# The DEEP LEARNING Revolution

**Gunter Roth**
**gunterr@nvidia.com**
SA EMEA
**October 2016**

# KEY DRIVERS FOR DEEP LEARNING

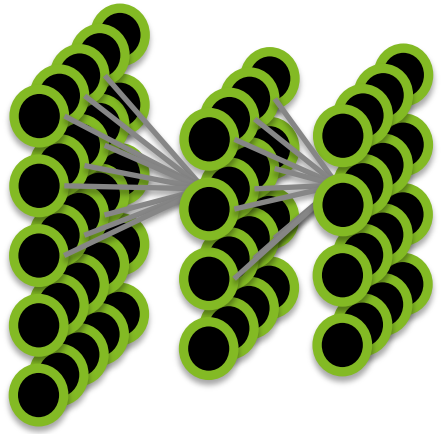
## Big Data

facebook — 350 millions images uploaded per day

Walmart — 2.5 Petabytes of customer data hourly

YouTube — 300 hours of video uploaded every minute

## Better Algorithms



## GPU Acceleration

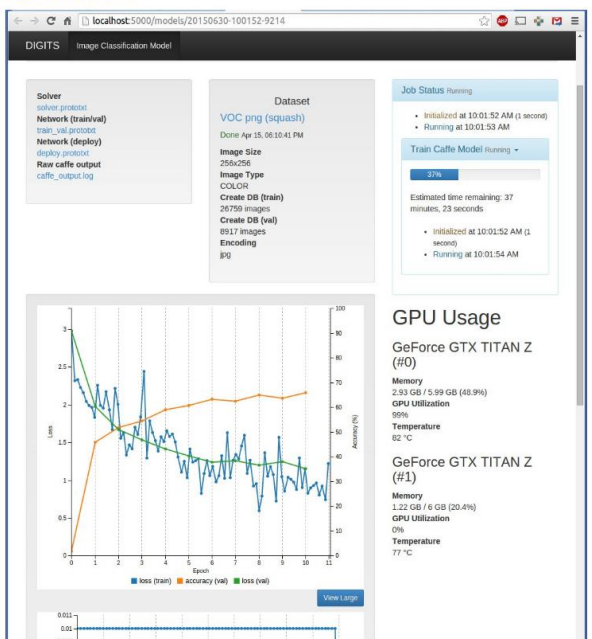

*"The Three Breakthroughs that have Finally Unleashed A.I. on the World"* — WIRED

# The DL Software stack

## NVIDIA DIGITS™

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

Learn more




## 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.

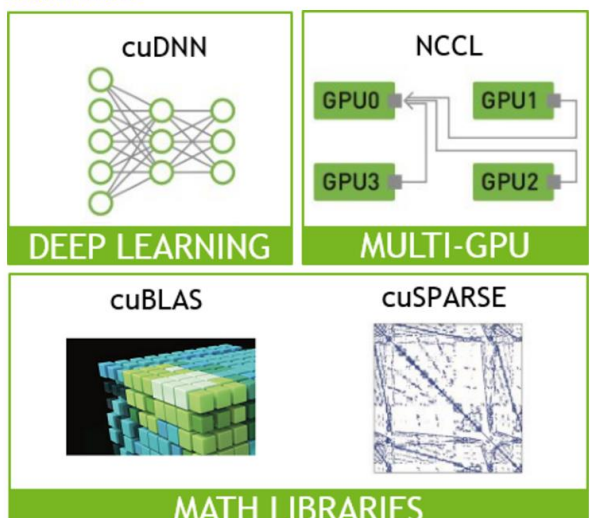
Learn more

Caffe
Chainer
DL4J
Deeplearning4j
MINERVA
mxnet
Purine
KERAS
Microsoft CNTK
MatConvNet
TensorFlow
theano
torch

