

# Artificial Intelligence Past, Present and Future

Grigory Sapunov  CTO / Intento

GoTo School on Machine Learning, Data Science and IoT  
Moscow 01.08.2016

# Overview

# AI

- Artificial Intelligence (AI) is a broad field of study dedicated to complex problem solving.
- Machine Learning (ML) is usually considered as a subfield of AI. ML is a data-driven approach focused on creating algorithms that has the ability to learn from the data without being explicitly programmed.
- Deep Learning (DL) is a subfield of ML focused on deep neural networks (NN) able to automatically learn hierarchical representations.

# TL;DR (The very short summary)

- Machine Learning (ML) is a very hot topic now and Deep Learning (DL) is dominating the talks about AI. DL outperformed humans in image recognition, approached the human level in speech recognition, and is now spreading into Natural Language Processing (NLP).
- But DL still have limitations and other subfields in AI still exist. Successful projects of the near future will probably be hybrid solutions.
- Hardware (mostly for ML/DL) is extremely important and it's getting hotter as well.

# Sections

1. History of AI
2. Machine Learning and Deep Learning
3. Knowledge Representation, Reasoning and Planning
4. Evolutionary Computations and Swarm Intelligence
5. Hardware
6. Artificial General Intelligence (AGI)
7. Initiatives

# History of AI

# Short History of AI

## 1. [1943-1955] The gestation of artificial intelligence

- McCulloch & Pitts: model of artificial neurons
- Hebb: simple updating rule for modifying the connection strengths between neurons (Hebbian learning)
- Two undergraduate students at Harvard, Marvin Minsky and Dean Edmonds, built the first neural network computer in 1950. The SNARL, as it was called, used 3000 vacuum tubes and a surplus automatic pilot mechanism from a B-24 bomber to simulate a network of 40 neurons.
- Alan Turing gave lectures on the topic as early as 1947 at the London Mathematical Society and articulated a persuasive agenda in his 1950 article "Computing Machinery and Intelligence." Therein, he introduced the Turing Test, machine learning, genetic algorithms, and reinforcement learning. He proposed the Child Programme idea, explaining "Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulated the child's?"

# Short History of AI

## 2. [1956] The birth of artificial intelligence

- John McCarthy moved to Dartmouth College. He convinced Minsky, Claude Shannon, and Nathaniel Rochester to help him bring together U.S. researchers interested in automata theory, neural nets, and the study of intelligence. They organized a two-month workshop at Dartmouth in the summer of 1956.
- *"We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."*
- The Logic Theorist reasoning program was able to prove most of the theorems in Chapter 2 of Russell and Whitehead's Principia Mathematica.

# Short History of AI

## 3. [1952-1969] Early enthusiasm, great expectations

- General Problem Solver (GPS). This program was designed from the start to imitate human problem-solving protocols. Within the limited class of puzzles it could handle, it turned out that the order in which the program considered subgoals and possible actions was similar to that in which humans approached the same problems.
- Geometry Theorem Prover, which was able to prove theorems that many students of mathematics would find quite tricky.
- A series of programs for checkers (draughts) that eventually learned to play at a strong amateur level.
- McCarthy defined the high-level language Lisp, which was to become the dominant AI programming language for the next 30 years.
- Minsky: microworlds, limited domain problems (calculus integration, geometric analogy problems, blocks world).
- *“Machines will be capable, within twenty years, of doing any work that a man can do.”* (Herbert Simon, 1965)

# Short History of AI

## 4. [1966-1973] A dose of reality

- Simon stated that within 10 years a computer would be chess champion, and a significant mathematical theorem would be proved by machine. These predictions came true (or approximately true) within 40 years rather than 10.
- (1957) Early machine translation efforts. It was thought initially that simple syntactic transformations based on the grammars of Russian and English, and word replacement from an electronic dictionary, would suffice to preserve the exact meanings of sentences.
- "***the spirit is willing but the flesh is weak***" translated as "**“водка хорошая, но мясо промухло”**" ("The vodka is good but the meat is rotten")
- The new back-propagation learning algorithms for multilayer networks that were to cause an enormous resurgence in neural-net research in the late 1980s were actually discovered first in 1969

# Short History of AI

## 5. [1974-1980] First AI winter

- (1966), a report by an advisory committee found that "there has been no machine translation of general scientific text, and none is in immediate prospect." All U.S. government funding for academic translation projects was canceled.
- (1969) Minsky and Papert's book Perceptrons proved that perceptrons could represent very little. Although their results did not apply to more complex, multilayer networks, research funding for neural-net research soon dwindled to almost nothing.
- (1973) Professor Sir James Lighthill was asked by the UK Parliament to evaluate the state of AI research in the United Kingdom. His report, now called the Lighthill report, criticized the utter failure of AI to achieve its "grandiose objectives." He concluded that nothing being done in AI couldn't be done in other sciences. He specifically mentioned the problem of "combinatorial explosion" or "intractability", which implied that many of AI's most successful algorithms would grind to a halt on real world problems and were only suitable for solving "toy" versions. British government ended support for AI research in all but two universities. Research would not revive on a large scale until 1983.

