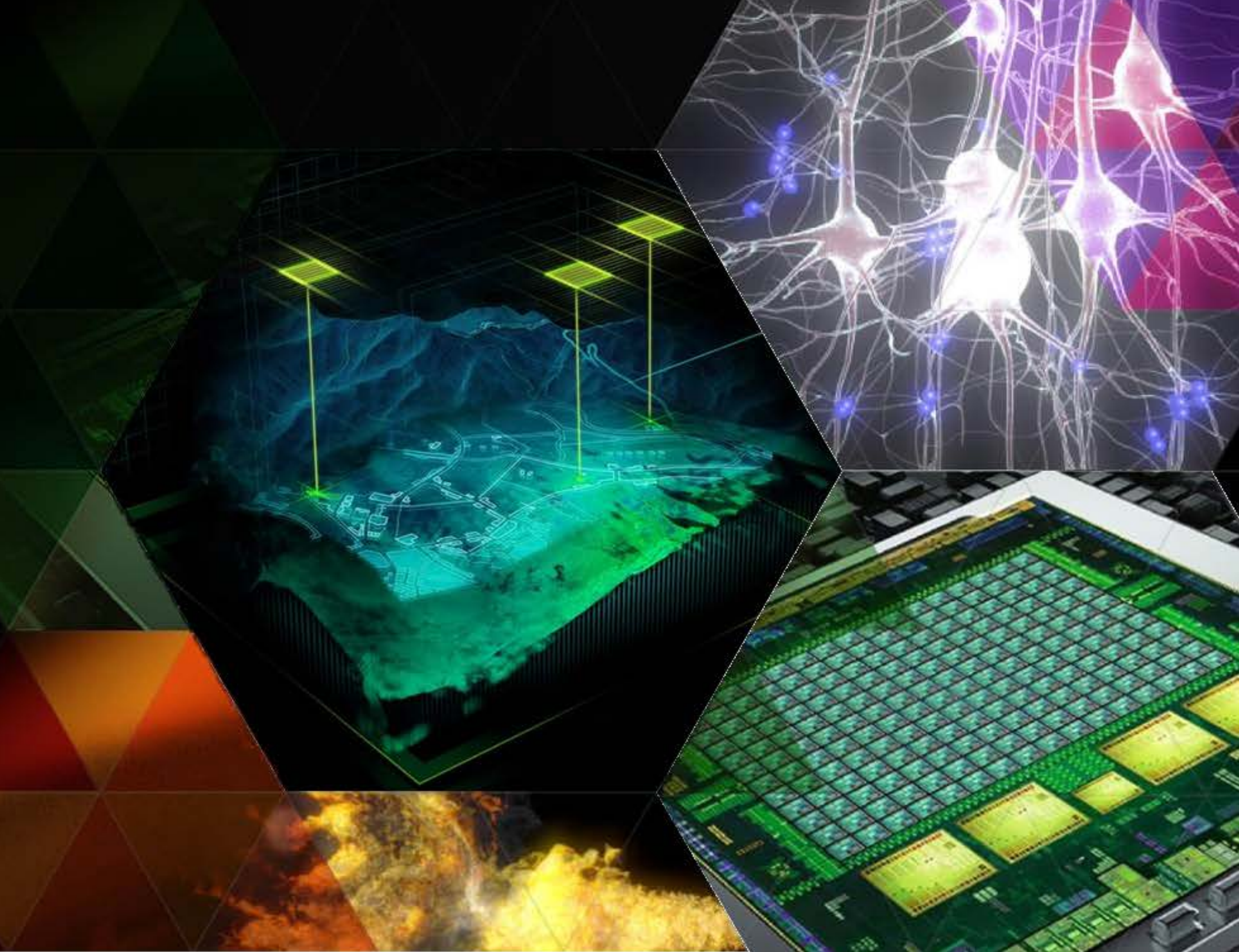
DEEP LEARNING FOR IMAGE CLASSIFICATION

GEOINT Training



Larry Brown Ph.D.
larryb@nvidia.com

June 2015

AGENDA

1 What is Deep Learning?
2 GPUs and Deep Learning
3 cuDNN and DiGiTS
4 Neural Network Motivation
5 Working with Deep Neural Networks
6 Using Caffe for Deep Learning
7 Summary – DL For GEOINT

What is Deep Learning?

DATA SCIENCE LANDSCAPE

Data Analytics

SQL Query

Machine Learning

Graph Analytics

Traditional Methods

Deep Neural Networks

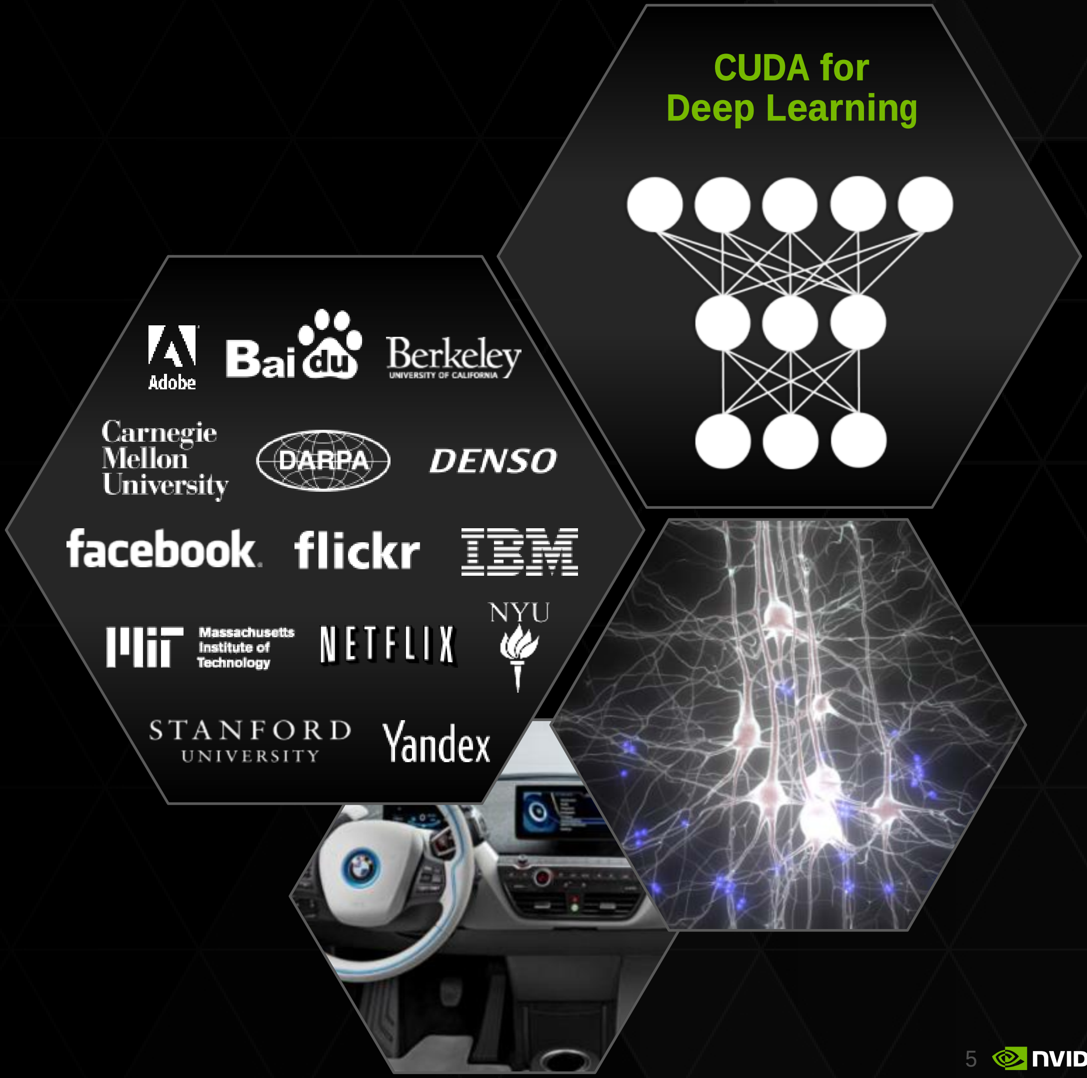
- Regression
- SVM
- Recommender systems

DEEP LEARNING & AI

“Machine Learning” is in some sense a rebranding of AI.

The focus is now on more specific, often 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 machine learning.

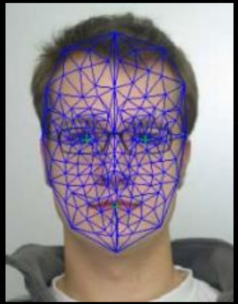


INDUSTRIAL USE CASES

...machine learning is pervasive

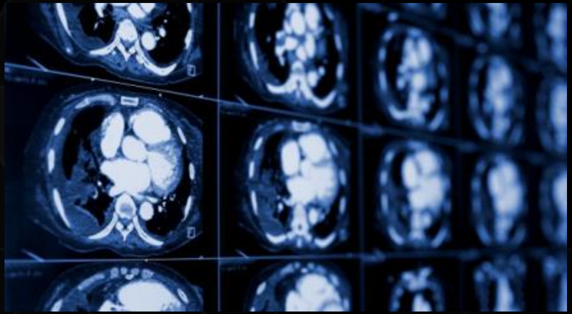
Social Media

Defense / Intelligence

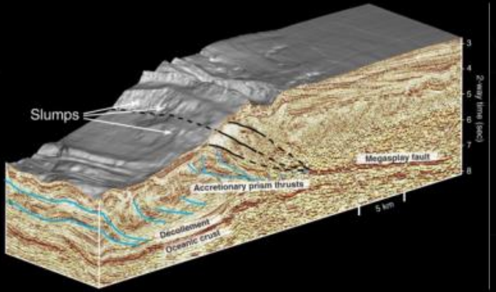


Consumer Electronics

Medical



Energy

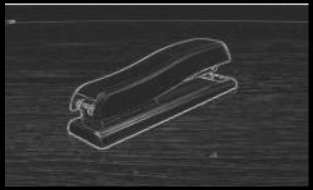


Media & Entertainment

TRADITIONAL ML – HAND TUNED FEATURES

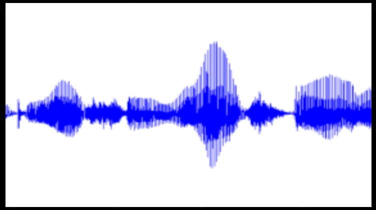
Images/video

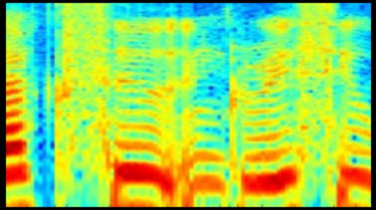

Image


Vision features


Detection

Audio

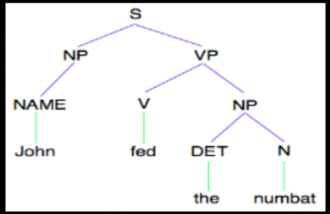

Audio


Audio features


Speaker ID

Text


Text



Text features

Text classification, Machine translation, Information retrieval, ...

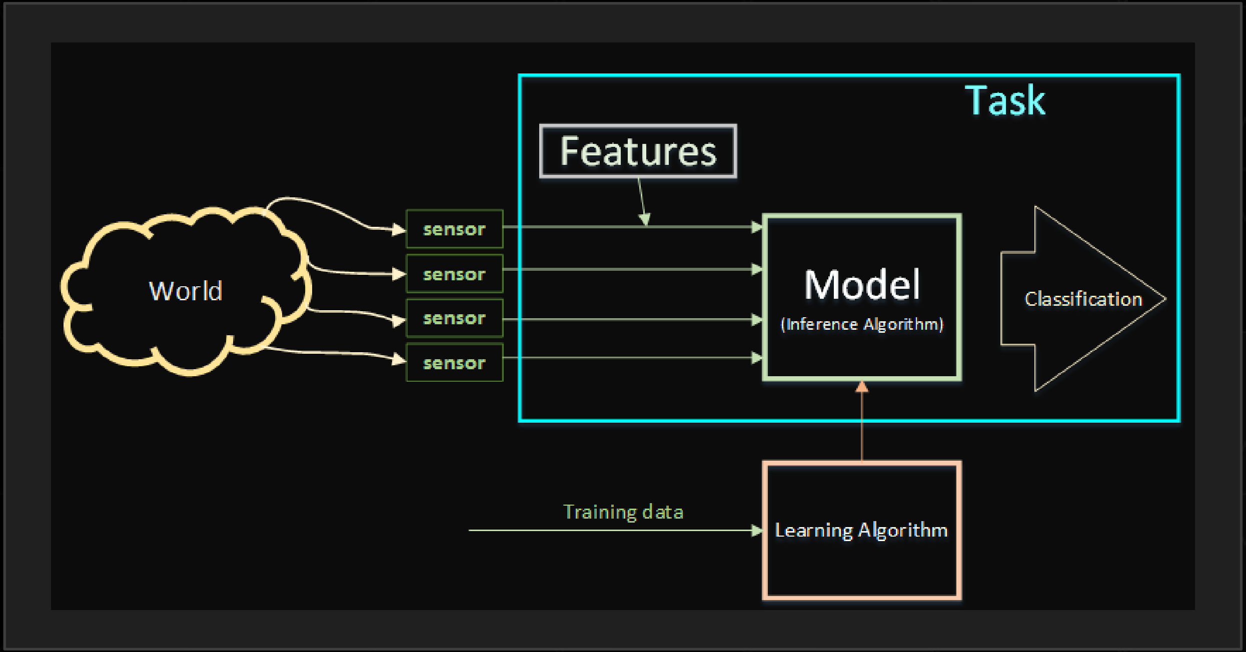
Slide courtesy of Andrew Ng, Stanford University

NVIDIA.

WHAT IS DEEP LEARNING?

