Statistical artificial intelligence

Ole Winther

Dept for Applied Mathematics and Computer Science
Technical University of Denmark (DTU)

Bioinformatics Centre/BRIC
University of Copenhagen (KU)

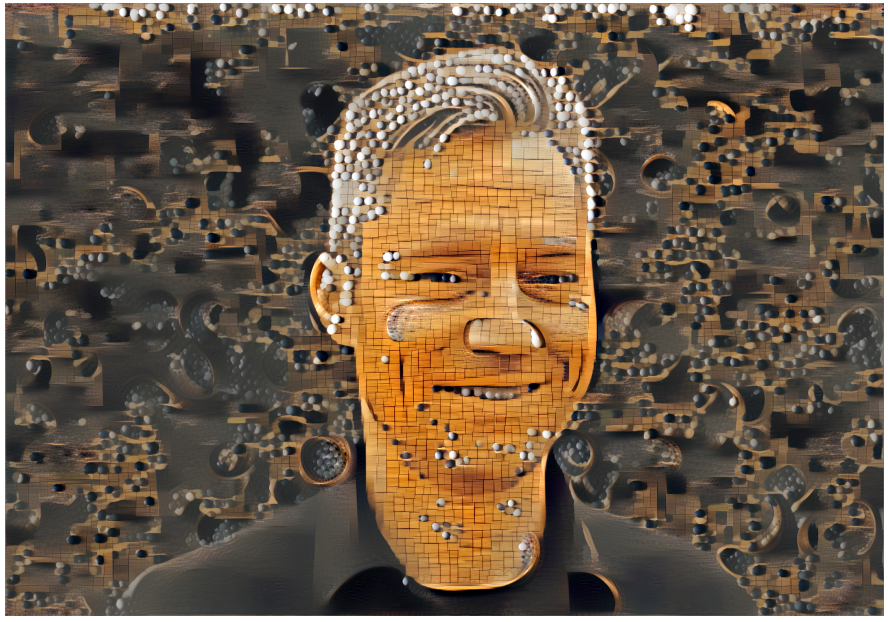
November 24, 2016

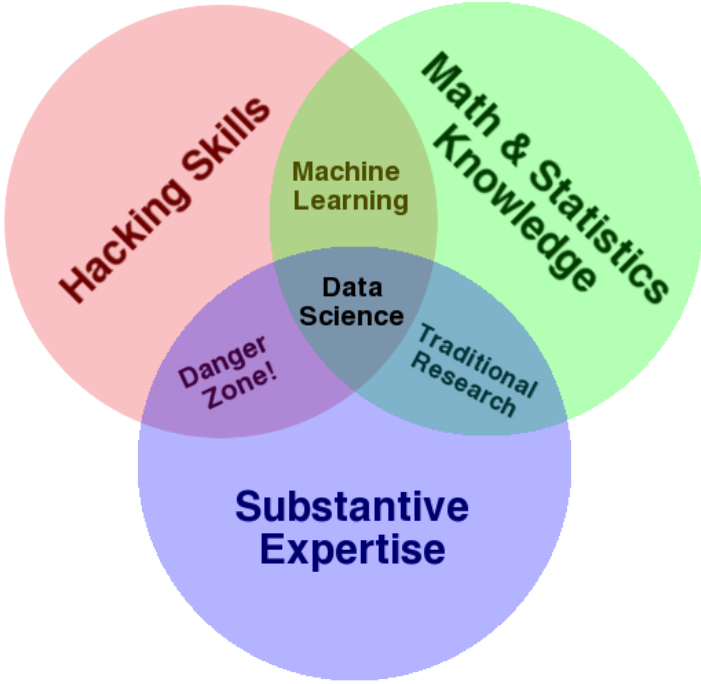
Objectives of talk

- What is statistical artificial intelligence?
- The deep learning hammer
- Case stories
- Demystifying DeepMind's reinforcement learning

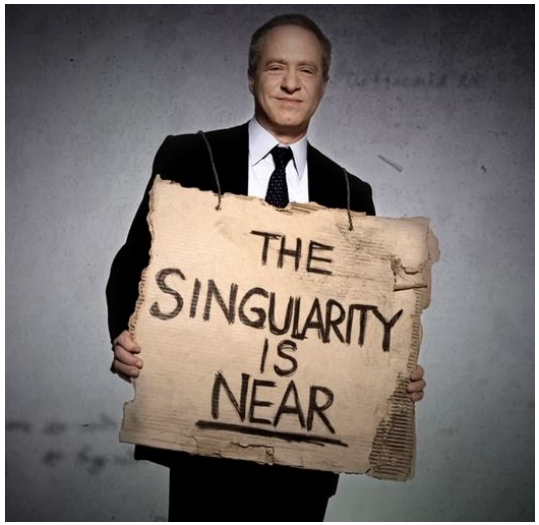
Objectives of talk

- What is statistical artificial intelligence?
- The deep learning hammer
- Case stories
- Demystifying DeepMind's reinforcement learning
- A bit about myself



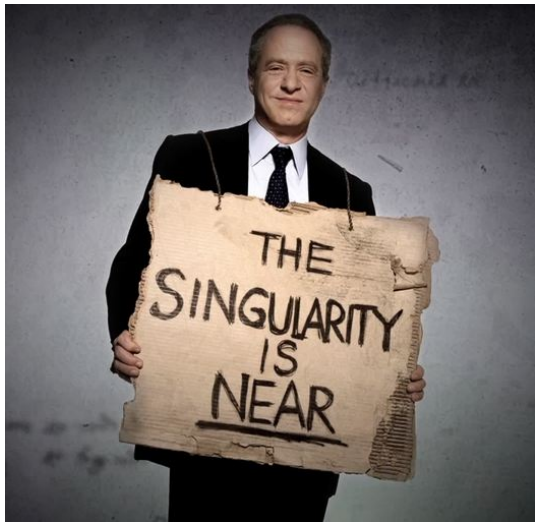


Are we heading towards the singularity?



kurzweilai.net

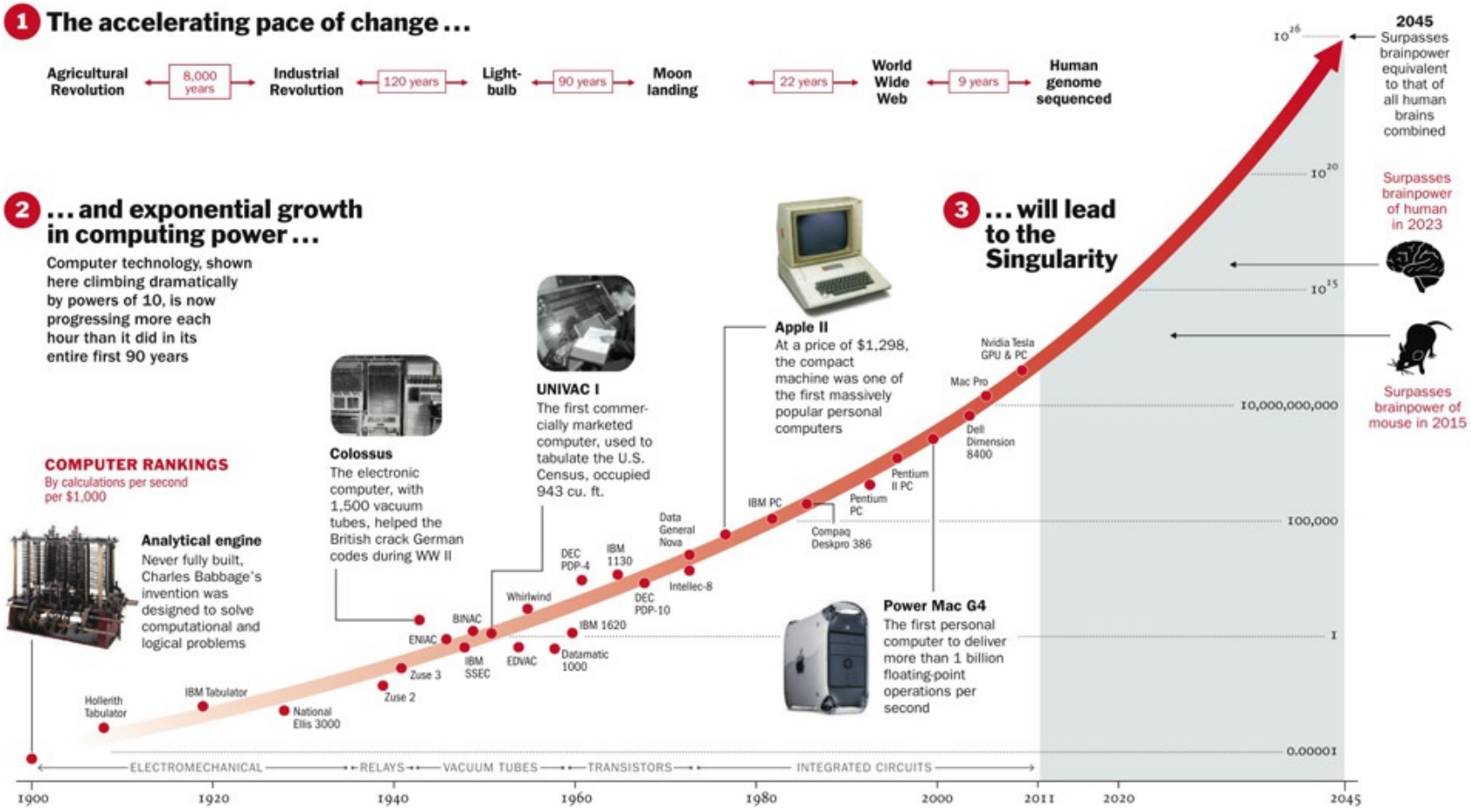
Are we heading towards the singularity?



kurzweilai.net

- Elon Musk at MIT AeroAstro Symp:
- If I were to guess at what our biggest existential threat is, it's probably that...
- With artificial intelligence, we are summoning the demon..
- Inofficial quotes (email to friend):
- The risk of something seriously dangerous happening is in the five year timeframe. 10 years at most,
- Unless you have direct exposure to groups like Deepmind, you have no idea how fast — it is growing at a pace close to exponential.
- mashable.com/2014/11/17/elon-musk-singularity/

Growth in computer power



Using the term artificial intelligence is new to me

Demis Hassabis, DeepMind CEO mission statement: “solving (general artificial) intelligence, and then using that to solve everything else”

Part 2:
Neural network primer

Major areas in AI

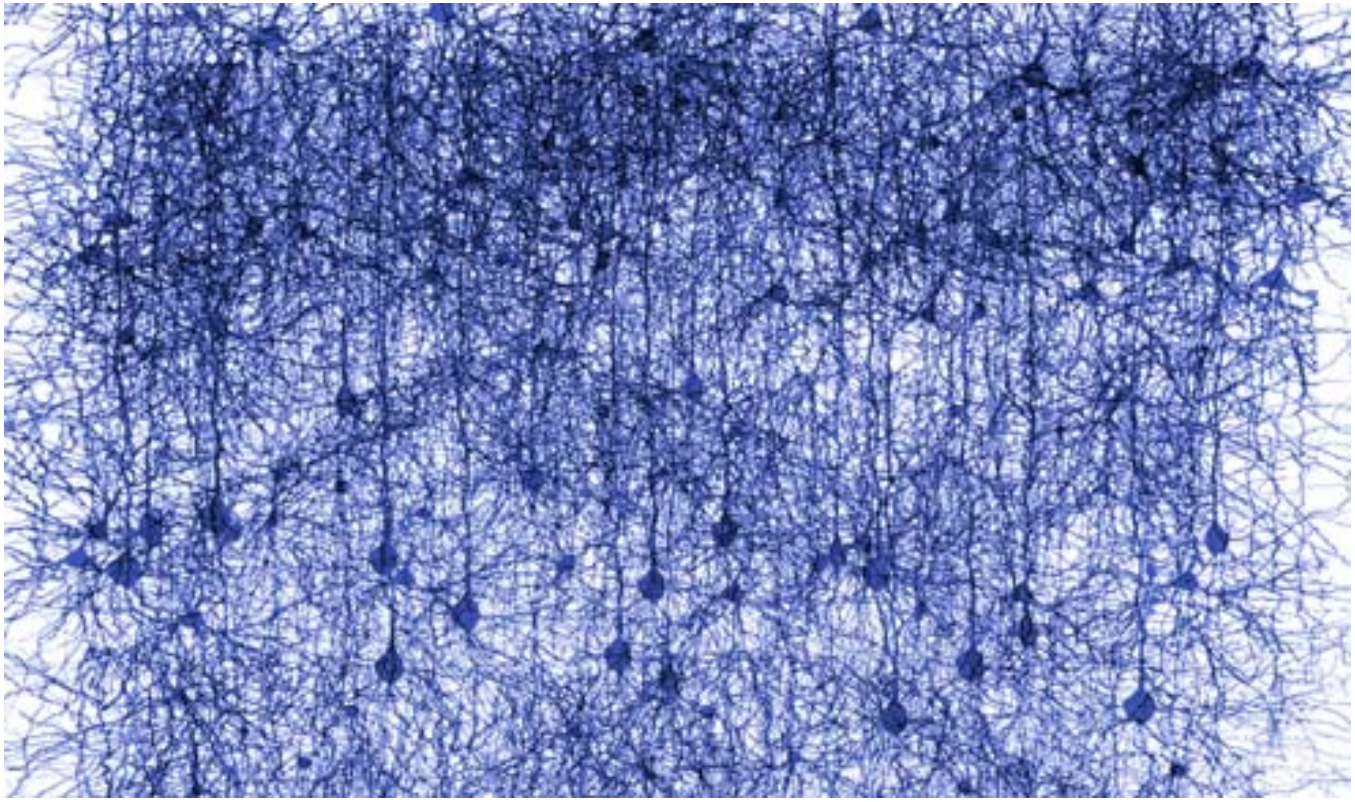
- Speech recognition
- Image classification
- Machine translation
- Question-answering
- Self-driving vehicles
- Dialogue systems
- General unsupervised learning

Major areas in AI

- Speech recognition
- Image classification
- Machine translation
- Question-answering
- Self-driving vehicles
- Dialogue systems
- General unsupervised learning

