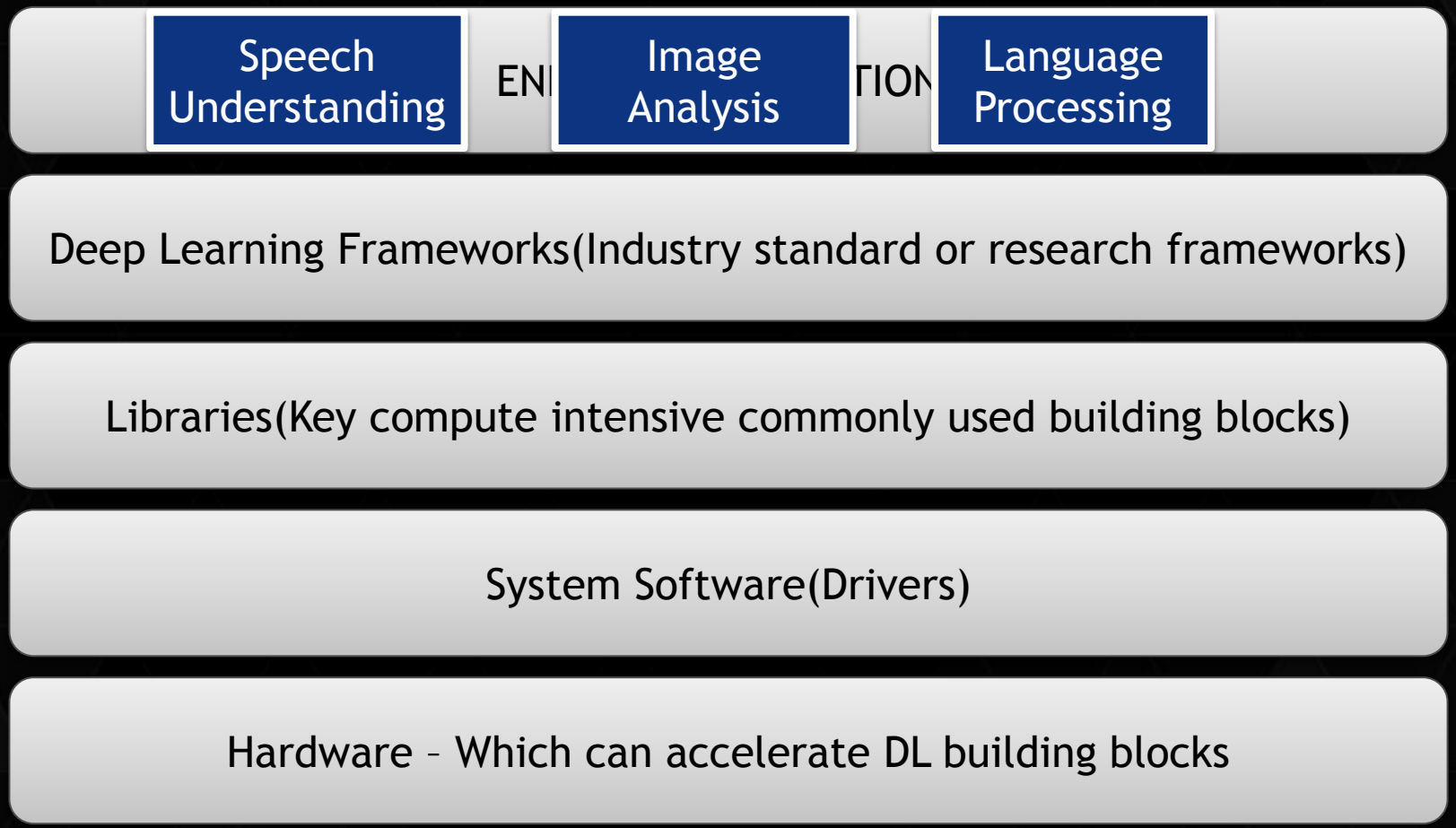
Batch Size	Training Time CPU	Training Time GPU	GPU Speed Up
64 images	64 s	7.5 s	8.5X
128 images	124 s	14.5 s	8.5X
256 images	257 s	28.5 s	9.0X

- ILSVRC12 winning model: “Supervision”
- 7 layers
- 5 convolutional layers + 2 fully-connected
- ReLU, pooling, drop-out, response normalization
- Implemented with Caffe
- Training time is for 20 iterations

