Fascinating Tales of a Strange Tomorrow

Julien Simon
Principal Technical Evangelist
julsimon@amazon.fr
@julsimon

amazon web services

Pop-up Loft
TEL AVIV

1956

Dartmouth Summer Research Project

John McCarthy

Coined the term “Artificial Intelligence”

Invented LISP (1958)

Received Turing Award (1971)

Forbidden Planet

Robbie the Robot

Predictions

- 1958 H. A. Simon and Allen Newell: "*within 10 years a digital computer will be the world's chess champion*" and "*within 10 years a digital computer will discover and prove an important new mathematical theorem*"
- 1965 H. A. Simon: "*machines will be capable, within 20 years, of doing any work a man can do*"
- 1967 Marvin Minsky: "*Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved*."
- 1970 Marvin Minsky: "*In from 3 to 8 years we will have a machine with the general intelligence of an average human being*"

https://en.wikipedia.org/wiki/History_of_artificial_intelligence

“It’s 2001. Where is HAL?”

Marvin Minsky
Co-founded the MIT AI lab (1959)
Advised Kubrick on 2001: A Space Odyssey (1968)
Received Turing Award (1969)

HAL 9000
HAL Laboratories (1992)

Meanwhile, on the US West Coast…

amazon web services

Millions of users… Mountains of data… Commodity hardware…
Bright engineers… Need to make money….

Gasoline waiting for a match!

The Machine Learning explosion

- 12/2004 - Google publishes Map Reduce paper
- 04/2006 - Hadoop 0.1

- 05/2009 – Yahoo sorts a Terabyte in 62 seconds
- The rest is history

Fast forward a few years

- ML is now a commodity
- Great, but still no HAL in sight
- Traditional Machine Learning doesn't work well with complex problems such as computer vision, computer speech or natural language processing
- Another AI winter, then?

Neural networks

- Through training, a neural network self-organizes and discovers features automatically: the more data, the better (unlike traditional ML)

- "Universal approximation machine" (Andrew Ng)
  - Artificial Intelligence is the New Electricity https://www.youtube.com/watch?v=21EiKfQYZXc

- Not new technology!
  - Perceptron (Rosenblatt, 1958)
  - Backpropagation (Werbos, 1975)

- They failed back then because
  - data sets were too small
  - computing power was not available

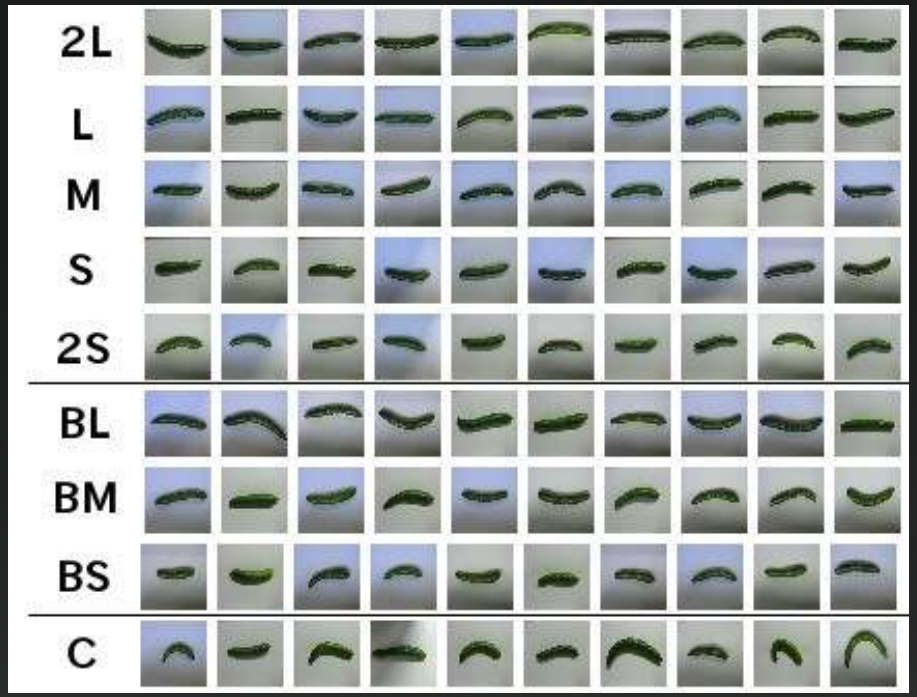
Why it's different this time

- Large data sets are available
  - Imagenet : 14M+ images http://www.image-net.org/
- GPUs and FPGAs deliver unprecedented amounts of computing power.
  - It's now possible to train networks that have hundreds of layers
- Scalability and elasticity are key assets for Deep Learning
  - Grab a lot of storage and compute resources for training, then release them
  - Using a DL model is lightweight: you can do it on a Raspberry Pi!

