Deep learning and its impact on engineering

Lyle Ungar
University of Pennsylvania

Deep learning is taking over

- **Machine vision**
  - Face/Object/Scene recognition
  - Self driving cars
- **Speech recognition (“speech to text”)**
  - Siri, Alexa …
- **Machine translation**
  - Google translate

Big Claims

“Big data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus.”

– McKinsey

Data Scientist: “The Sexiest Job of the 21st Century”

- Davenport and Patil, Harvard Business Review 2012

All machine learning is optimization

$\hat{y} = f(\boldsymbol{x}; \boldsymbol{\theta})$

$\text{argmin}_\theta \; ||y - \hat{y}||$

So what's new?

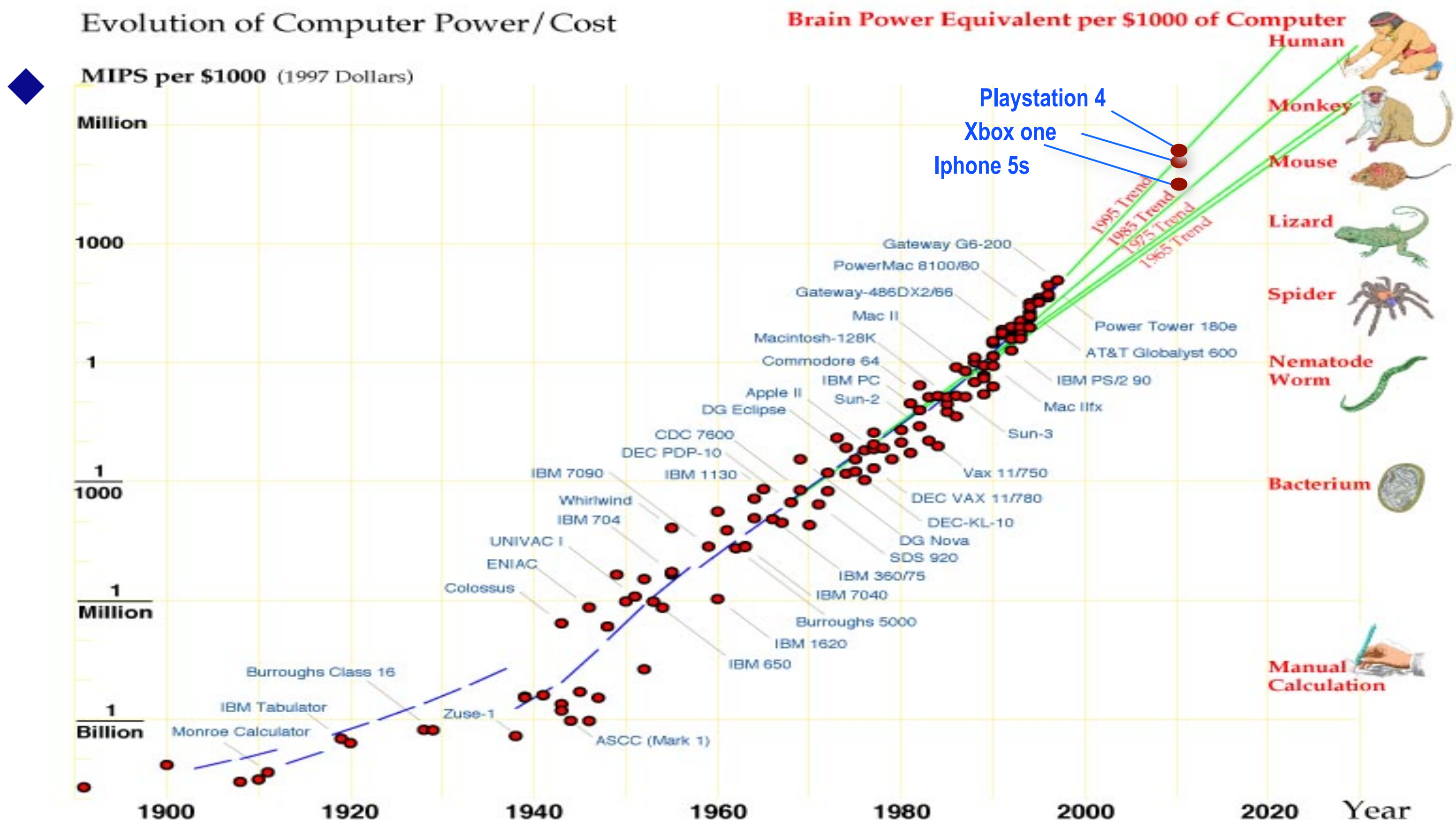
(Slightly) different loss functions

(Slightly) different optimization methods

Different, flexible, functional forms for *f*

Lots more CPU/GPU

Increasing computer power



Hans Moravec

The unreasonable effectiveness of data

- **Scene completion using millions of photographs**
  - J Hays, AA Efros - Communications of the ACM, 2008

Flexible model forms

$\hat{y} = f(\mathbf{x}; \boldsymbol{\theta})$

X	**y**
Web page, ad	**Click on ad?**
Past purchases….	**NPV**
Facebook posts	**Age, Sex, Personality, …**

Male or female?

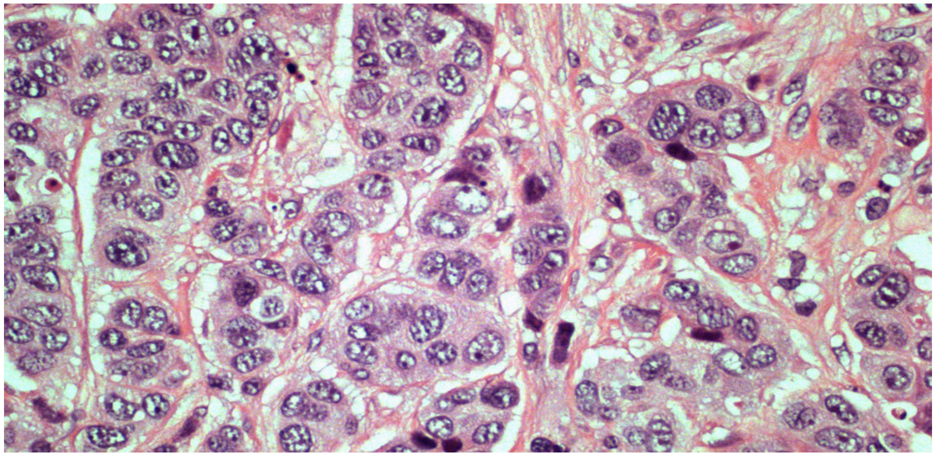
wwbp.org

Male or female?

wwbp.org

Flexible model forms

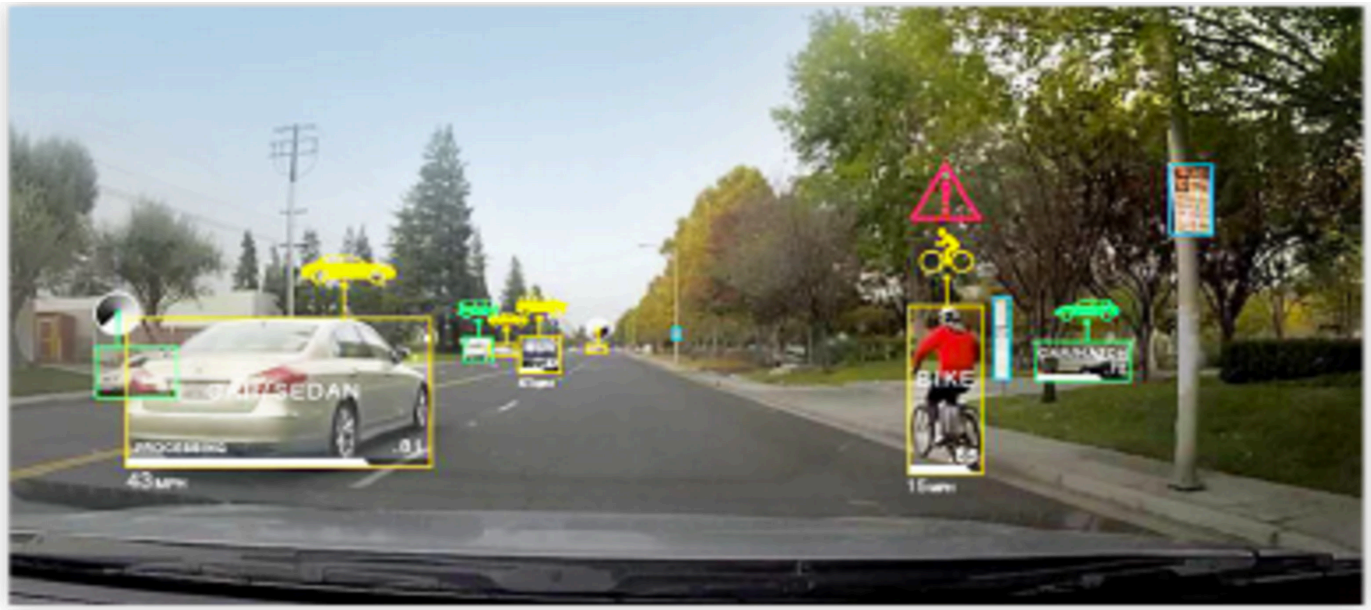
X	y
biopsy image	Cancer present?