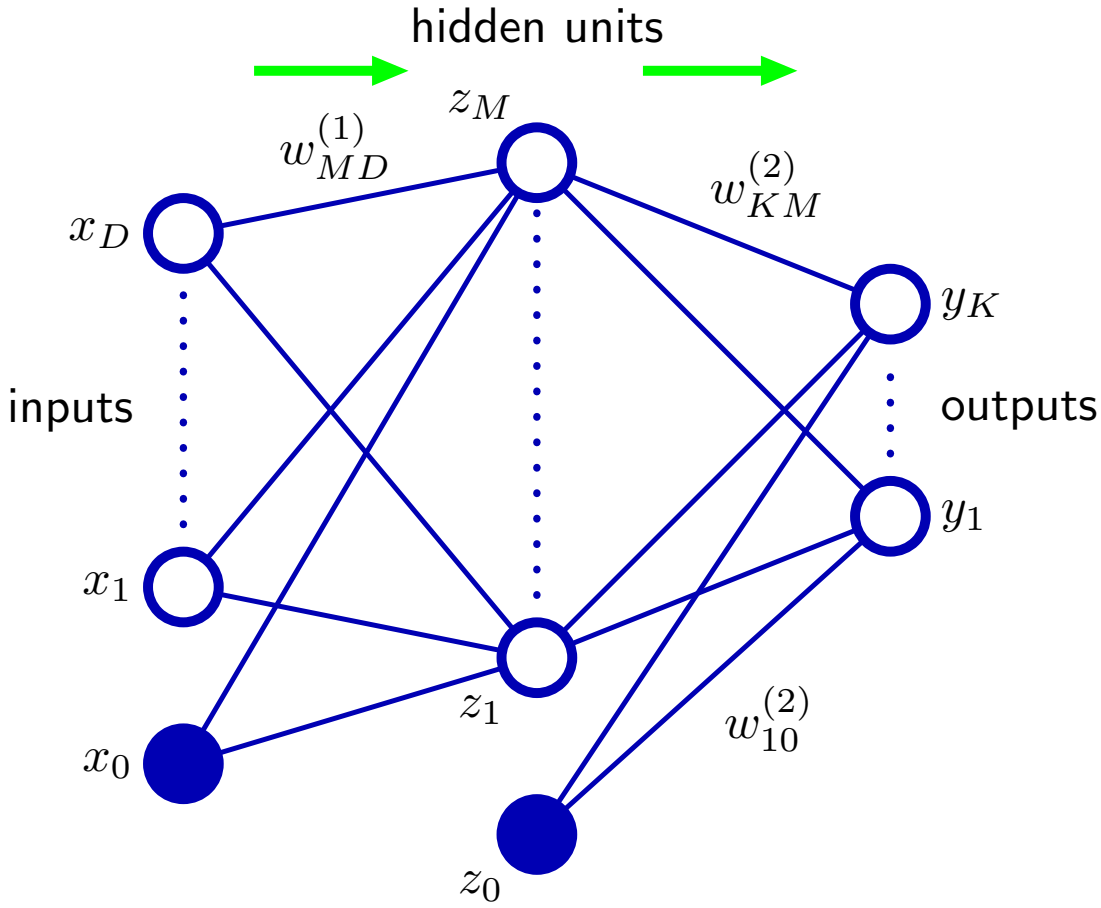
Deep learning

- Feedforward neural networks
- Convolutional neural networks (CNN) - images
- Recurrent neural networks - sequences



Approx. 10^{11} neurons and 10^{14} synapses in a human brain

Feed forward neural networks



$$y_k(\mathbf{x}, \mathbf{w}) = \sigma\left(\sum_{j=0}^{M} w_{kj}^{(2)} \underbrace{f\left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i\right)}_{z_j}\right)$$

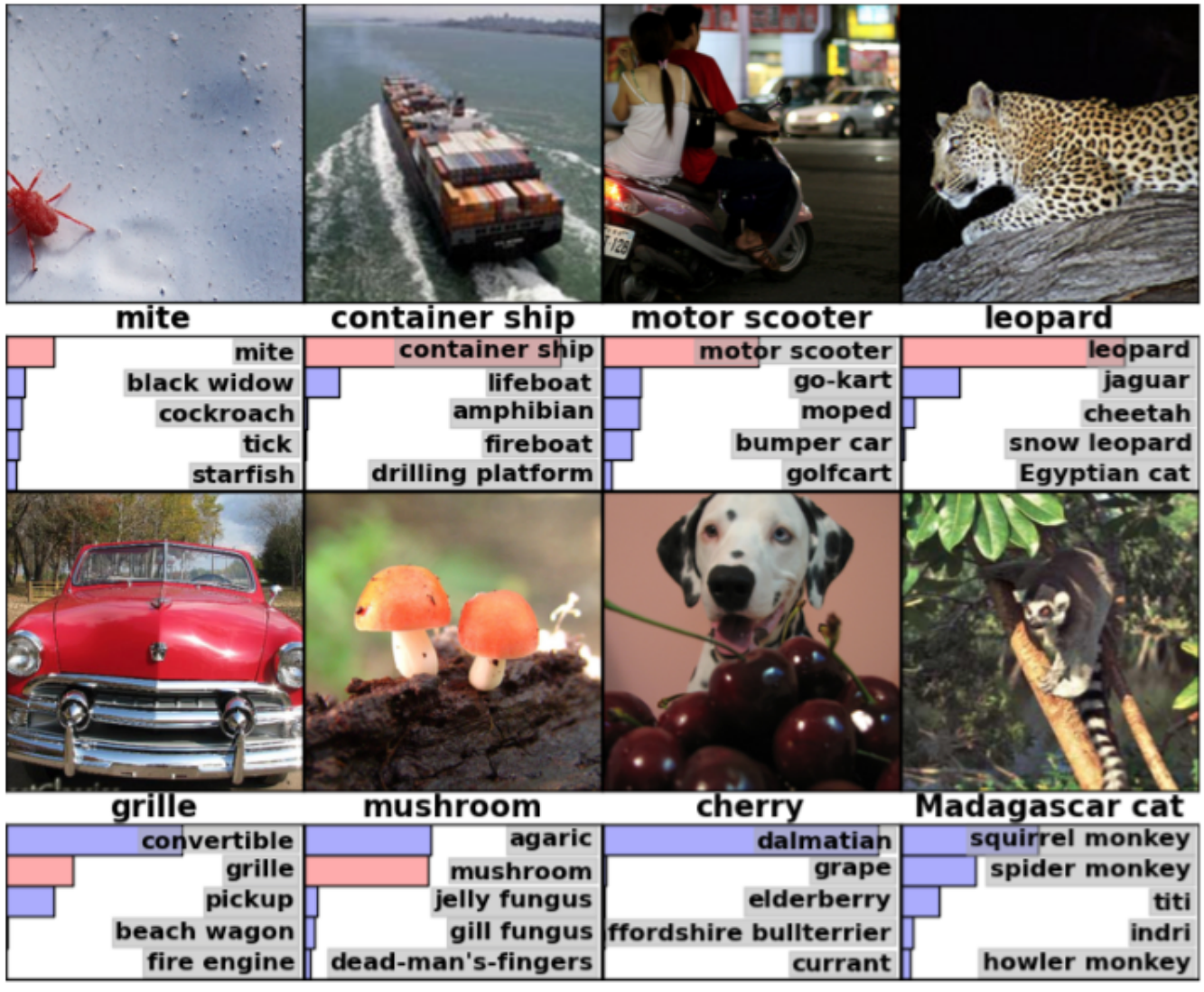
Part 3:

The deep learning revolution - some cases

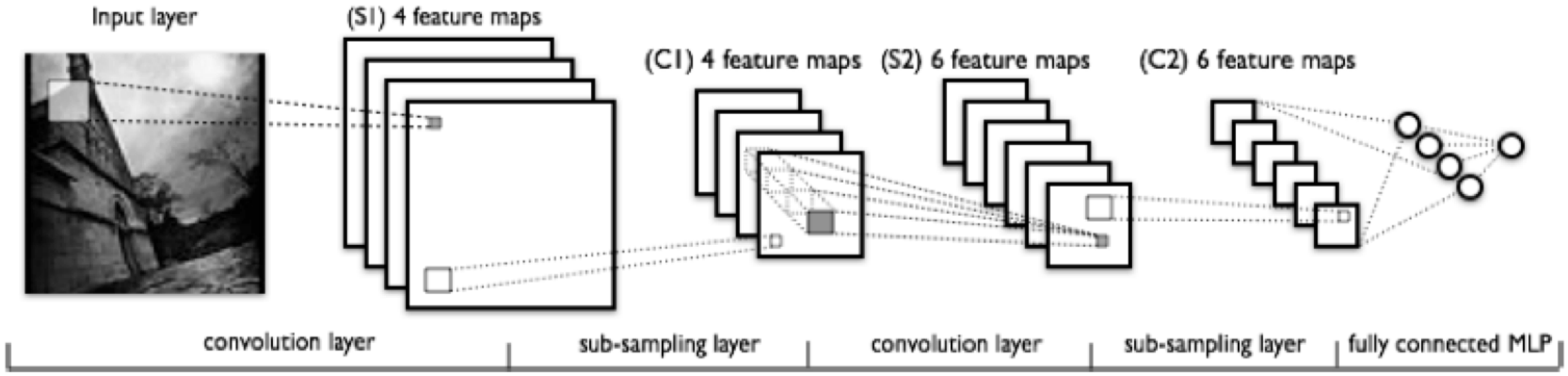
Achilles' heel of traditional AI: Perception in natural environment

xkcd.com/1425

ImageNet - image classification



Convolutional neural networks



$$\begin{bmatrix} 10 & 0 & -10 \\ 0 & 0 & 0 \\ -10 & 0 & 10 \end{bmatrix}$$

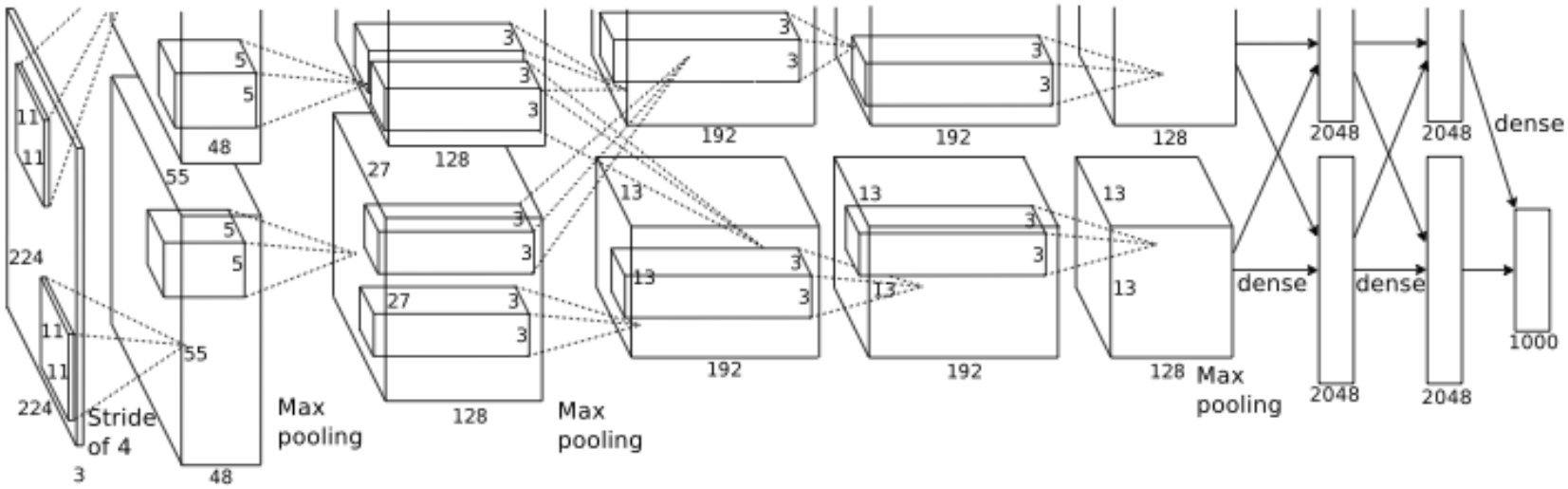
Feature engineering vs engineered models

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

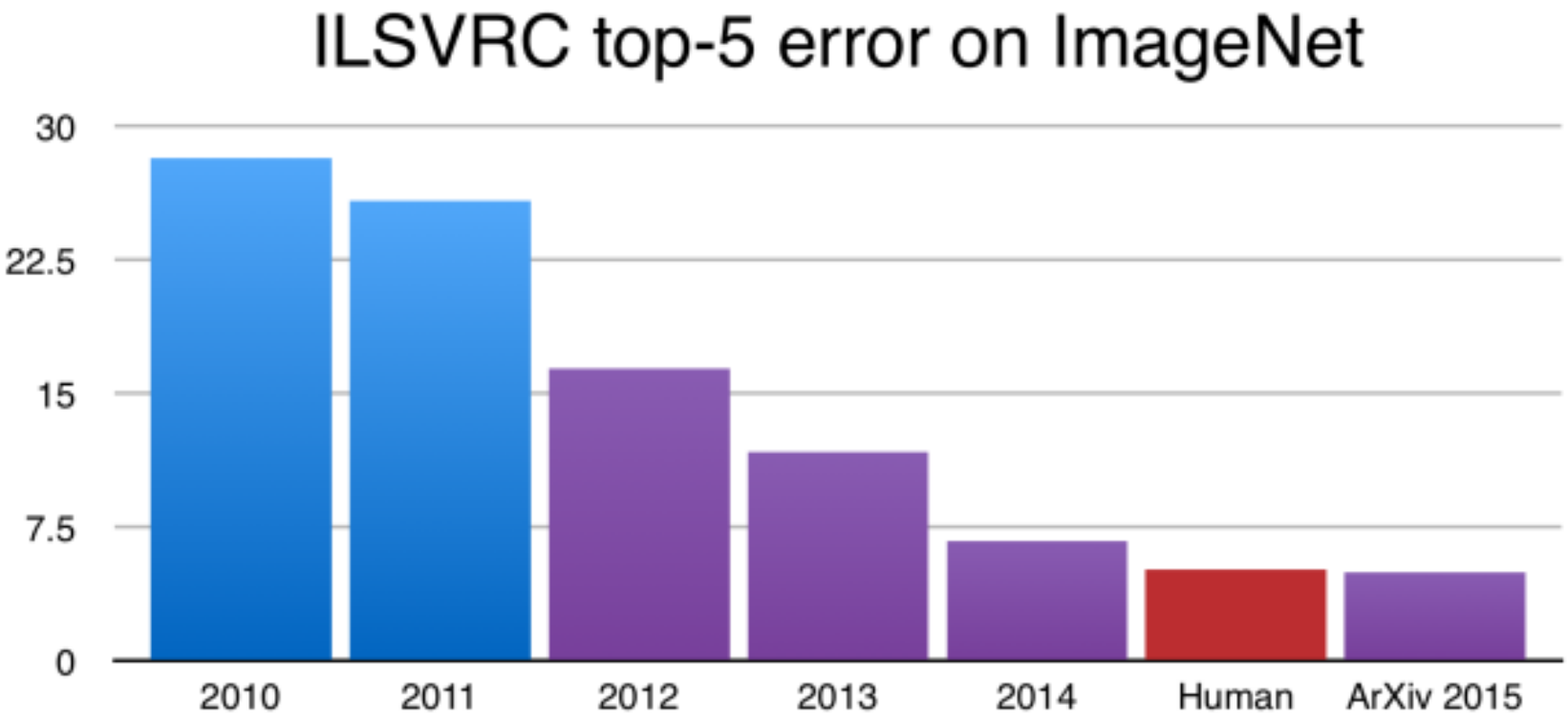
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