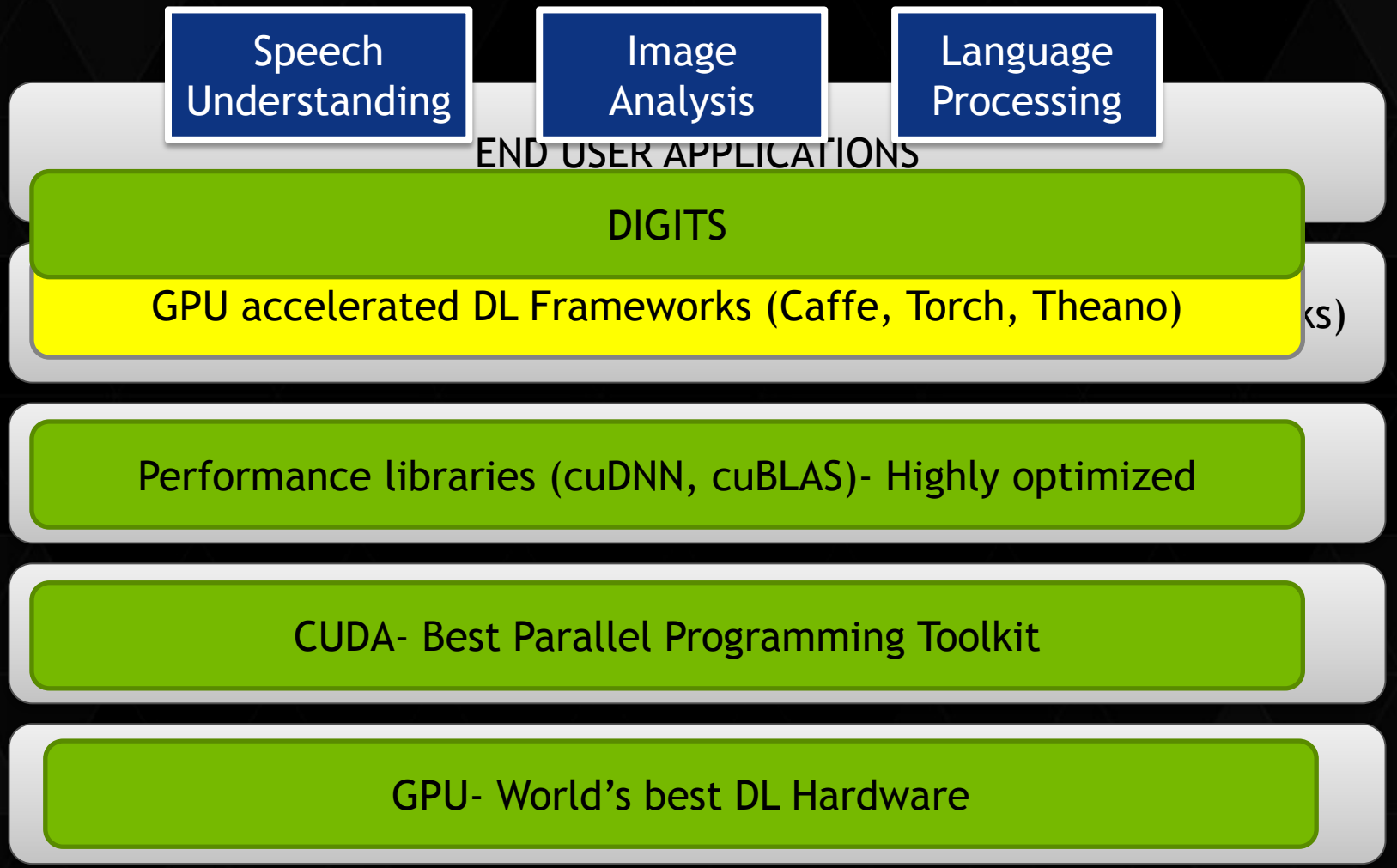
- Dual 10-core Ivy Bridge CPUs
- 1 Tesla K40 GPU
- CPU times utilized Intel MKL BLAS library
- GPU acceleration from CUDA matrix libraries (cuBLAS)