Flexible model forms

X　　　　　　　　y

Camera image　　　Objects in it



nvidia

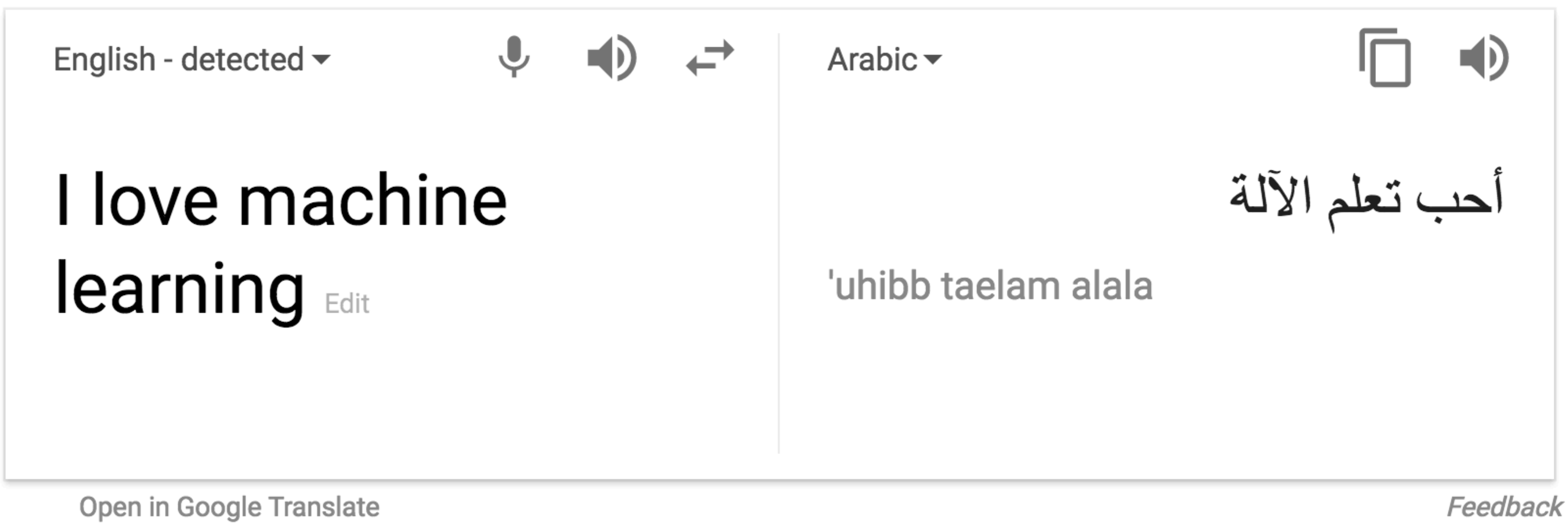
Flexible model forms

X

English sentence

y

Translation



Artificial Neural Nets

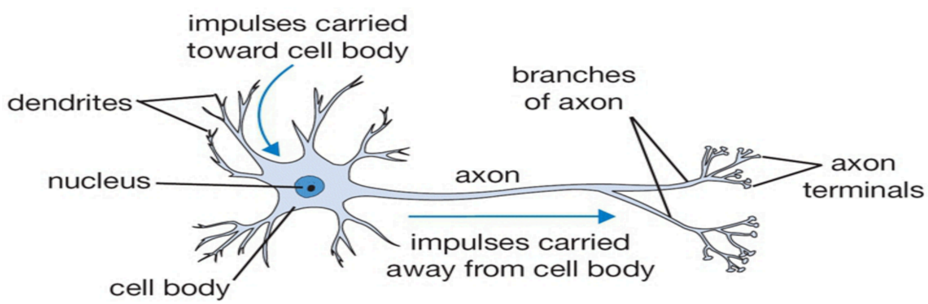
- **Non-parametric**
  - Or, technically, semi-parametric
  - Flexible model form
- **Used when there are vast amounts of data**
  - Hence popular (again) now
- **Deep networks**
  - Idea: representation should have *many* different levels of abstraction

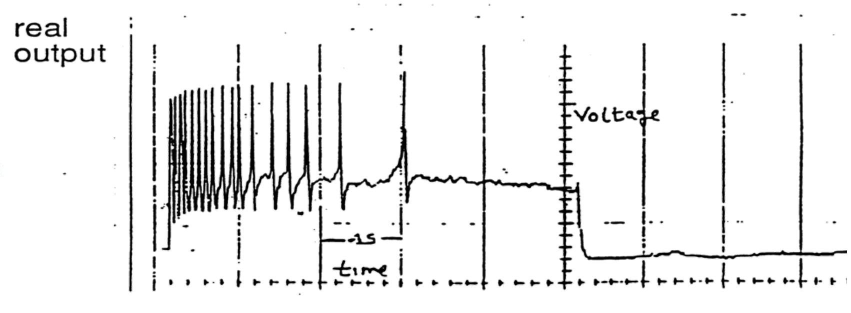
Neural Nets can be

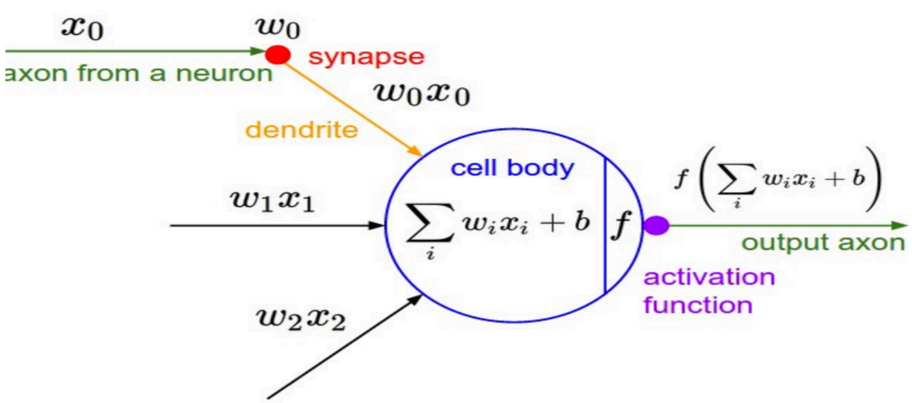
- **Supervised**
  - Generalizes *logistic regression* to a semi-parametric form
- **Unsupervised**
  - Generalizes *PCA* to a semi-parametric form
- **Reinforcement**

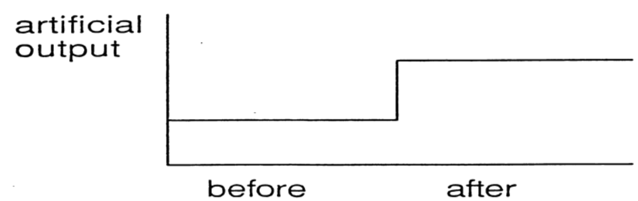
Neural nets often have built in structure

“Real” and Artificial neuron







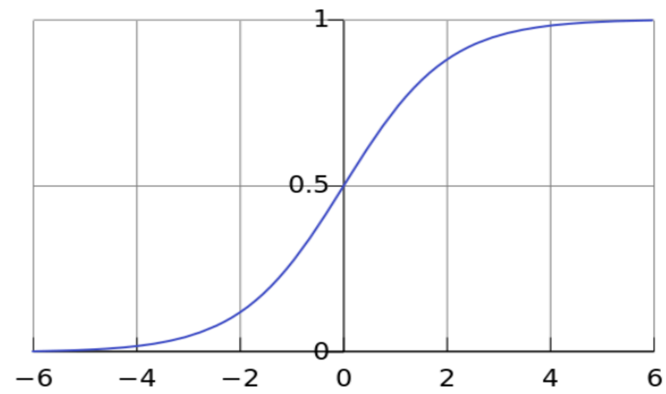


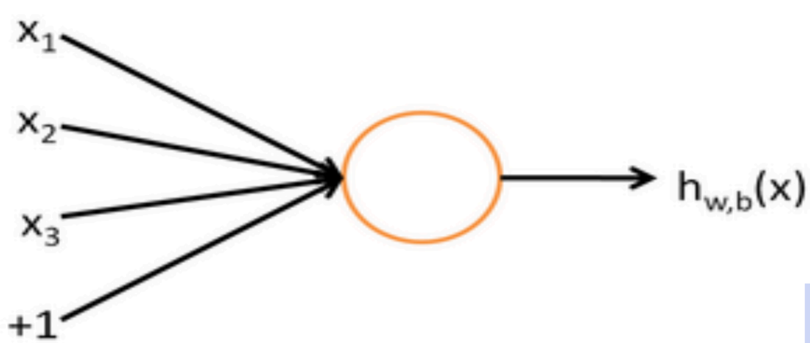
http://cs231n.github.io/neural-networks-1/