Systems that learn to recognize objects that are important, without us telling the system explicitly what that object is ahead of time

- Key components
  - Task
  - Features
  - Model
  - Learning Algorithm



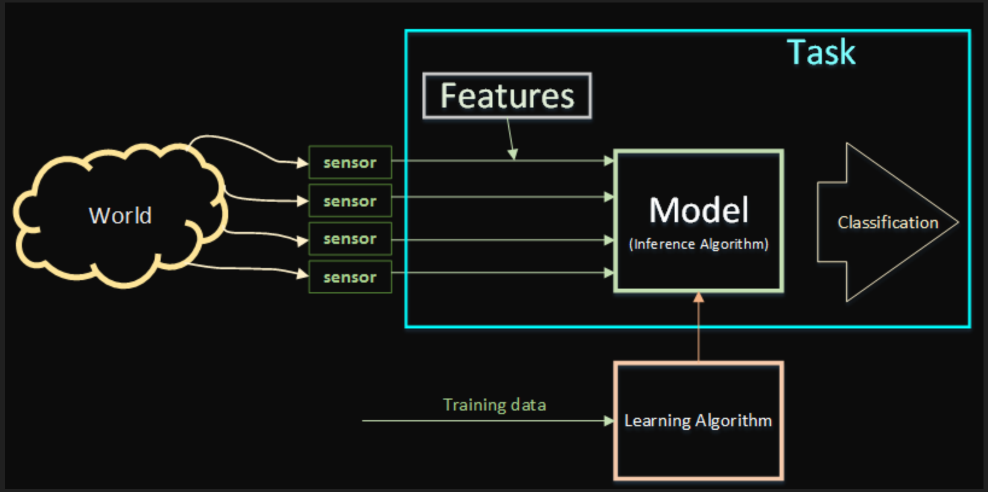
THE PROMISE OF MACHINE LEARNING

ML Systems Extract Value From Big Data

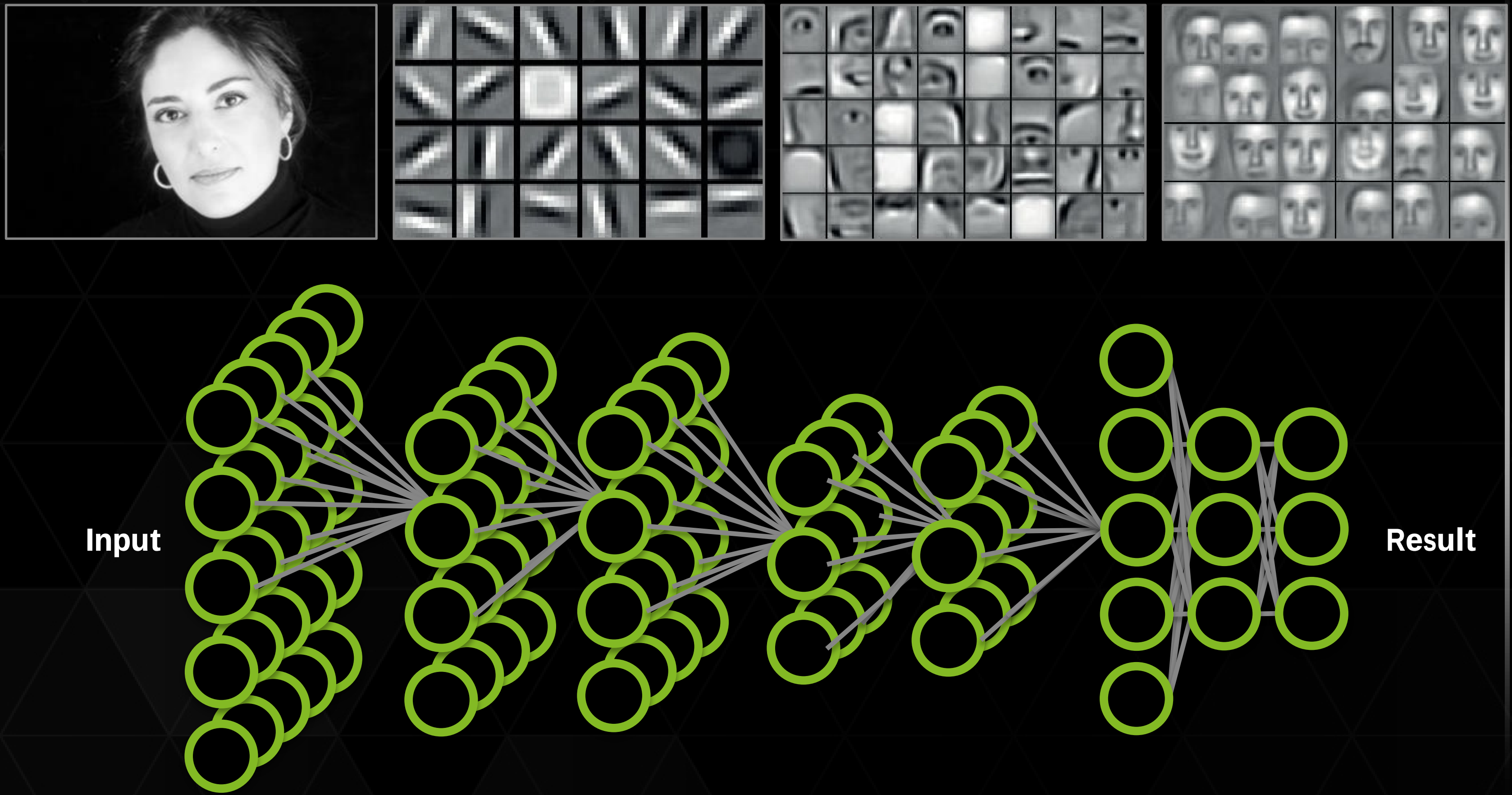
facebook | 350 millions images uploaded per day

Walmart | 2.5 Petabytes of customer data hourly

YouTube | 100 hours of video uploaded every minute



WHAT MAKES DEEP LEARNING DEEP?



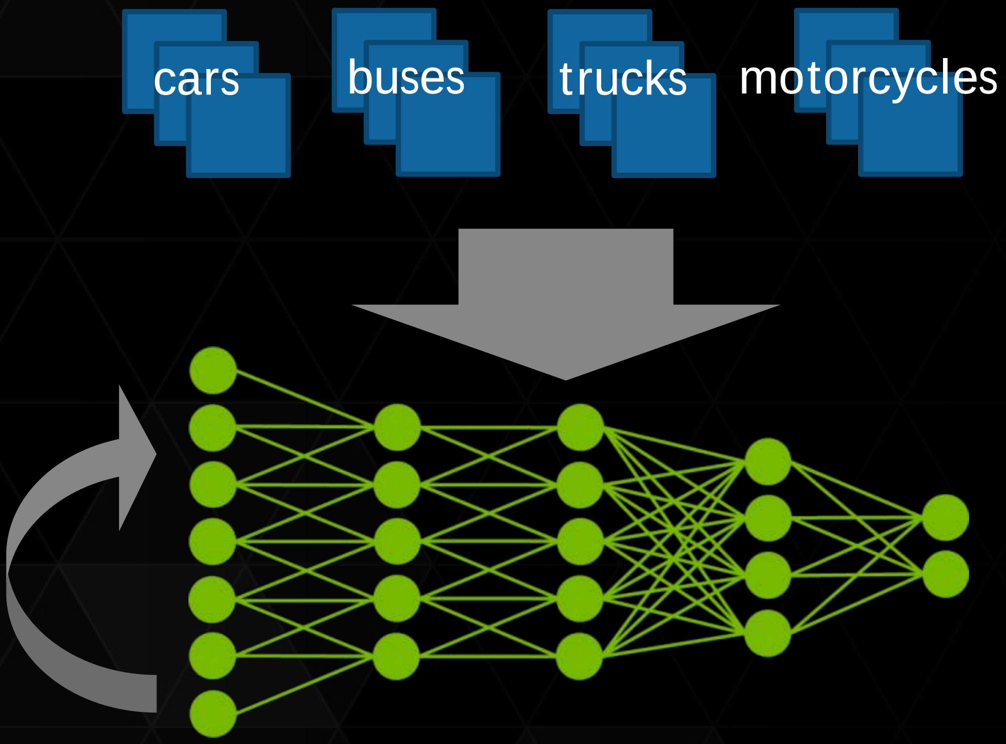
Today's Largest Networks

~10 layers
1B parameters
10M images
~30 Exaflops
~30 GPU days

Human brain has trillions of parameters - only 1,000 more.

IMAGE CLASSIFICATION WITH DNNS

Training



Inference

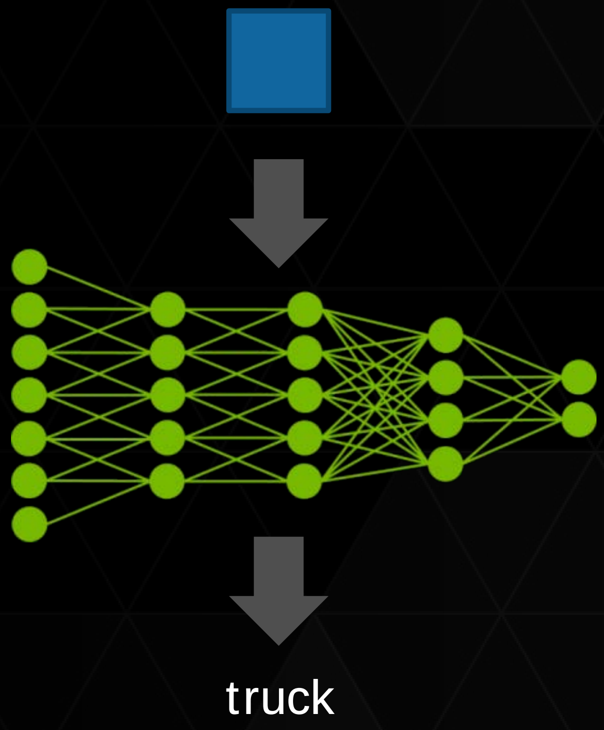
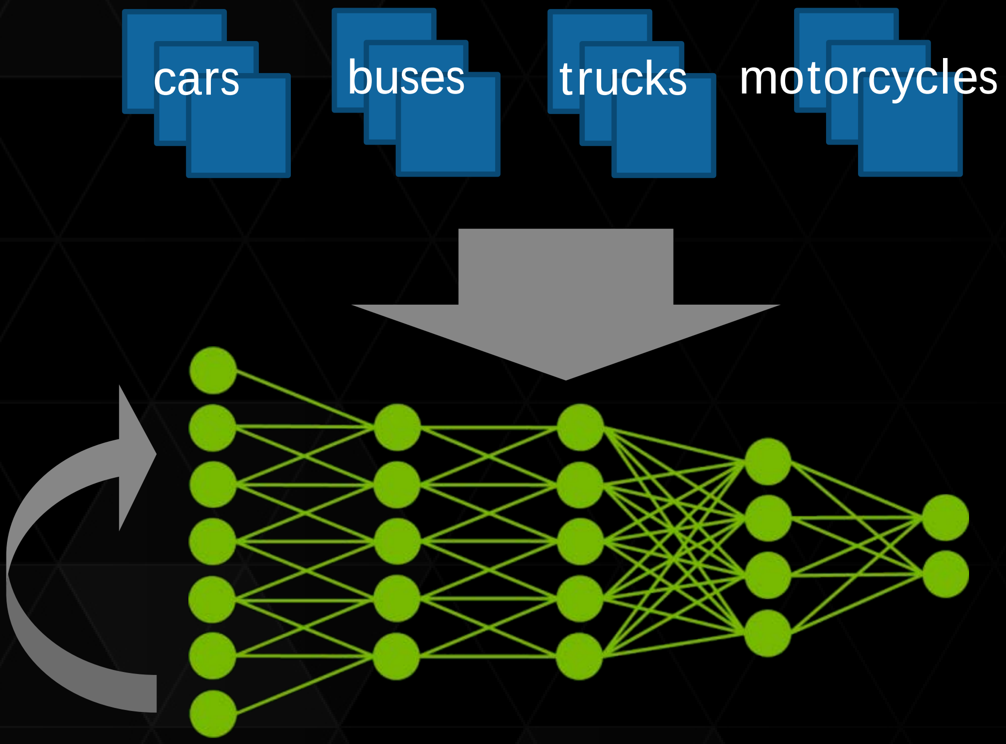


IMAGE CLASSIFICATION WITH DNNS

Training



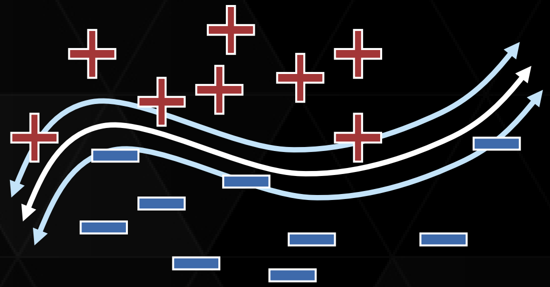
- Typical training run
  - Pick a DNN design
  - Input 100 million training images spanning 1,000 categories
  - *One week* of computation
- Test accuracy
  - If bad: modify DNN, fix training set or update training parameters

DEEP LEARNING ADVANTAGES

Deep Learning

- Don't have to figure out the features ahead of time.
- Use same neural net approach for many different problems.
- Fault tolerant.
- Scales well.

Support Vector Machine



Linear classifier

Regression

Decision Trees

Bayesian

Clustering

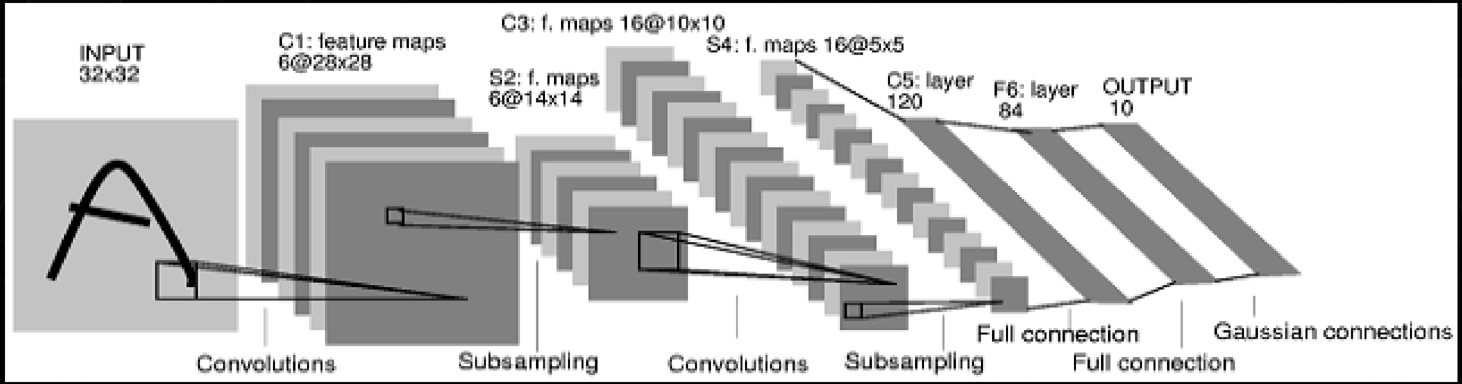
Association Rules

CONVOLUTIONAL NEURAL NETWORKS

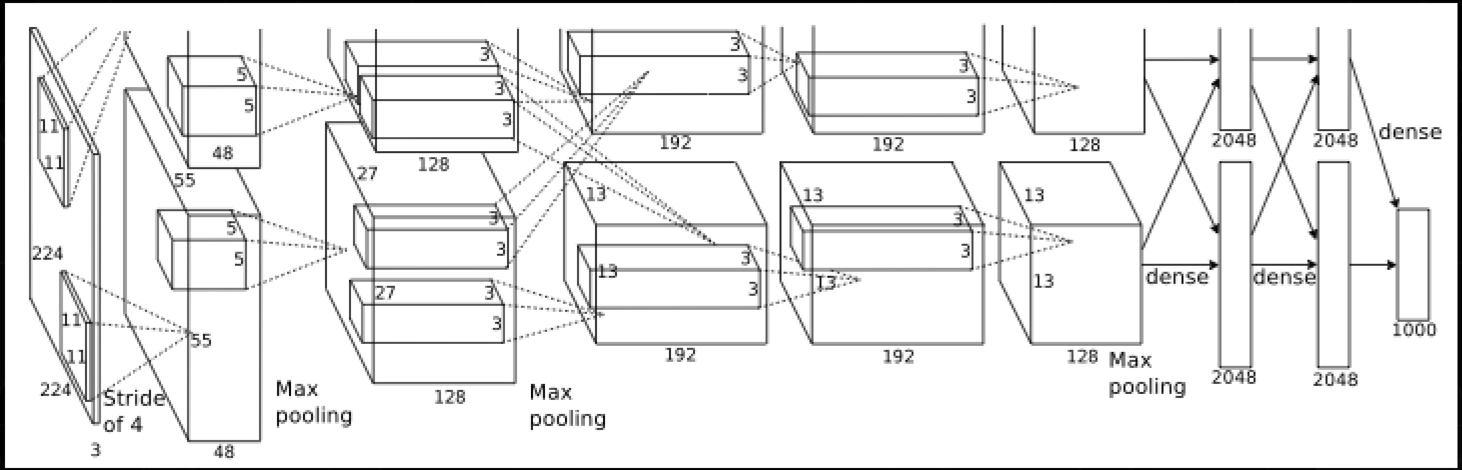
- Biologically inspired.
- Neuron only connected to a small region of neurons in layer below it called the *receptive field*.
- A given layer can have many convolutional filters/kernels. Each filter has the same weights across the whole layer.
- Bottom layers are convolutional, top layers are fully connected.
- Generally trained via supervised learning.
  - Supervised
  - Unsupervised
  - Reinforcement

...ideal system automatically switches modes...