www.cs.toronto.edu/~fritz/absps/imagenet.pdf

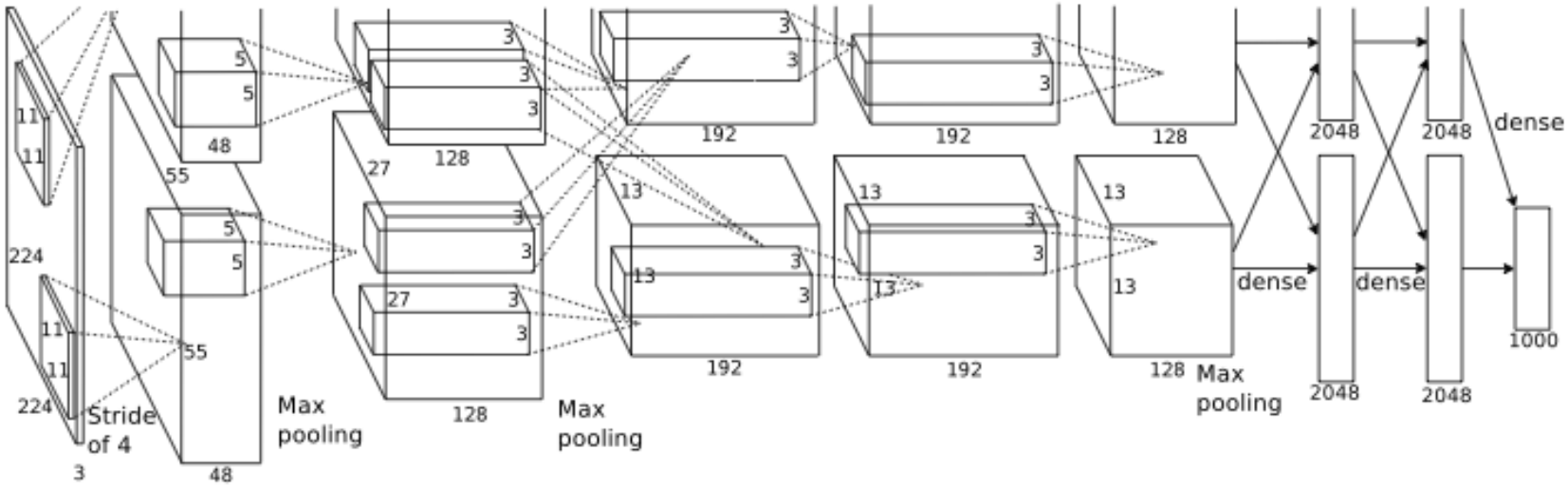
Imagenet classification challenge



AlexNet - A Krizhevsky et al. (2012) won with huge margin (16.4% error compared to 26.2%) by deep learning.
Soon everyone started using deep learning and GPUs.

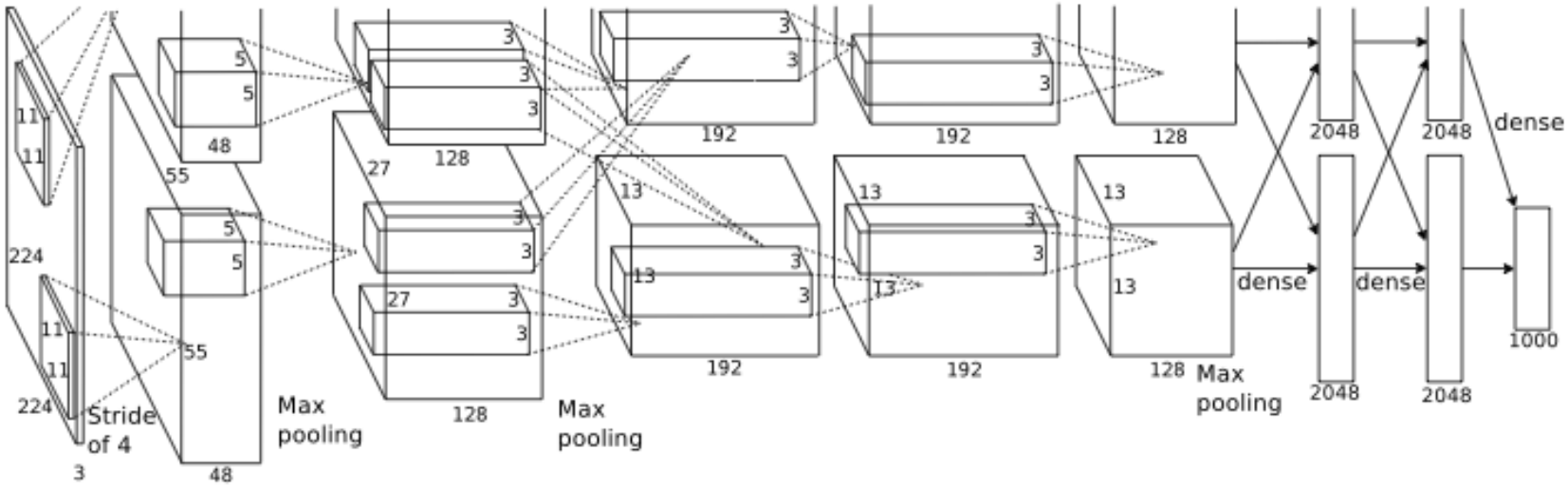
We need bigger brains

- AlexNet (2012): 16.4% error, 8 layers, 1.4 Gflop

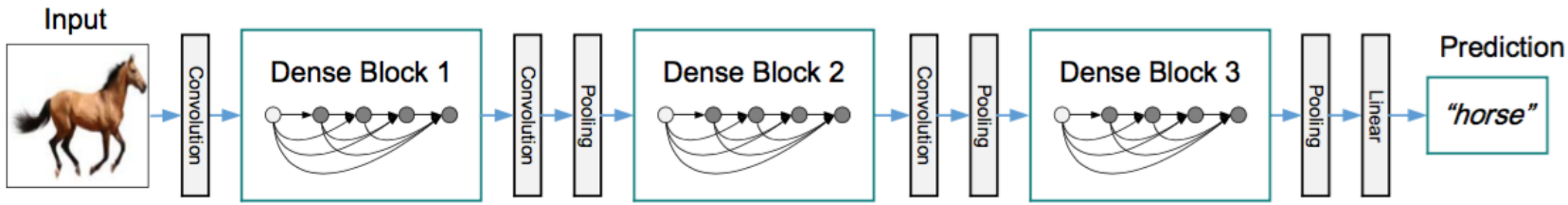


We need bigger brains

- AlexNet (2012): 16.4% error, 8 layers, 1.4 Gflop



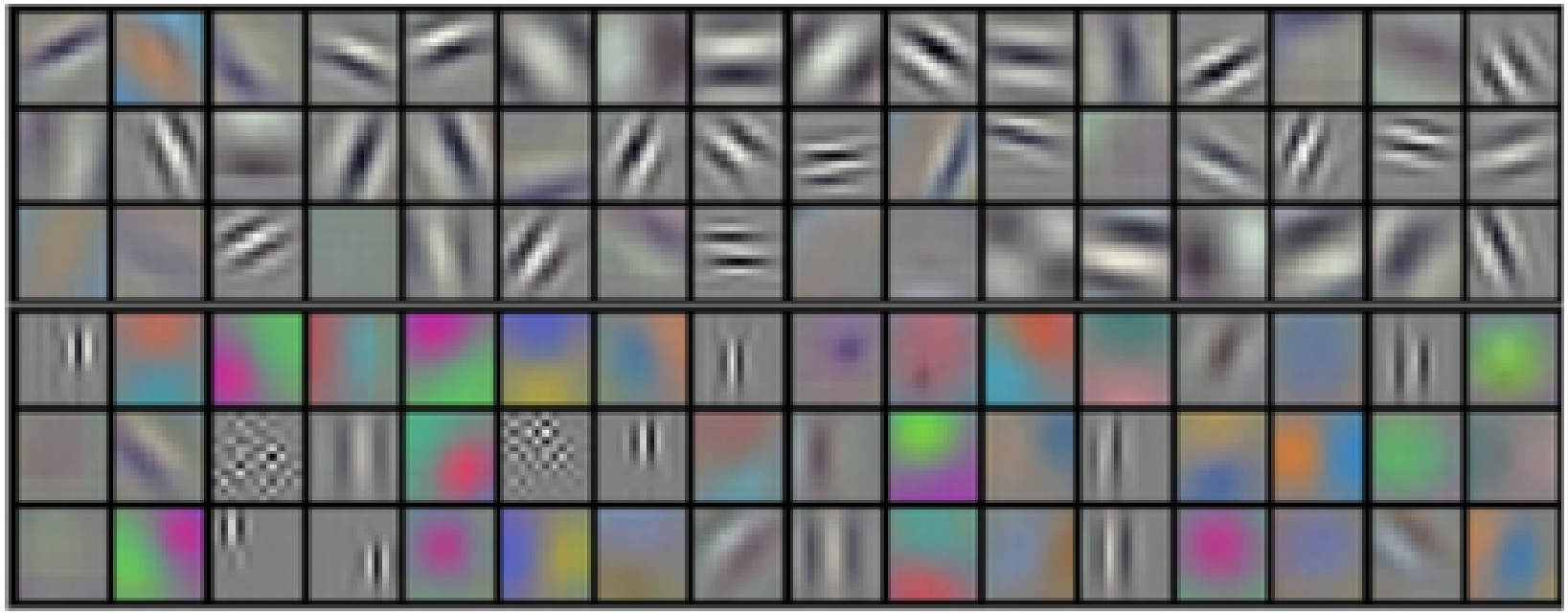
- ResNet (2016): 3.5% error, 152 layers, 22.6 Gflop.



- (This is a so-called DenseNet and not a ResNet.)
- Source: Source Jen-Hsun Huang, CEO NVIDIA, GTC Europe, 2016

All filters are learned from training data

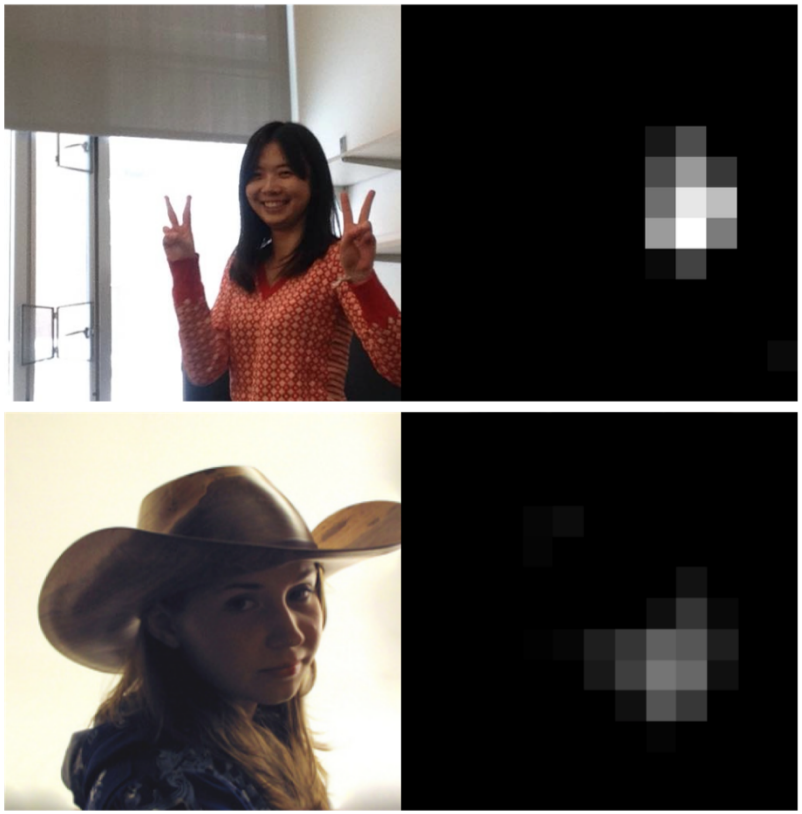
- First layer filters



- `www.cs.toronto.edu/~fritz/absps/imagenet.pdf`

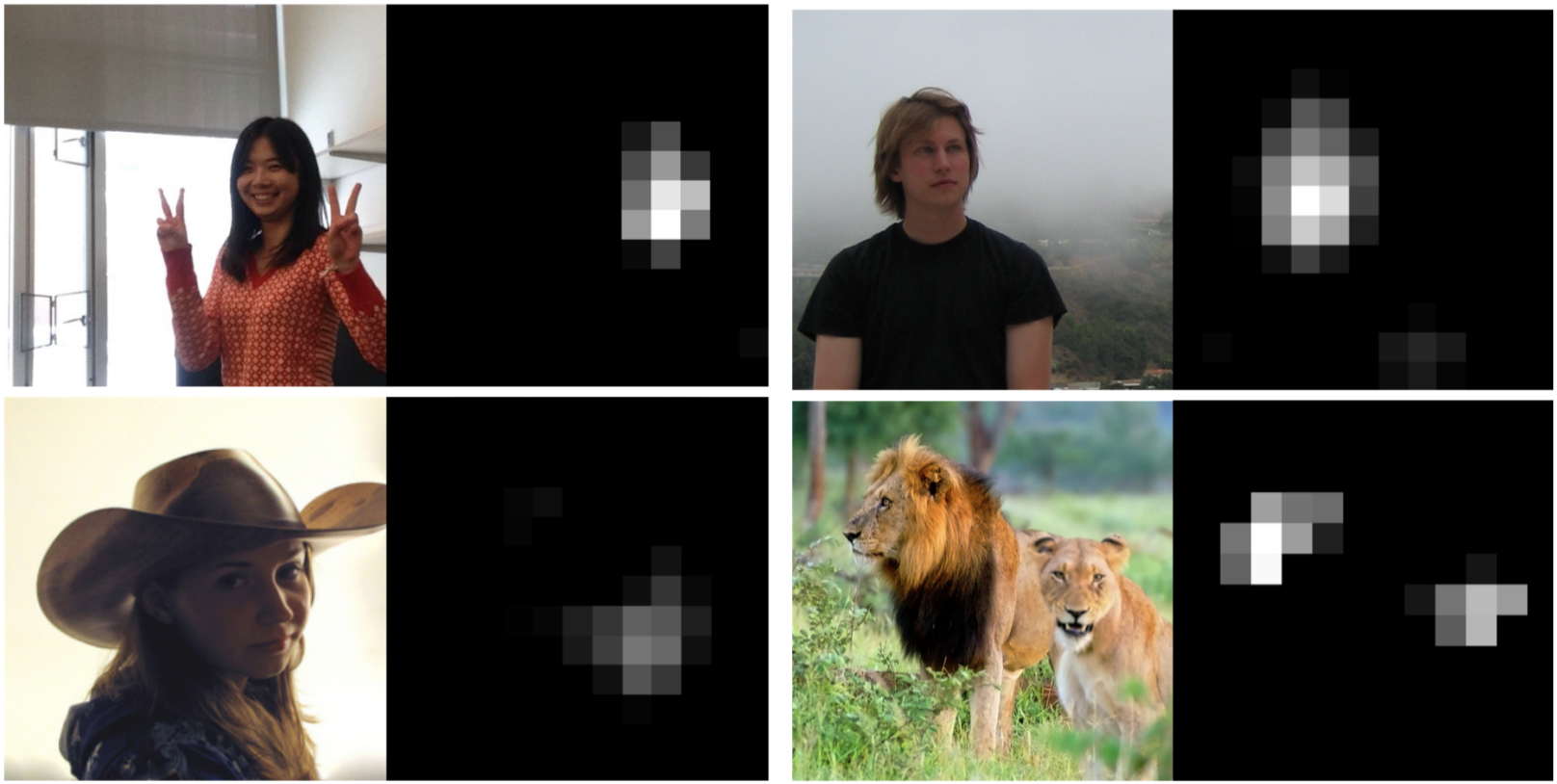
Emergent higher level abstractions

- Look at output of filter in 5th layer!