One neuron does logistic regression

$$h_{w,b}(x) = f(w^{\top}x + b)$$

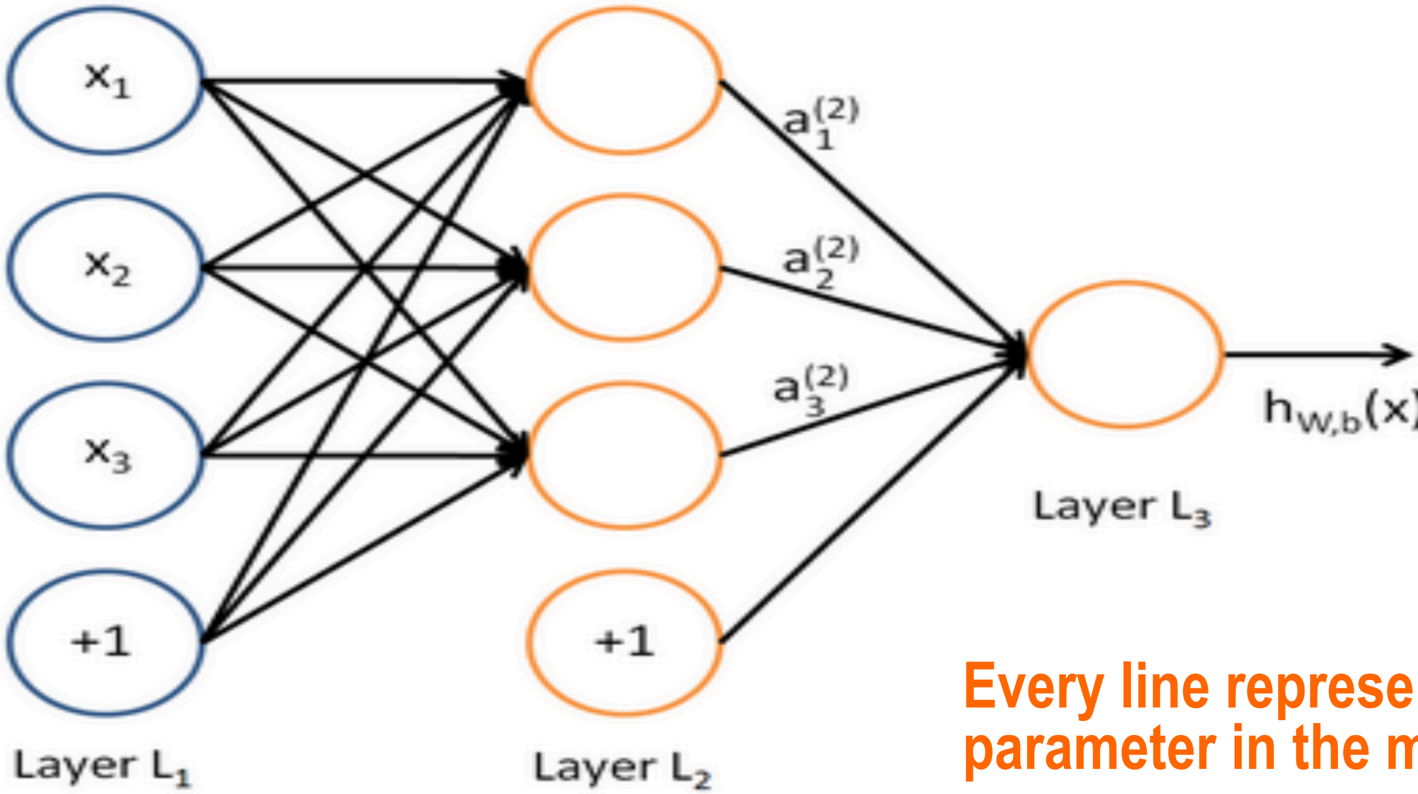
$$f(z) = \frac{1}{1+e^{-z}}$$





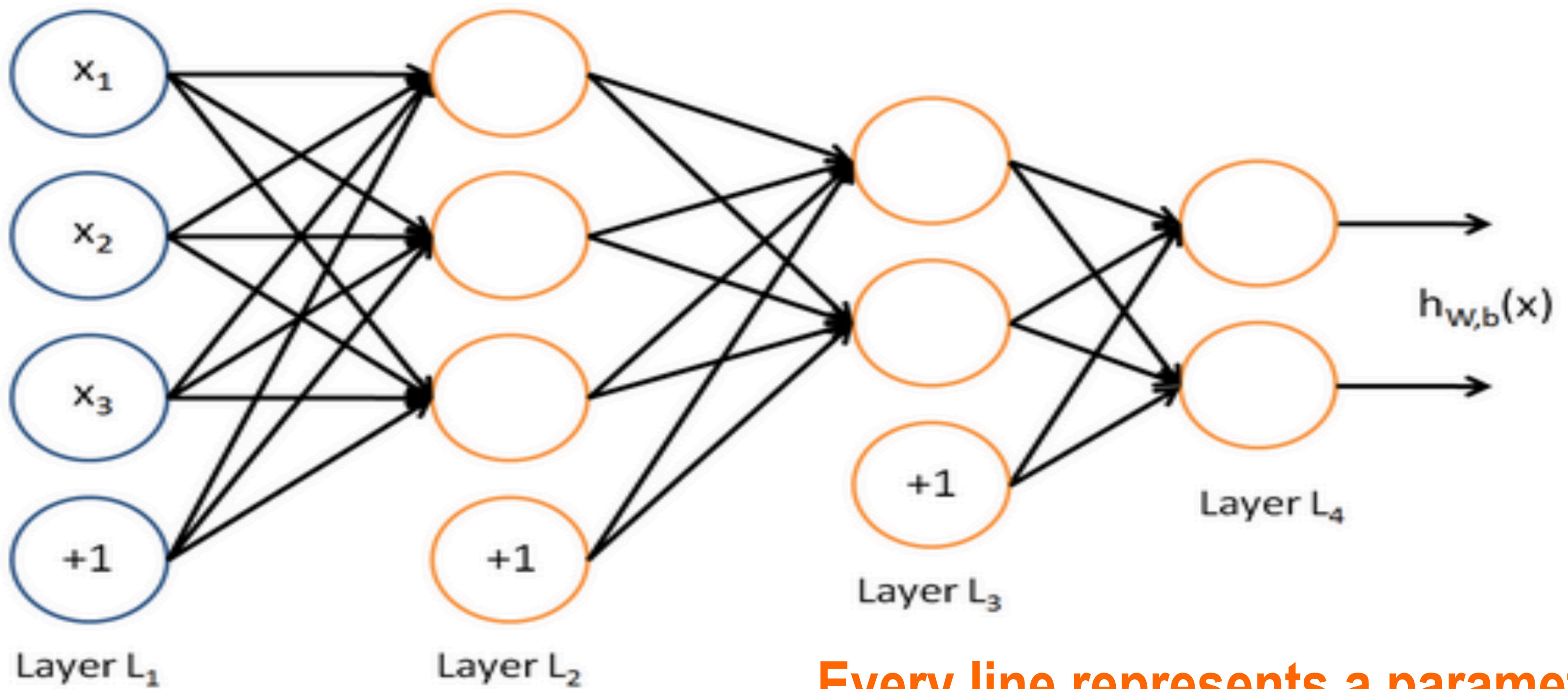
Socher and Manning tutorial

Neural nets stack logistic regressions



Every line represents a parameter in the model

Neural nets stack logistic regressions



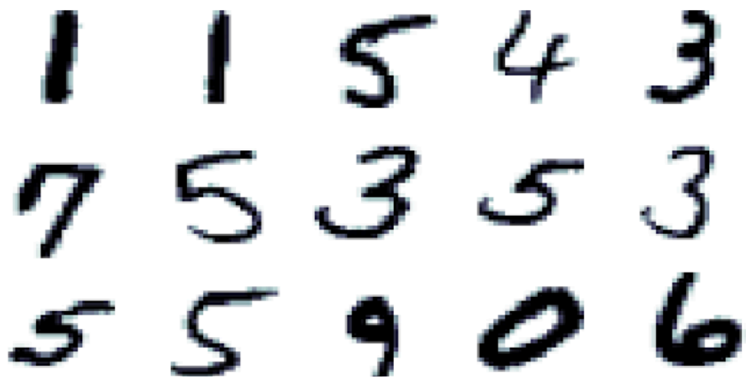
Every line represents a parameter in the model, estimated using gradient descent

ANNs do pattern recognition

- **Map input "percepts" to output categories or actions**
  - Image of an object → what it is
  - Image of a person → who it is
  - Picture → caption describing it
  - Board position → probability of winning
  - A word → the sound of saying it
  - Sound of a word → the word
  - Sequence of words in English → their Chinese translation

MNIST

- Classify 28x28 images of handwritten digits
- **Train:** 50,000
- **Test:** 10,000



Error (%)	Method	Reference
5.0	KNN	Lecun et al. (1998)
3.6	1k RBF + linear classifier	Lecun et al. (1998)
1.6	2-layer NN	Simard et al. (2003)
1.53	boosted stumps	Kegl et al. (2009)
1.4	SVM	Lecun et al. (1998)
0.79	DNN	Srivastava (2013)
0.45	conv-DNN	Goodfellow et al. (2013)
0.21	conv-DNN	Wi et al. (2013)

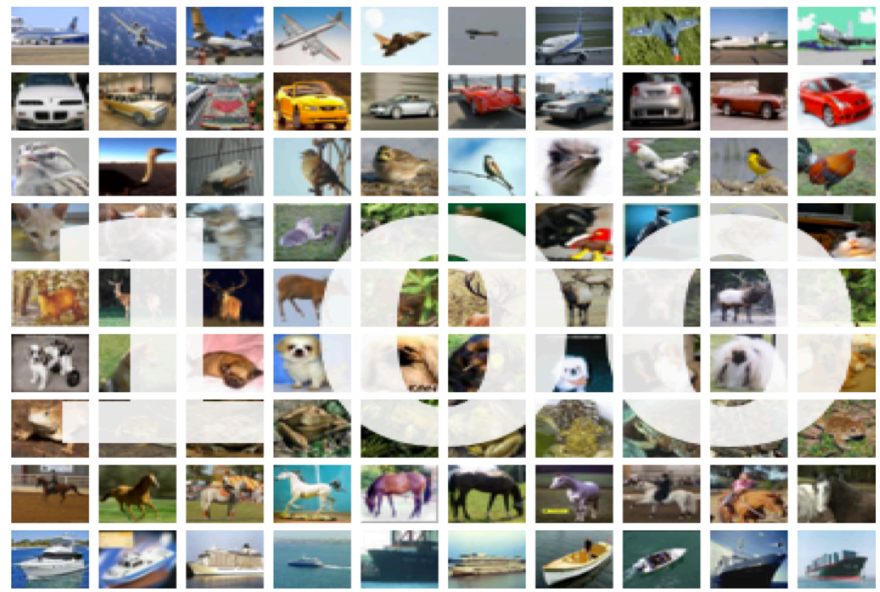
Street View House Numbers

- Classify 32x32 color images of digits
- Digits taken from housenumbers in Google Street View
- **Train:** 604,388
- **Test:** 26,032

Error (%)	Method	Reference
36.7	WDCH	Netzer et al. (2011)
15	HOG	Netzer et al. (2011)
9.4	KNN	Netzer et al. (2011)
2.47	conv-DNN	Goodfellow et al. (2013)
2	Human	Netzer et al. (2013)
1.92	conv-DNN	Lee et al. (2015)

CIFAR-100

- Classify 32x32 color images into 100 classes
- Images taken from TinyImages dataset at MIT
- **Train:** 50,000
- **Test:** 10,000



Error (%)	Method	Reference
43.77	SVM	Jia et al. (2012)
39.20	OMP	Lin and Kung (2014)
38.57	conv-DNN	Goodfellow et al. (2013)
36.18	DNN	Srivastava and Alakhutdinov (2015)
34.57	conv-DNN	Lee et al. (2015)

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
Ilya Sutskever
Geoffrey Hinton

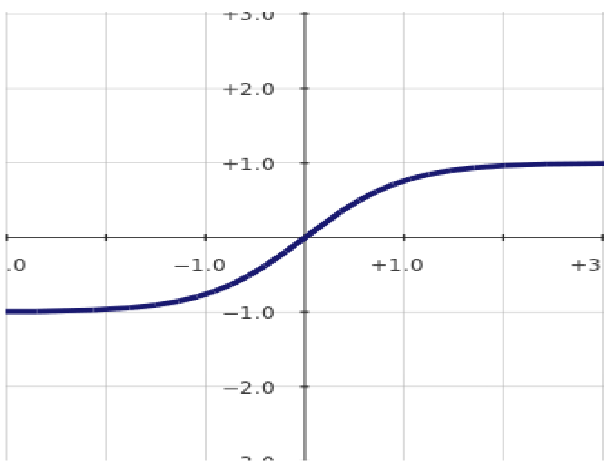
University of Toronto
Canada

“AlexNet” 2012

Neurons

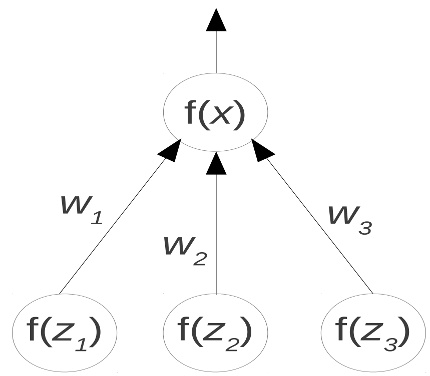
Traditional: sigmoidal e.g. logistic function

f(x) = tanh(x)



Hyperbolic tangent

Very bad (slow to train)

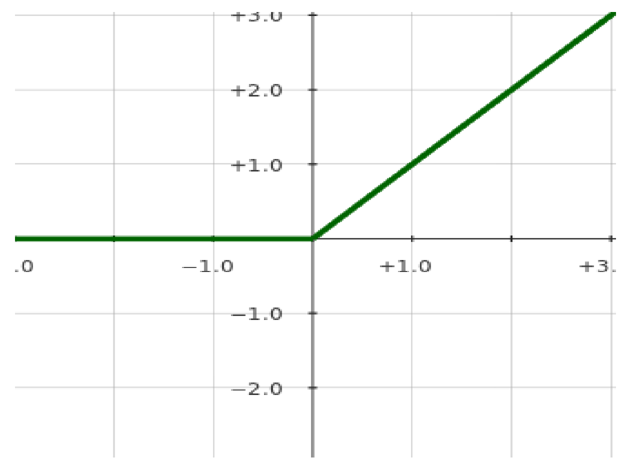


$x = w_1 \mathrm{f}(z_1) + w_2 \mathrm{f}(z_2) + w_3 \mathrm{f}(z_3)$

x is called the total input to the neuron, and f(x) is its output

But one can use any nonlinear function

f(x) = max(0, x)

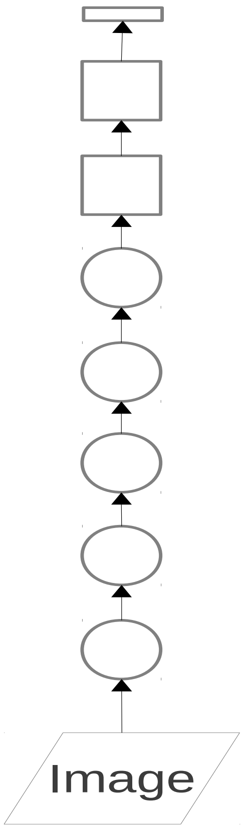


Rectified Linear Unit (ReLU)