CONVOLUTIONAL NETWORKS BREAKTHROUGH



Y. LeCun et al. 1989-1998 : Handwritten digit reading



A. Krizhevsky, G. Hinton et al. 2012 : Imagenet classification winner

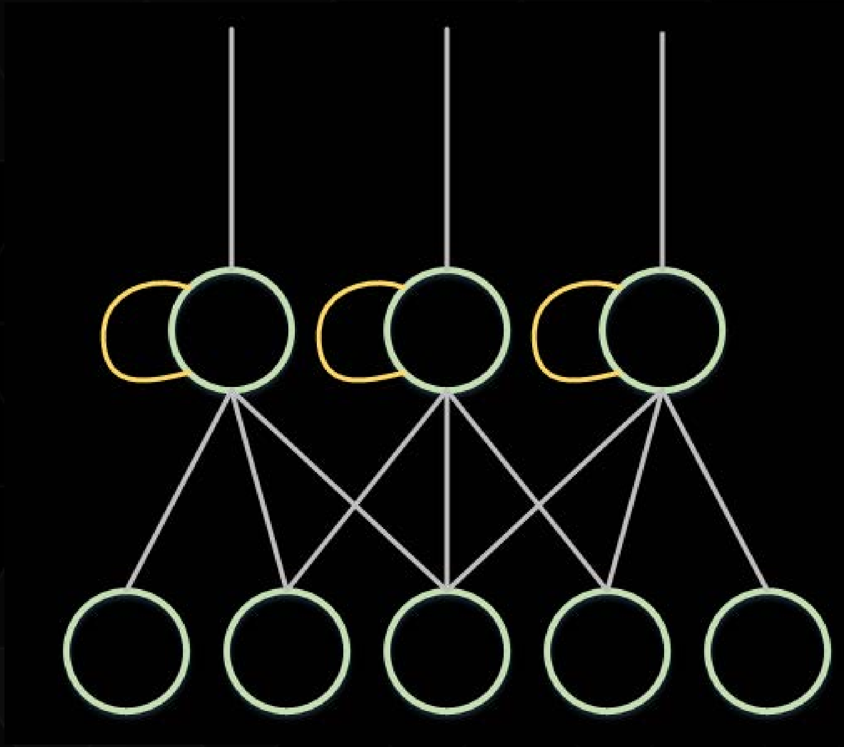
CNNS DOMINATE IN PERCEPTUAL TASKS

- 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

RECURRENT NEURAL NETWORK - RNN

AKA: “LSTM”

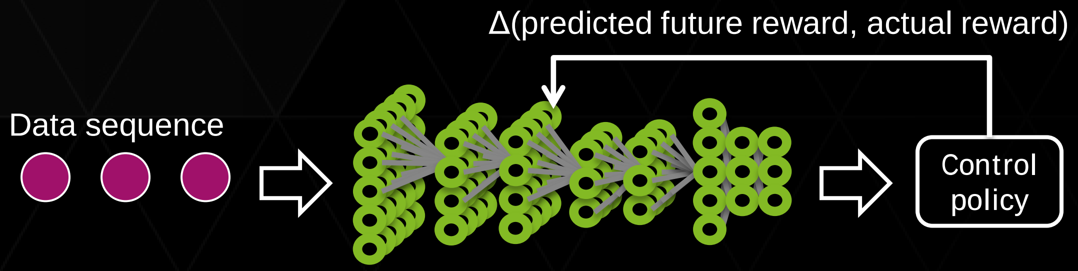


- Remembers prior state.
- Good for sequences.

- Predict next character given input text.
- Translate sentence between languages.
- Generate a caption for an image.

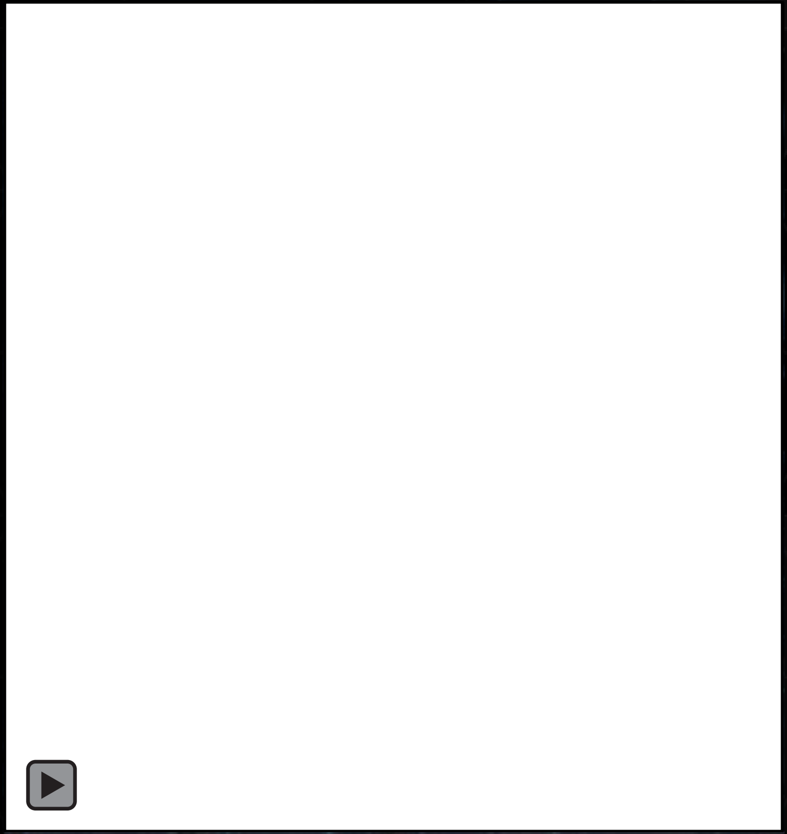
SENSOR / PLATFORM CONTROL

Reinforcement learning



Applications

- Sensor tasking
- Autonomous vehicle navigation



[11] Google DeepMind in Nature

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 ML Techniques

Deep Neural Networks

Compute Density

GPUs

ML systems extract value from Big Data

GEOINT ANALYSIS WORKFLOW

TODAY

Big Data

NUMBERS

IMAGES VIDEOS
SOUNDS TEXT

BOTTLENECK

Metadata filters

Noisy content

Human perception

Near perfect perception

Mission focused analysis

VISION

Big Data

NUMBERS

IMAGES VIDEOS
SOUNDS TEXT

DL based machine perception

Near human level perception

Content based filters

Mission relevant content

Mission focused analysis

GPUs and Deep Learning

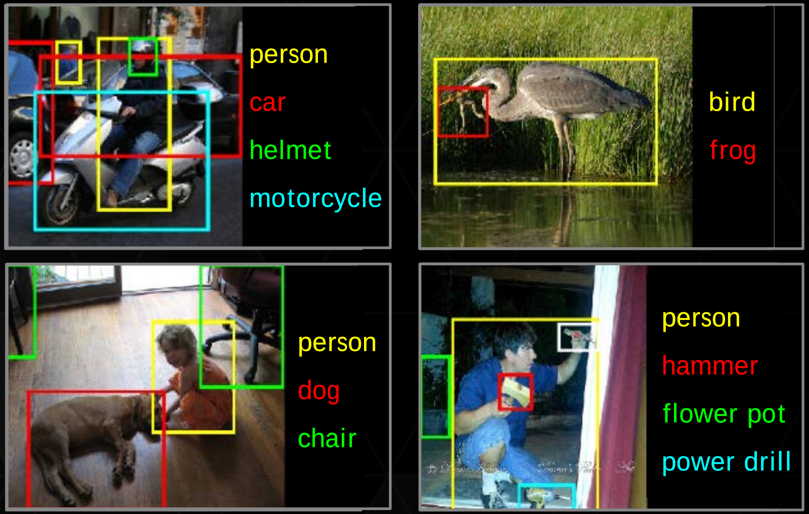
GPUs — THE PLATFORM FOR DEEP LEARNING

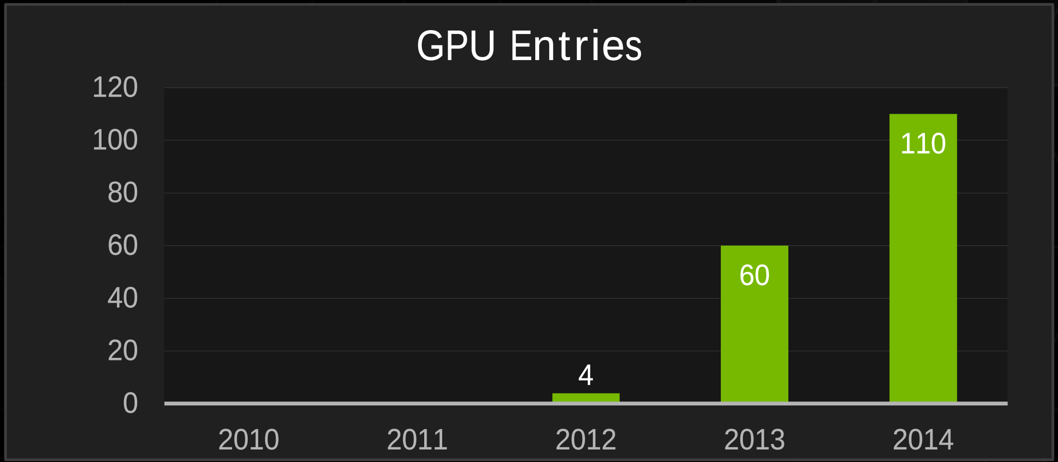
Image Recognition Challenge

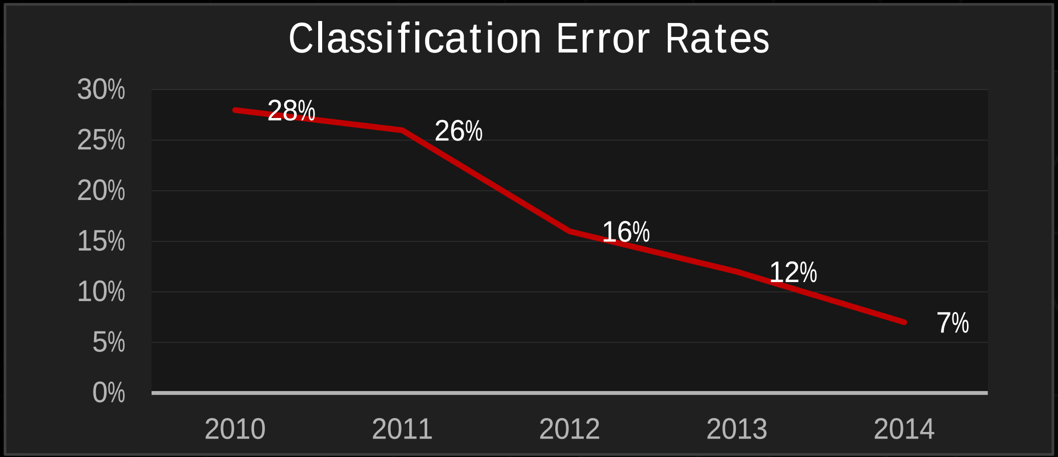
1.2M training images • 1000 object categories

Hosted by

IMAGENET







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

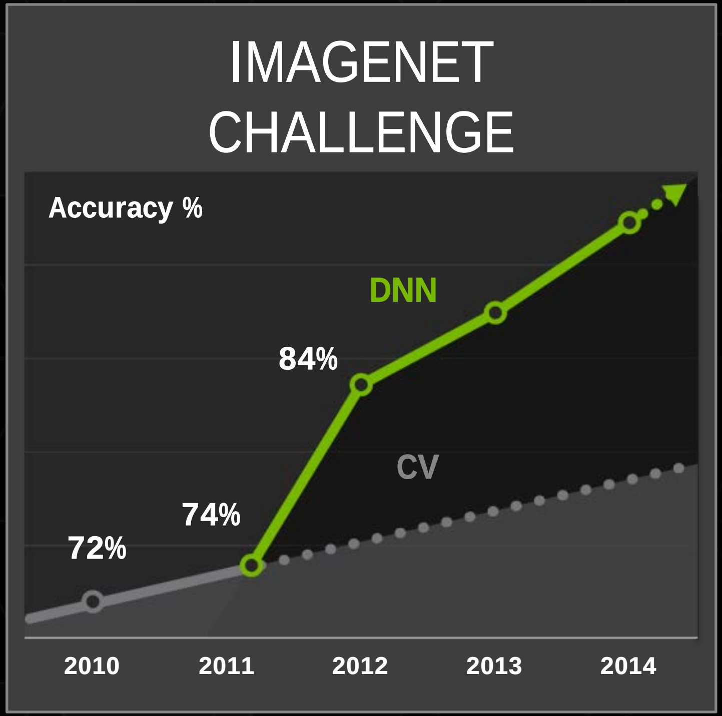
1,000 CPU Servers
2,000 CPUs • 16,000 cores

600 kWatts
$5,000,000

STANFORD AI LAB

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

4 kWatts
$33,000



“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

GOOGLE KEYNOTE AT GTC 2015

The theme of deep learning carried through our guest keynotes. Jeff Dean, senior fellow at Google, described how the company is using GPU-powered deep neural networks to bring greater levels of intelligence to image, text, and speech recognition. He also highlighted work done by the recently acquired Deep Mind. Using Atari video games, the researchers trained a network to not just classify, but take actions in an environment. Ultimately, the network beat a series of games and the work earned the cover of *Nature* magazine.

"We love GPU cards. We just use a lot of them."

— Jeff Dean, Google

GOOGLE USES DEEP LEARNING FOR UNDERSTANDING

What are all these numbers?

What are all these words?