## 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.

Learn more



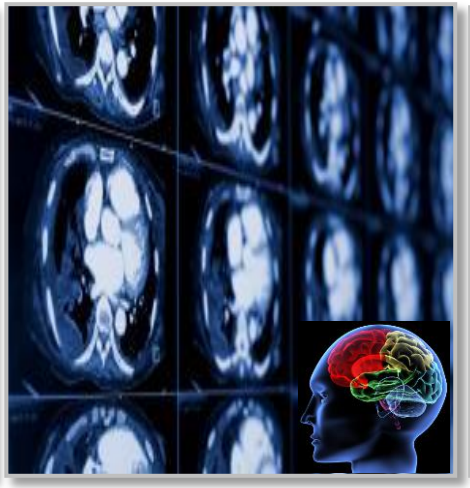

Source: https://developer.nvidia.com/deep-learning-getting-started

# EVERY INDUSTRY WANTS DEEP LEARNING

## Cloud Service Provider


- 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

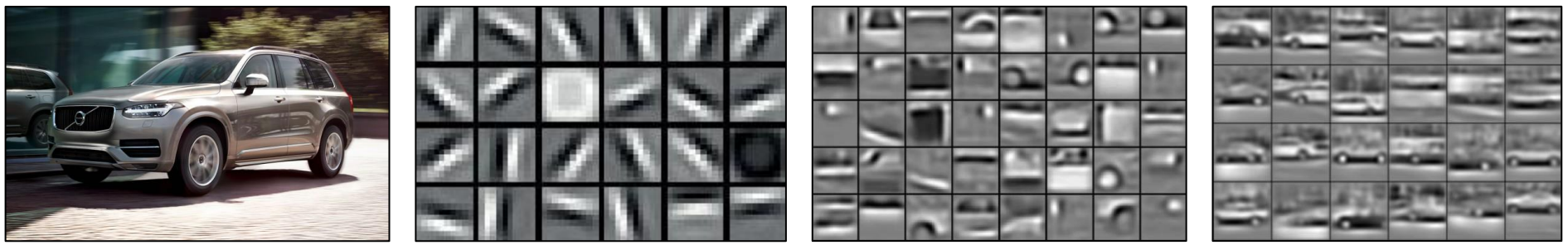
## Security & Defense

- Face recognition
- Video surveillance
- Cyber security

## Autonomous Machines


- Pedestrian detection
- Lane tracking
- Recognize traffic sign

# CONVOLUTIONAL NEURAL NETWORKS



Image

“Volvo XC90”

*Image source: “Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks” ICML 2009 & Comm. ACM 2011. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.*

# Long short-term memory (LSTM)

Hochreiter (1991) analysed vanishing gradient "*LSTM falls out of this almost naturally*"