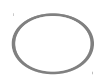
Very good (quick to train)

"AlexNet" 2012

Overview of our model



- **Deep**: 7 hidden “weight” layers
- **Learned**: all feature extractors initialized at white Gaussian noise and learned from the data
- Entirely supervised
- **More data = good**

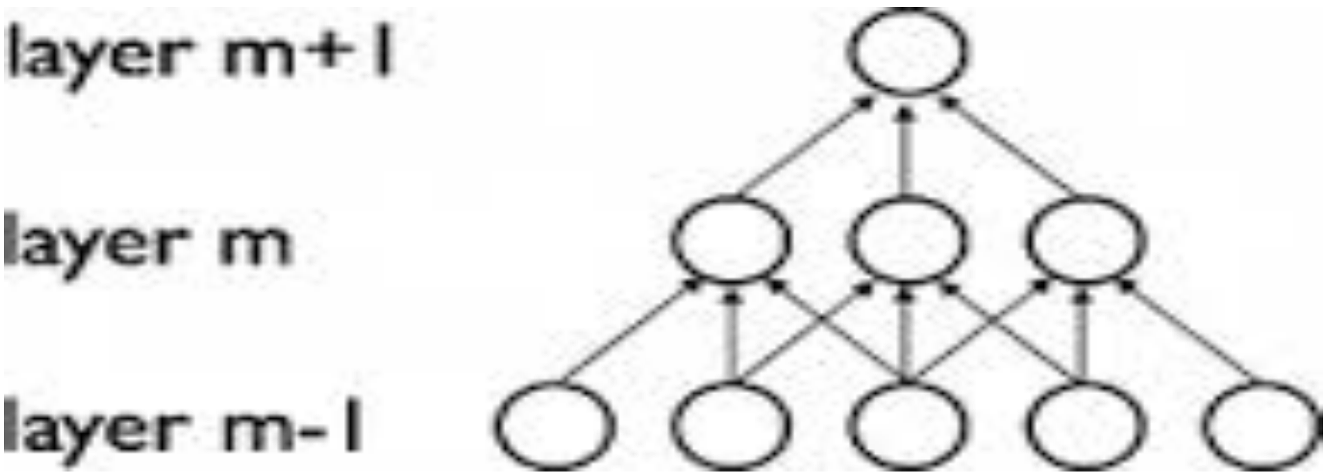


Convolutional layer: convolves its input with a bank of 3D filters, then applies point-wise non-linearity

Fully-connected layer: applies linear filters to its input, then applies point-wise non-linearity

“AlexNet” 2012

Local receptive fields

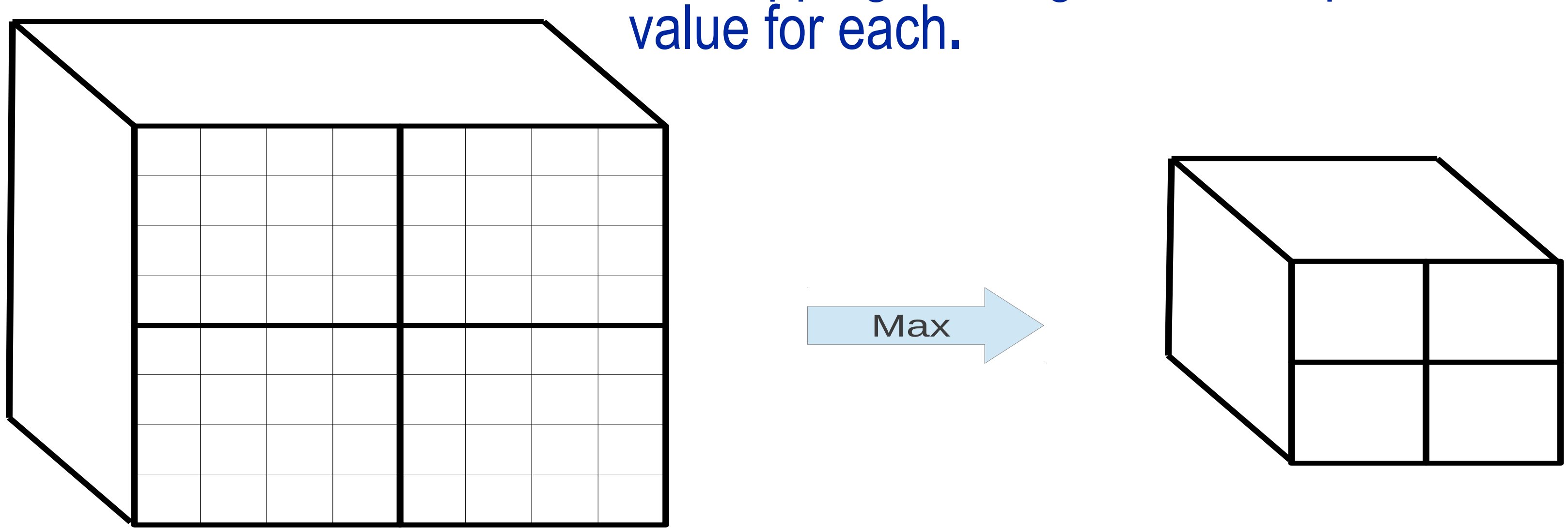


In vision, a neuron may only get inputs from a limited set of "nearby" neurons

"AlexNet" 2012

Local pooling

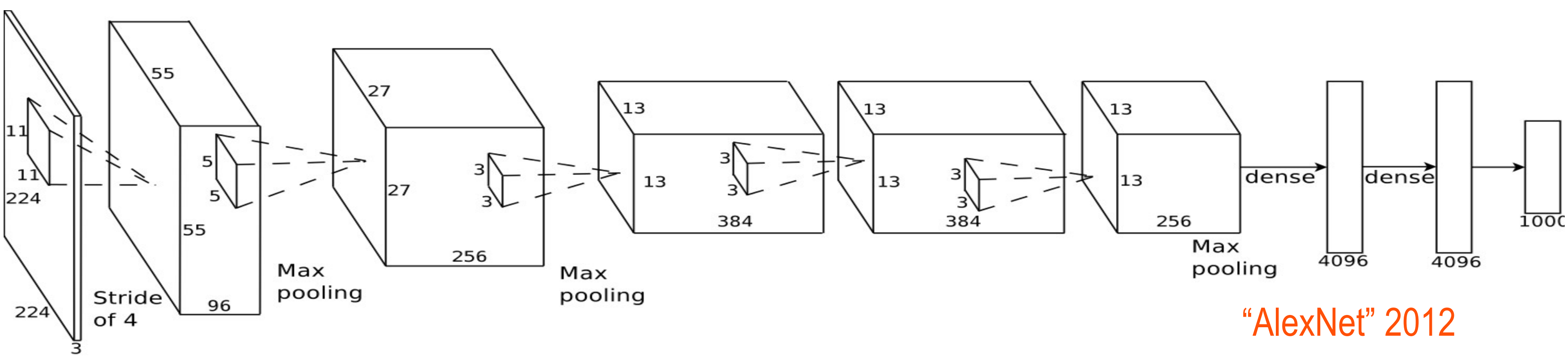
Max-pooling partitions the input image into non-overlapping rectangles and outputs the maximum value for each.



Reduces the computational complexity
Provides translation invariance.

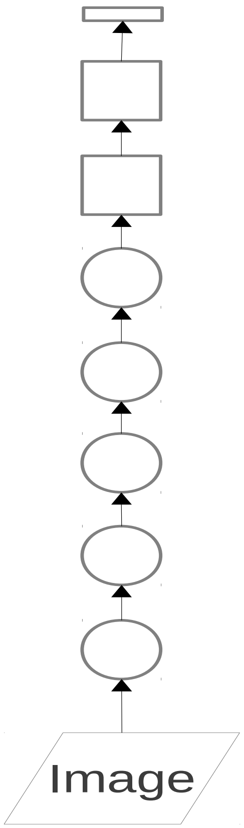
Our model

- Max-pooling layers follow first, second, and fifth convolutional layers
- The number of neurons in each layer is given by 253440, 186624, 64896, 64896, 43264, 4096, 4096, 1000

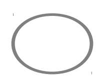


“AlexNet” 2012

Overview of our model



- Trained with stochastic gradient descent on two NVIDIA GPUs for about a week
- 650,000 neurons
- 60,000,000 parameters
- 630,000,000 connections
- **Final feature layer:** 4096-dimensional



Convolutional layer: convolves its input with a bank of 3D filters, then applies point-wise non-linearity

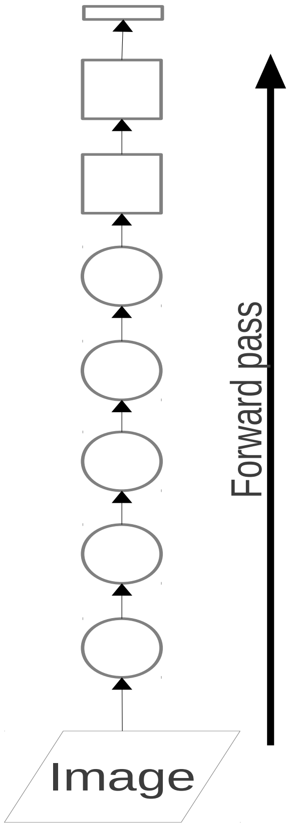
Fully-connected layer: applies linear filters to its input, then applies point-wise non-linearity

“AlexNet” 2012

Local convolutional filters

Fully-connected filters

Training



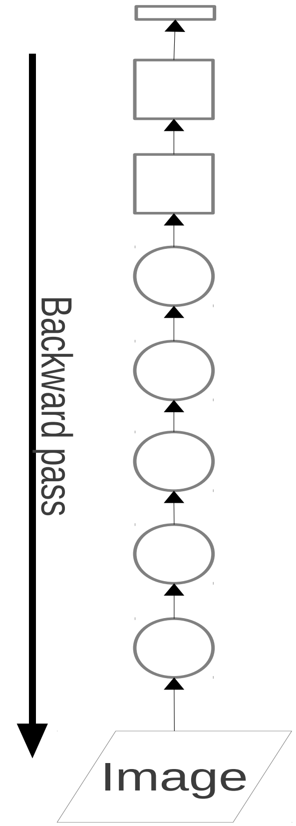
Using stochastic gradient descent and the *backpropagation algorithm* (just repeated application of the chain rule)

One output unit per class

x_i = total input to output unit i

$$f(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{1000} \exp(x_j)}$$

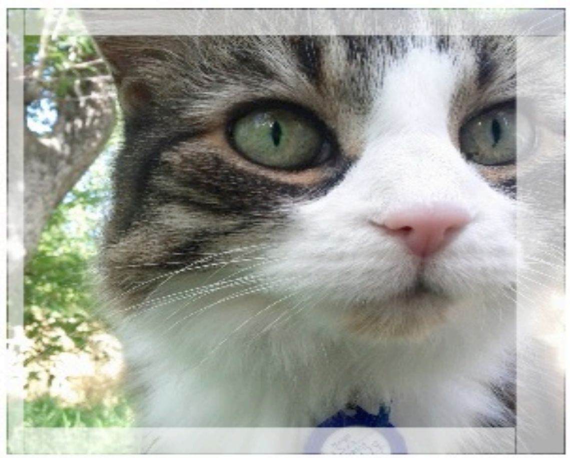
We maximize the log-probability of the correct label, $\log f(x_t)$