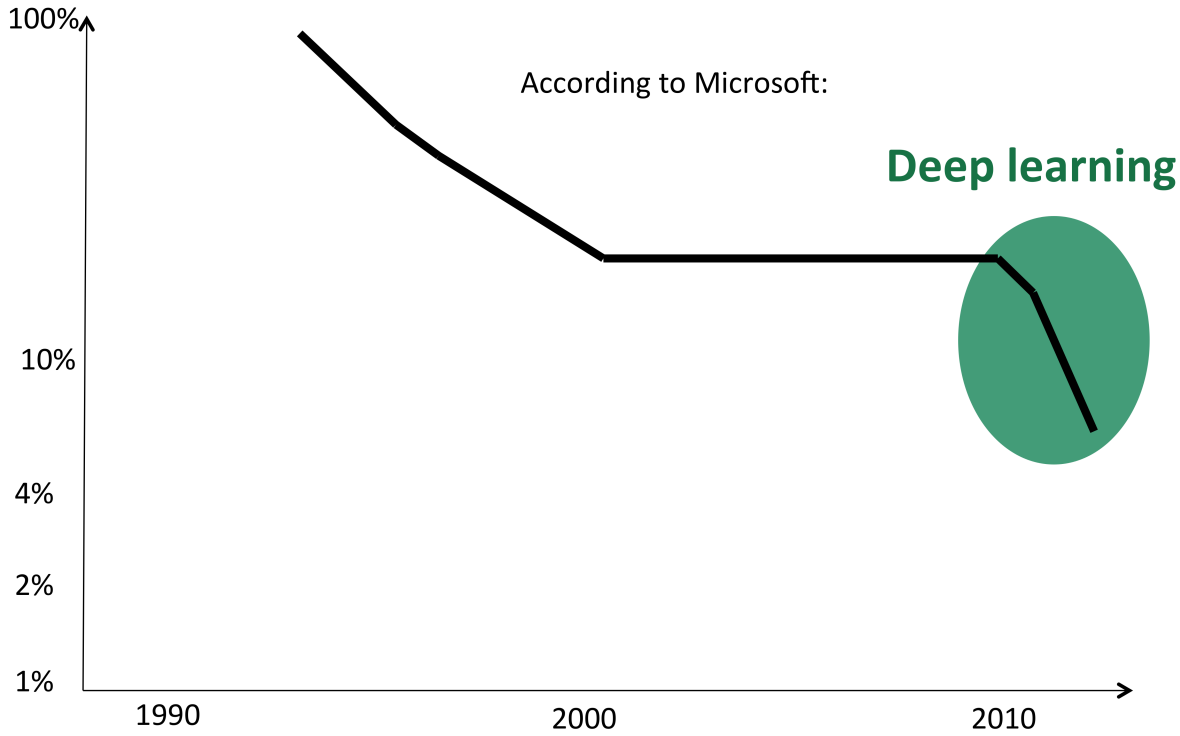
Emergent higher level abstractions

- Look at output of filter in 5th layer!



Yosinski et. al., ICML, google: deepvis

Speech recognition breakthrough



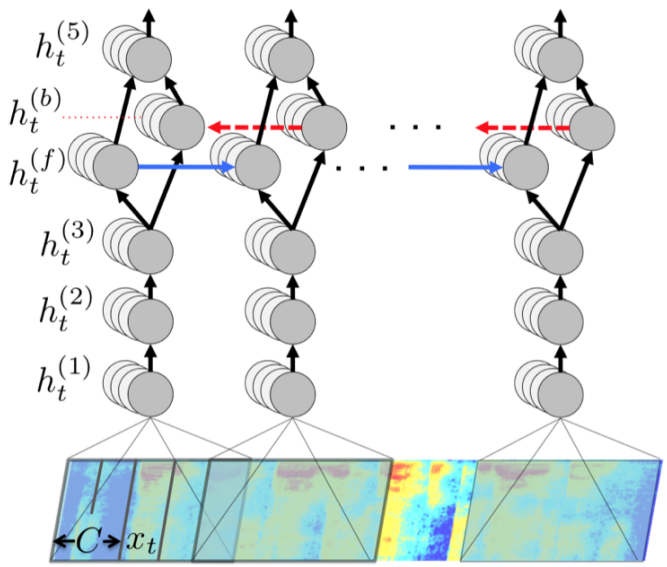
Plot from Yoshua Bengio

Recurrent neural networks – DeepSpeech

DeepSpeech: Scaling up end-to-end speech recognition

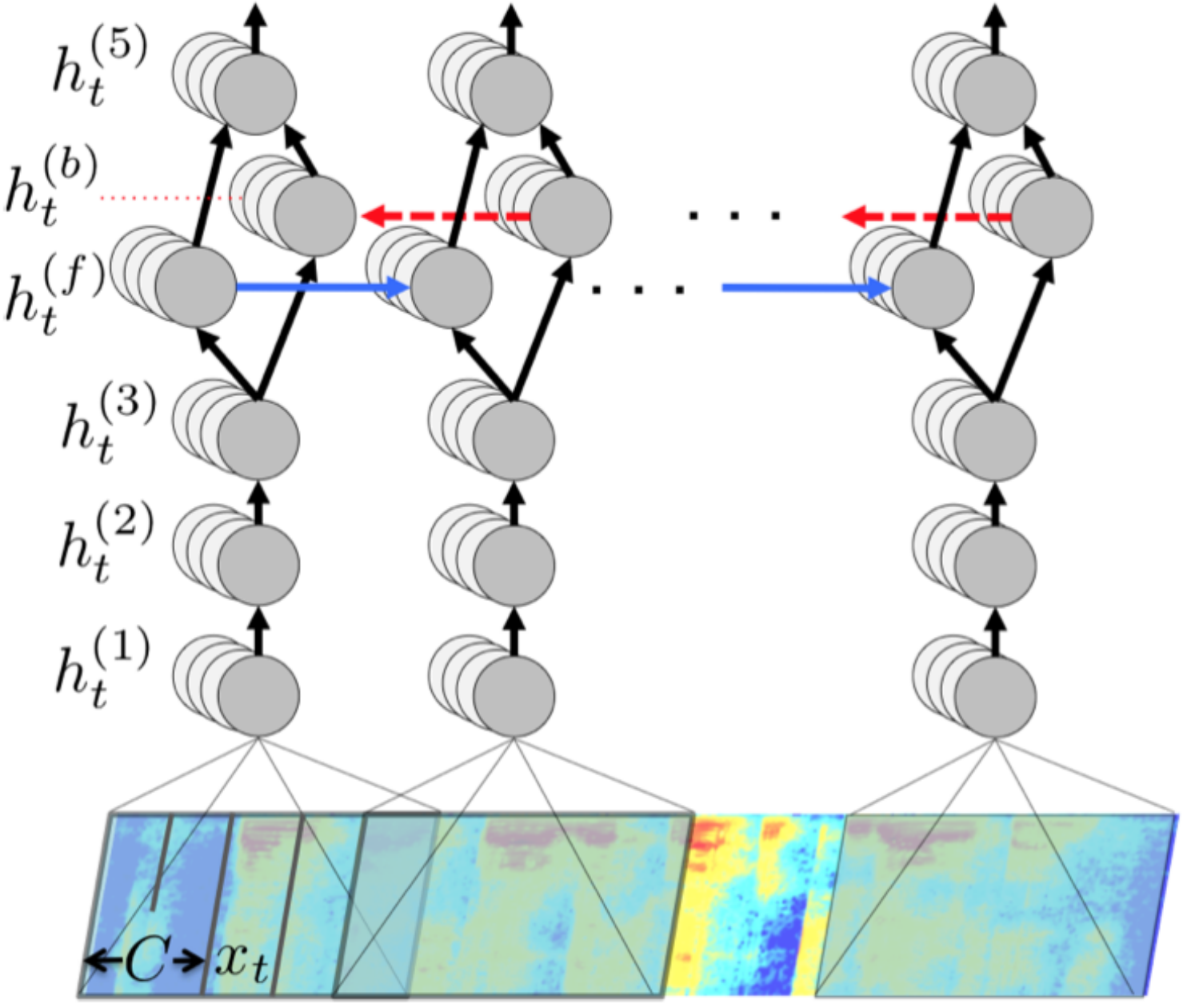
Awni Hannun,* Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Y. Ng

Baidu Research – Silicon Valley AI Lab

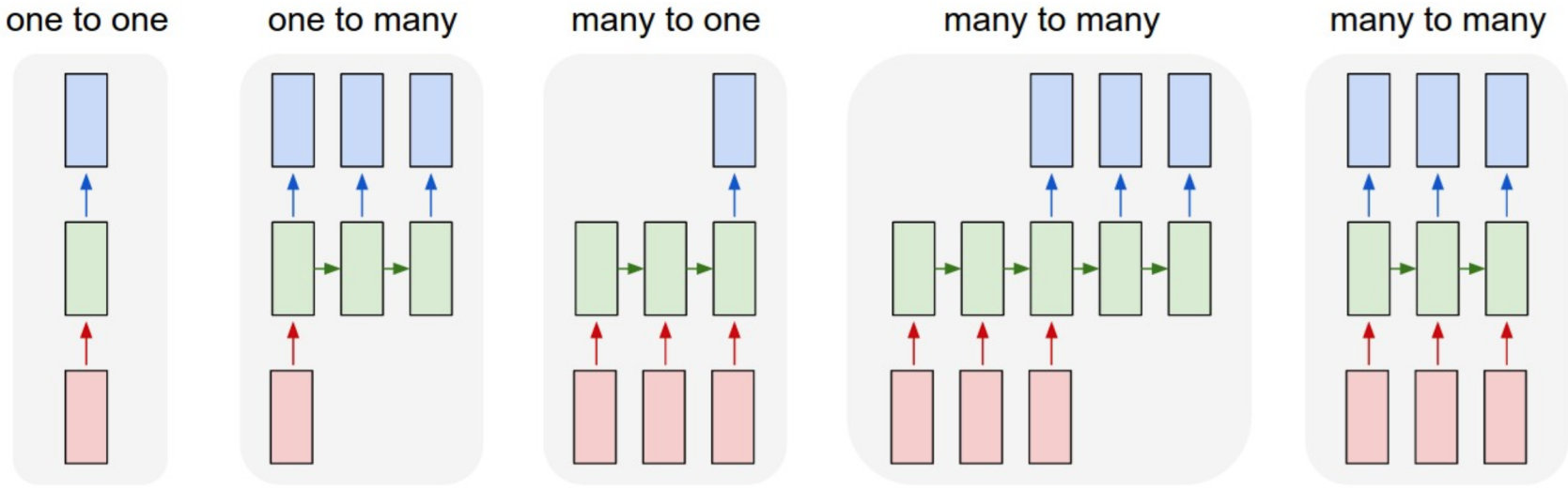


Deep speech 2: ∼human level performance + realtime server

DeepSpeech



Overview recurrent achitectures



- Vanilla mode, no RNN.
- E.g. image classification

- Sequence output
- E.g. image captioning

- Sequence input
- E.g. sentiment analysis

- Sequence input and output (*encoder-decoder*, *sequence-to-sequence*)
- E.g. translation, question answering

- Synced sequence input and output
- E.g. label each video frame

From Andrej Karpathy blog: The Unreasonable Effectiveness of Recurrent Neural Networks