Large-Scale Deep Learning For Building Intelligent Computer Systems, Jeff Dean (Google), http://www.ustream.tv/recorded/60071572

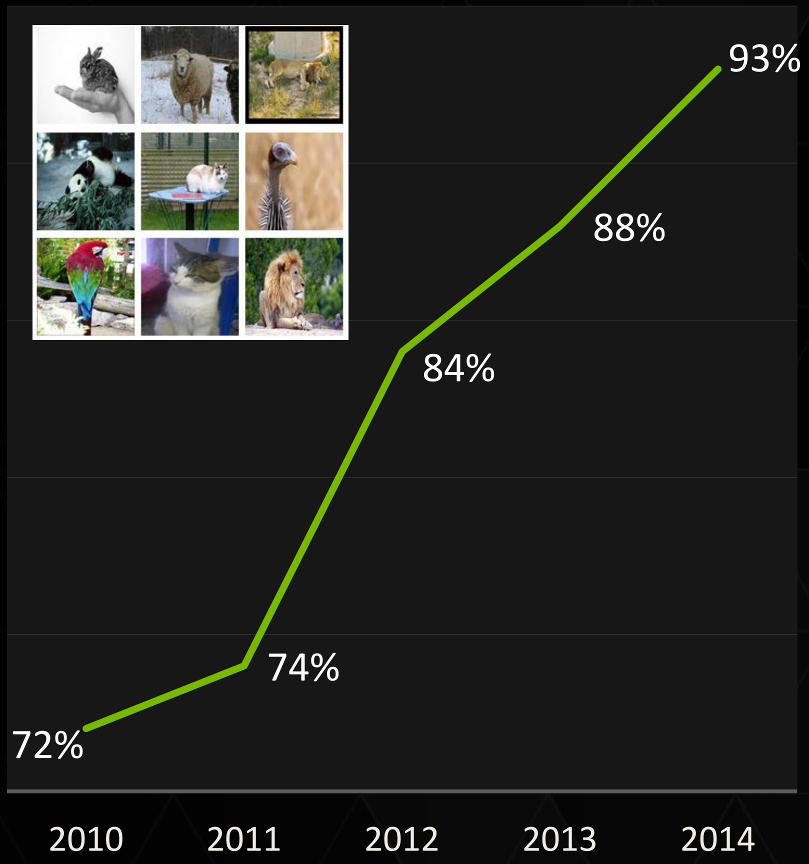
WHY ARE GPUs GOOD FOR DEEP LEARNING?

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

GPUs deliver --

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

ImageNet Challenge Accuracy



GPU ACCELERATION

Training A Deep, Convolutional Neural Network

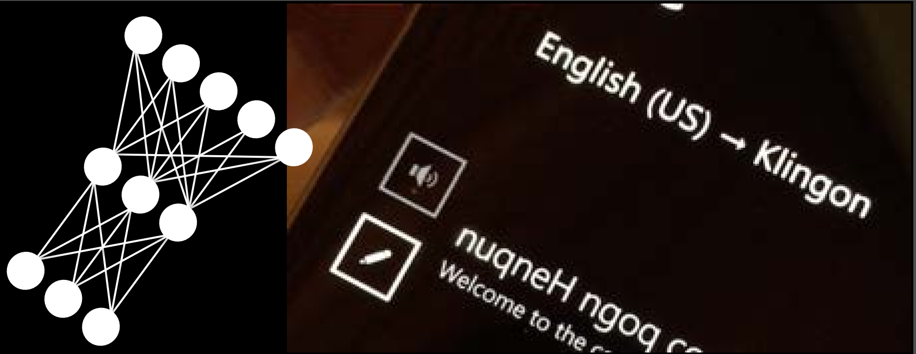
Batch Size	Training Time CPU	Training Time GPU	GPU Speed Up
64 images	64 s	7.5 s	8.5X
128 images	124 s	14.5 s	8.5X
256 images	257 s	28.5 s	9.0X

- ILSVRC12 winning model: "Supervision"
- 7 layers
- 5 convolutional layers + 2 fully-connected
- ReLU, pooling, drop-out, response normalization
- Implemented with Caffe

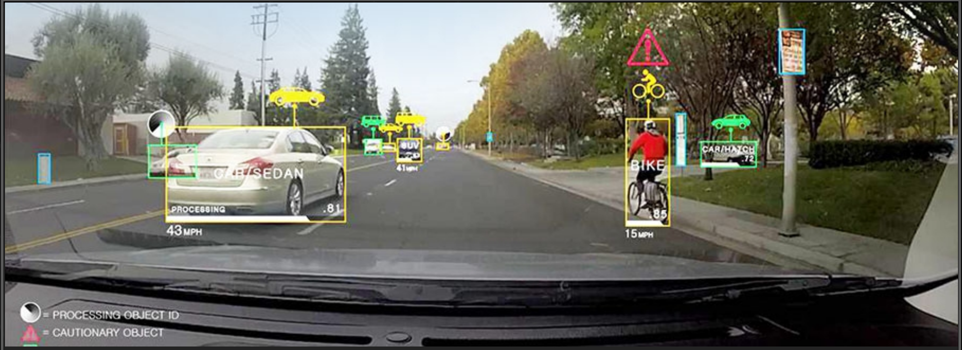
- Dual 10-core Ivy Bridge CPUs
- 1 Tesla K40 GPU
- CPU times utilized Intel MKL BLAS library
- GPU acceleration from CUDA matrix libraries (cuBLAS)

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

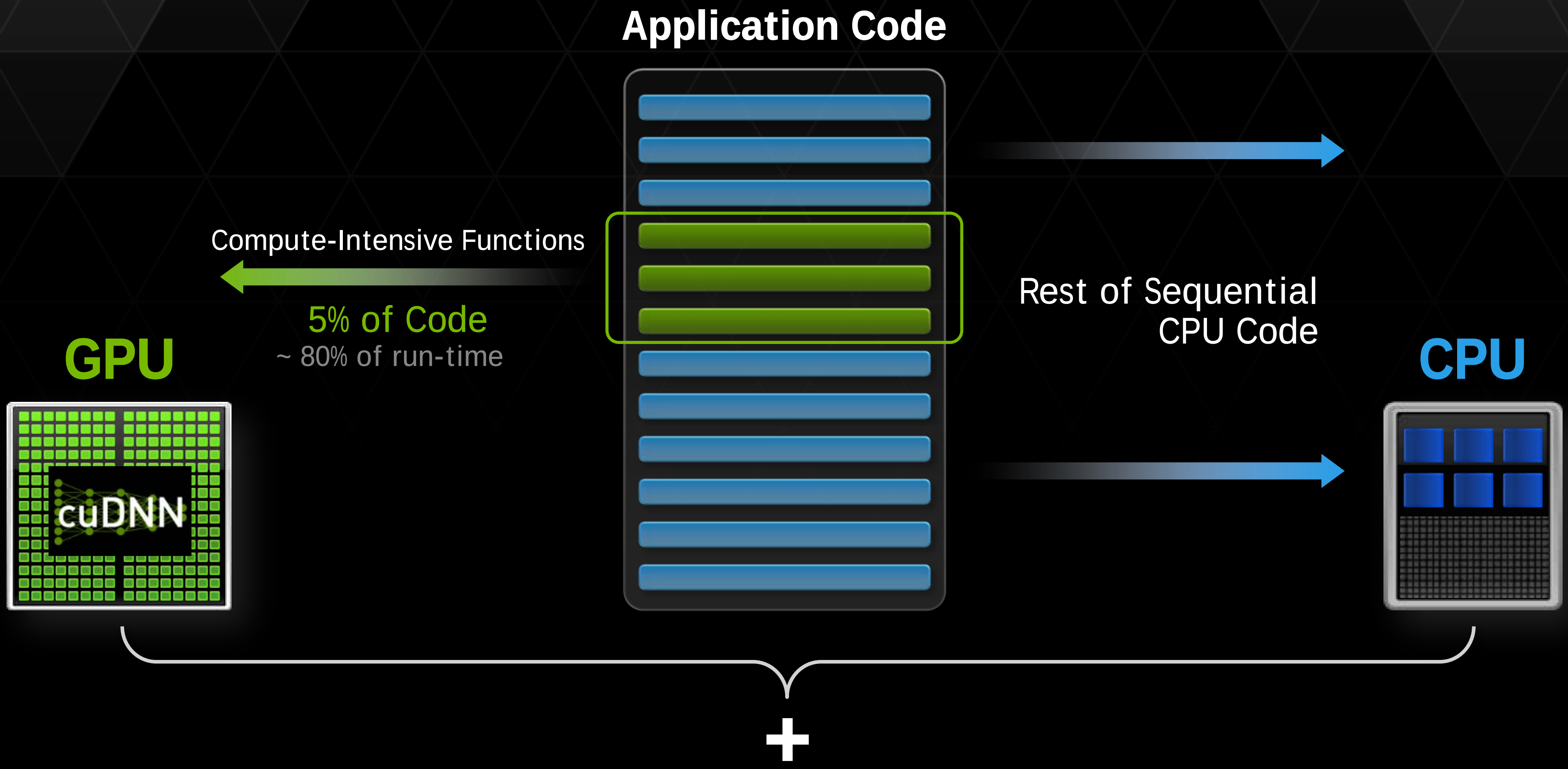
GPU-ACCELERATED DEEP LEARNING FRAMEWORKS

	CAFFE	TORCH	THEANO	CUDA-CONVNET2	KALDI
Domain	Deep Learning Framework	Scientific Computing Framework	Math Expression Compiler	Deep Learning Application	Speech Recognition Toolkit
cuDNN	2.0	2.0	2.0	--	--
Multi-GPU	In Progress	In Progress	In Progress	✓	✓(nnet2)
Multi-CPU	✗	✗	✗	✗	✓(nnet2)
License	BSD-2	GPL	BSD	Apache 2.0	Apache 2.0
Interface(s)	Text-based definition files, Python, MATLAB	Python, Lua, MATLAB	Python	C++	C++, Shell scripts
Embedded (TK1)	✓	✓	✗	✗	✗

http://developer.nvidia.com/deeplearning

cuDNN

HOW GPU ACCELERATION WORKS



WHAT IS cuDNN?

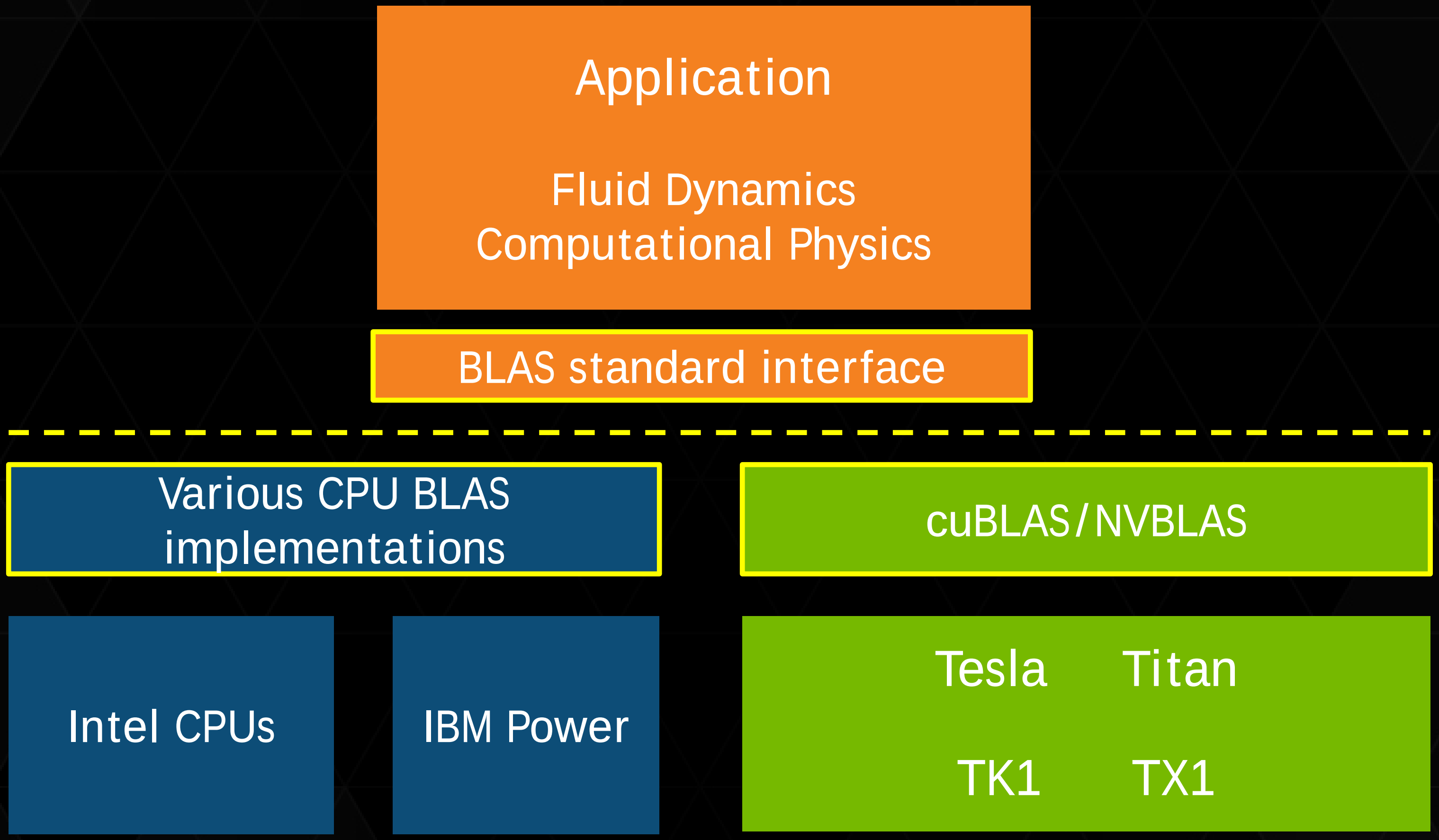
cuDNN is a library of primitives for deep learning

Applications

Programming Languages	Libraries cuDNN	OpenACC Directives
Maximum Flexibility	"Drop-in" Acceleration	Easily Accelerate Applications

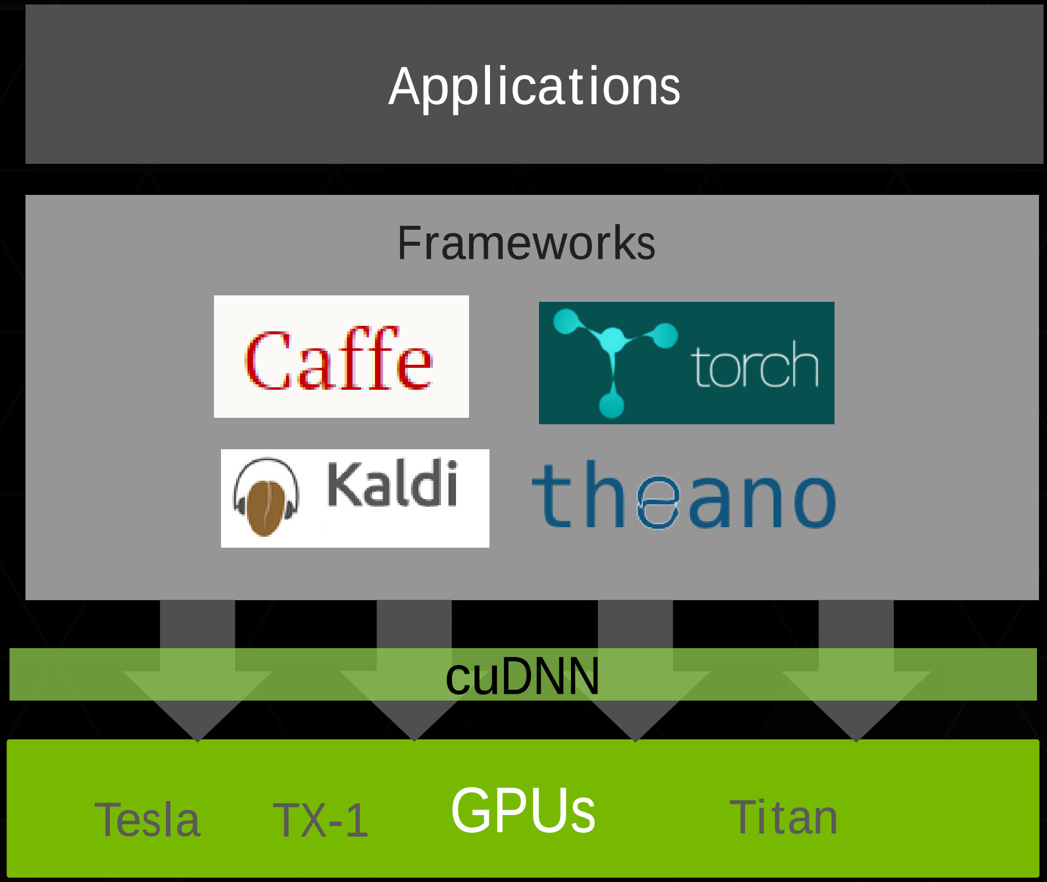
ANALOGY TO HPC

cuDNN is a library of primitives for deep learning



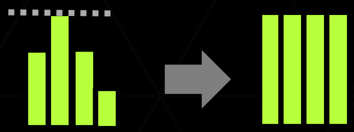
DEEP LEARNING WITH cuDNN

cuDNN is a library of primitives for deep learning

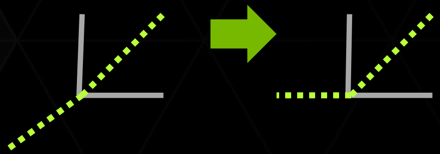


cuDNN ROUTINES

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



- Activation - Pointwise non-linear function



CONVOLUTIONS – THE MAIN WORKLOAD

- Very compute intensive, but with a large parameter space

1 Minibatch Size
2 Input feature maps
3 Image Height
4 Image Width
5 Output feature maps
6 Kernel Height
7 Kernel Width
8 Top zero padding
9 Side zero padding
10 Vertical stride
11 Horizontal stride