"AlexNet" 2012

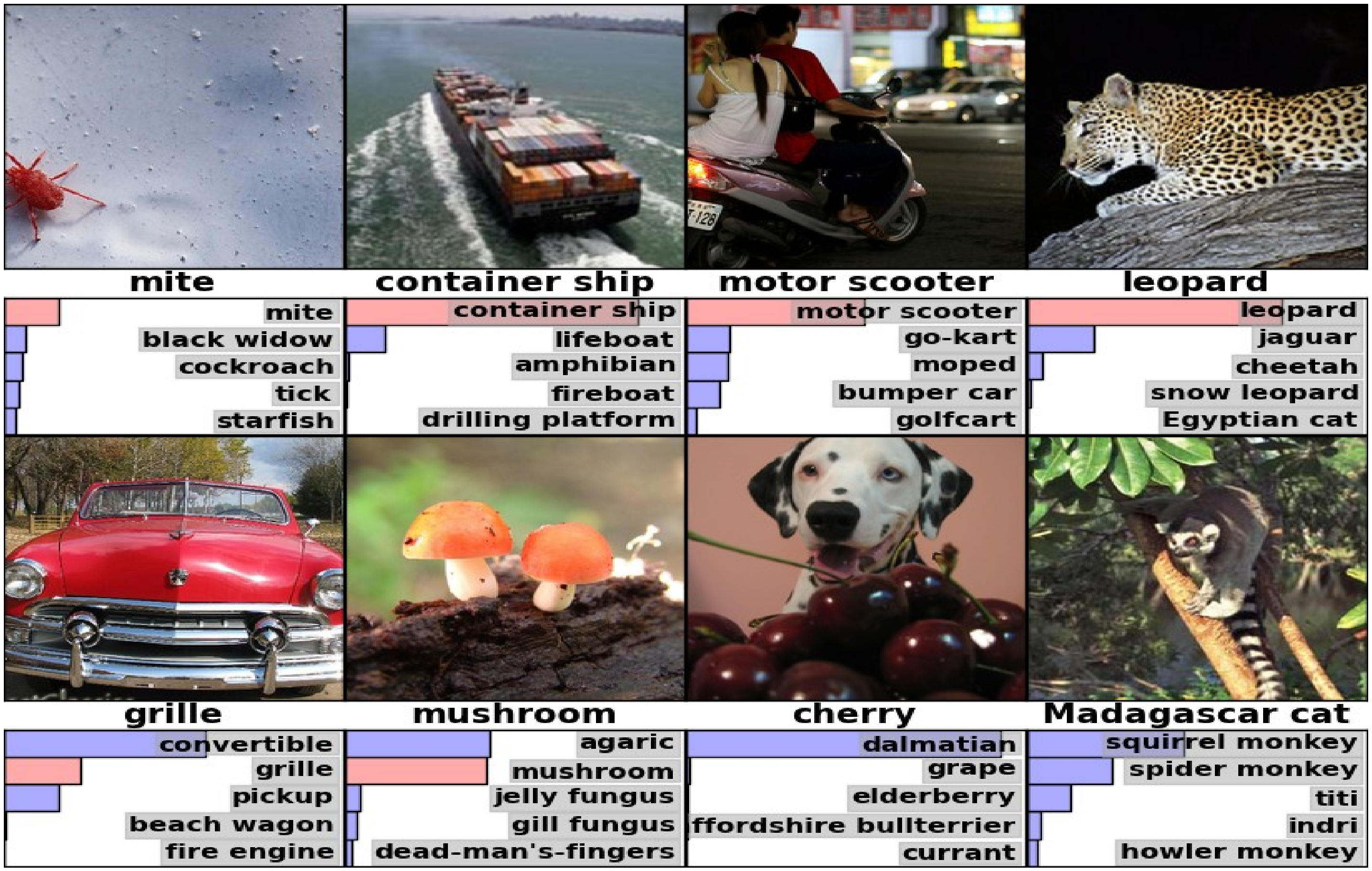
Data augmentation

- Our neural net has 60M real-valued parameters and 650,000 neurons
- It overfits a lot. Therefore we train on 224x224 patches extracted randomly from 256x256 images, and also their horizontal reflections.