DL software landscape

HOW TO WRITE APPLICATIONS USING DL



HOW NVIDIA IS HELPING DL STACK



GPU-ACCELERATED DEEP LEARNING FRAMEWORKS

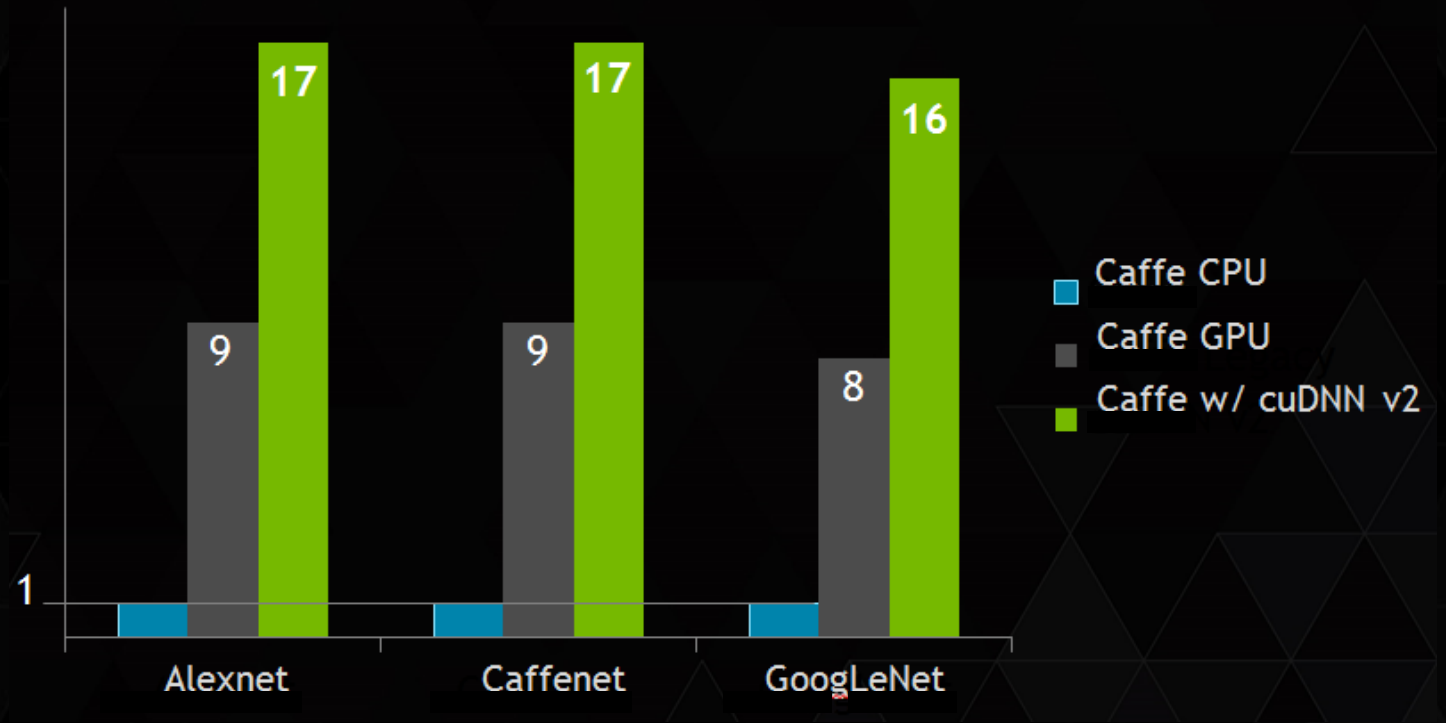
	CAFFE	TORCH	THEANO	KALDI
Domain	Deep Learning Framework	Scientific Computing Framework	Math Expression Compiler	Speech Recognition Toolkit
cuDNN	2.0	2.0	2.0	--
Multi-GPU	via DIGITS 2	In Progress	In Progress	✓ (nnet2)
Multi-CPU	✗	✗	✗	✓ (nnet2)
License	BSD-2	GPL	BSD	Apache 2.0
Interface(s)	Command line, Python, MATLAB	Lua, Python, MATLAB	Python	C++, Shell scripts
Embedded (TK1)	✓	✓	✗	✗