# Short History of AI

## 6. [1969-1979] Knowledge-based systems

- Use domain-specific knowledge instead of general-purpose search mechanism trying to string together elementary reasoning steps to find complete solutions (weak method). Domain-specific knowledge allows larger reasoning steps and can more easily handle typically occurring cases in narrow areas of expertise.
- The DENDRAL program to solve the problem of inferring molecular structure from the information provided by a mass spectrometer.
- Expert systems proliferate.
- MYCIN to diagnose blood infections. With about 450 rules, MYCIN was able to perform as well as some experts, and considerably better than junior doctors.

# Short History of AI

## 7. [1980-present] AI becomes an industry

- The first successful commercial expert system, RI, began operation at the Digital Equipment Corporation (McDermott, 1982). The program helped configure orders for new computer systems; by 1986, it was saving the company an estimated \$40 million a year. By 1988 DEC's AI group had 40 expert systems deployed, with more on the way.
- DuPont had 100 in use and 500 in development, saving an estimated \$10 million a year.
- Nearly every major U.S. corporation had its own AI group and was either using or investigating expert systems.
- In 1981, the Japanese announced the "Fifth Generation" project, a 10-year plan to build intelligent computers running Prolog.
- In response, the United States formed the Microelectronics and Computer Technology Corporation (MCC) as a research consortium designed to assure national competitiveness. In both cases, AI was part of a broad effort, including chip design and human-interface research.
- In Britain, the Alvey report reinstated the funding that was cut by the Lighthill report.

# Short History of AI

## 8. [1986-present] The return of neural networks

- In the mid-1980s at least four different groups reinvented the back-propagation learning algorithm first found in 1969.
- These so-called connectionist models of intelligent systems were seen by some as direct competitors both to the symbolic models promoted by Newell and Simon and to the logicist approach of McCarthy and others.
- The current view is that connectionist and symbolic approaches are complementary, not competing. As occurred with the separation of AI and cognitive science, modern neural network research has bifurcated into two fields, one concerned with creating effective network architectures and algorithms and understanding their mathematical properties, the other concerned with careful modeling of the empirical properties of actual neurons and ensembles of neurons.

# Short History of AI

## 9. [1987-early 1990s] Second AI winter

- The collapse of the Lisp machine market in 1987. Specialized computers, called Lisp machines, were optimized to process the programming language Lisp, the preferred language for AI. Workstations by companies like Sun Microsystems offered a powerful alternative to LISP machines and companies like Lucid offered a LISP environment for this new class of workstations.
- By the early 90s, the earliest successful expert systems proved too expensive to maintain. They were difficult to update, they could not learn, they were "brittle" (i.e., they could make grotesque mistakes when given unusual inputs), and they fell prey to problems (such as the qualification problem). Expert systems proved useful, but only in a few special contexts.
- By 1991, the impressive list of Fifth generation computer project (Japan) goals penned in 1981 had not been met. Funding cuts in american projects as well.

# Short History of AI

## 10. [1987-present] AI adopts the scientific method

- The isolationism is currently being abandoned. There is a recognition that machine learning should not be isolated from information theory, that uncertain reasoning should not be isolated from stochastic modeling, that search should not be isolated from classical optimization and control, and that automated reasoning should not be isolated from formal methods and static analysis.
- To be accepted, hypotheses must be subjected to rigorous empirical experiments, and the results must be analyzed statistically for their importance (vs. ad hoc method). It is now possible to replicate experiments by using shared repositories of test data and code.
- Machine learning and Data mining as separate fields.

# Short History of AI

## 11. [1995-present] The emergence of intelligent agents

- One of the most important environments for intelligent agents is the Internet. AI systems have become so common in Web-based applications that the "-bot" suffix has entered everyday language.
- Realization that the previously isolated subfields of AI might need to be reorganized somewhat when their results are to be tied together. AI has been drawn into much closer contact with other fields, such as control theory and economics, that also deal with agents.
- Artificial General Intelligence or AGI looks for a universal algorithm for learning and acting in any environment.

# Short History of AI

## 12. [2001-present] The availability of very large data sets

- Throughout the 60-year history of computer science, the emphasis has been on the algorithm as the main subject of study. But some recent work in AI suggests that for many problems, it makes more sense to worry about the data and be less picky about what algorithm to apply.
- More data gives significant improvement of quality with the same algorithm. Here is one of rationales behind the big data trend.
- Many works suggest that the “knowledge bottleneck” in AI—the problem of how to express all the knowledge that a system needs—may be solved in many applications by learning methods rather than hand-coded knowledge engineering, provided the learning algorithms have enough data to go on.

<b>Year</b>	<b>Breakthroughs in AI</b>	<b>Datasets (First Available)</b>	<b>Algorithms (First Proposed)</b>
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka “The Extended Book” (1991)	Negascout planning algorithm (1983)
2005	Google’s Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google’s GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google’s Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
<b>Average No. of Years to Breakthrough:</b>		<b>3 years</b>	<b>18 years</b>

# Short History of AI

## 13. [2010-present] Deep-learning revolution

- Significant advances in machine learning, especially deep learning (neural networks)
- Speech recognition and Computer vision is dominated by deep learning.
- More details in the next section.