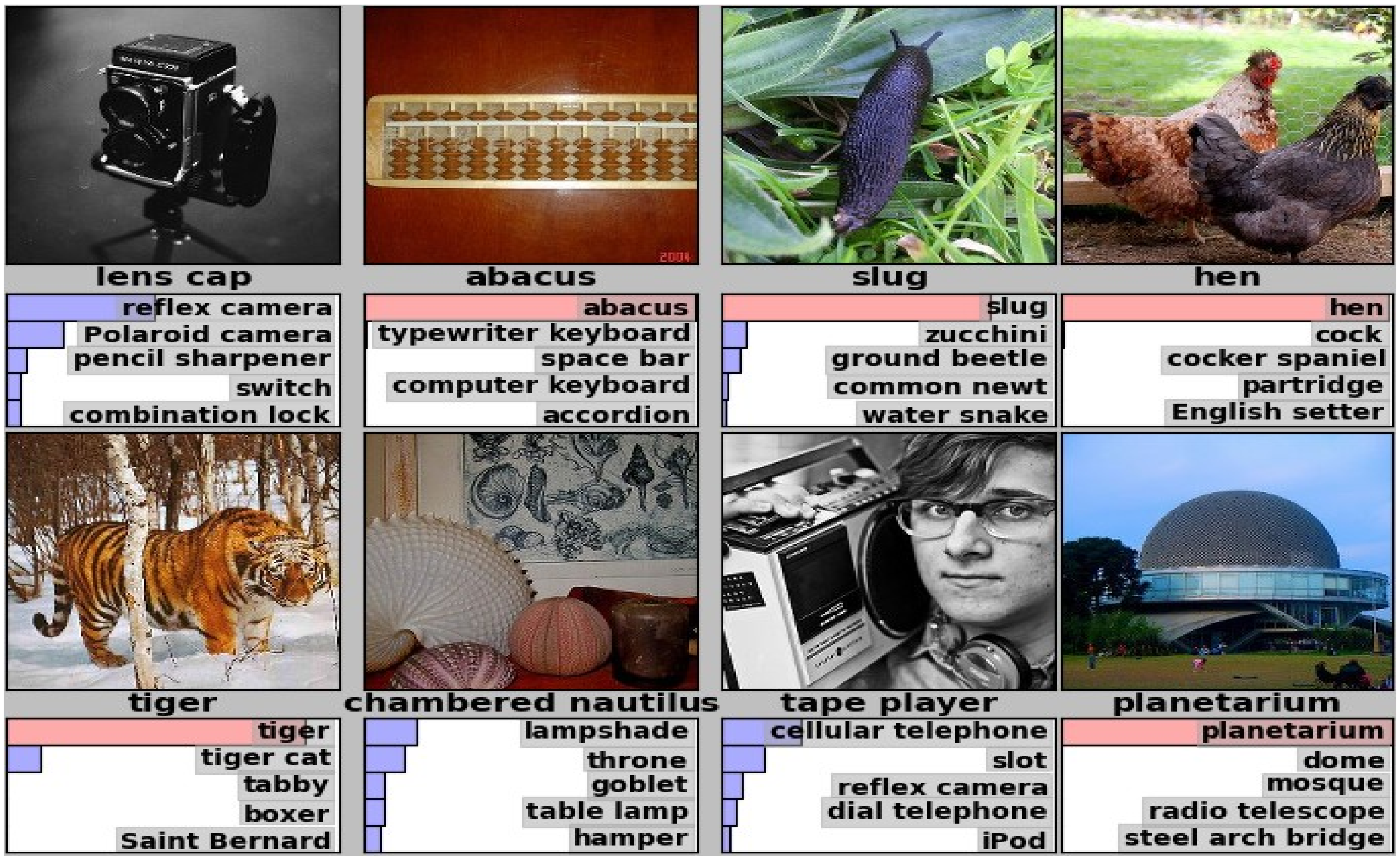


"AlexNet" 2012

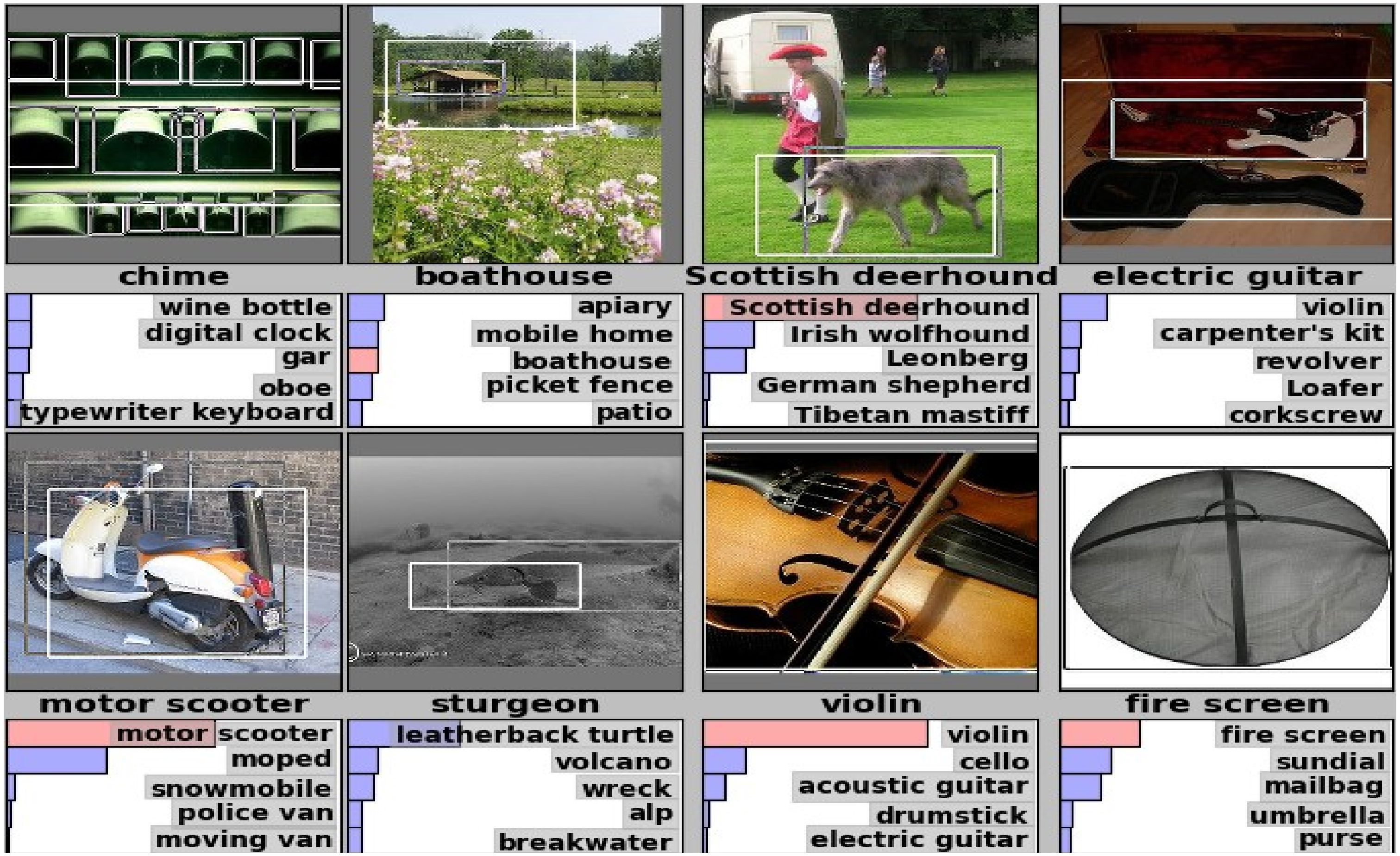
Validation classification



Validation classification

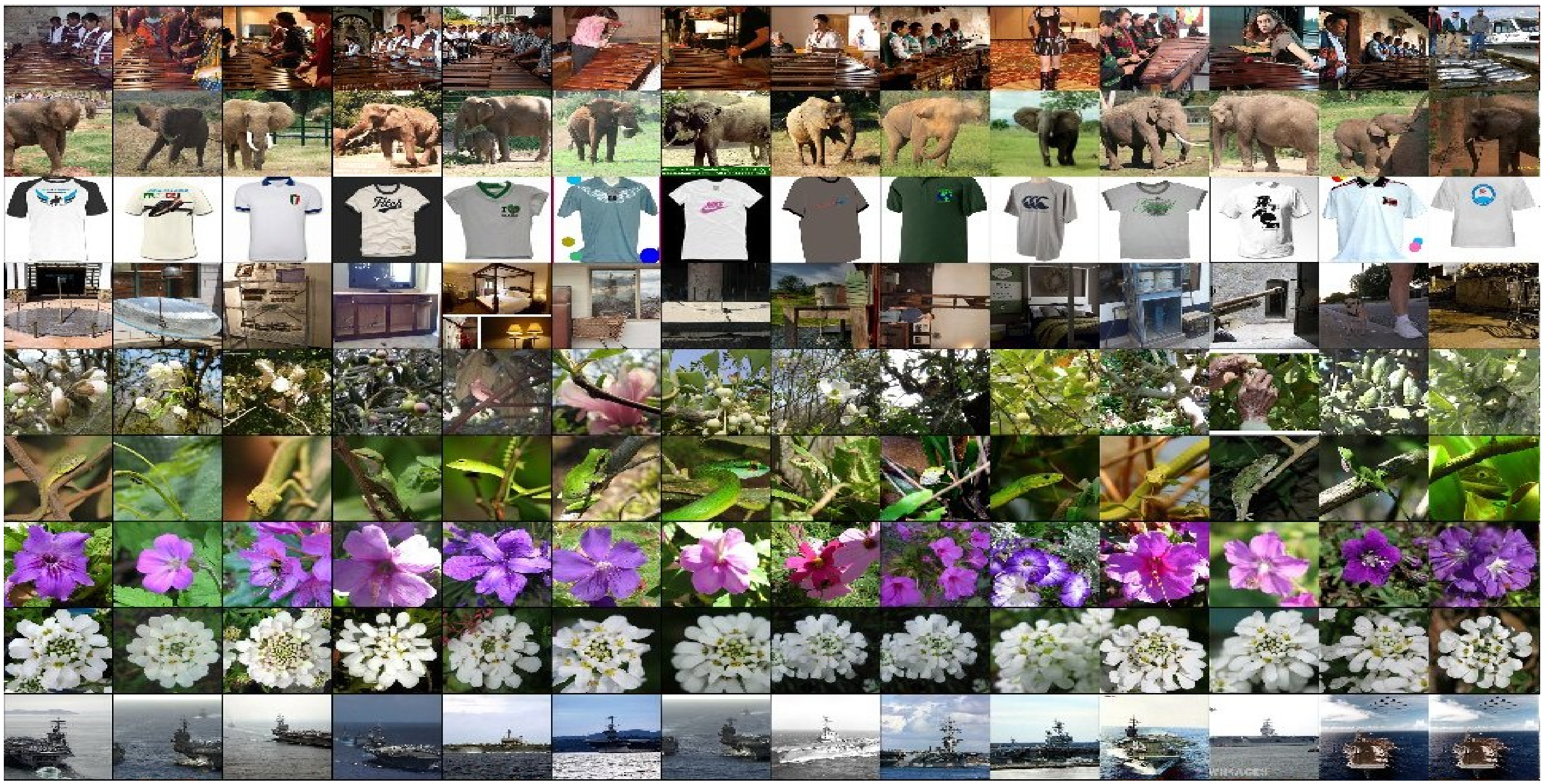


Validation localizations



Retrieval experiments

First column contains query images from ILSVRC-2010 test set, remaining columns contain retrieved images from training set.



Now used for image search; Benefit: good generalization

Both recognized as “meal”

Jeff Dean, google

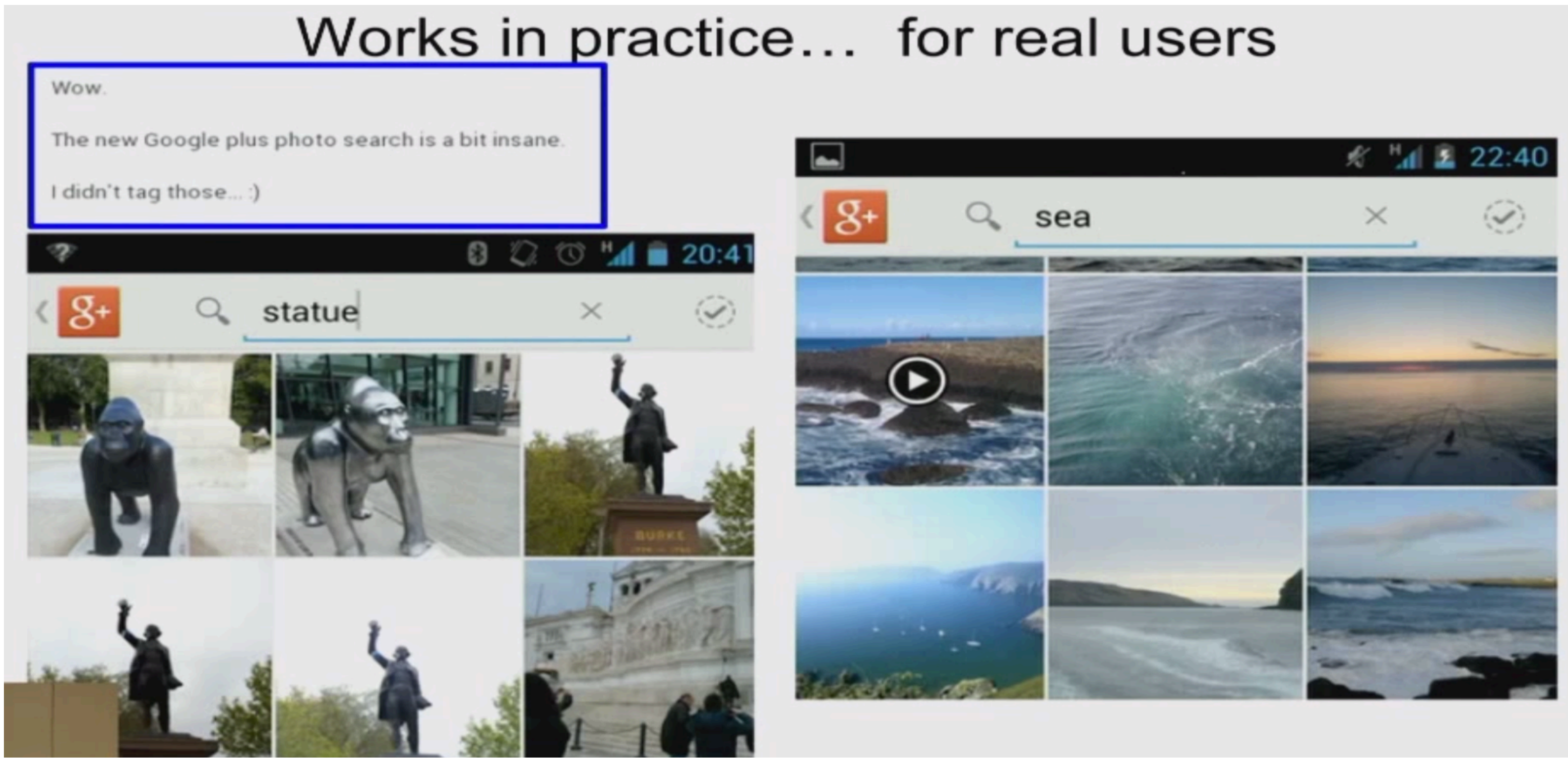
Sensible errors (sometimes)

“snake”

“dog”