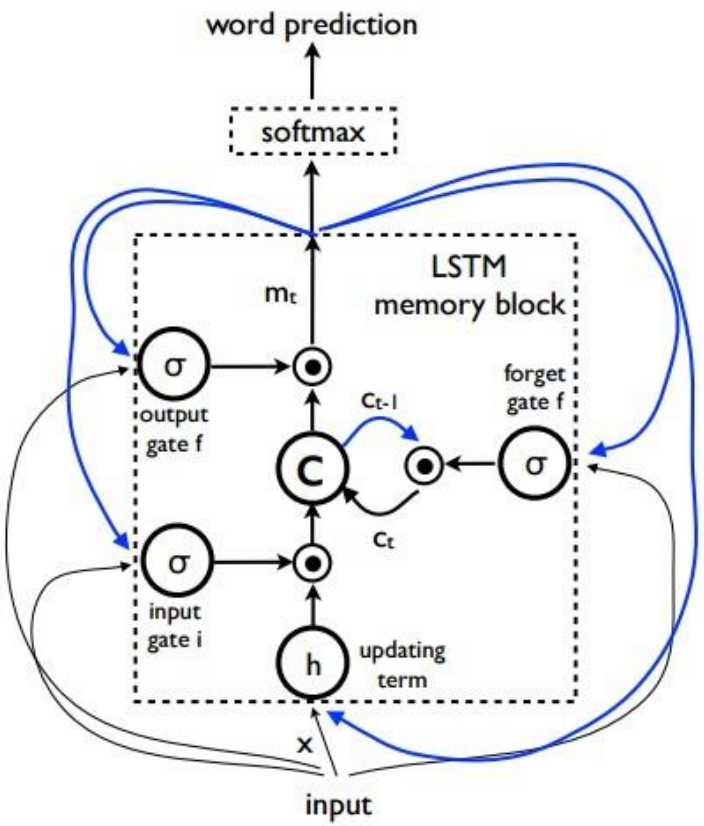
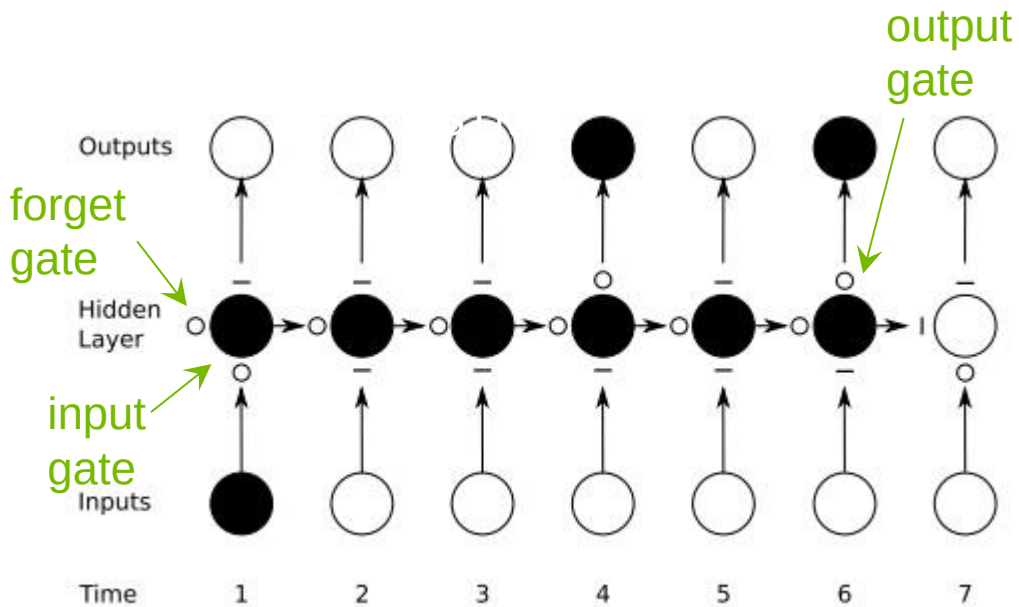



*Fig from Vinyals et al, Google April 2015 NIC Generator*

**Training via backprop unfolded in time**




Long time dependencies are preserved until input gate is closed (-) and forget gate is open (O)

**Gates control importance of the corresponding activations**

*Fig from Graves, Schmidhuber et al, Supervised Sequence Labelling with RNNs*

# THE NEXT STEP – NATURAL LANGUAGE PROCESSING

CNN + RNN

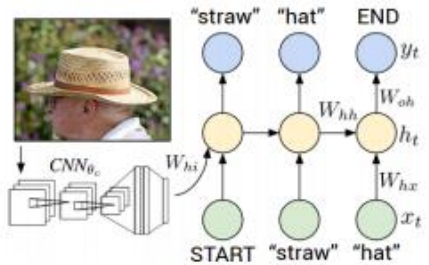



Figure 4. Diagram of our multimodal Recurrent Neural Network generative model. The RNN takes a word, the context from previ-

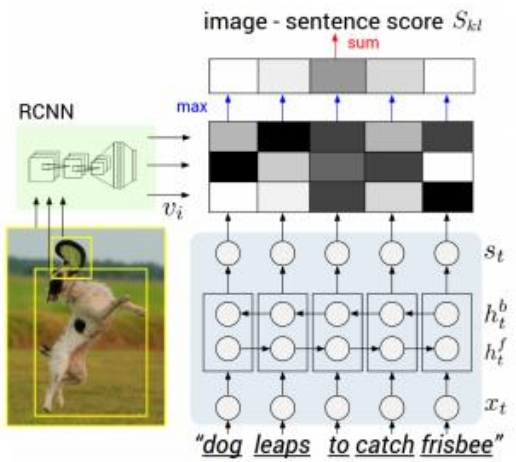



Figure 3. Diagram for evaluating the image-sentence score $S_{kl}$. Object regions are embedded with a CNN (left). Words (enriched by their context) are embedded in the same multimodal space with a BRNN (right). Pairwise similarities are computed with inner products (magnitudes shown in grayscale) and finally reduced to image-sentence score with Equation 8.