I for one welcome our new Deep Learning Overlords

amazon web services

Sorting cucumbers in Japan



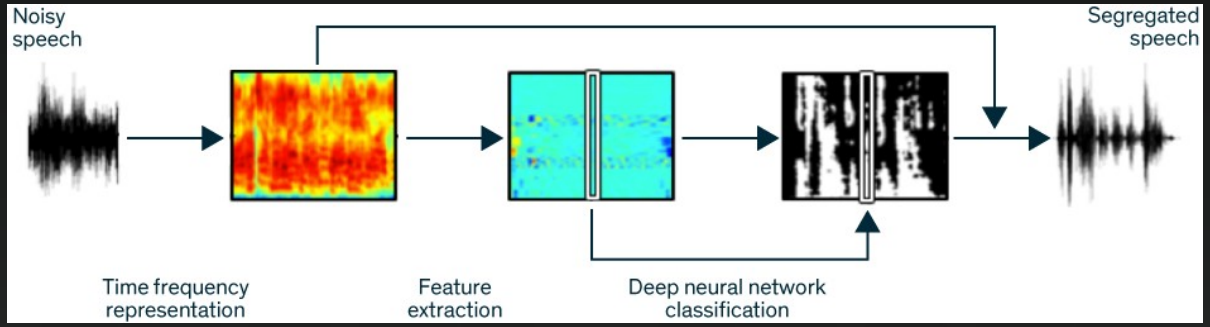
https://cloud.google.com/blog/big-data/2016/08/how-a-japanese-cucumber-farmer-is-using-deep-learning-and-tensorflow

Detecting plant diseases

https://blogs.nvidia.com/blog/2016/12/13/ai-fights-plant-disease/
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5032846/

amazon web services

Improving hearing aids



http://spectrum.ieee.org/consumer-electronics/audiovideo/deep-learning-reinvents-the-hearing-aid

Flipping burgers

Flippy

http://misorobotics.com/

amazon web services

Amazon Echo

Now what?

amazon web services

Dive as deep as you need to

AWS GPU Instances

- g2 (2xlarge, 8xlarge)
  - 32 vCPUs, 60 GB RAM
  - 4 NVIDIA K520 GPUs
  - 16 GB of GPU memory, 6144 CUDA cores
- p2 (xlarge, 8xlarge, 16xlarge)
  - 64 vCPUs, 732 GB RAM
  - 16 NVIDIA GK210 GPUs
  - 192 GB of GPU memory, 39936 CUDA cores
  - 20 Gbit/s networking

EC2 Instance Type	Total
g2.2xlarge	$0.65/hr
g2.8xlarge	$2.60/hr
p2.8xlarge	$7.20/hr
p2.xlarge	$0.90/hr
p2.16xlarge	$14.40/hr

https://aws.amazon.com/blogs/aws/new-g2-instance-type-with-4x-more-gpu-power/
https://aws.amazon.com/blogs/aws/new-p2-instance-type-for-amazon-ec2-up-to-16-gpus/
https://aws.amazon.com/ec2/Elastic-GPUs/

AWS Deep Learning AMI

- Deep Learning Frameworks – 5 popular Deep Learning Frameworks (mxnet, Caffe, Tensorflow, Theano, and Torch) all prebuilt and pre-installed
- Pre-installed components – Nvidia drivers, cuDNN, Anaconda, Python2 and Python3
- AWS Integration – Packages and configurations that provide tight integration with Amazon Web Services like Amazon EFS (Elastic File System)
- Amazon Linux & Ubuntu

https://aws.amazon.com/about-aws/whats-new/2017/03/deep-learning-ami-release-v2-0-now-available-for-amazon-linux/

mxnet

mxnet resources

http://mxnet.io/
https://github.com/dmlc/mxnet
https://github.com/dmlc/mxnet-notebooks

http://www.allthingsdistributed.com/2016/11/mxnet-default-framework-deep-learning-aws.ht