http://developer.nvidia.com/deeplearning

All three frameworks covered in the associated “Intro to DL” hands-on lab

cuDNN V2 - PERFORMANCE

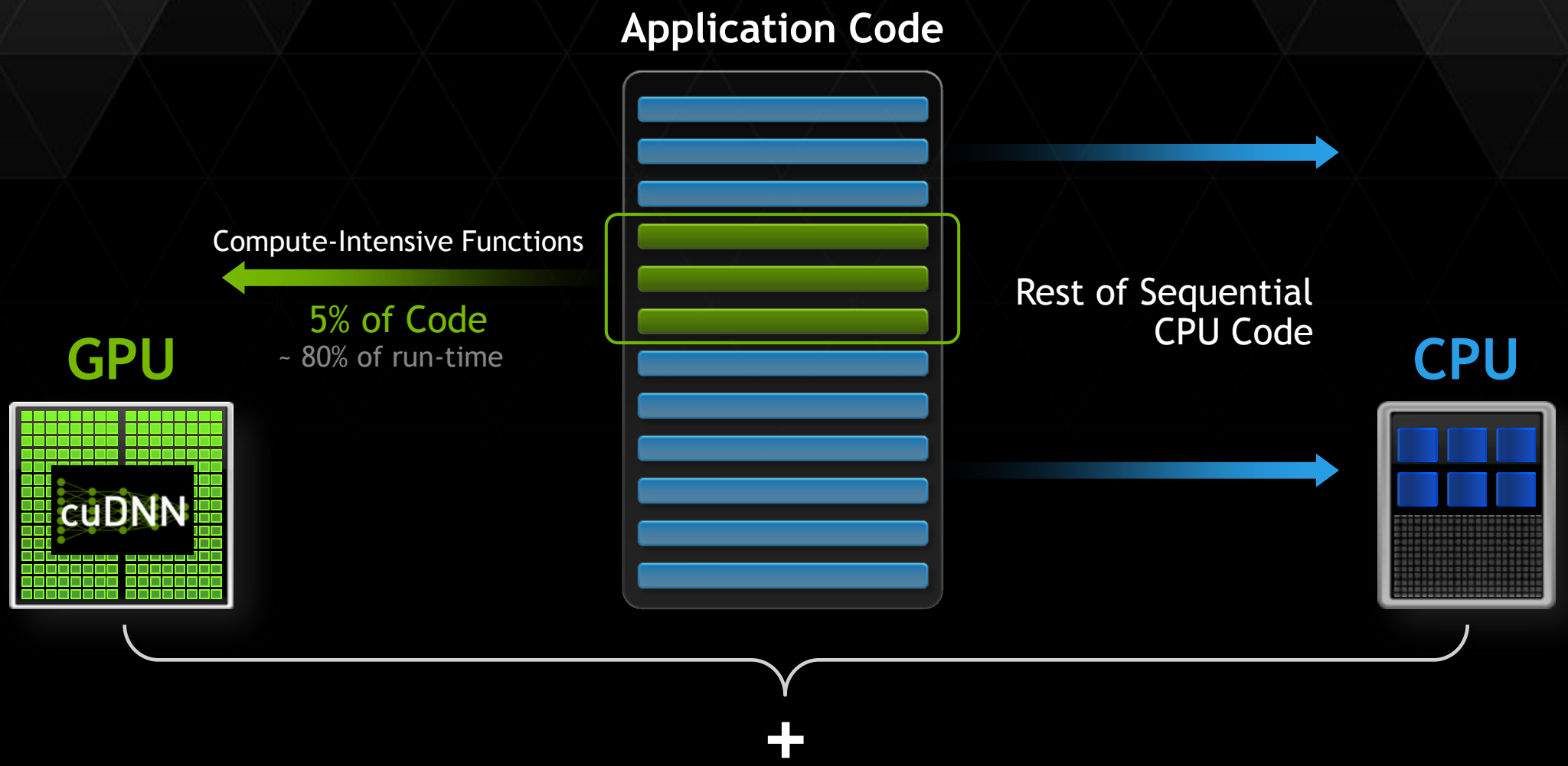
v3 coming soon



CPU is 16 core Haswell E5-2698 at 2.3 GHz, with 3.6 GHz Turbo

GPU is NVIDIA Titan X

HOW GPU ACCELERATION WORKS