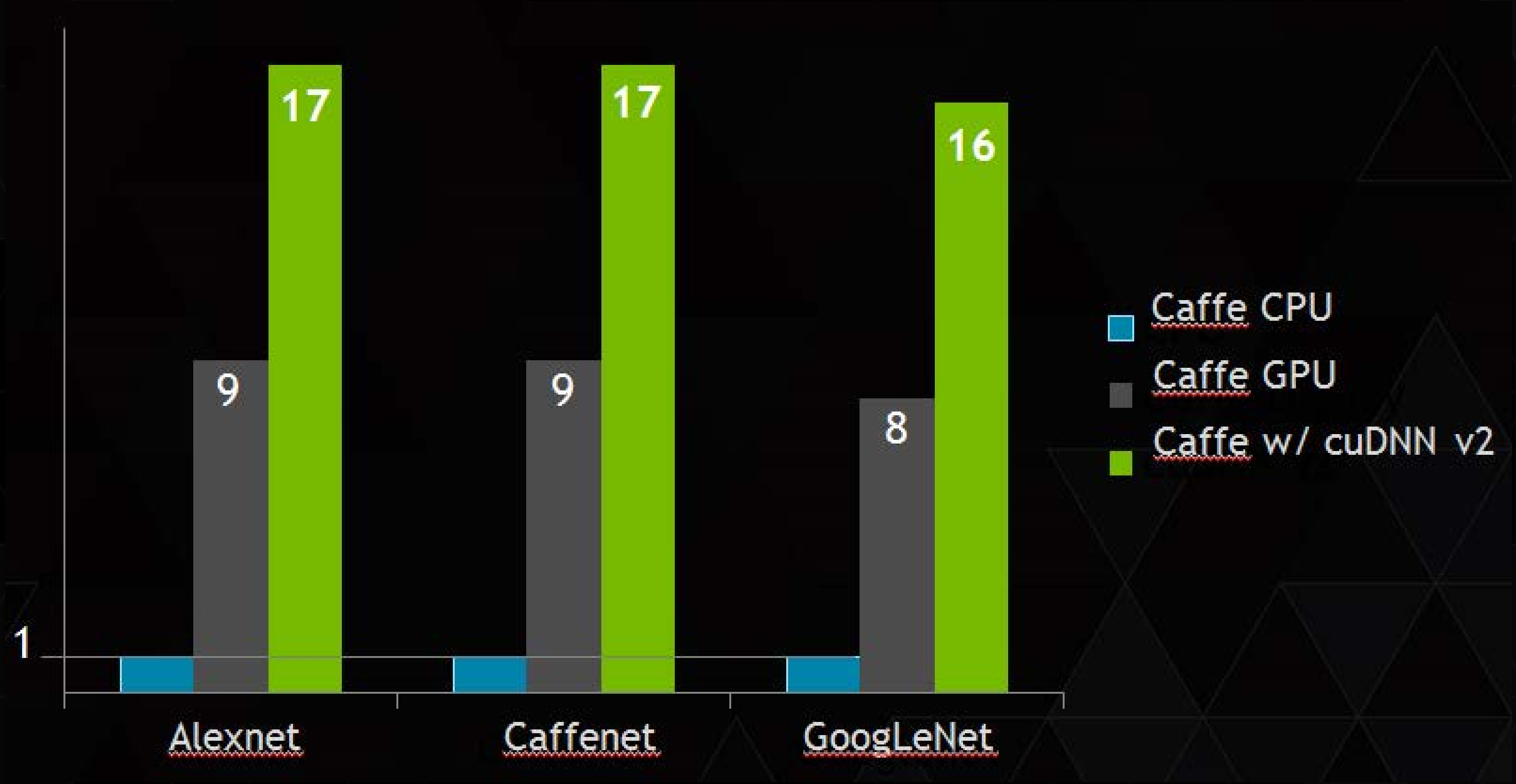
- Layout and configuration variations

- Other cuDNN routines have straightforward implementations

EXAMPLE — OVERFEAT LAYER 1

```
/* Allocate memory for Filter and ImageBatch, fill with data */
cudaMalloc( &ImageInBatch , ... );
cudaMalloc( &Filter , ... );
...

/* Set descriptors */
cudnnSetTensor4dDescriptor( InputDesc, CUDNN_TENSOR_NCHW, 128, 96, 221, 221);
cudnnSetFilterDescriptor( FilterDesc, 256, 96, 7, 7 );
cudnnSetConvolutionDescriptor( convDesc, InputDesc, FilterDesc,
    pad_x, pad_y, 2, 2, 1, 1, CUDNN_CONVOLUTION);

/* query output layout */
cudnnGetOutputTensor4dDim(convDesc, CUDNN_CONVOLUTION_FWD, &n_out, &c_out, &h_out, &w_out);

/* Set and allocate output tensor descriptor */
cudnnSetTensor4dDescriptor( &OutputDesc, CUDNN_TENSOR_NCHW, n_out, c_out, h_out, w_out);
cudaMalloc(&ImageBatchOut, n_out * c_out * h_out * w_out * sizeof(float));

/* launch convolution on GPU */
cudnnConvolutionForward( handle, InputDesc, ImageInBatch, FilterDesc, Filter, convDesc,
    OutputDesc, ImageBatchOut, CUDNN_RESULT_NO_ACCUMULATE);
```

CUDNN V2 - PERFORMANCE



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

GPU is NVIDIA Titan X

CUDNN EASY TO ENABLE

Caffe

- Install cuDNN on your system
- Download CAFFE
- In CAFFE `Makefile.config`
  - uncomment `USE_CUDNN := 1`
- Install CAFFE as usual
- Use CAFFE as usual.

torch

- Install cuDNN on your system
- Install Torch as usual
- Install `cudnn.torch` module
- Use `cudnn` module in Torch instead of regular nn module.
- `cudnn` module is API compatable with standard nn module.
  Replace `nn` with `cudnn`

CUDA 6.5 or newer required

DiGiTS

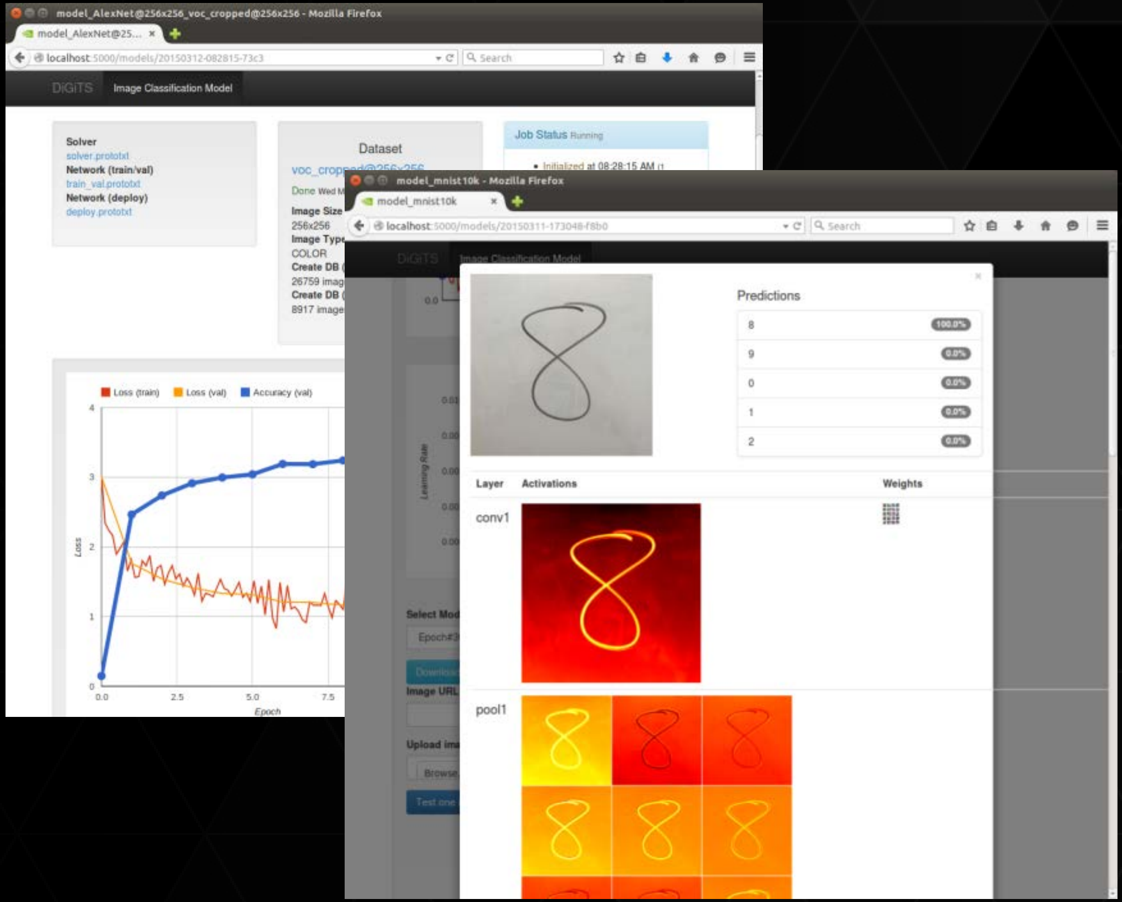
Deep Learning GPU Training System

DIGITS

Interactive Deep Learning GPU Training System

Data Scientists & Researchers:

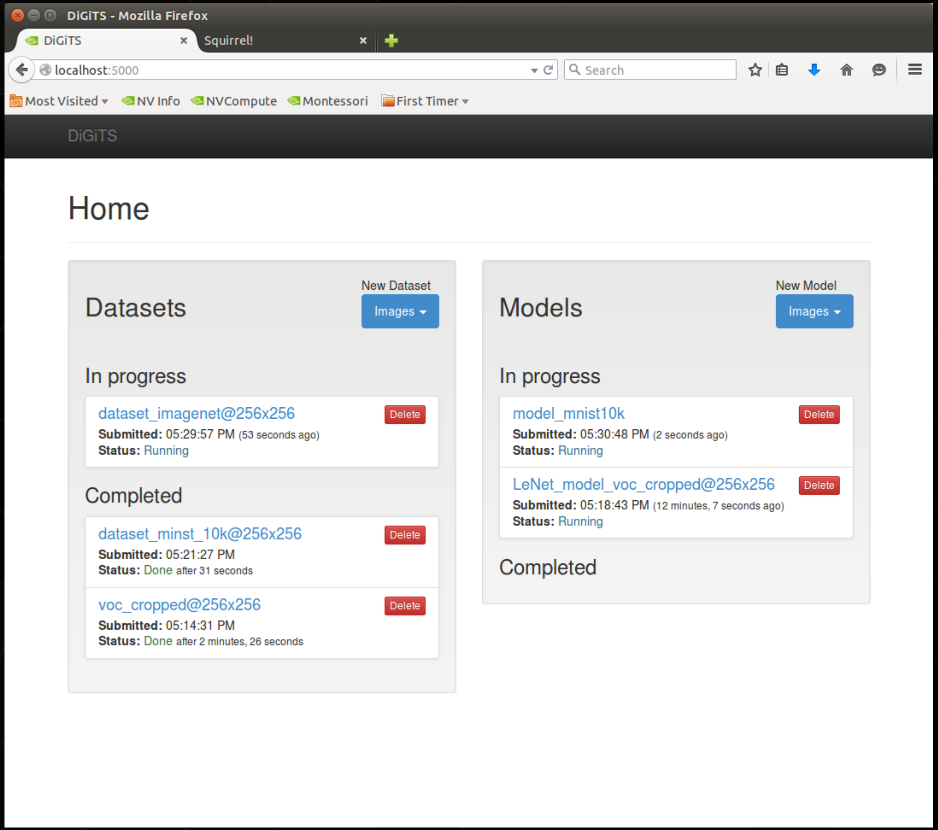
- Quickly design the best deep neural network (DNN) for your data
- Visually monitor DNN training quality in real-time
- Manage training of many DNNs in parallel on multi-GPU systems



developer.nvidia.com/digits

DIGITS

Deep Learning GPU Training System



- Available at developer.nvidia.com/digits
- Free to use
- v1.0 supports classification on images
- Future versions: More problem types and data formats (video, speech)

(Also available on Github for advanced developers)