https://github.com/awslabs/deeplearning-cfn

mxnet demo

Deep Learning AMI on p2.16xlarge
Training and predicting the MNIST data set

amazon web services

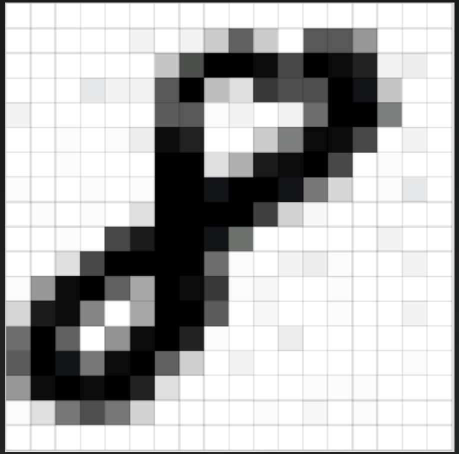
MNIST dataset

http://yann.lecun.com/exdb/mnist/

70,000 handwritten digits

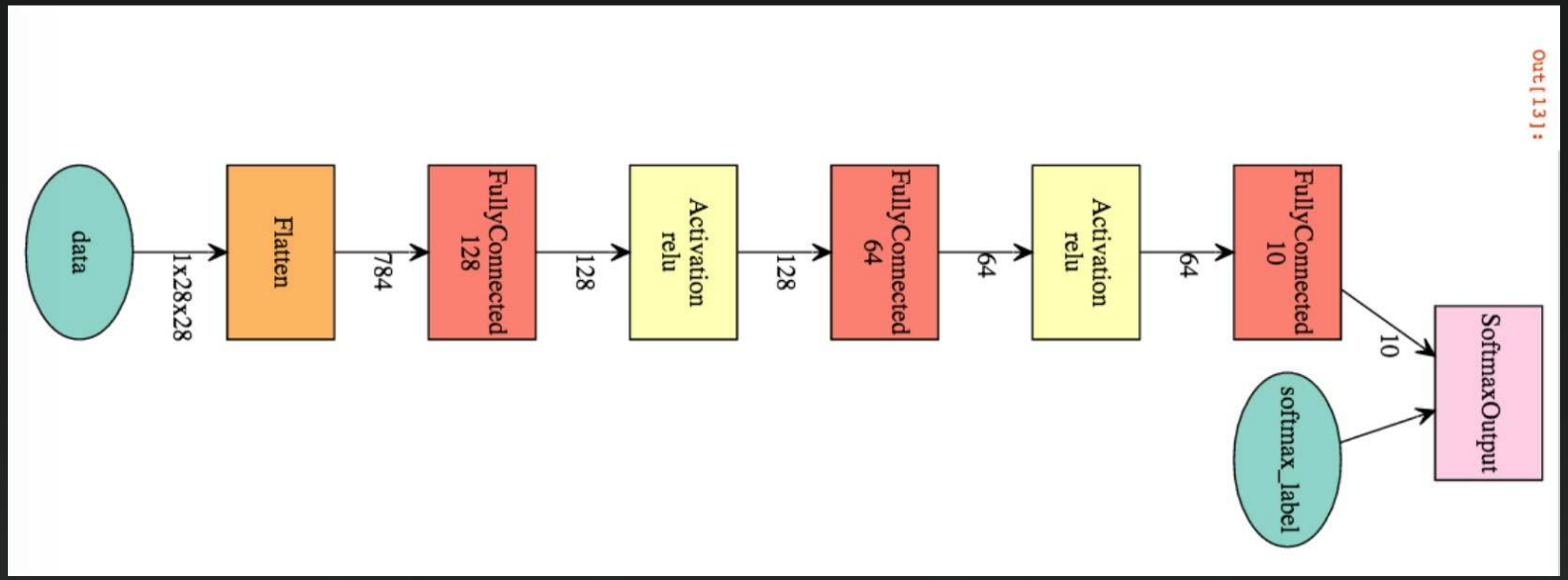
28x28 pixels

Greyscale (0 to 255)