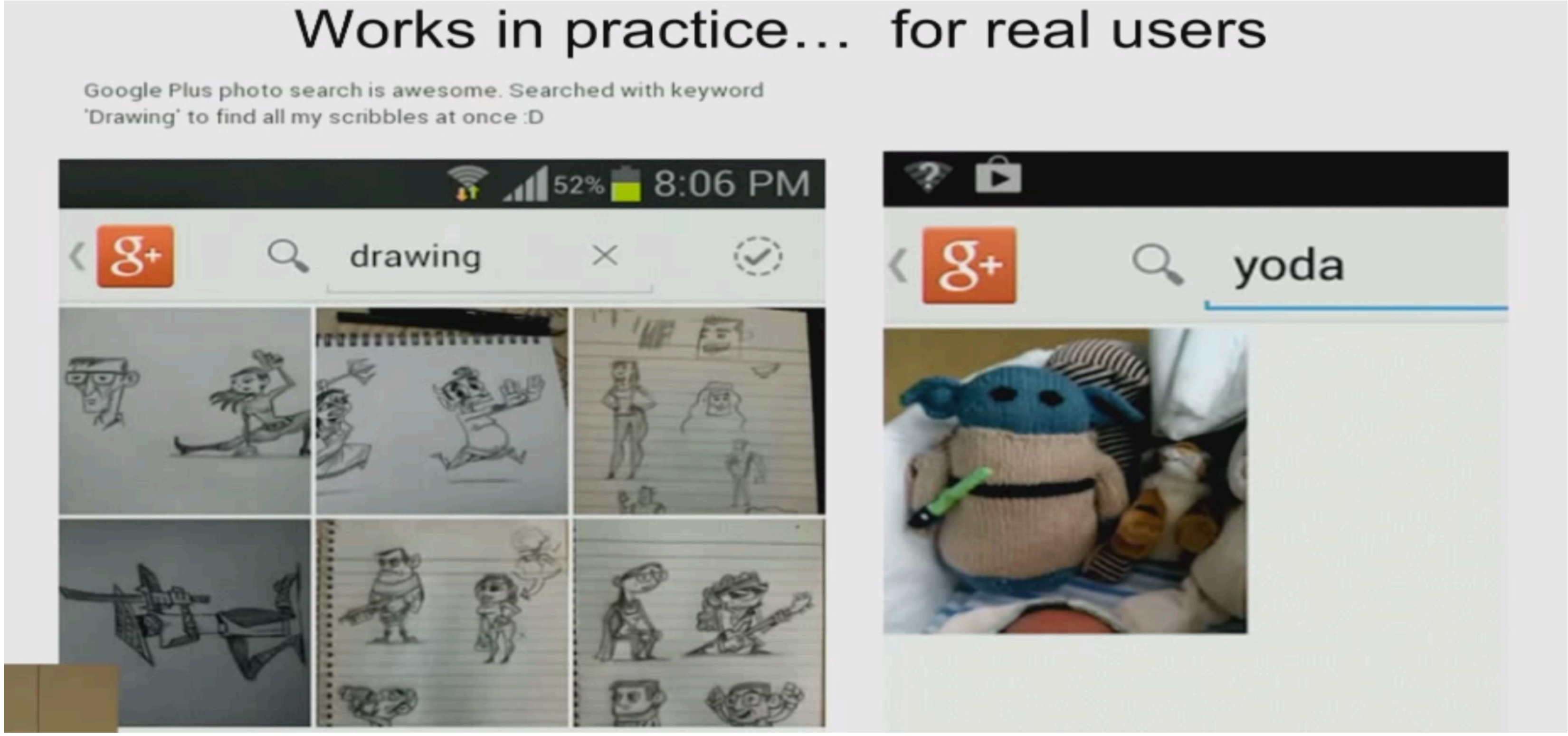
Jeff Dean, google

Now used for image search



Jeff Dean, google

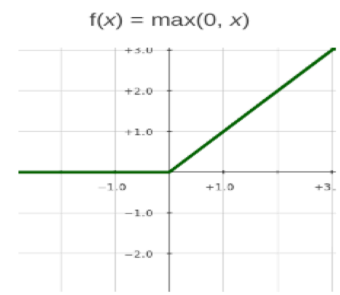
Now used for image search



Jeff Dean, google

Modern deep nets

- **Often use rectified linear units (RLUs)**
  - Less problems of saturation than logistic



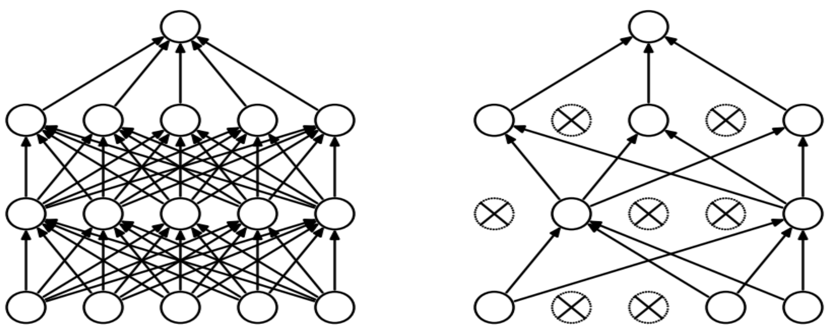
- **Use a variety of loss functions**
  - Log likelihood (uses *softmax*)

$$p(y = j|x) = \frac{e^{w_j^\top x}}{\sum_k e^{w_k^\top x}}$$

- **Can be very deep**
- **Solved with mini-batch gradient descent**
- **Regularized using L_2 penalty plus "dropout"**
  - and partial convergence and ..

Dropout

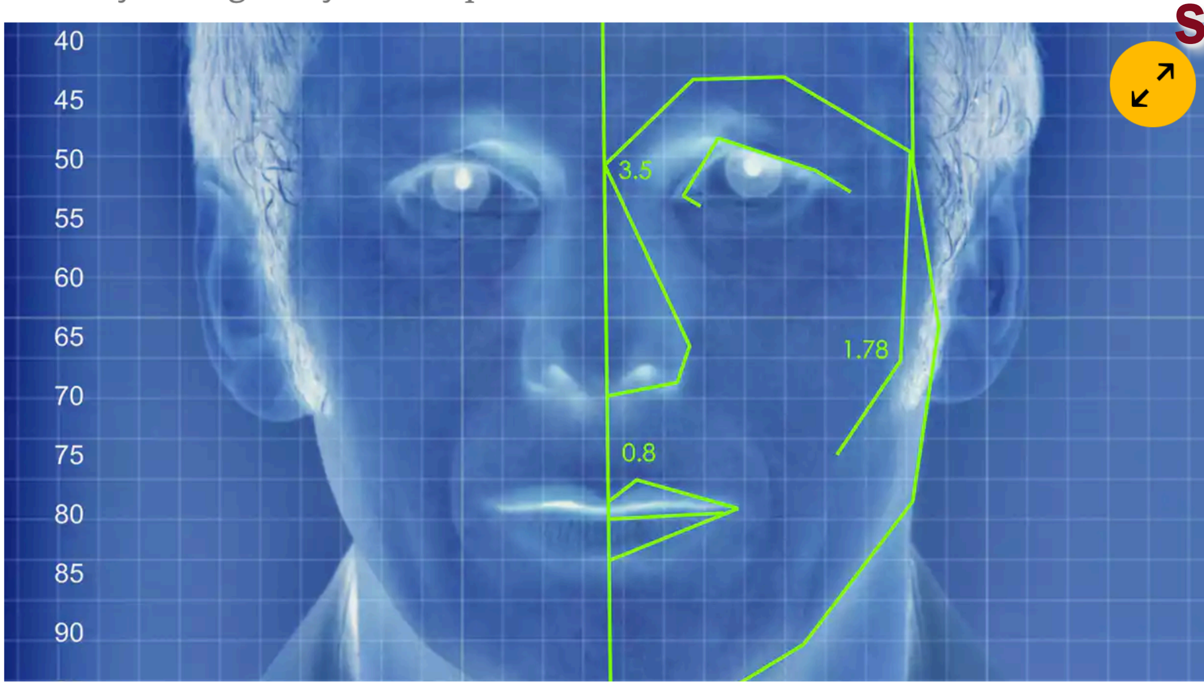
- Randomly (temporarily) remove a fraction *p* of the nodes (with replacement)
  - Usually p = 1/2
- Repeatedly doing this samples (in theory) over exponentially many networks
  - Bounces the network out of local minima
- For the final network use all the weights but shrink them by *p*



(a) Standard Neural Net (b) After applying dropout.

New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions



Deep neural networks are more accurate than humans at detecting sexual orientation from facial images

Michal Kosinski & Yilun Wang 2017

https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph

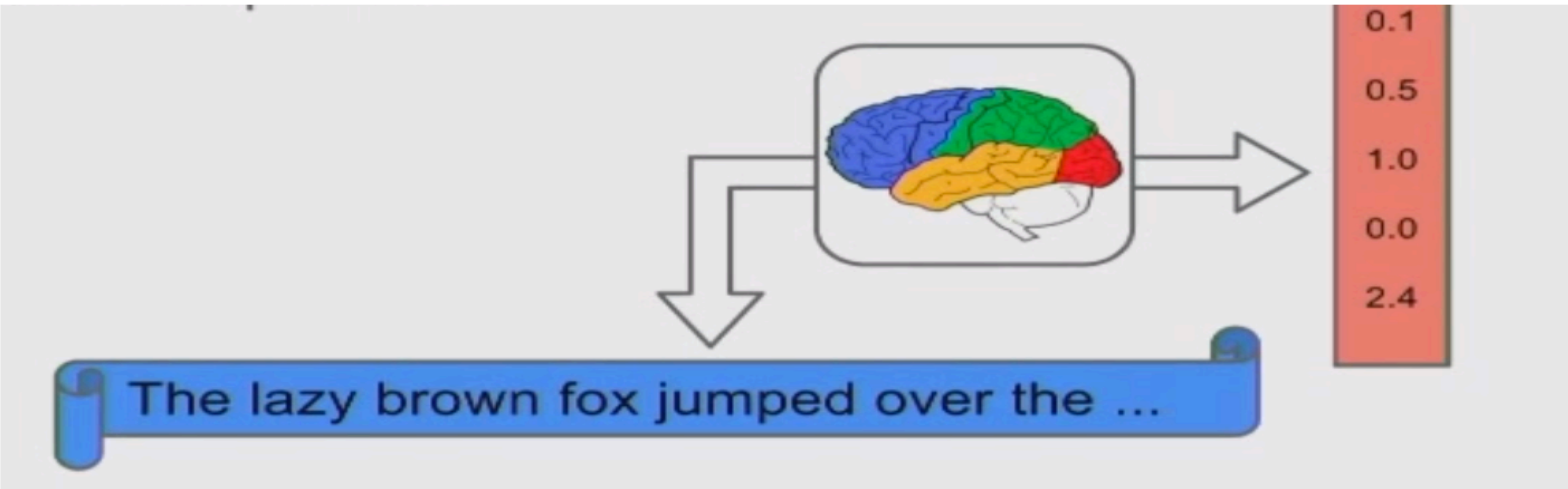
Detecting sexual orientation – semi-supervised learning

- Download images and labels from a dating site
  - where people declare their sexual orientation
  - keep images with a single “good” Caucasian face
- Use pretrained CNN to compute ~ 4,000 ‘scores’/image
  - VGG-Face was trained on 2.6 million faces
- Use logistic regression on PCA of the scores to predict orientation

Recurrent Neural Nets

- **Generalize Hidden Markov Models (HMMs)**
- **Predict the next observation given the past observations**
- **Or can map one sequence to another sequence**
  - **An encoder**
    - sentence (sequence of words) to vector
  - **A decoder**
    - vector to sentence (sequence of words)