"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

"little girl is eating piece of cake."

"baseball player is throwing ball in game."

"woman is holding bunch of bananas."

"black cat is sitting on top of suitcase."

"a young boy is holding a baseball bat."

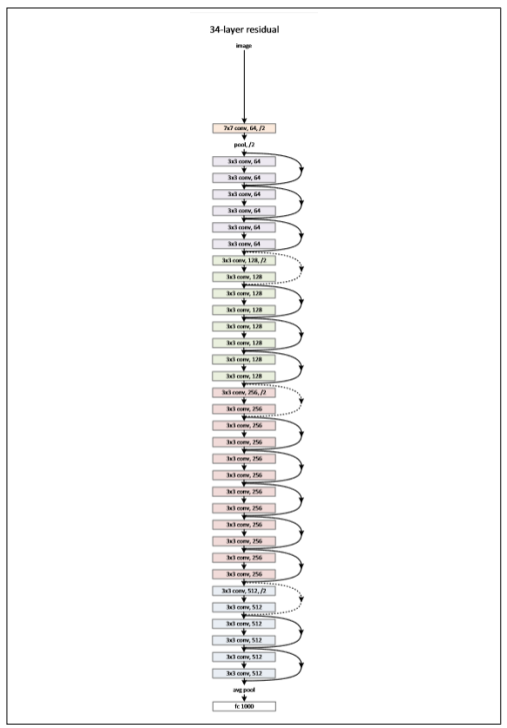
"a cat is sitting on a couch with a remote control."

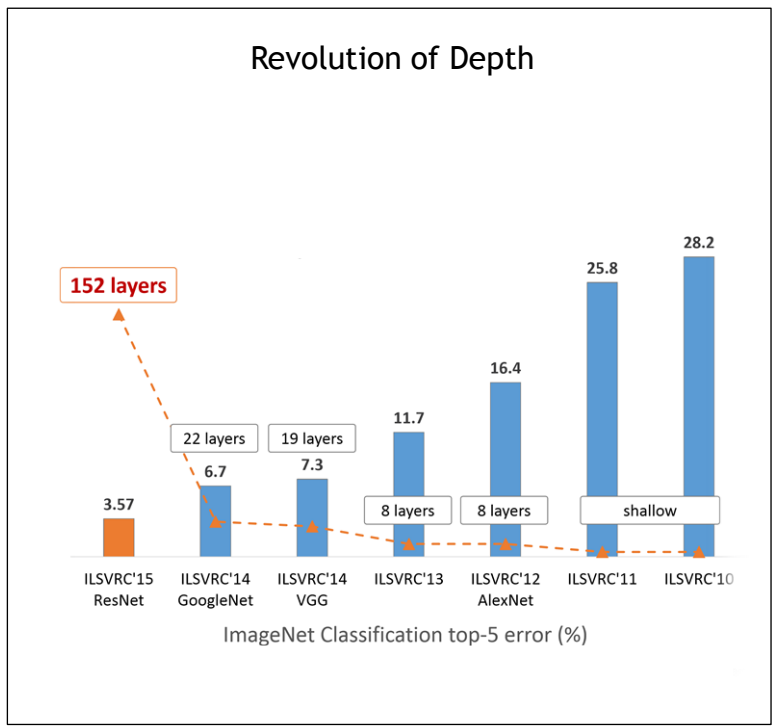
"a woman holding a teddy bear in front of a mirror."