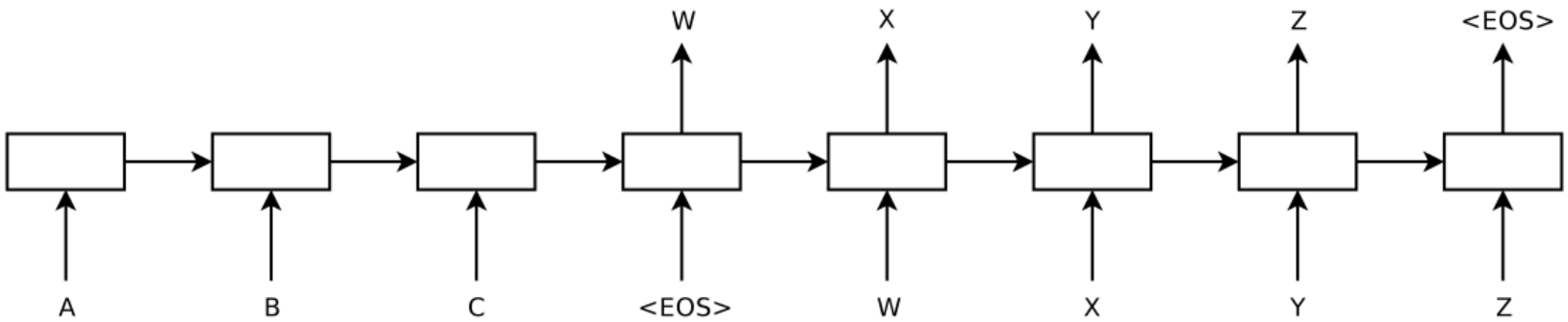
Encoder-decoder - machine translation

Sequence to Sequence Learning with Neural Networks

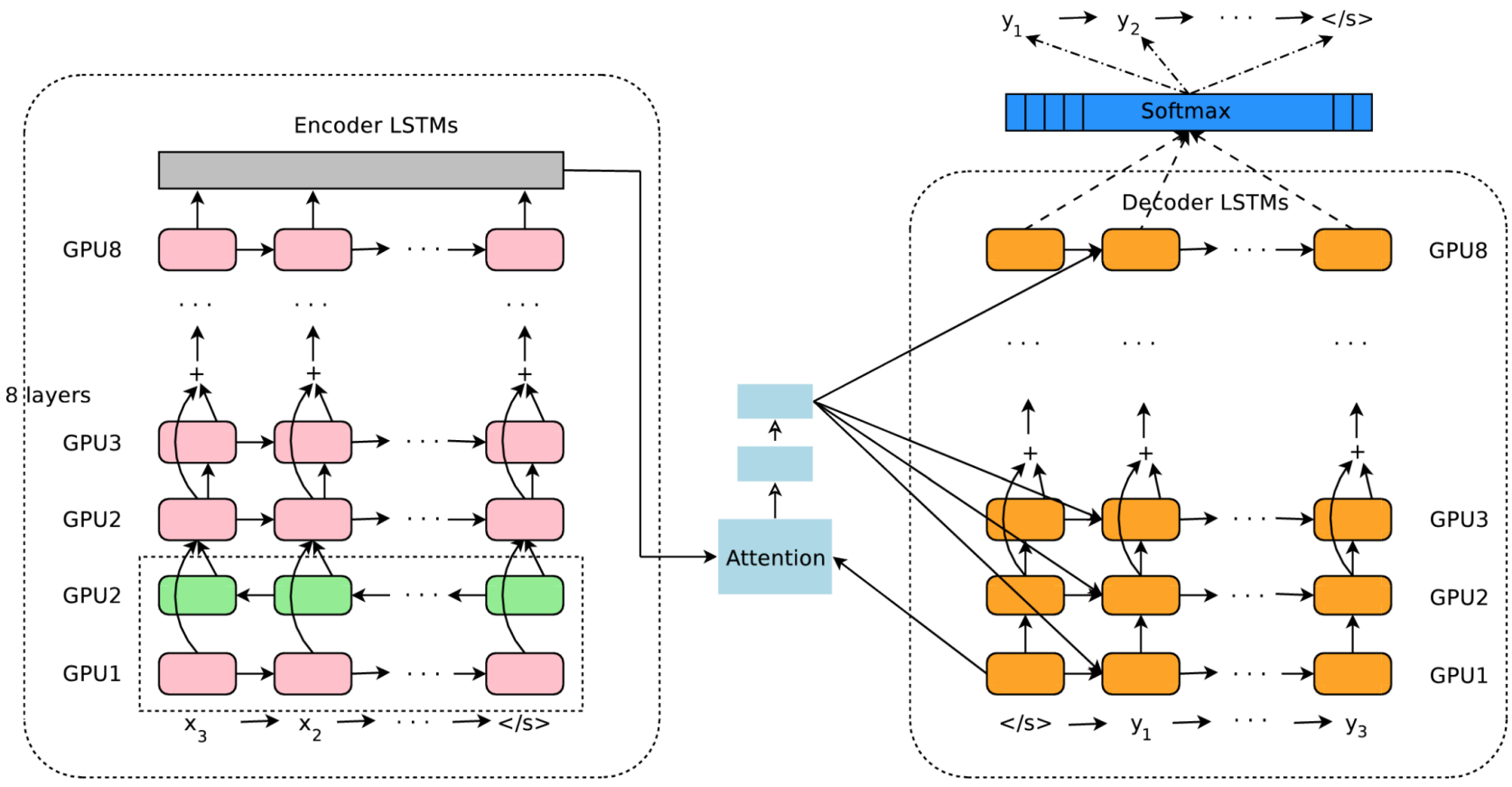
Ilya Sutskever
Google
ilyasu@google.com

Oriol Vinyals
Google
vinyals@google.com

Quoc V. Le
Google
qvl@google.com



We need bigger brains II



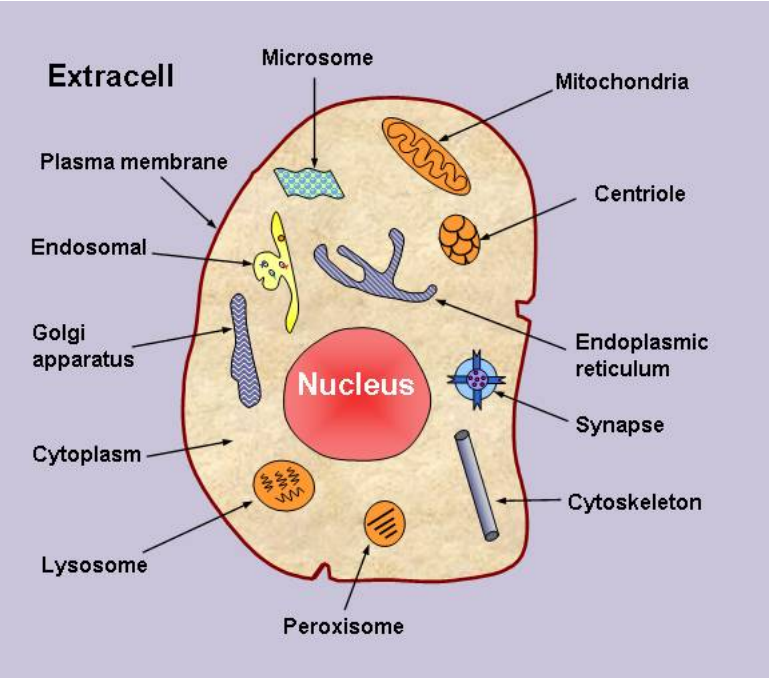
Wu et. al., Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, 2016.

Part 4:

Deep data science and (big) data

Deep learning – DTU and KU research group

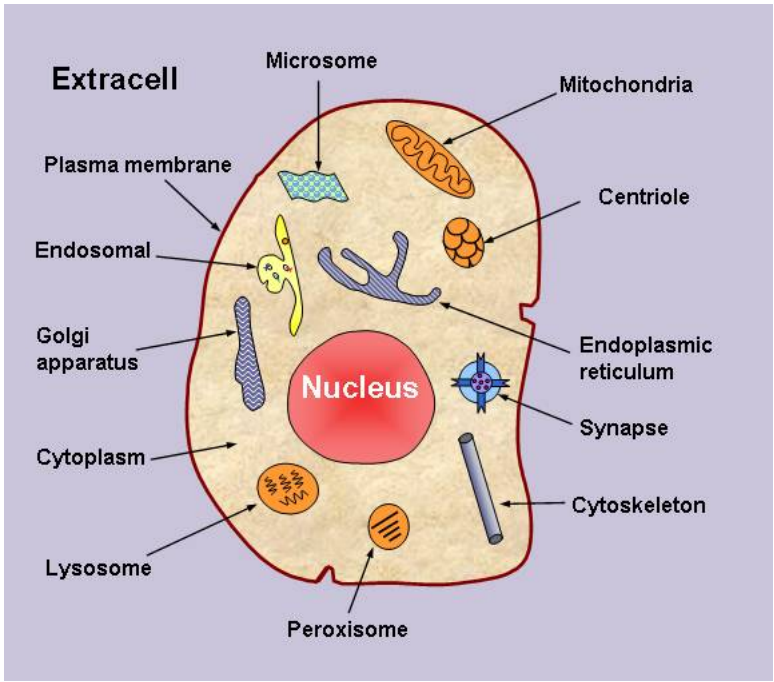
- End-to-end!
- Structured data - sequences+
- Bioinformatics



- Information retrieval - search in `findzebra`

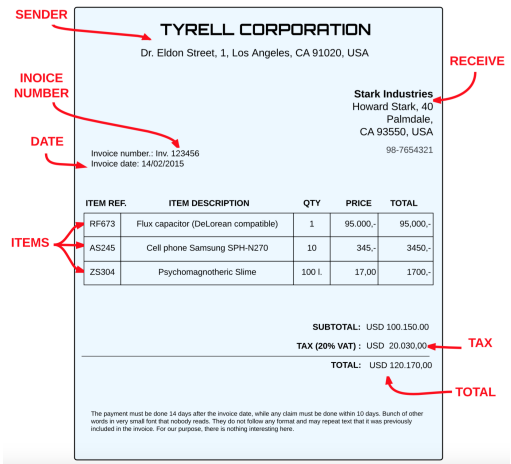
Deep learning – DTU and KU research group

- End-to-end!
- Structured data - sequences+
- Bioinformatics



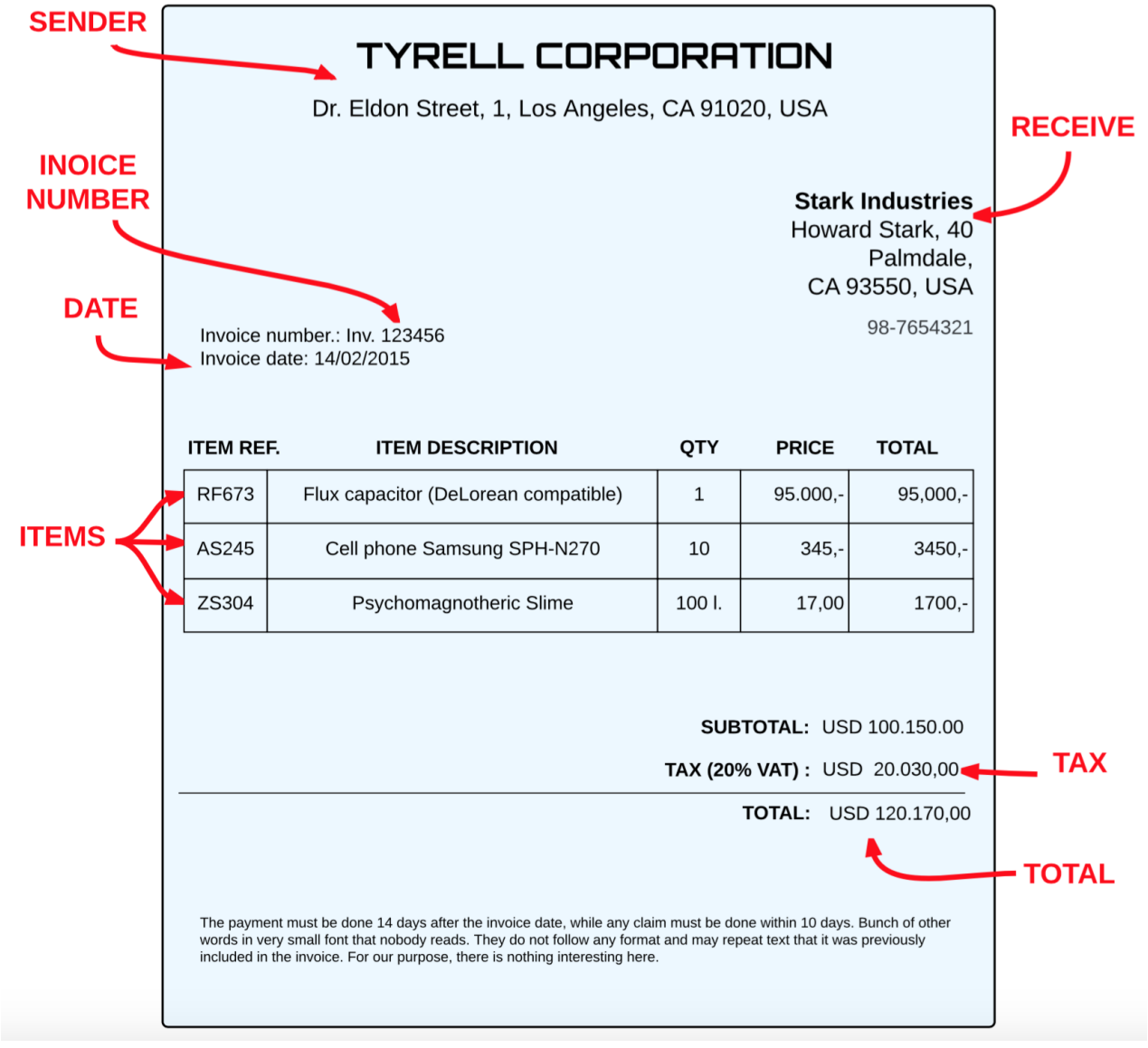
- Information retrieval - search in `findzebra`

- Green tech – Siemens Windpower and `greengoenergy.com`
- Document interpretation - `tradeshift.com`



- Variational un- and semi-supervised learning

Document interpretation with `tradeshift.com`



FindZebra search - Ellen's case story

For 25 years, Ellen struggled to find a diagnosis for the multitude of debilitating symptoms that seemed to increase year after year.

- Her symptoms included muscle cramps, intense headaches, rapid weight gain, fatigue, edema, intolerance to heat, excessive sweating, joint pain, tingling in her hands and feet, frequent bone fractures, acid reflux, intense anxiety and panic attacks, high blood pressure, high cholesterol, high blood sugar, sleep apnea, menstrual irregularities, peripheral vision loss and double vision.
- Source: `http://www.uptodate.com/`
- Any suggestions?