HOW DO YOU GET DIGITS

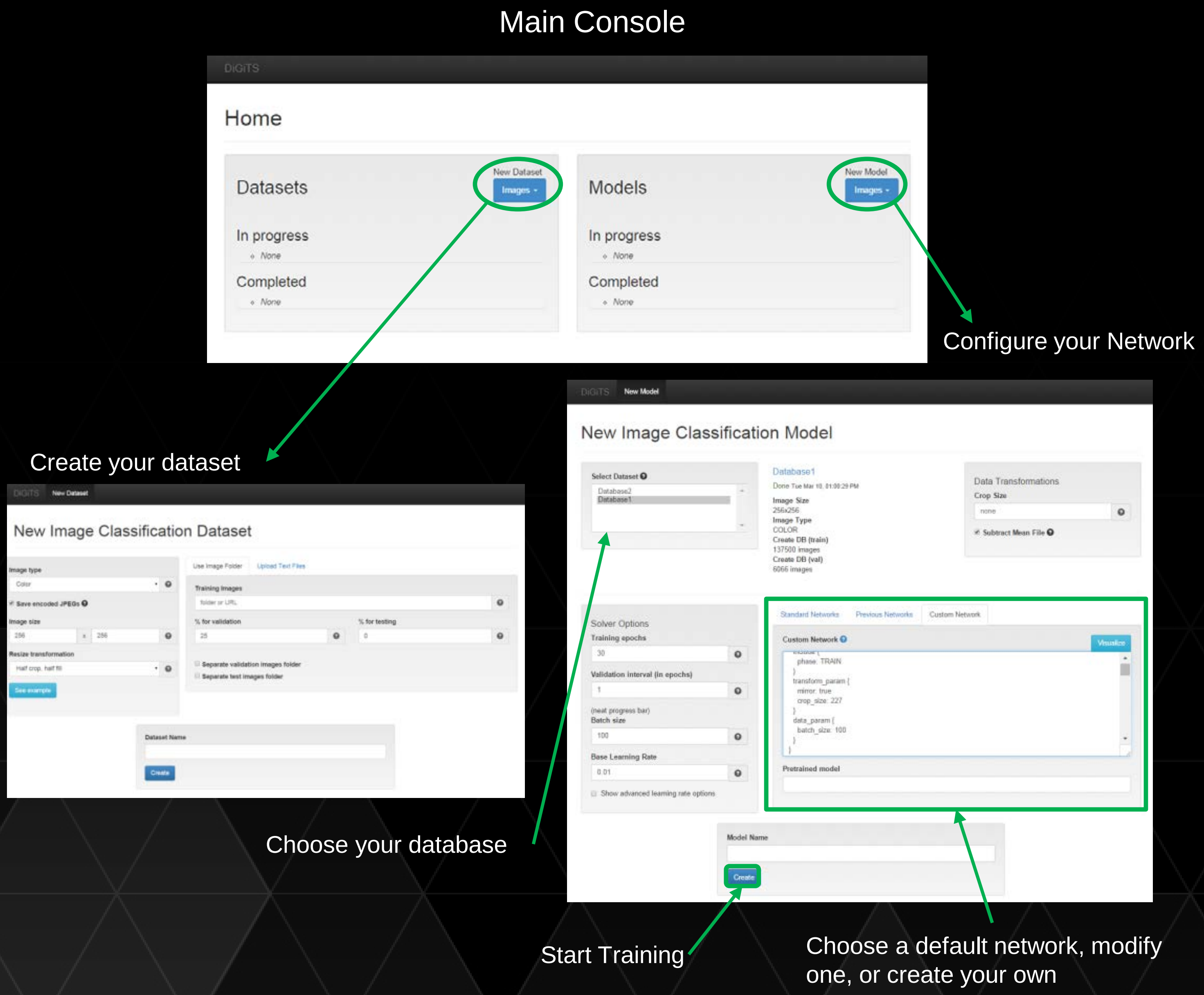
- Two options
  - Download DIGITS from developer.nvidia.com/digits
  - Download the source code from GitHub.com – www.github.com/nvidia/digits
  - Launch with one command "python digits-devserver"

DIGITS Workflow

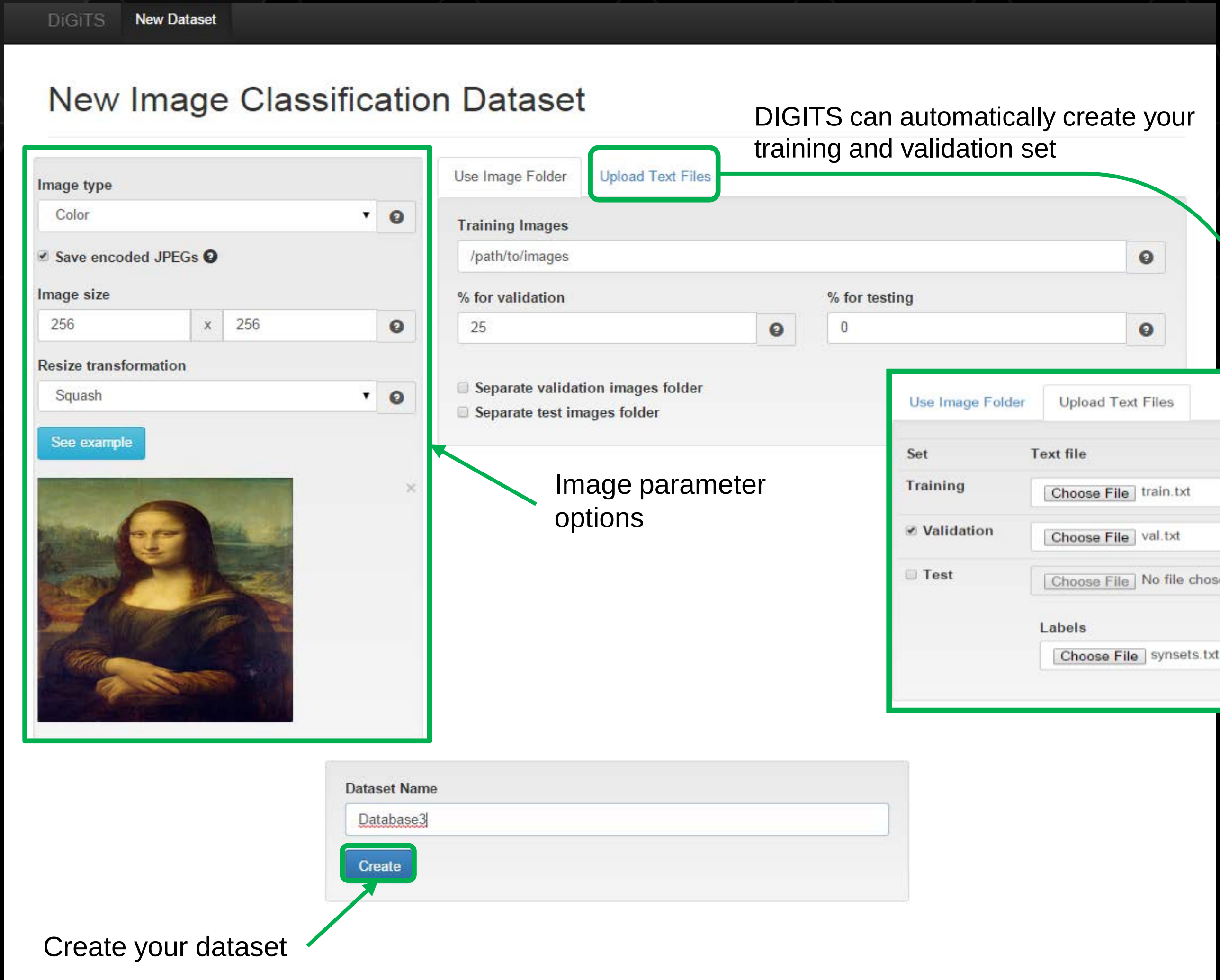
Create your database

Configure your model

Start training



CREATE THE DATABASE



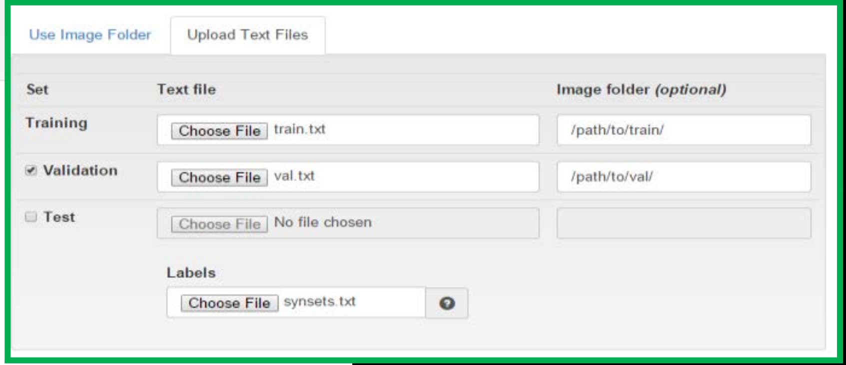
DIGITS can automatically create your training and validation set

Image parameter options

Create your dataset

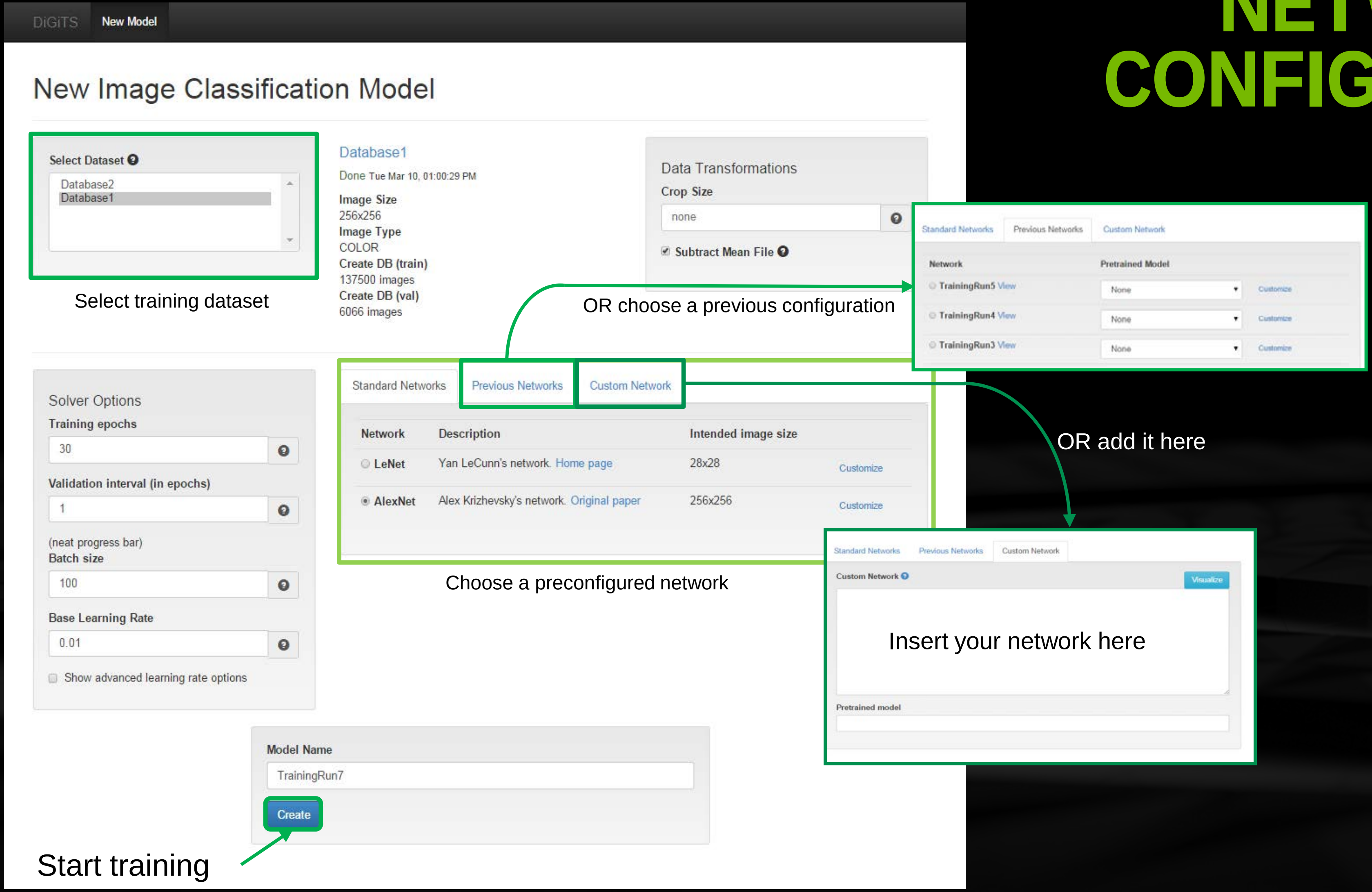
OR

Insert the path to your train and validation set



OR use a URL list

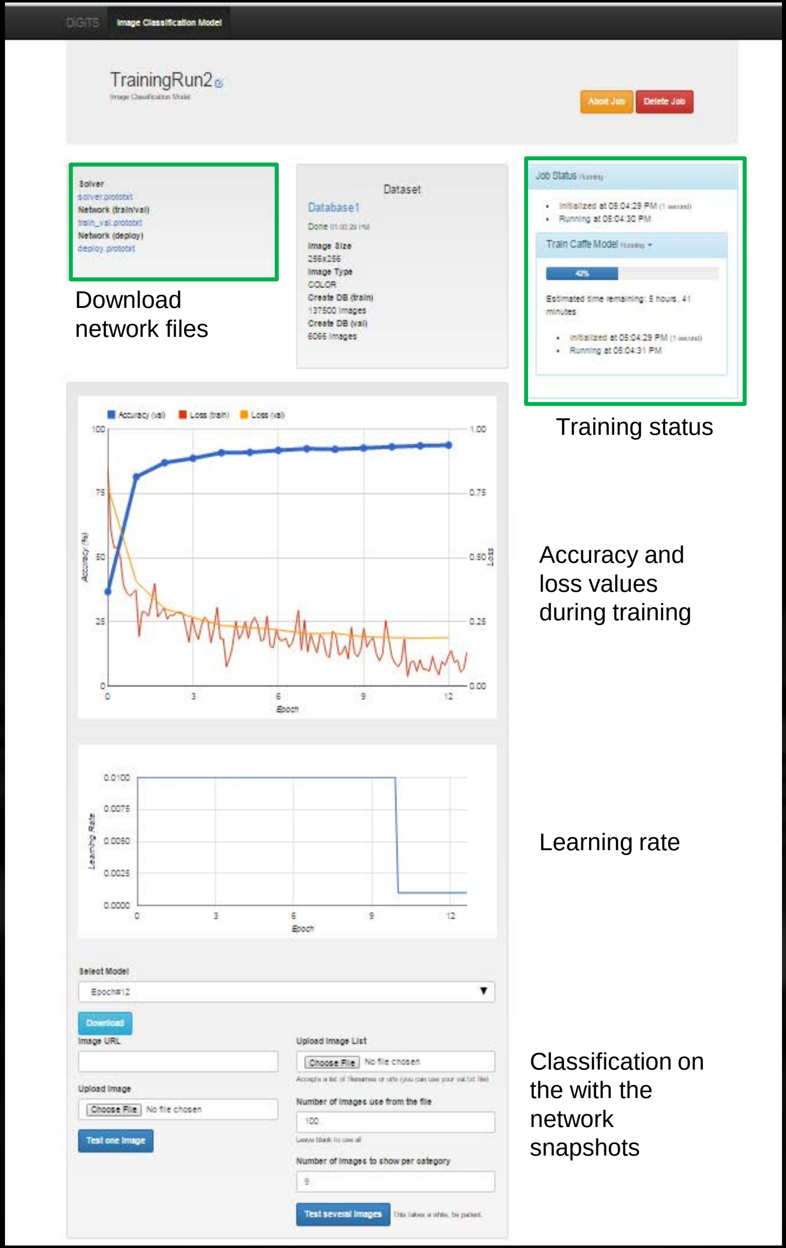
NETWORK CONFIGURATION



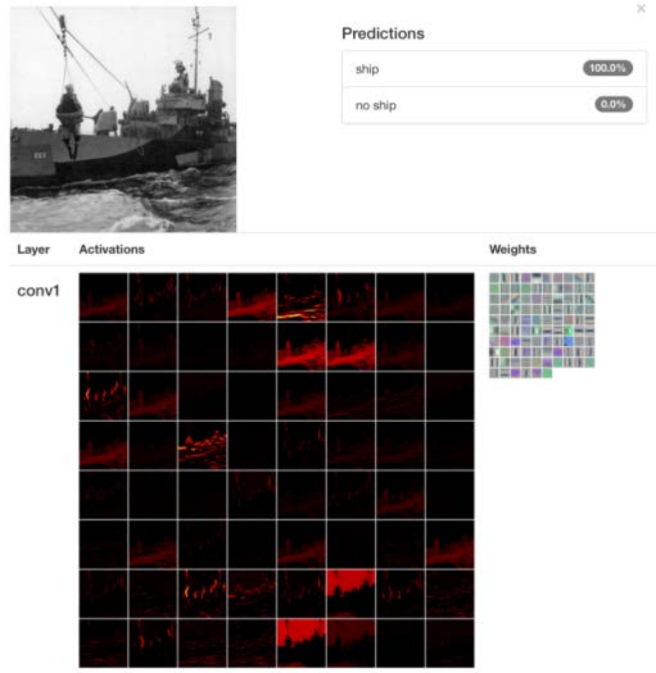
DIGITS

Visualize DNN performance in real time

Compare networks



Classification



Neural Network Motivation

NEURAL NETWORK MOTIVATION

"One learning algorithm" hypothesis

Auditory & Somatosensory cortex can learn to see.