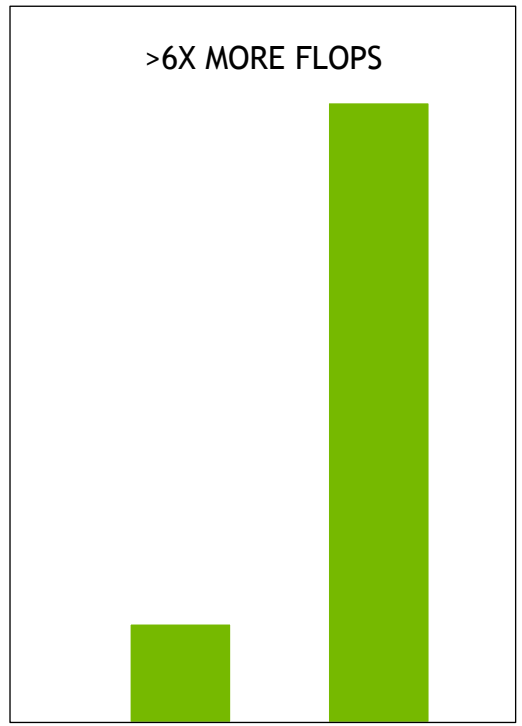
"a horse is standing in the middle of a road."

# GOOGLE DEEPMIND ALPHAGO CHALLENGE


# MICROSOFT: “SUPER DEEP NETWORKS”




Microsoft Deep ResNet
http://arxiv.org/pdf/1512.03385v1.pdf

# Deep Learning Hardware

# END-TO-END PRODUCT FAMILY

## TRAINING

### FULLY INTEGRATED DL SUPERCOMPUTER

DGX-1

### DESKTOP

Titan X

### HYPERSCALE

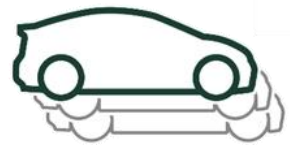
Tesla M40
Tesla P40

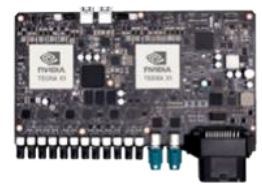
## INFERENCE

### DATA CENTER

Tesla M4
Tesla P4

### AUTOMOTIVE





Drive PX

### EMBEDDED

Jetson TX1

# 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

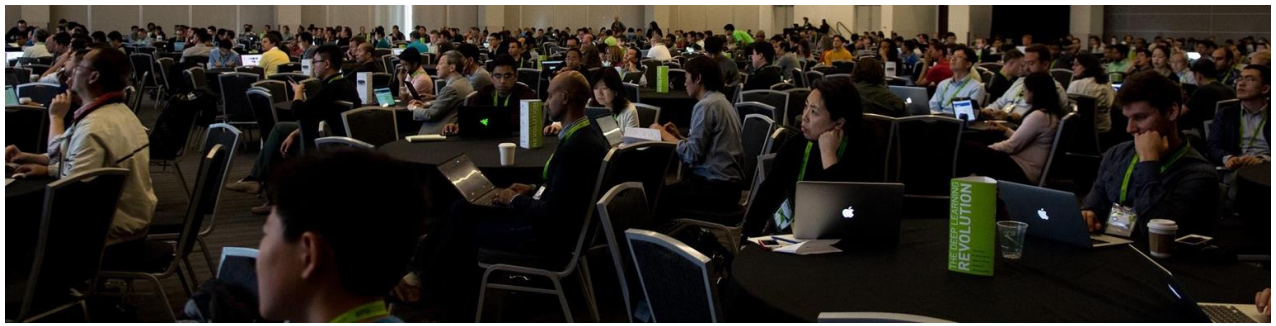
7 TB SSD Deep Learning Cache

Dual 10GbE, Quad IB 100Gb

3RU – 3200W

# NVIDIA Deep Learning Institute

Hands-on Training for Data Scientists and Software Engineers



Training organizations and individuals to solve challenging problems using Deep Learning

On-site workshops and online courses presented by certified experts

Covering complete workflows for proven application use cases

Image classification, object detection, natural language processing, recommendation systems, and more

**www.nvidia.com/dli**

# DQN: deep Q-learning network

## REINFORCEMENT LEARNING

### Mastering Breakout

# Pieter Abbeel

gym.openai.com

# Questions?
# Bon appetit!

gunterr@nvidia.com

NVIDIA