FindZebra search — findzebra.com

"When you hear hoofbeats behind you, don't expect to see a zebra"

fever, anterior mediastinal mass and central necrosis

Search

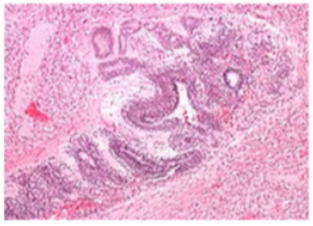
Filters

Group by: disease gene

- Mediastinal tumor
- Desmoplastic small round cell … (2)
- Thymoma (2)
- Adult-onset Still's disease (3)
- Large-cell lymphoma (3)
- Follicular lymphoma
- Periodic fever, familial, autoso… (2)
- Thymic carcinoma
- TNF receptor associated periodic…
- Japanese encephalitis
- Periodic fever, familial, autosomal…

Mediastinal tumor

Retrieved: 28-09-2014
Source: WIKIPEDIA (Original article)

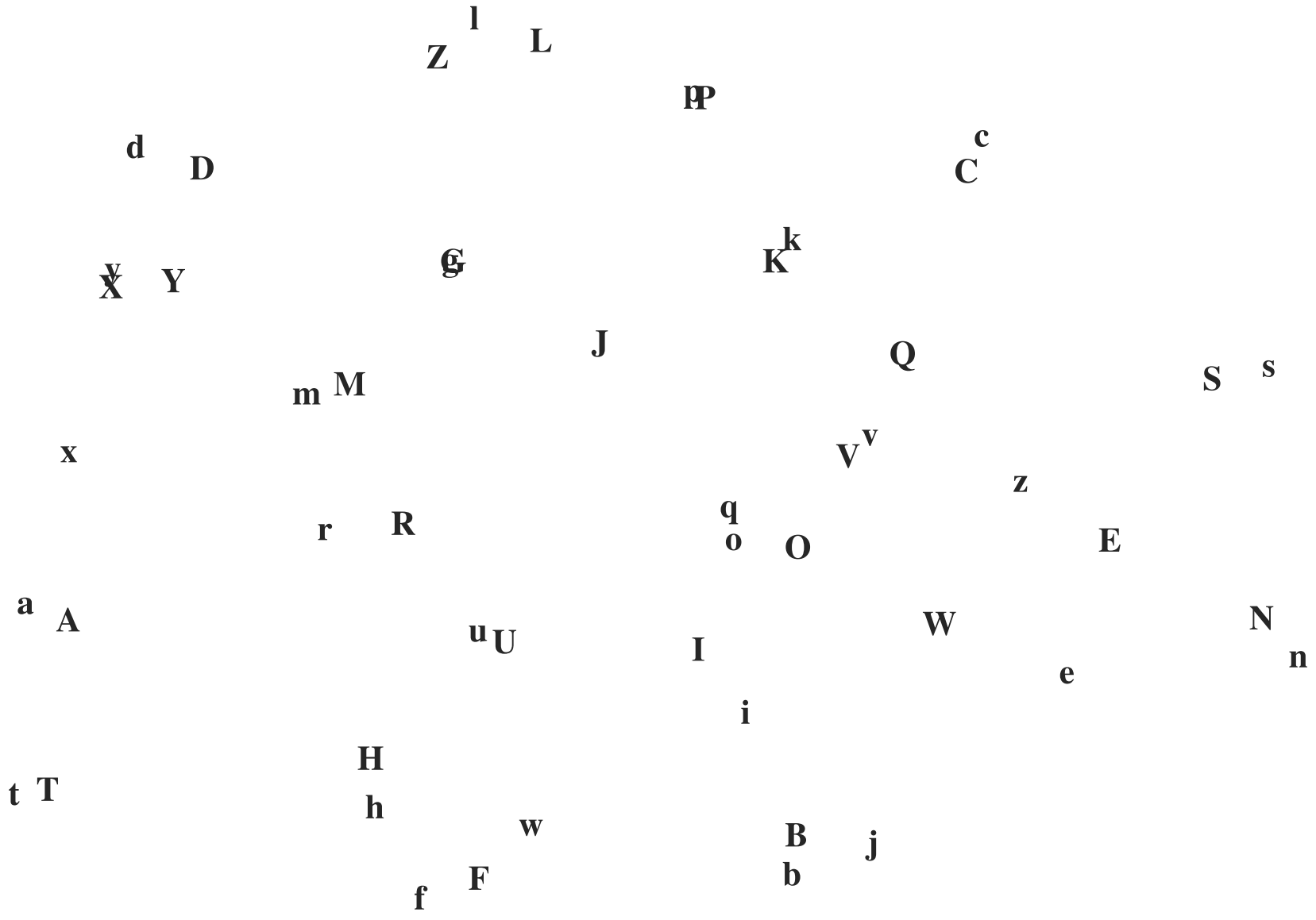


The mediastinum is the cavity that separates the lungs from the rest of the chest. It contains the heart, esophagus, trachea, thymus, and aorta. The mediastinum has three main parts: the anterior mediastinum (front), the middle mediastinum, and the posterior mediastinum (back). The most common mediastinal masses are neurogenic tumors (20% of mediastinal tumors), usually found in the posterior mediastinum, followed by thymoma (15-20%) located in the anterior mediastinum. Masses in the anterior portion of the mediastinum can include thymoma, lymphoma, pheochromocytoma, germ cell tumors including teratoma, thyroid tissue, and parathyroid lesions. Masses in this area are more likely to be malignant than those in other compartments. Masses in the posterior portion of the mediastinum tend to be neurogenic in origin, and in adults tend to be of neural sheath origin including neurilemomas and neurofibromas. Lung cancer typically spreads to the lymph nodes in the mediastinum.

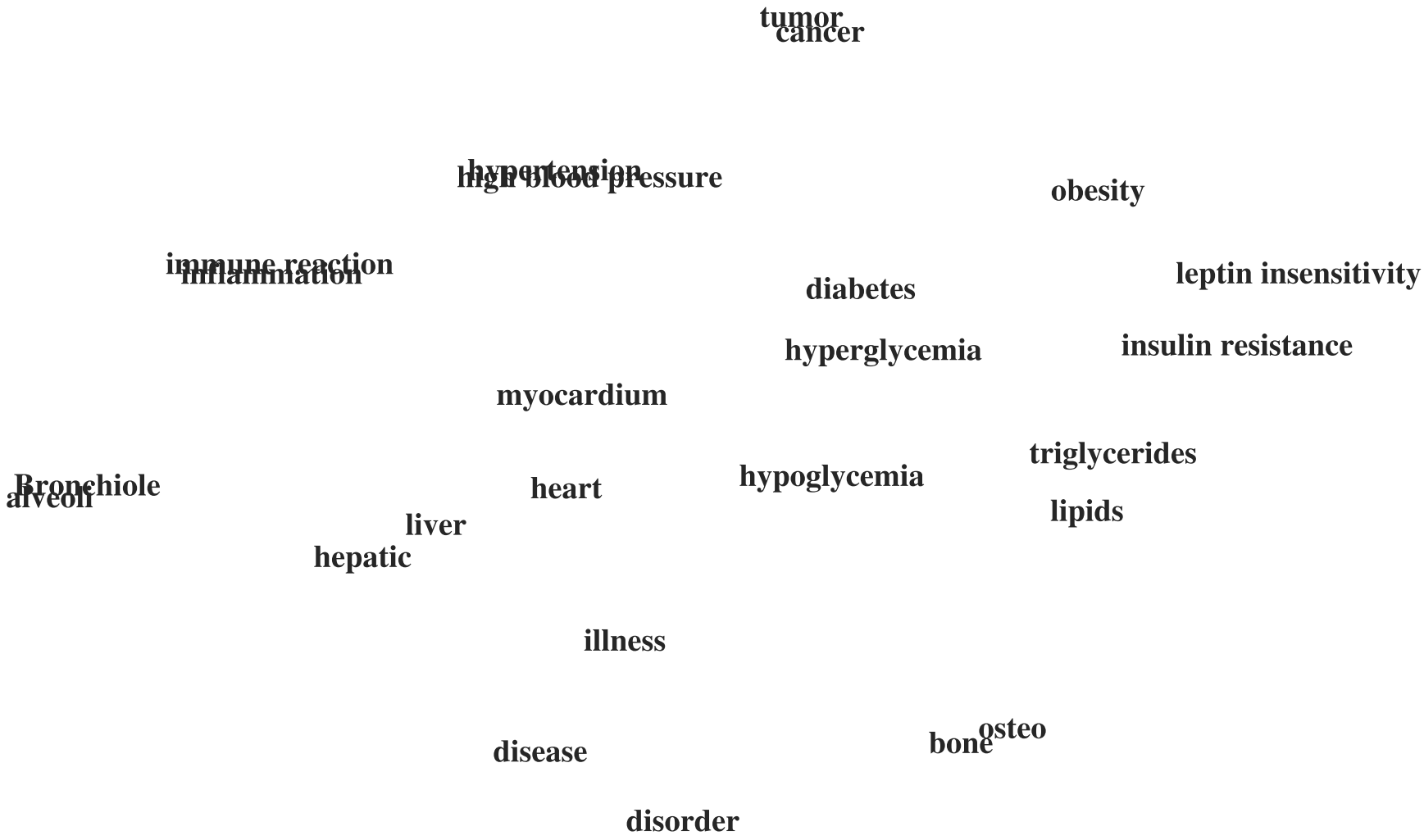
Diagnosis

In several editions of Physical Diagnosis, concerning mediastinal tumors the author writes: According to Christian1 the mediastinal neoplasms which are neither so rare nor so obscure as to make diagnosis practically impossible are:

Visualizing latent representation for characters



Visualizing latent representation for words



Example queries

Query	Disease	Combined Rank	Chars2Words2Doc Rank	Solr Rank
- syndrome characterized by the following tumors: medullary thyroid, adrenomedullary, parathyroid	men type 2	1	1	1
- diagnosis associated with abdominal obesity, elevated triglycerides, low hdl, hypertension, and elevated glucose	metabolic syndrome	1	1	1
- cystic lung disease in women of childbearing age	lymphangioleiomyomatosis	1	1	1
- 8 months old, male, progressive signs of respiratory distress, tachypnea, pulmonary hypertension, tortuosity of aortic arch, facial dysmorphisms	arterial tortuosity syndrome	1	1	1
- eponym for acute pericarditis 2 weeks after myocardial infarction	dressler's syndrome	1	2	1
- diagnosis for otalgia, ear discharge, and pruritus	otitis externa	1	2	1
- hemolytic anemia associated with igm antibodies directed against rbc i/i antigens	cold agglutinin disease	1	2	1
- recurrent unilateral retro-orbital headache with remissions and relapses	cluster headache	1	2	1
- most likely diagnosis causing generalized pruritus in the setting of inflammatory bowel disease	primary sclerosing cholangitis	1	7	1
- diagnosis associated with anion gap acidosis, osmolar gap, and urinary calcium oxalate crystals	ethylene glycol poisoning	1	9	1
- diagnosis when platelet count > 600,000/Âµl in absence reactive thrombocytosis	essential thrombocythemia	2	1	2
- autism spectrum disorder characterized by social impairment and repetitive behaviors but normal language and cognition development (final jeopardy)	asperger syndrome	2	1	2
- cancer associated with hepatitis c	hepatocellular carcinoma	2	1	10
- diagnosis for regular tachycardia associated with intermittent jugular cannon waves	ventricular tachycardia	2	1	21
- myocardial complication in patients who receive mantle-field radiation for hodgkin's disease	myocardial infarction	2	1	23
- most common cause of goiter worldwide	iodine deficiency	3	4	2
- personality disorder characterized by unstable interpersonal relationships, impulsiveness, and self-destructive behaviors	borderline personality	3	5	2
- most common type of paroxysmal reentrant supraventricular tachycardia	atrioventricular nodal reentrant tachycardia	3	5	3
- tropical zoonosis associated with hemorrhage, abdominal pain, hepatosplenomegaly and this finding	leptospirosis	3	20	2
- most consistent radiographic finding associated with ankylosing spondylitis	sacroiliitis	3	46	2
- "signature injury" of soldiers deployed to iraq and afghanistan	traumatic brain injury	3	69	2
- periodic headache classically found in male smokers	cluster headache	3	408	2
- agent causing conjunctivitis and umbilicated papular lesions on lid margins	molluscum contagiosum	3	774	2
- syndrome characterized by acute swelling and pain in the legs and arms that evolves into a brawny, indurated edema in patients who have performed unaccustomed physical exertion	eosinophilic fasciitis	3	1009	2
- hereditary condition suggested by genital edema in a child riding a bicycle	hereditary angioedema	4	2	40
- characterized by lack of expression of cd55 and cd59 on erythrocytes	paroxysmal nocturnal hemoglobinuria	5	53	3
- gene mutation associated with familial adenomatous polyposis	apc gene	6	4	17
- eponym applied to a painless nodule in the palmar aponeurosis	dupuytren contracture	7	7	4
- diagnosis classically associated with low-back pain and stiffness that worsens at night and improves with physical activity or a hot shower	ankylosing spondylitis	8	4	26
- diagnosis associated with hepatitis, negative viral serology, psychiatric symptoms and family history of hepatitis	wilson's disease	8	7	19
- most common cause of goiter and hypothyroidism in usa	hashimoto disease	10	6	7
- name of childhood illness caused by parvovirus b 19	erythema infectiosum	14	7	14

Results

Metric / *Model*	MRR	Recall@10	Recall@20
Chars2Doc CNN (Small)	0.195	0.319	0.387
Chars2Doc BLSTM (Small)	0.174	0.343	0.407
Chars2Words2Doc BLSTM (Large)	0.327	0.548	**0.665**
Solr	**0.355**	**0.585**	0.657

Results

Metric / *Model*	MRR	Recall@10	Recall@20
Chars2Doc CNN (Small)	0.195	0.319	0.387
Chars2Doc BLSTM (Small)	0.174	0.343	0.407
Chars2Words2Doc BLSTM (Large)	0.327	0.548	**0.665**
Solr	**0.355**	**0.585**	0.657

- SOLR and C2W2D make different errors (Recall@20)

		Chars2Words2Doc	
		+	-
Solr	+	0.540	0.117
	-	0.125	0.218

Results

Metric / *Model*	MRR	Recall@10	Recall@20
Chars2Doc CNN (Small)	0.195	0.319	0.387
Chars2Doc BLSTM (Small)	0.174	0.343	0.407
Chars2Words2Doc BLSTM (Large)	0.327	0.548	**0.665**
Solr	**0.355**	**0.585**	0.657

- SOLR and C2W2D make different errors (Recall@20)

		Chars2Words2Doc	
		+	-
Solr	+	0.540	0.117
	-	0.125	0.218

- Simple combination:
  MRR=**0.373**, Recall@10=**0.657** and Recall@20=**0.738**

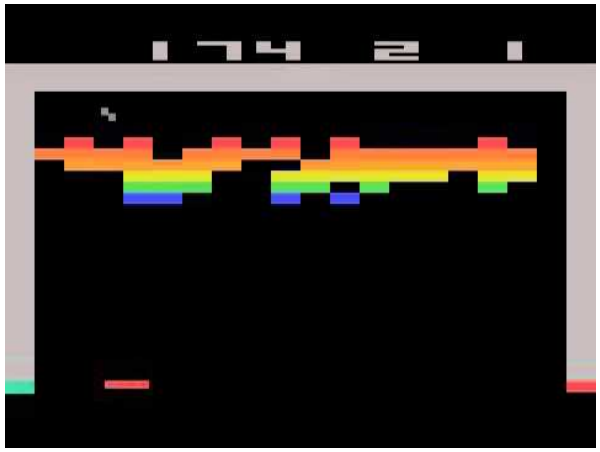
Part 5:

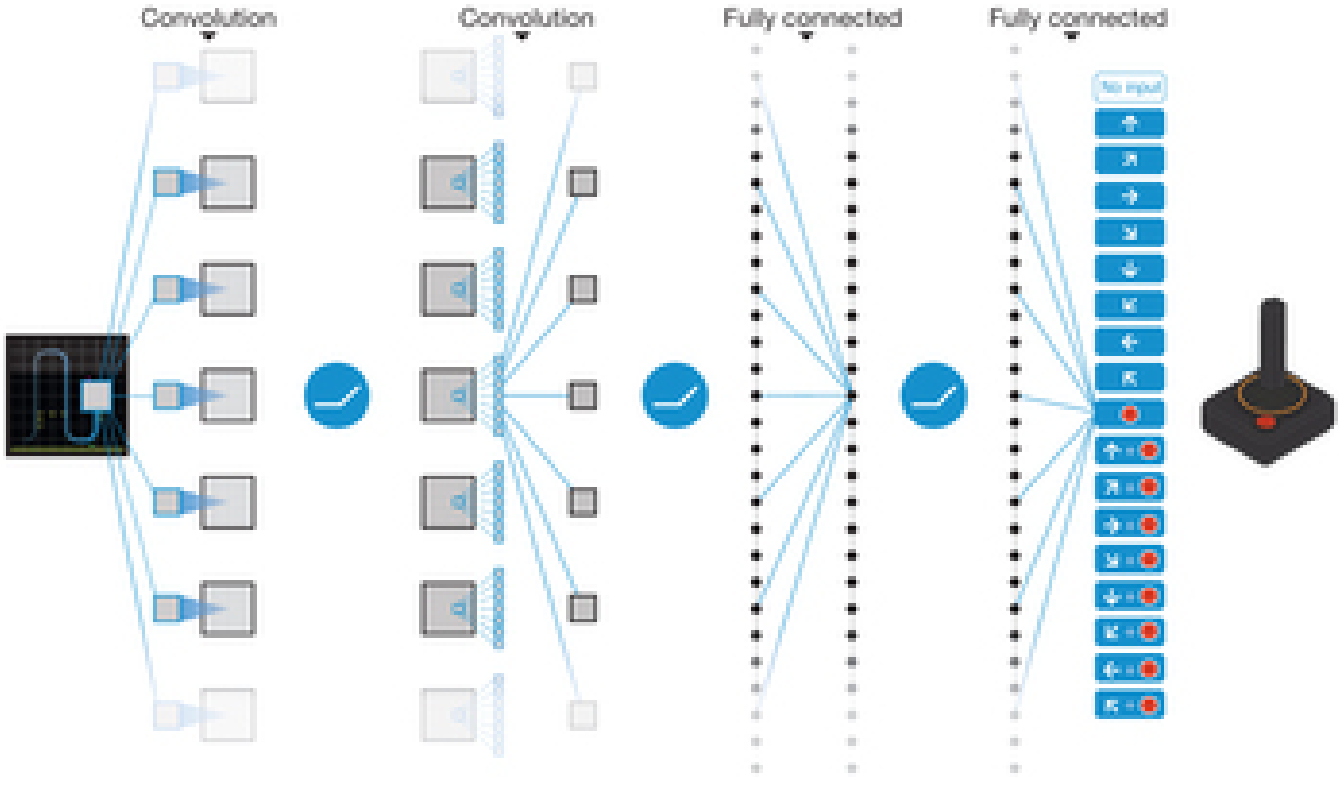
Attempt to demystify DeepMind's reinforcement learning

Reinforcement learning

- An agent learning to take actions in an environment so as to maximize some notion of cumulative reward.
- Actions taken now will affect future reward.

Deep Q learning on Atari games



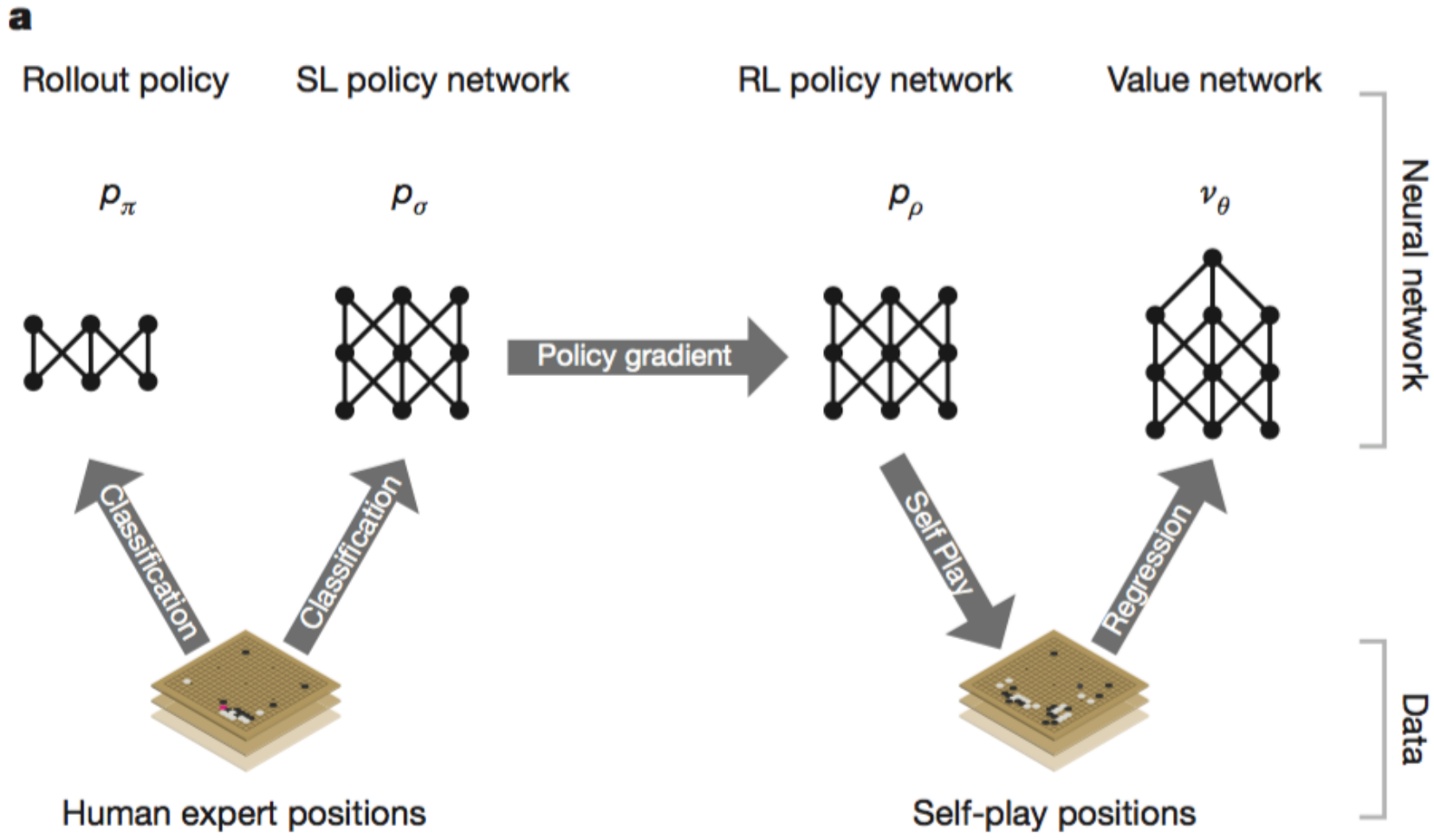


AlphaGo

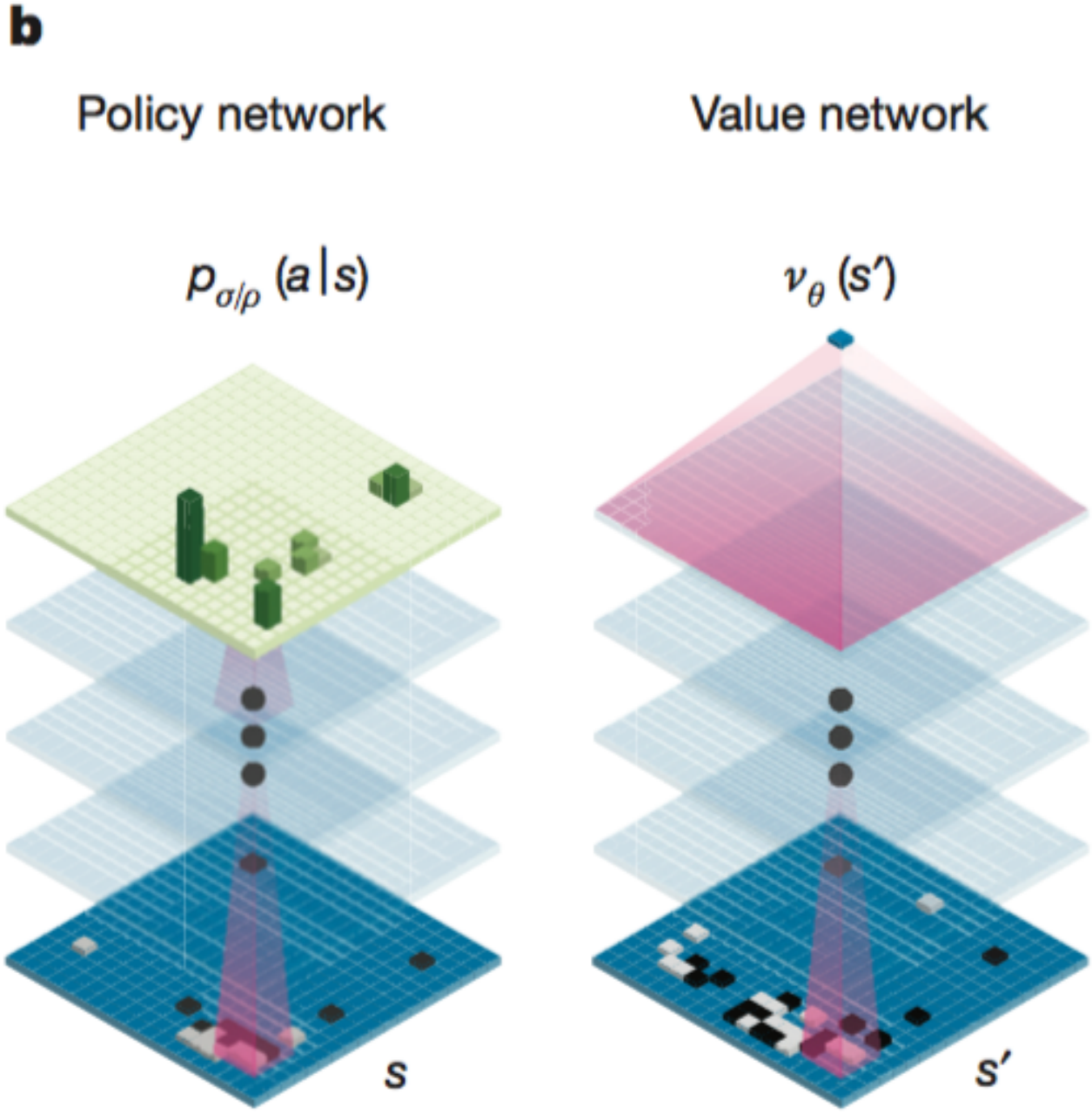
AlphaGo

- s = state = game position
- Game breadth: b
- Game depth. d
- Complexity: b^d for calculating $v^*(s)$
- Chess: $b \approx 35$, $d \approx 80$
- Go: $b \approx 250$, $d \approx 150$

Rollout policy, SL and RL policy and value networks



Policy and value networks



Step 1: supervised learning (SL) policy $p_\sigma(a|s)$

- a = action
- s = state
- We have large database of expert games.
- Train classifier to imitate expert moves (s, a):

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial\sigma}$$

- Supervised learning (SL) policy

Step 2: reinforcement learning (RL) policy $p_\rho(a|s)$

- The policy network $p_\rho(a|s)$ plays against (a younger version of itself).
- Record whether it wins/losses: $z_t = r(T) = +1/-1$
- Train classifier to imitate expert moves (s, a):

$$\Delta\rho \propto \frac{\partial \log p_\rho(a_t|s_t)}{\partial\rho} z_t$$

- Better than SL policy.

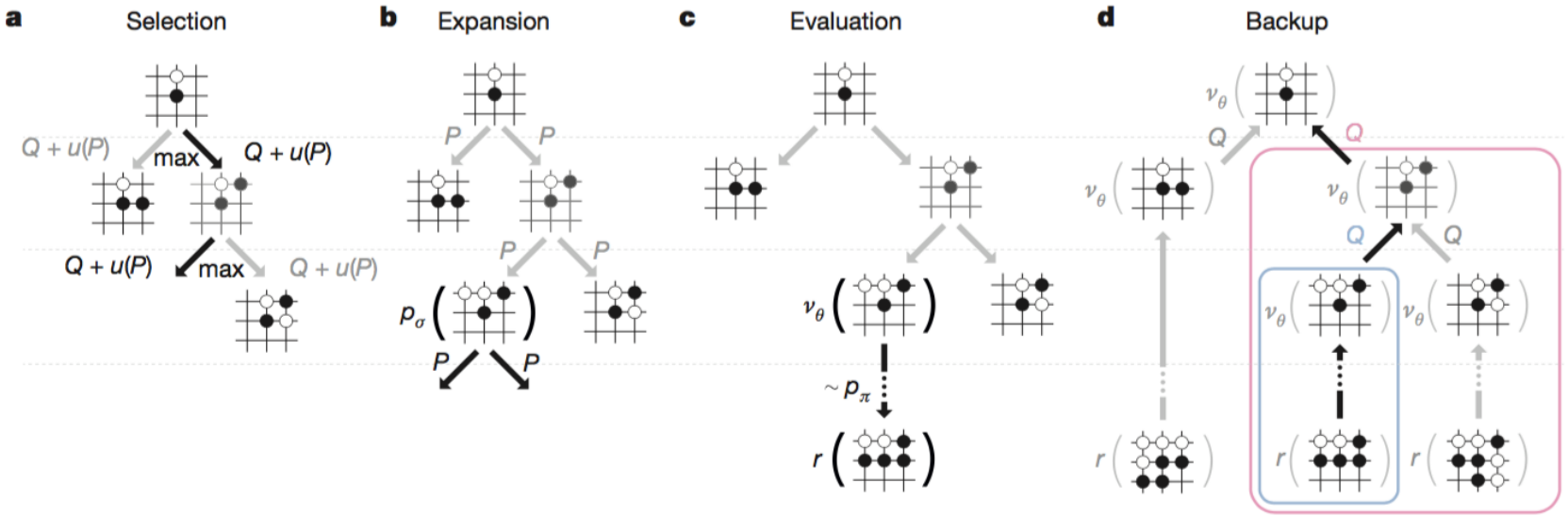
Step 3: RL value function $\nu_\theta(s)$

- Remember that value function $v^*(s)$ is unknown.
- We can try to learn the value function of our RL policy network by a network: $\nu_\theta(s)$:

$$\Delta\theta \propto \frac{\partial \log \nu_\theta(s)}{\partial\theta}(z - \nu_\theta(s))$$