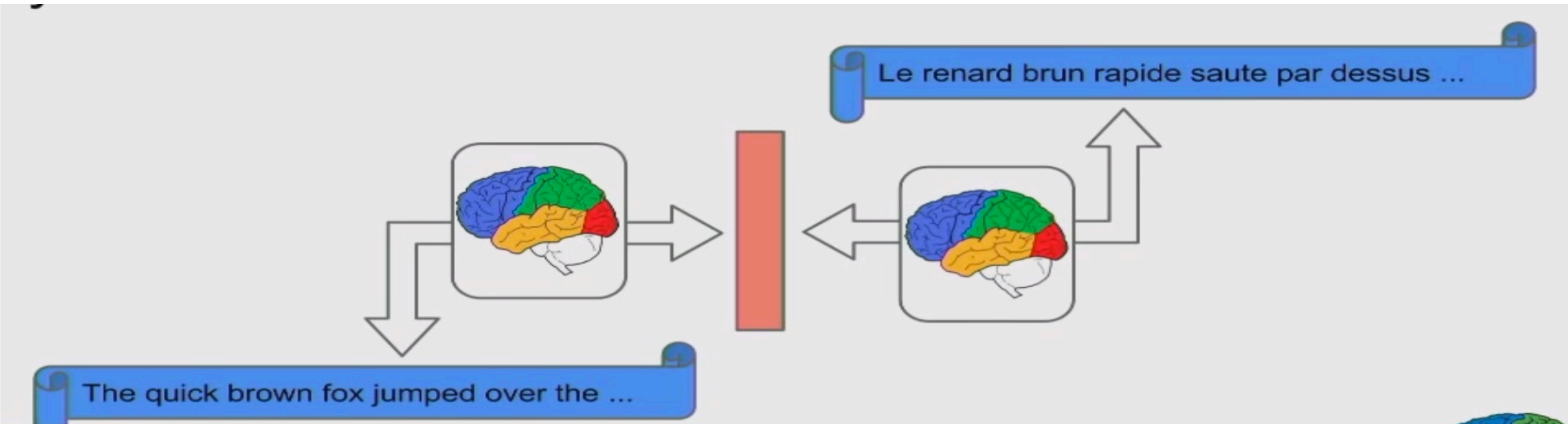
LSTM encodes a sentence



Jeff Dean, google

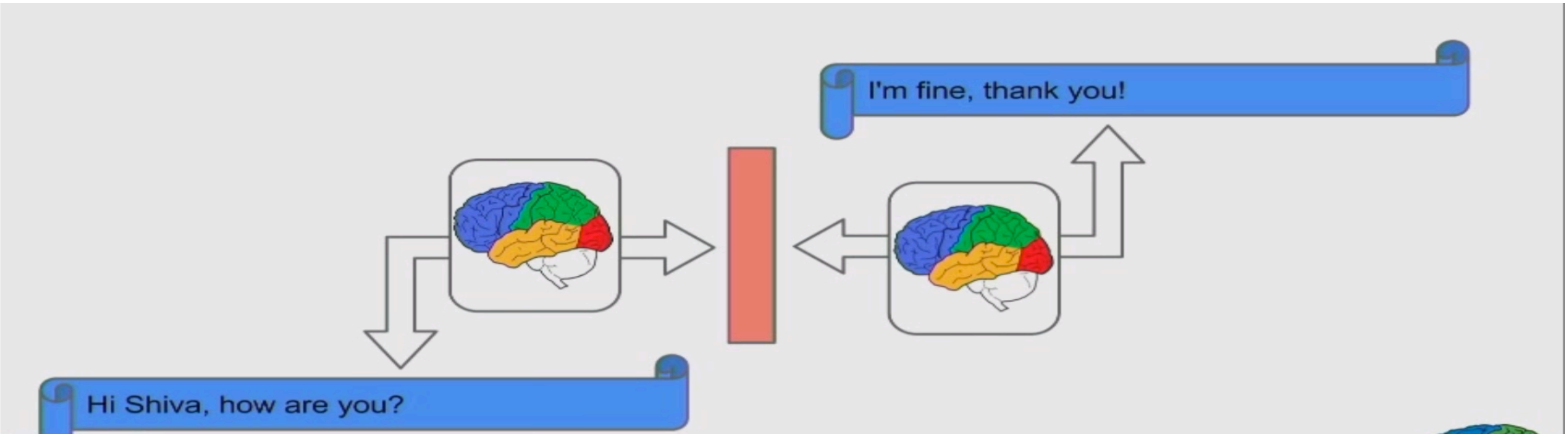
https://www.youtube.com/watch?v=90-S1M7Ny_o&spfreload=1

Encode and Decode = translate



Jeff Dean, google

… or a chatbot



Jeff Dean, google

It works (sort of)

User> hello , i have a problem with my machine

BrainStop> hi

BrainStop> this is shiva .

BrainStop> how are you doing today ?

User> hi shiva , how are you ?

BrainStop> i ' m fine , thank you !

BrainStop> how may i assist you today ?

User> i want to access using vpn

BrainStop> currently the computer is connected to corp network ?

User> no

BrainStop> check this solution :

BrainStop> https : / / INTERNAL_HOSTNAME_REDACTED / articles / connecting - to - disconnecting - from - ssl - vpn # heading - 4

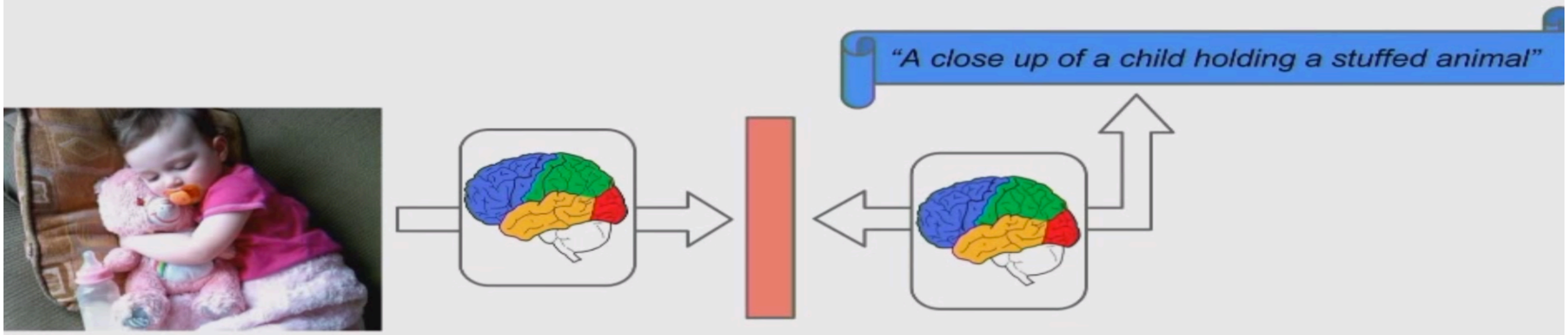
User> thanks , bye !

BrainStop> thank you .

Jeff Dean, google

Can map images to words

Connect sequence and image models, you get a captioning system

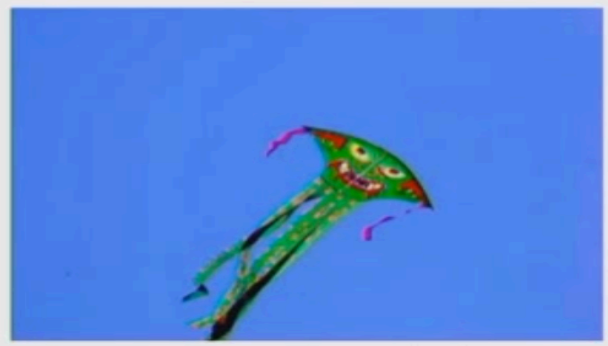


It works (sort of)

A man holding a tennis racquet on a tennis court.

Two pizzas sitting on top of a stove top oven

A group of young people playing a game of Frisbee



A man flying through the air while riding a snowboard

Jeff Dean, google

Reinforcement learning

- **Train a model to take actions that maximize a ‘reward’**
  - Instead of predicting a response
- **Learn to play go**
- **Learn to play a video game by trial and error**
  - Given only the pixels on the screen
- **Now used for reducing energy consumption in data centers.**

Deep learning in engineering

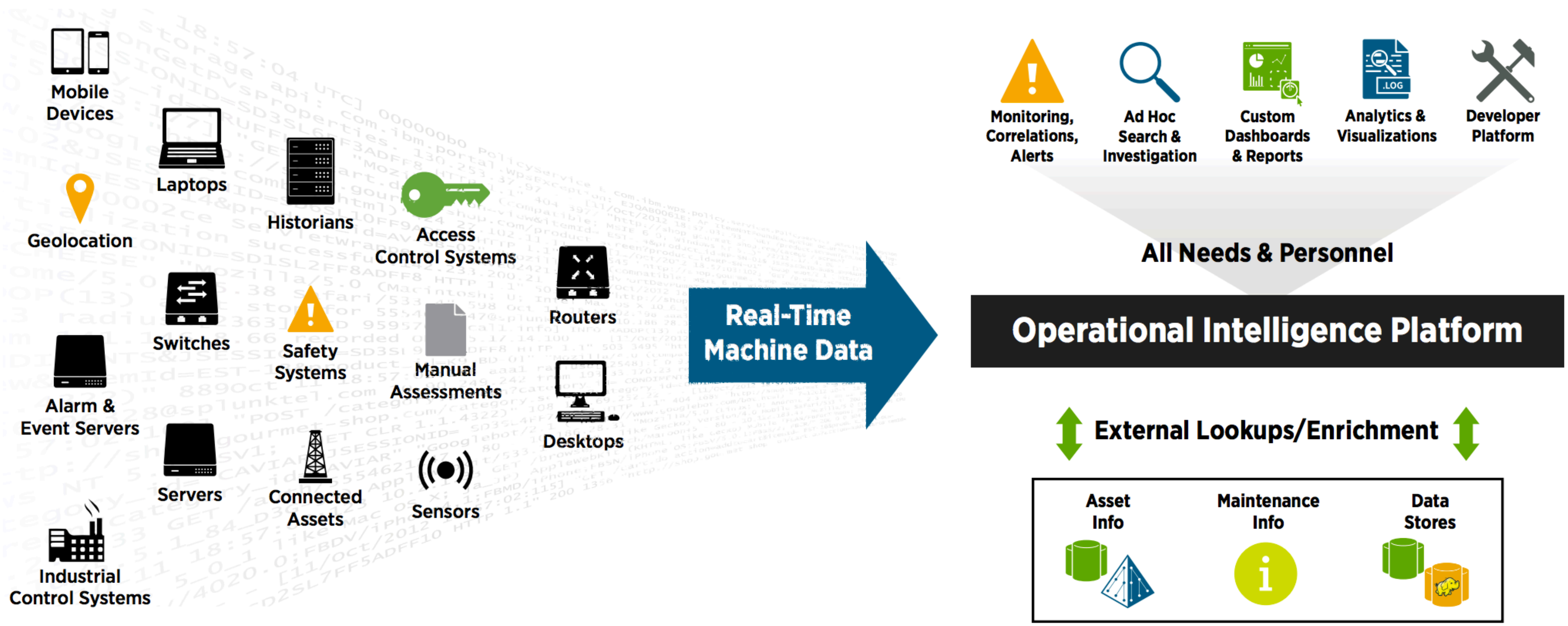
- **Robotics**
- **Soft sensors**
  - Viscosity, corrosion, photodegradation, …
- **Demand estimation**
  - Power usage, sales …

splunk> listen to your data™

2017:
$8.5 billion

2012 IPO:
$3.3 billion

Splunk



Mobile Devices
Laptops
Historians
Geolocation
Access Control Systems
Switches
Safety Systems
Manual Assessments
Routers
Alarm & Event Servers
Desktops
Servers
Connected Assets
Sensors
Industrial Control Systems
Real-Time Machine Data
Monitoring, Correlations, Alerts
Ad Hoc Search & Investigation
Custom Dashboards & Reports
Analytics & Visualizations
Developer Platform
All Needs & Personnel
Operational Intelligence Platform
External Lookups/Enrichment
Asset Info
Maintenance Info
Data Stores

Take-aways

- **Neural nets are just *very* flexible models**
  - with some structure imposed
  - and lots of regularization
- **They have revolutionized machine vision, speech recognition, translation, ⋯**
  - And soon engineering?
- **Training by example**
  - not by modeling or programming