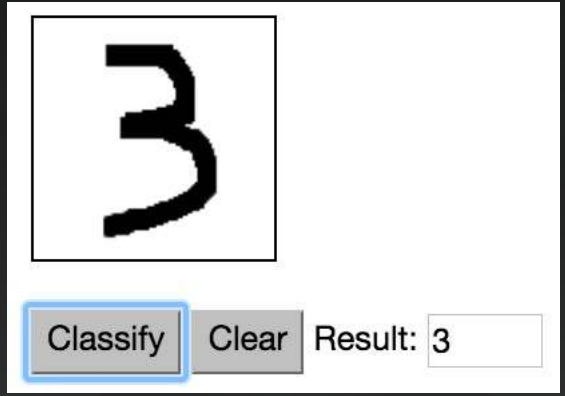
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0
0	0	1	0	0	0	41	160	250	255	235	162	255	238	206	11	13	0
0	0	0	16	9	9	150	251	45	21	184	159	154	255	233	40	0	0
10	0	0	0	0	0	145	146	3	10	0	11	124	253	255	107	0	0
0	0	3	0	4	15	236	216	0	0	38	109	247	240	169	0	11	0
1	0	2	0	0	0	253	253	23	62	224	241	255	164	0	5	0	0
6	0	0	4	0	3	252	250	228	255	255	234	112	28	0	2	17	0
0	2	1	4	0	21	255	253	251	255	172	31	8	0	1	0	0	0
0	0	4	0	163	225	251	255	229	120	0	0	0	0	0	11	0	0
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0
3	79	242	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0
125	255	141	0	87	244	255	208	3	0	0	13	0	1	0	1	0	0
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	1	3	0
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0
6	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

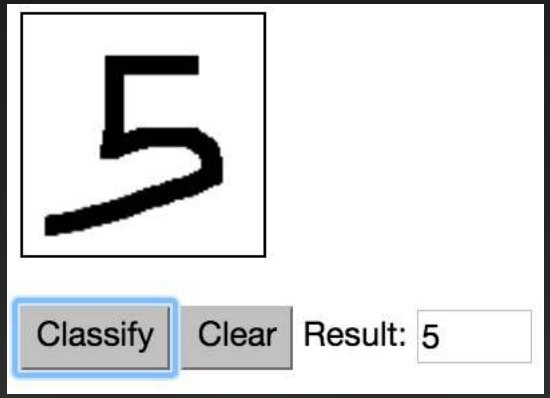
Multilayer perceptron



Train and test

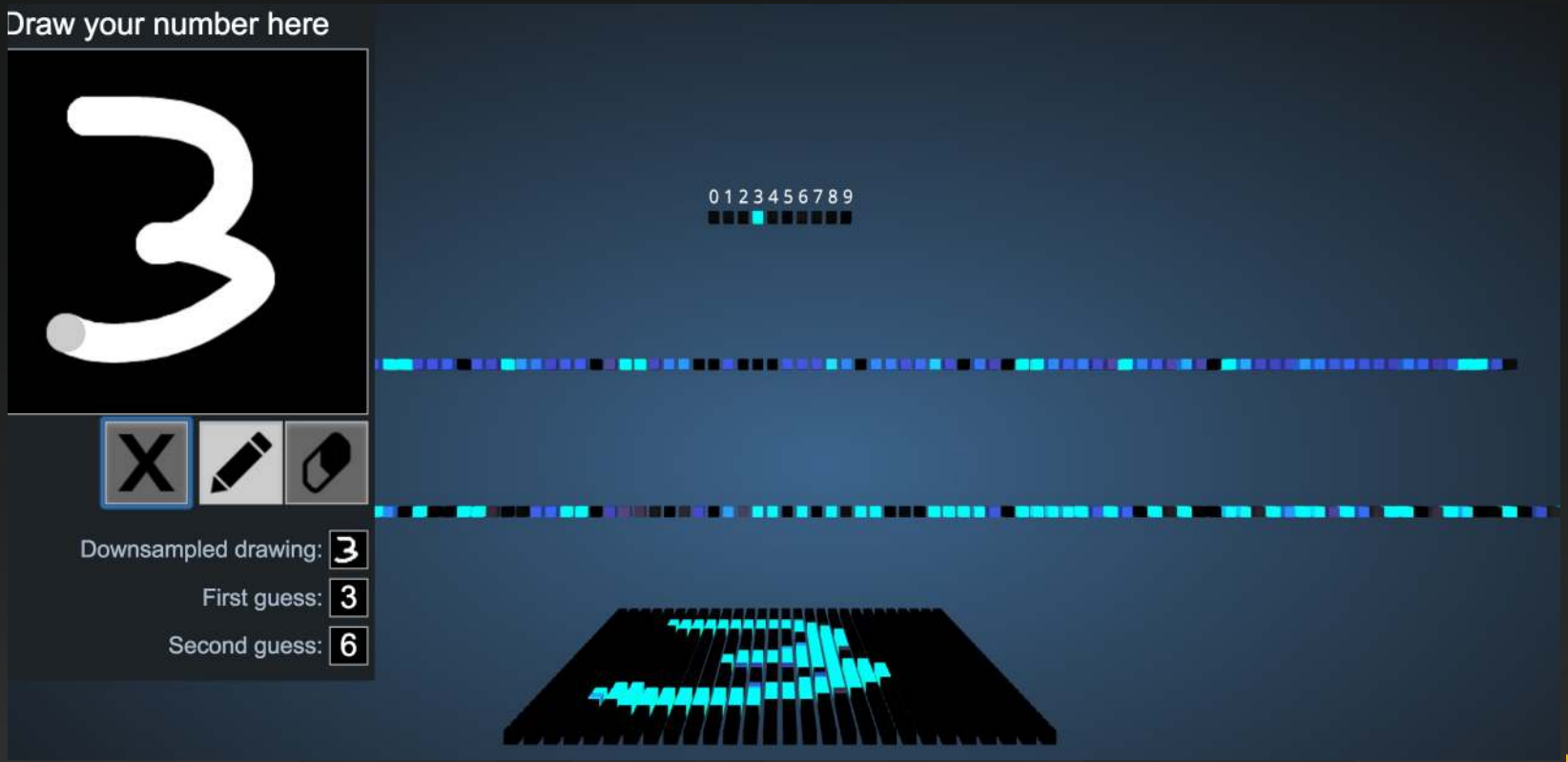
```
INFO:root:Epoch[9] Batch [5600]    Speed: 6475.51 samples/sec    Train-accuracy=0.987500
INFO:root:Epoch[9] Batch [5800]    Speed: 6541.05 samples/sec    Train-accuracy=0.988000
INFO:root:Epoch[9] Batch [6000]    Speed: 6481.38 samples/sec    Train-accuracy=0.988000
INFO:root:Epoch[9] Resetting Data Iterator
INFO:root:Epoch[9] Time cost=9.317
INFO:root:Epoch[9] Validation-accuracy=0.963600
```