We can connect any sensor to any part of the brain, and the brain figures it out.

See with your tongue

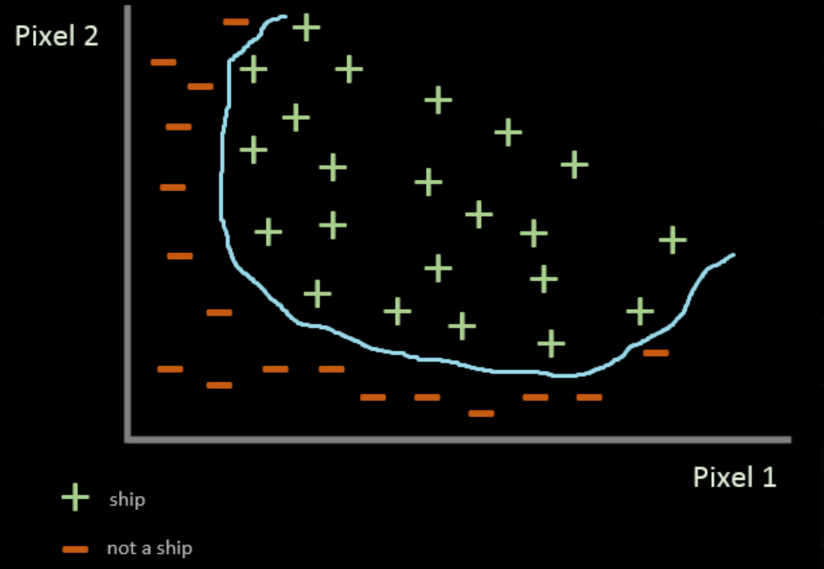
Adding sense of direction

Echolocation

NEURAL NETS SCALE EASIER

Why use neural nets? Consider computer vision...

When the decision space is non-linear, and the number of features is very large.

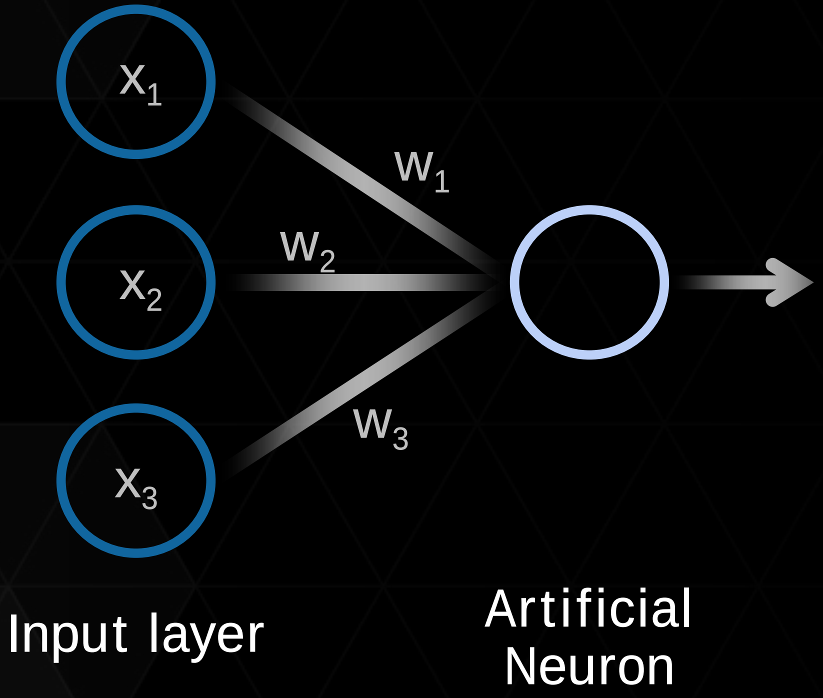


256 x 256 image = 65536 pixels (x3 for color)

Quadratic features (x1 * x2) - over 4 billion!

WHAT'S IN A NEURON?

Artificial neuron is modeled as a "Logistic Unit".

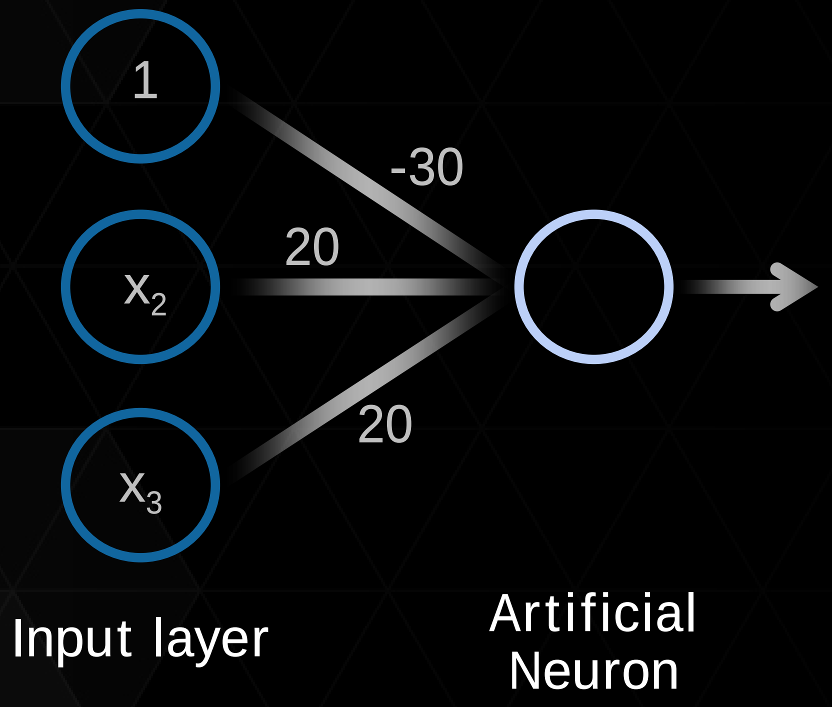


$$z = x_1 w_1 + x_2 w_2 + x_3 w_3$$

$$\text{Activation} = \frac{1}{1 + e^{-z}}$$

NEURONS CAN COMPUTE

Artificial neuron can compute logical operations like AND OR



x_2	x_3	Activation
0	0	0
1	0	0
0	1	0
1	1	1

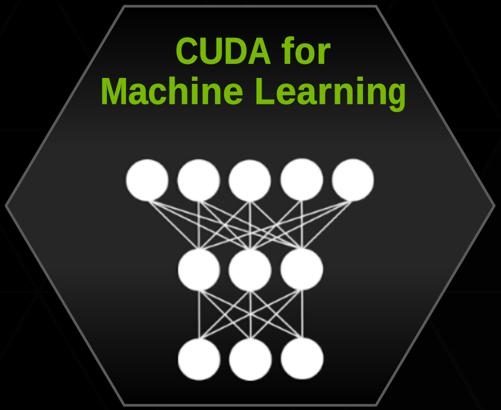
DEEP VERSUS TWO-LAYER NETWORKS

Theory says two fully-connected layers can solve any problem.

G. Cybenko - Approximation by Superpositions of a Sigmoidal Function, *Mathematics of Control, Signals and Systems*, 1989

"In theory, there is no difference between theory and practice.
In practice, there is."

- More memory versus more time.
- Few functions can be computed in two layers without an exponentially large look-up table.
- Using more than 2 steps can reduce memory by an exponential factor.

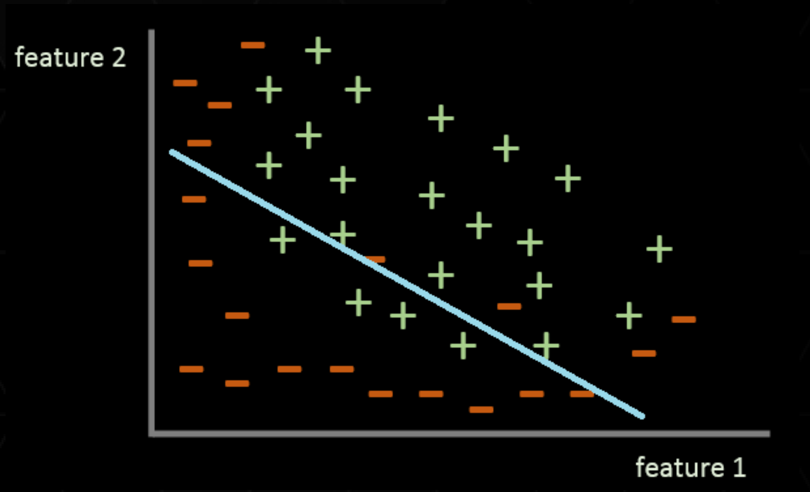


Working with Deep Neural Networks

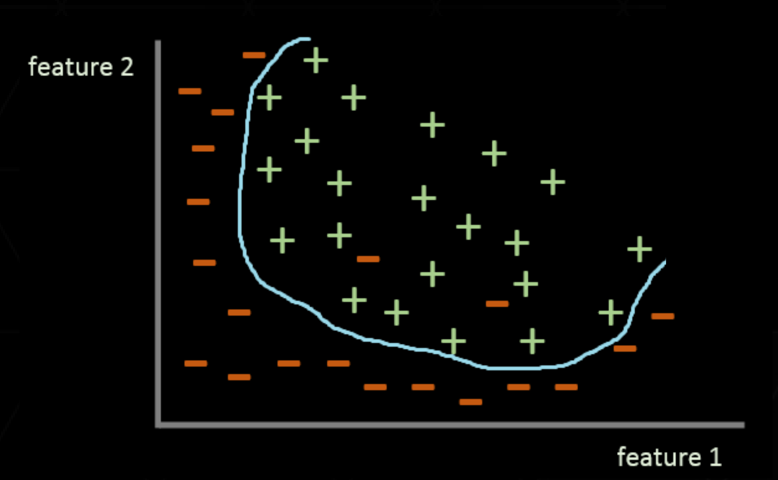
OVERFITTING & UNDERFITTING

Important terminology...

High Bias



Underfitting



Just right

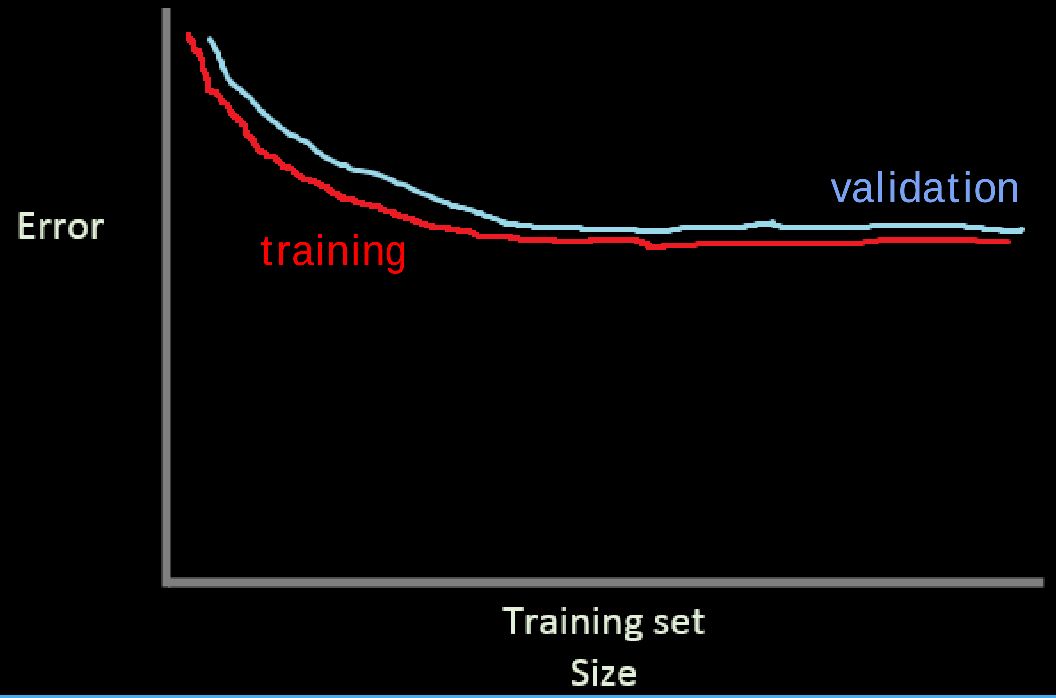
High Variance

Overfitting

LEARNING CURVE

Underfitting example

High Bias



Actions

- Increase size of neural network.
- Reduce "lambda" / "weight decay" (regularization)

LEARNING CURVE

Overfitting example

High Variance

Actions

- Get more data / examples – “Augmentation”
- Reduce network size / parameters – “Dropout”
- Increase “lambda” / “weight decay” (regularization)

DATA AUGMENTATION

Augmentation expands your dataset

- Mirror images
- Distorted / blurred
- Rotations
- Color changes

NEURAL NETWORK GUIDANCE

1. Use Data Augmentation.
2. Start with well-known network.
3. Initialize weights with small random values.
4. Ensure accuracy improving as network is being trained.
5. Plot learning curves to diagnose under / over fitting.

NEURAL NETWORK STRENGTH

Using a large/complex neural network implies Low Bias.

Using a large data set implies Low Variance.