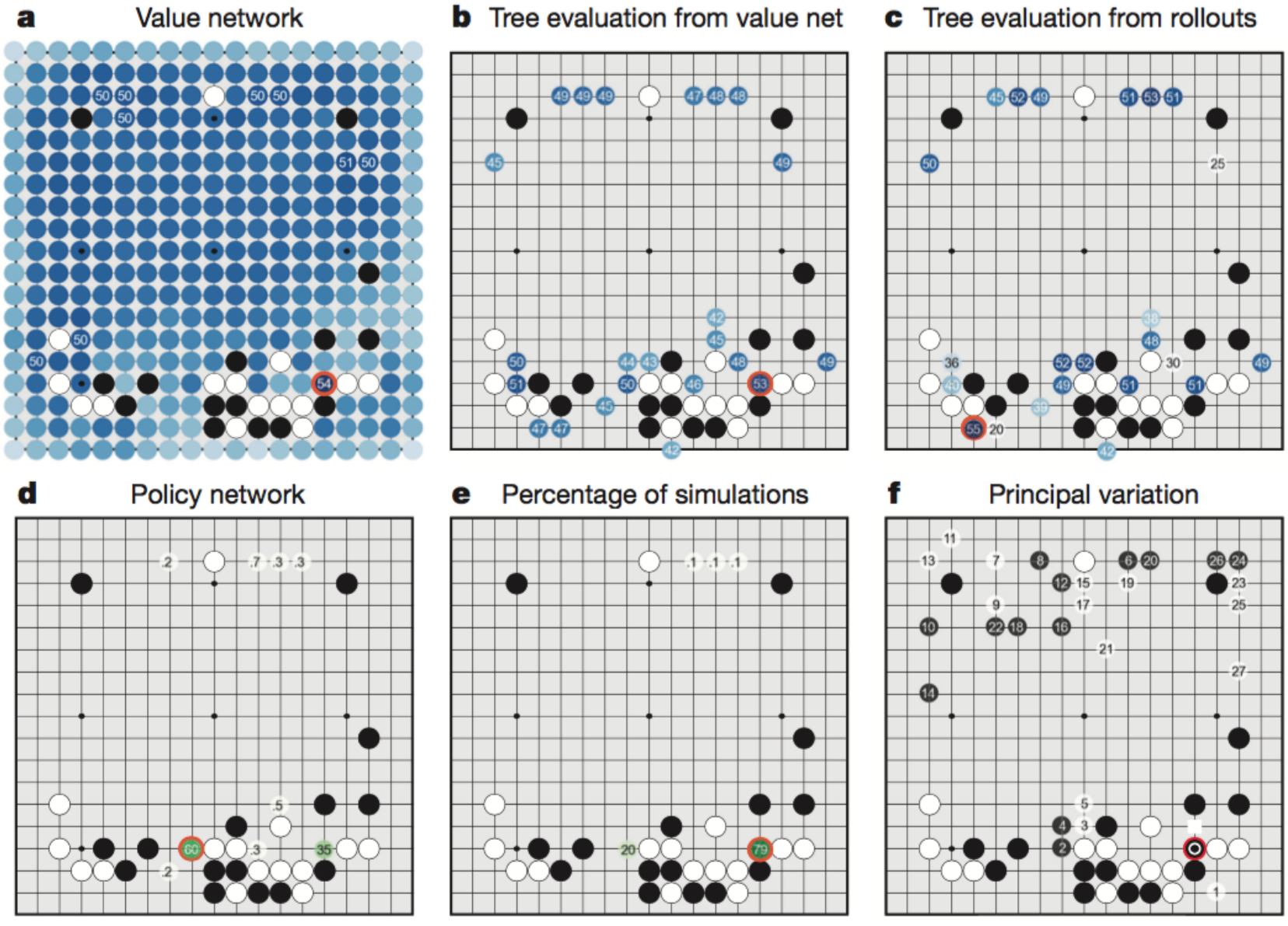
- We can use this as an ingredient in a Monte Carlo tree search
- to score positions s

Step 4: Monte Carlo tree search



- SL (and not RL) policy network used for to propose moves in tree search
- SL better for exploration!
- RL value function used for scoring positions.

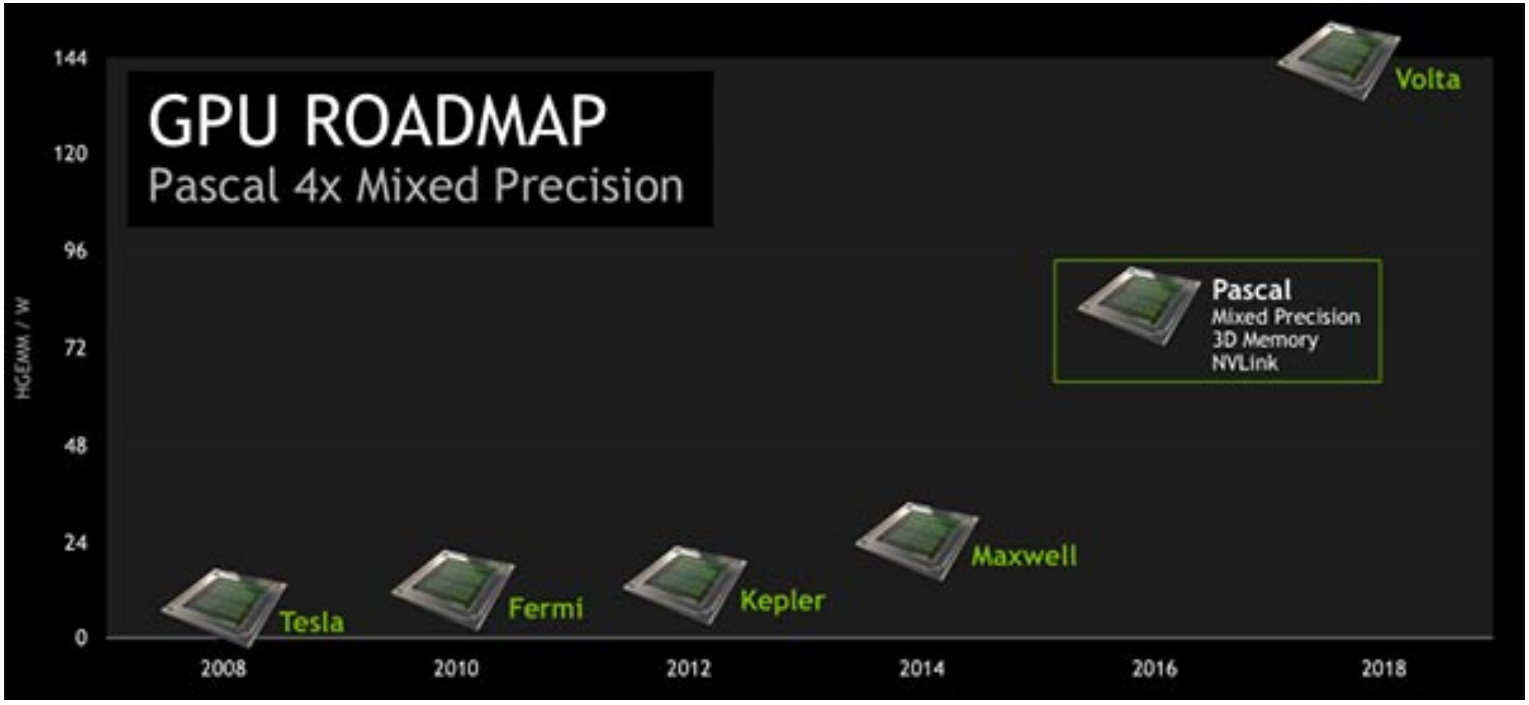
An example


a Value network
b Tree evaluation from value net
c Tree evaluation from rollouts
d Policy network
e Percentage of simulations
f Principal variation

Where humans excel & RL mostly fails

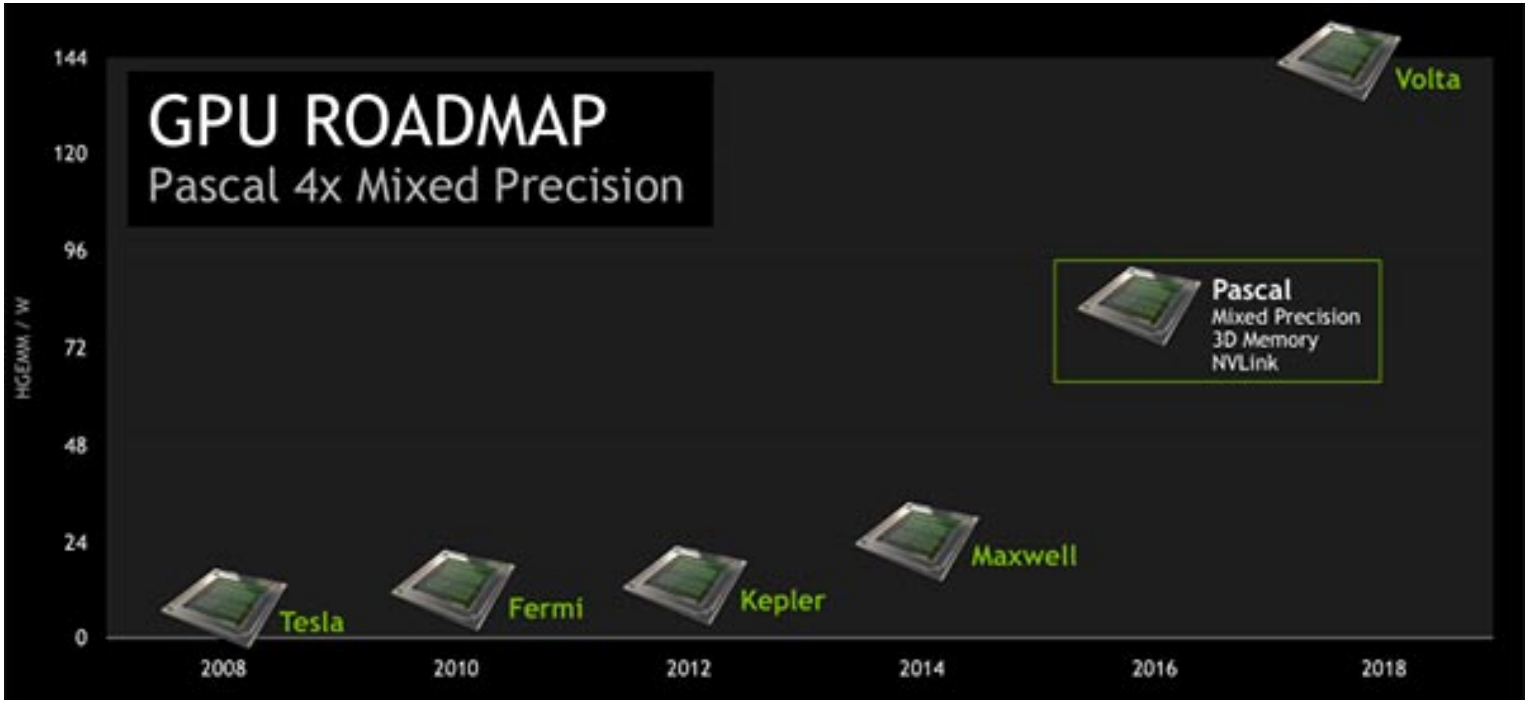
Summary

- Expect fast progress in coming years!



Summary

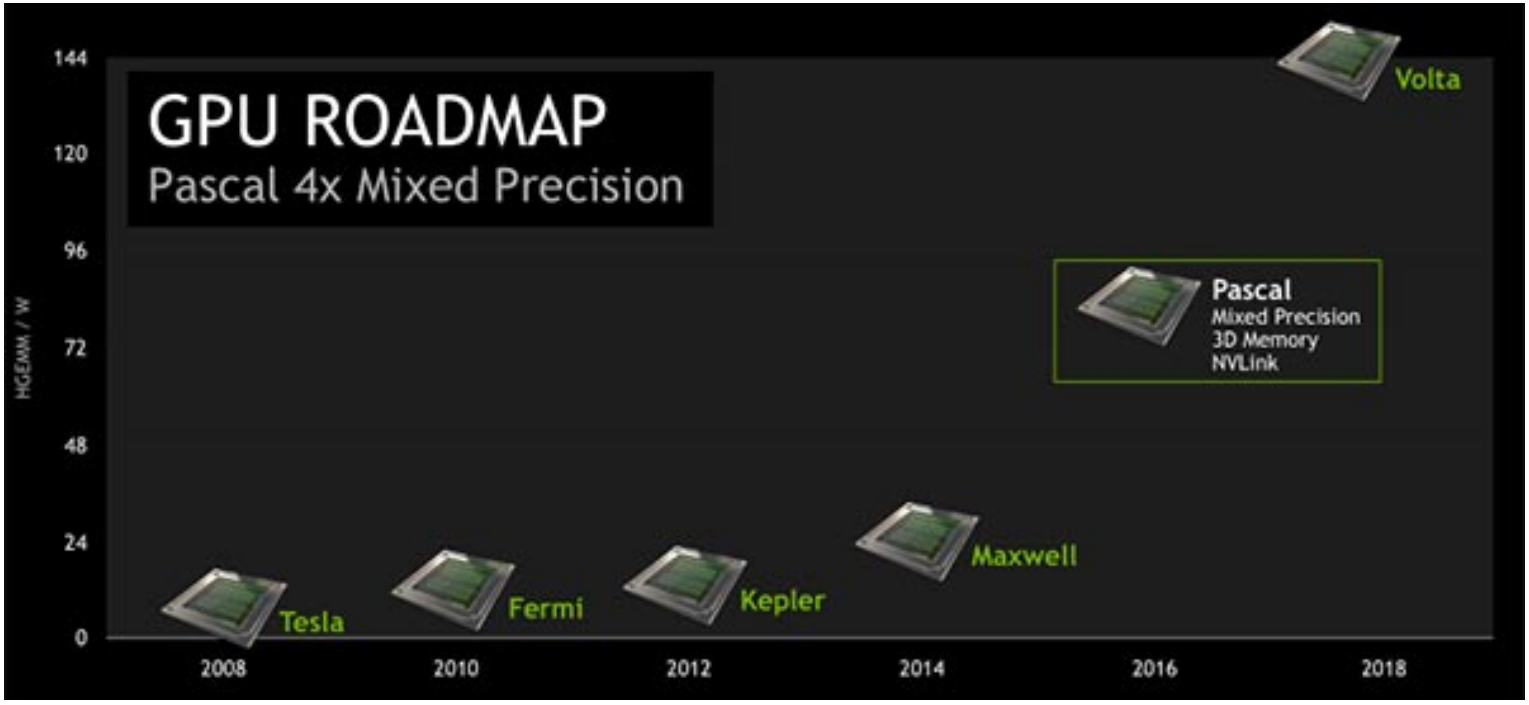
- Expect fast progress in coming years!



- Deep learning will dominate data-rich natural perceptional data settings: vision, speech, text, ...

Summary

- Expect fast progress in coming years!



- Deep learning will dominate data-rich natural perceptional data settings: vision, speech, text, ...
- But not for small n large p.
- Deep learning needs lots of data whereas
- natural learning is often one-shot.
- Reinforcement learning is difficult!