Web visualization

http://scs.ryerson.ca/~aharley/vis/fc/



Now the hard questions…

- Can my business benefit from Deep Learning?
  - DL : “solving the tasks that are easy for people to perform but hard to describe formally”
- Should I build my own network?
  - Do I have the expertise?
  - Do I have enough time, data & compute to train it?
- Or should I use a pre-trained model?
  - How well does it fit my use case?
  - On what data was it trained?
- Or should I use a high-level service?

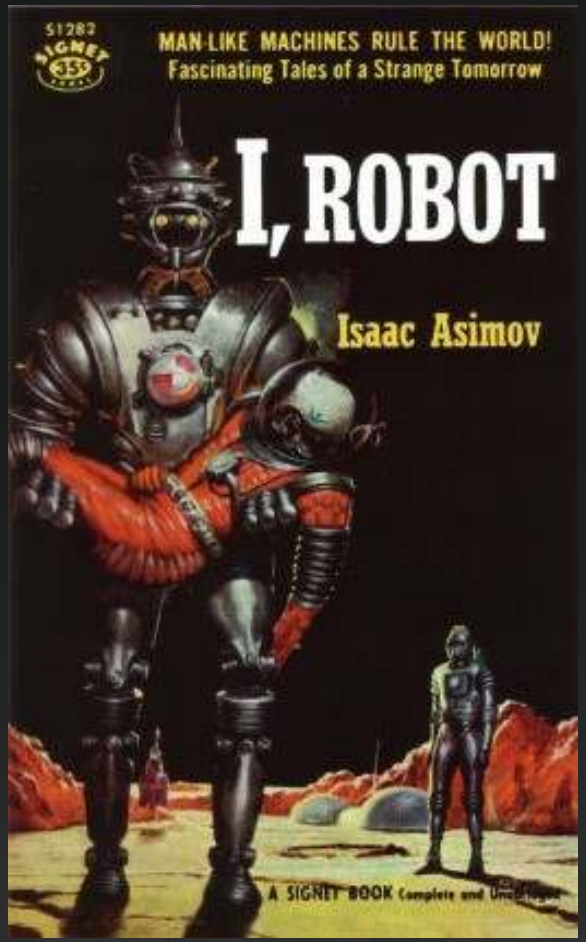
- Same questions as ML years ago ☺

Science catching up with Fiction ?

October 2014: Tesla Autopilot

October 2015: 30,000 robots in Amazon Fulfillment Centers

May 2016: AI defeats Lee Sedol, Go world champion



Will man-like machines rule the world?

Who knows?

Whatever happens, these will be fascinating tales of a strange tomorrow.

Thank You

Julien Simon
julsimon@amazon.fr
@julsimon

Your feedback is important to us!

amazon web services

Pop-up Loft
TEL AVIV