Neural Networks + Big Data = Good ✓ Stuff

Using Caffe for Deep Learning

LEARNING A BIT MORE WITH CAFFE

Caffe

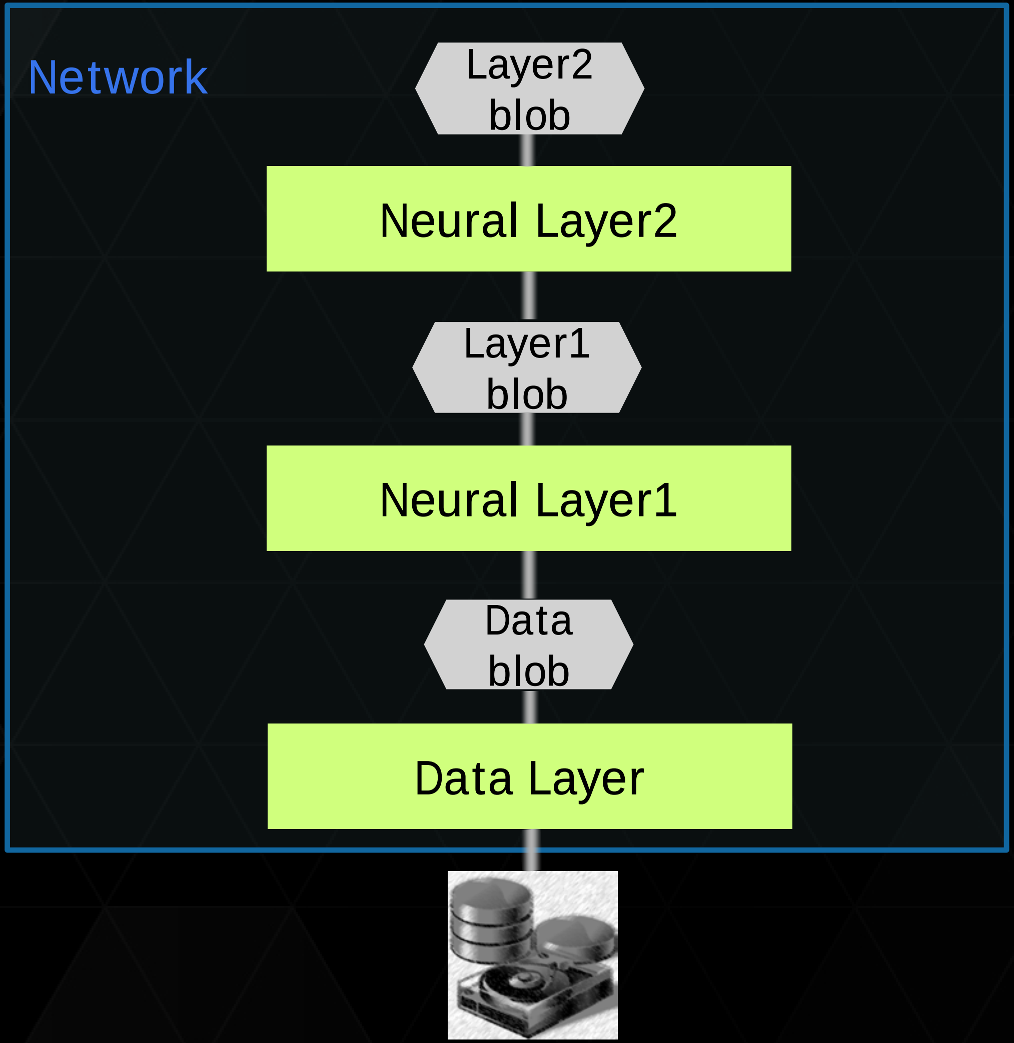
Let’s learn a bit more about DNNs by learning a bit about Caffe.

Caffe was developed at UC Berkeley.

We’ll learn about layer types, and how to think about neural network architecture.

Though we’ll use Caffe as our working example, these concepts are useful in general.

NETWORKS, LAYERS & BLOBS



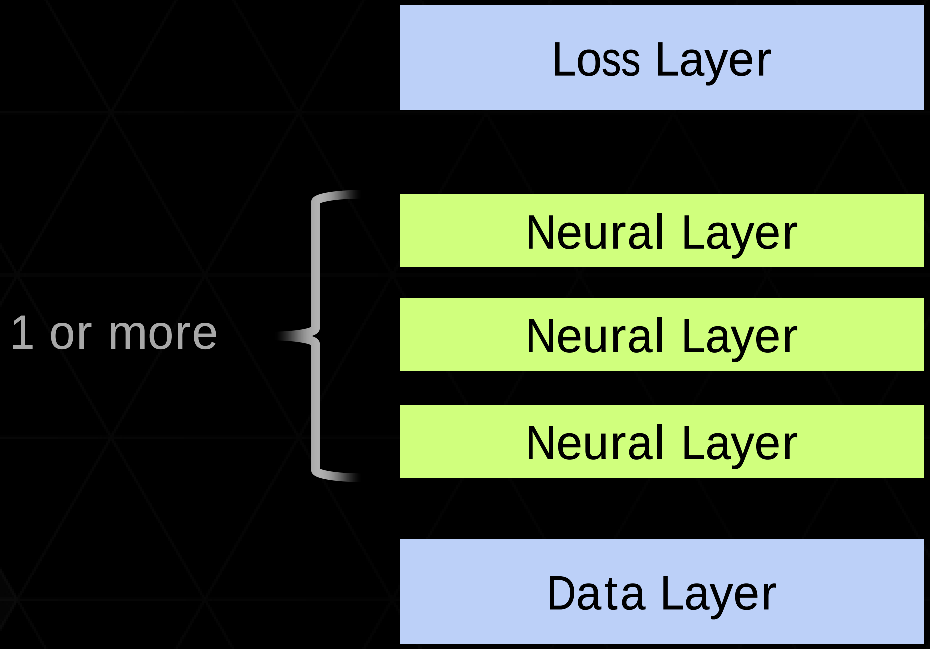
Blob – describes data
- batch of images
- model parameters

Layer - computation

Caffe

OVERALL NETWORK STRUCTURE

Ignoring blobs here...



Caffe

CAFFE MODELS DEFINED IN PLAINTEXT

```
name: "LogReg"
layer {
  name: "mnist"
  type: "Data"
  top: "data"
  top: "label"
  data_param {
    source: "input_leveldb"
    batch_size: 64
  }
}
layer {
  name: "ip"
  type: "InnerProduct"
  bottom: "data"
  top: "ip"
  inner_product_param {
    num_output: 2
  }
}
layer {
  name: "loss"
  type: "SoftmaxWithLoss"
  bottom: "ip"
  bottom: "label"
  top: "loss"
}
```

Caffe

CAFFE NEURAL LAYERS

1 or more

Neural Layer

Neural Layer

Neural Layer

Convolution

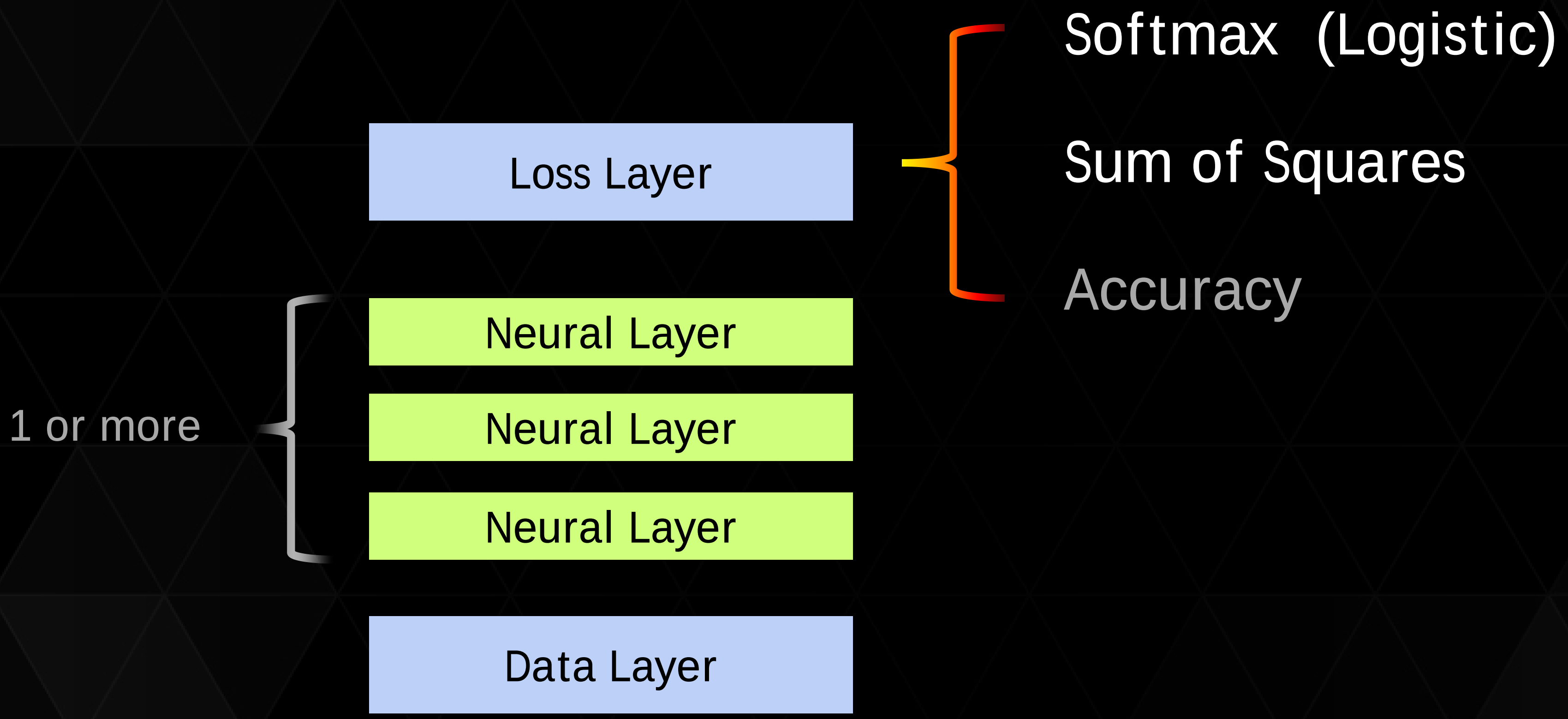
Inner Product = Fully Connected

Pooling

Local Response Normalization

Caffe

CAFFE “LOSS” LAYERS



Summary – Deep Learning for GEOINT

DEEP LEARNING AS GEOINT FORCE MULTIPLIER

- Managing Big Data
  - Real-time near-human level perception at web-scale
- Data exploration and discovery
  - Semantic and similarity based search
  - Dimensionality reduction
  - Transfer learning
- Model sharing
  - Compact model representations
  - Models can be fine-tuned based on multiple analysts feedback

SUMMARY - DL FOR GEOINT

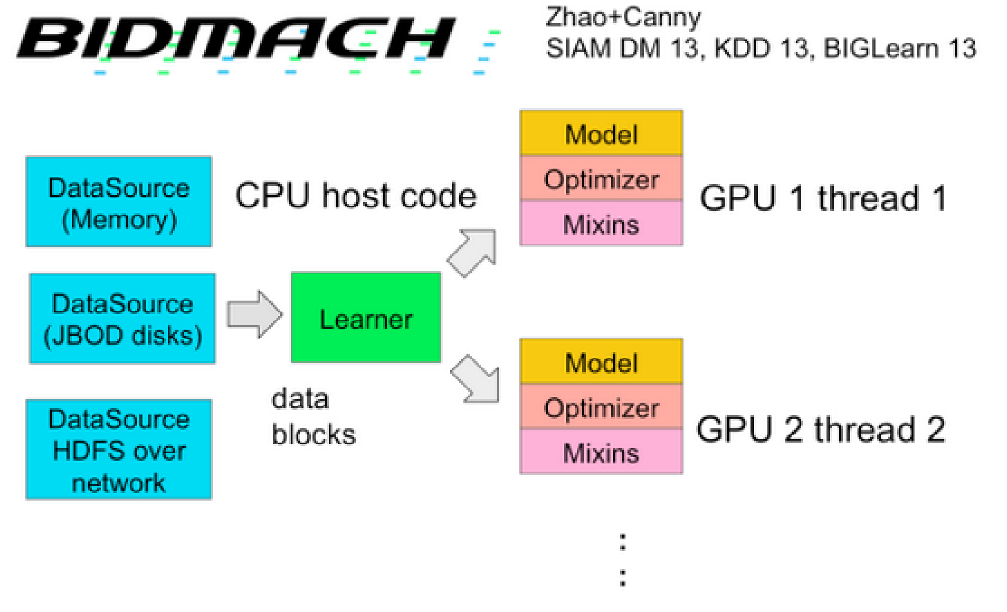
Deep Learning

- Adaptable to many varied GEOINT workflows and deployments scenarios
- Available to apply in production and R&D today
- Approachable using open-source tools and libraries

Machine Learning and Data Analytics

TRADITIONAL MACHINE LEARNING

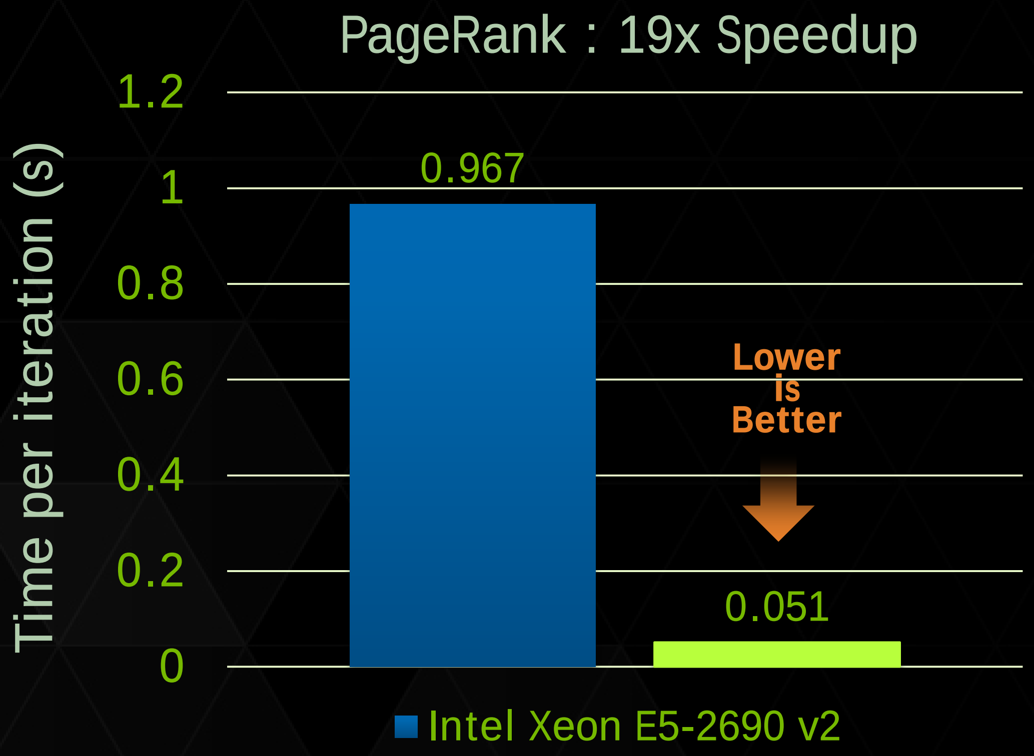
For your many non-DL applications…



- Interactive environment for easily building and deploying ML systems.
- Holds records for performance on many common ML tasks, on single nodes or clusters.
- Uses Scala. Feels like SciPy or Matlab.

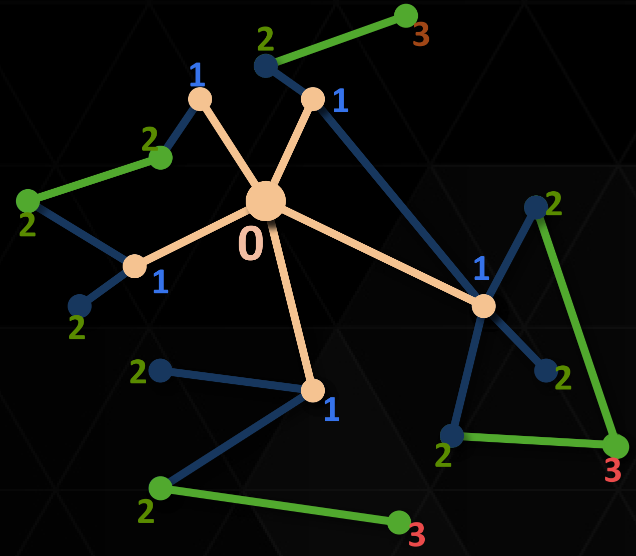
GPU ACCELERATION FOR GRAPH ANALYTICS

- Comms & Social networks
- Cyber pattern recognition
- Shortest path finding



1 GPU vs 60 Nodes

280x vs optimized Spark
1440x vs Spark
3420x vs Hadoop



NVIDIA®

Thank you!