cuDNN ROUTINES

- Convolutions – 80-90% of the execution time
- Pooling - Spatial smoothing

- Activations - Pointwise non-linear function

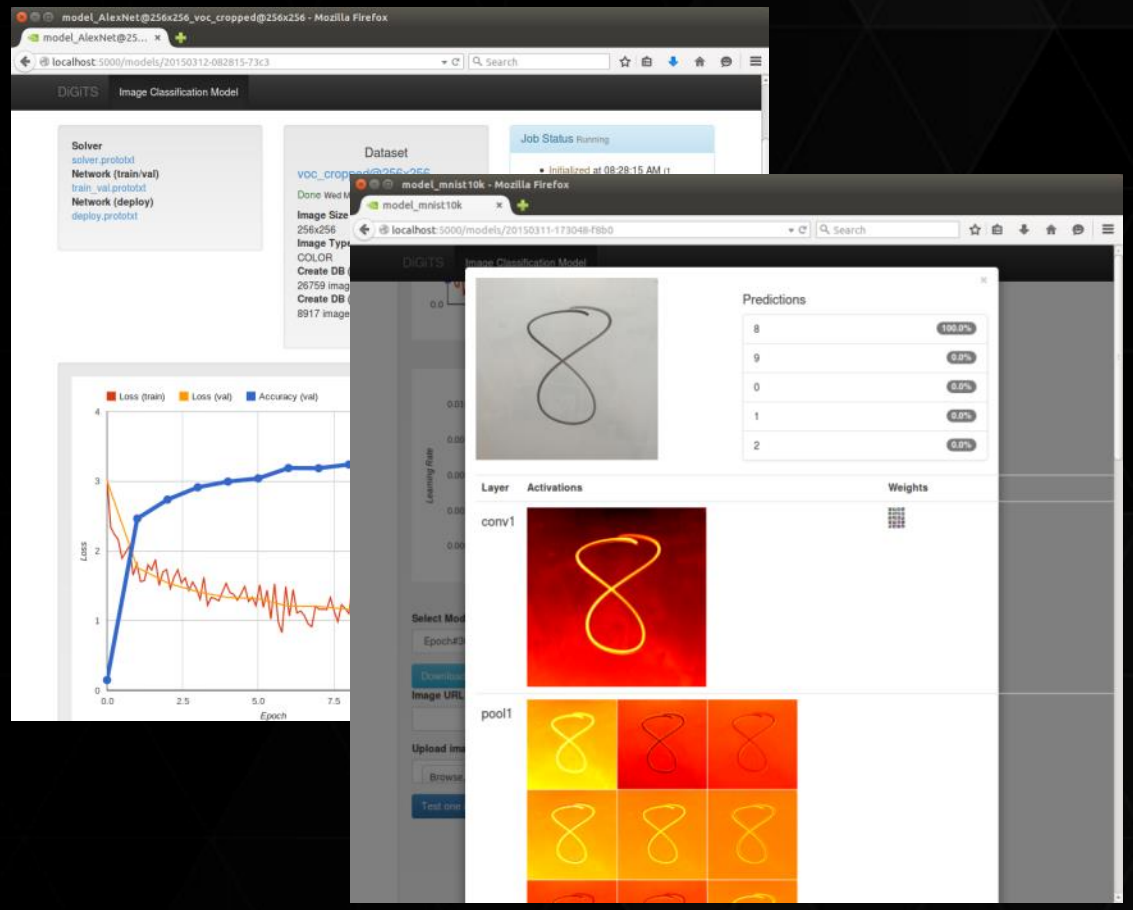
https://developer.nvidia.com/cudnn

DIGITS

Interactive Deep Learning GPU Training System

Data Scientists & Researchers:

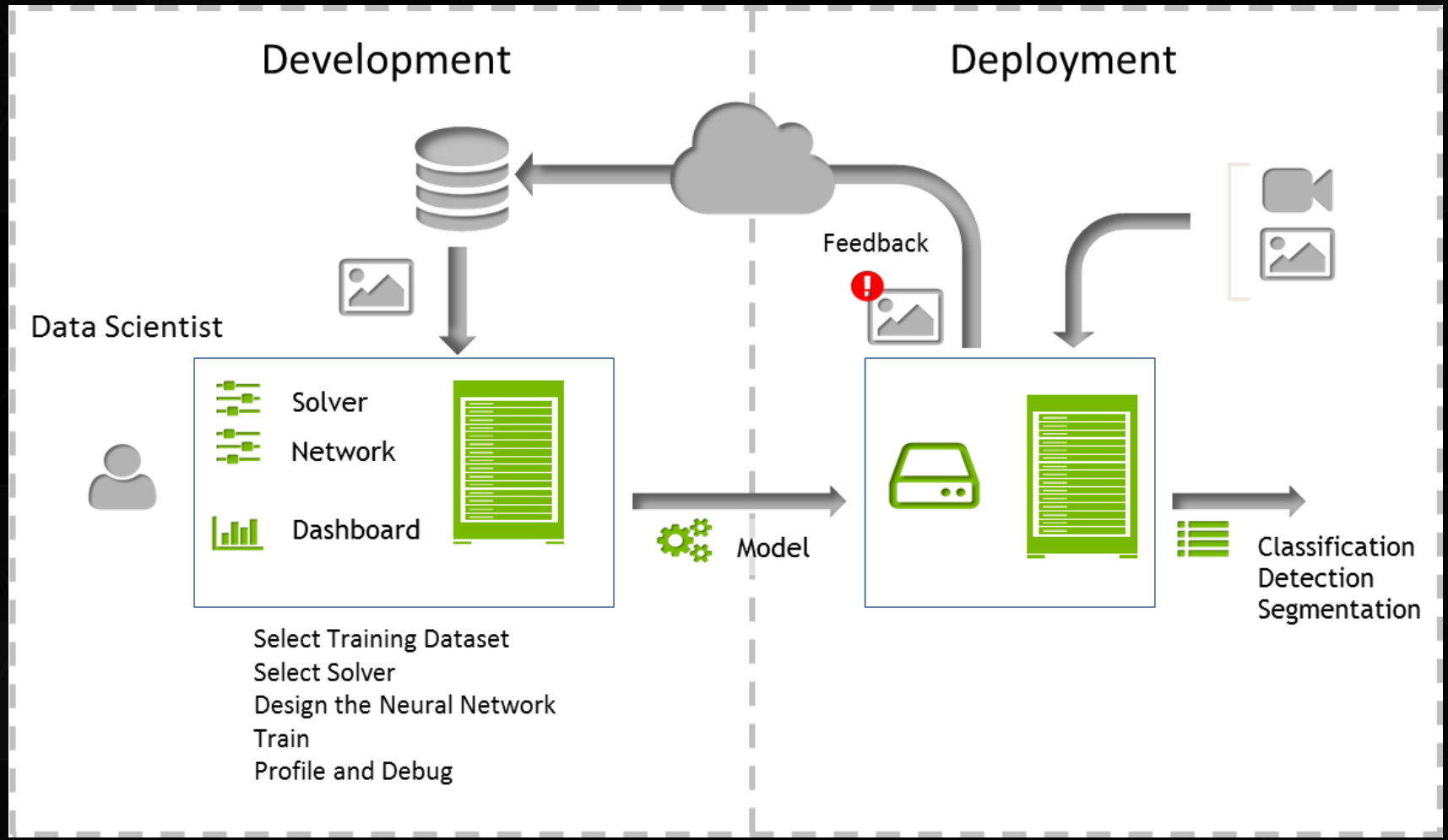
- Quickly design the best deep neural network (DNN) for your data
- Visually monitor DNN training quality in real-time
- Manage training of many DNNs in parallel on multi-GPU systems
- DIGITS 2 - Accelerate training of a single DNN using multiple GPUs



https://developer.nvidia.com/digits

DL deployment

DEEP LEARNING DEPLOYMENT WORKFLOW



DEEP LEARNING LAB SERIES SCHEDULE

- 7/22 Class #1 - Introduction to Deep Learning
- 7/29 Office Hours for Class #1

- 8/5 Class #2 - Getting Started with DIGITS interactive training system for image classification
- 8/12 Office Hours for Class #2

- 8/19 Class #3 - Getting Started with the Caffe Framework
- 8/26 Office Hours for Class #3

- 9/2 Class #4 - Getting Started with the Theano Framework
- 9/9 Office Hours for Class #4

- 9/16 Class #5 - Getting Started with the Torch Framework
- 9/23 Office Hours for Class #5

- More information available at developer.nvidia.com/deep-learning-courses

HANDS-ON LAB

1. Create an account at nvidia.qwiklab.com
2. Go to “Introduction to Deep Learning” lab at bit.ly/dlnvlab1
3. Start the lab and enjoy!

- Only requires a supported browser, no NVIDIA GPU necessary!
- Lab is free until end of Deep Learning Lab series