# Machine Learning & Deep Learning

# Classical ML: Tasks

Typical ML tasks are:

1. **Classification**: predict a class of an object (cat/dog from an image, male/female from user's web activity, spam/ham from an email contents, etc)
2. **Regression**: predict a continuous value for an object (sales in the next month, price of a house with specific features, energy consumption for a household, etc)
3. **Clustering**: group similar objects together (find user groups different by their behaviour on the site, cluster customers into meaningful groups, etc)
4. ...

# Classical ML: Tasks (cont.)

Typical ML tasks are:

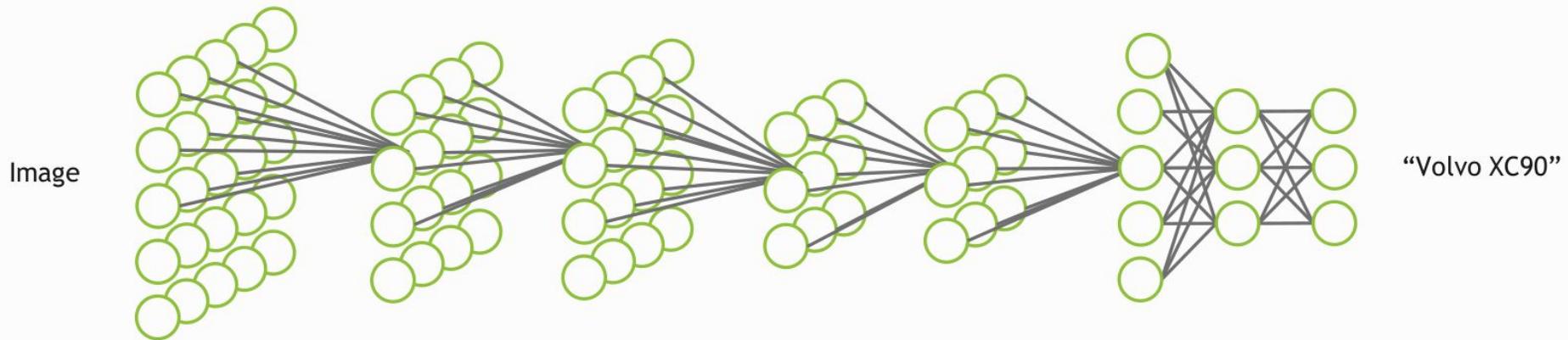
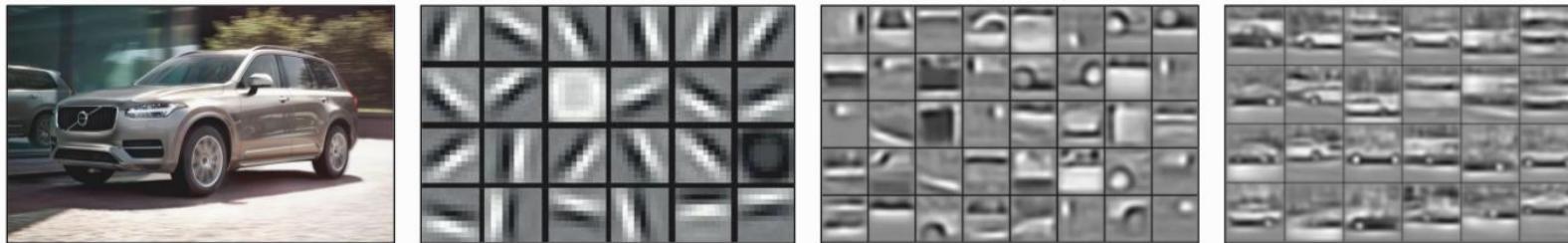
4. **Dimensionality reduction:** “compress” the data from a high-dimensional representation into a lower-dimensional one (useful for visualization or as an internal transformation for other ML algorithms)
5. **Ranking:** arrange a list of objects to maximize some utility (i.e. by click probability, or by the relevance to a query)
6. **Recommendations:** filter a small subset of objects from a large collection and recommend them to a user (increase sales and client loyalty, time spent on a service, cross-sale/up-sale, etc).

# Classical ML: Features

1. Well established field for rather simple tasks on meaningful features (object representations)
  - a. Much data are already useful features themselves: age, number of purchases, average order value, length, duration, weight, ...
  - b. Some fields have developed useful features to work with (SIFT, HoG, etc for images; MFCC for sound; TF\*IDF for words)
  - c. Recent advances in deep learning allowed to do much better here [see next slides]
2. Problem when you have no good features for your task
  - a. A large part of world is quantized but have no easily extractable inherent semantics: pixels, audio frames, characters and texts (partially) [deep learning aims to solve it]
  - b. Maybe even larger part of world is not quantized at all: too few biodata, sensor data, much of our life experiences are still analog and do not present in the digital world (for example, what we know about people emotion states? Only emoticons, manually chosen tags and photos (very partially))

# Deep Learning/Neural Nets: Overview

The idea of deep learning is the ability to learn hierarchical representations of features (and for images they do resemble some neuroscience data on how does visual system work).



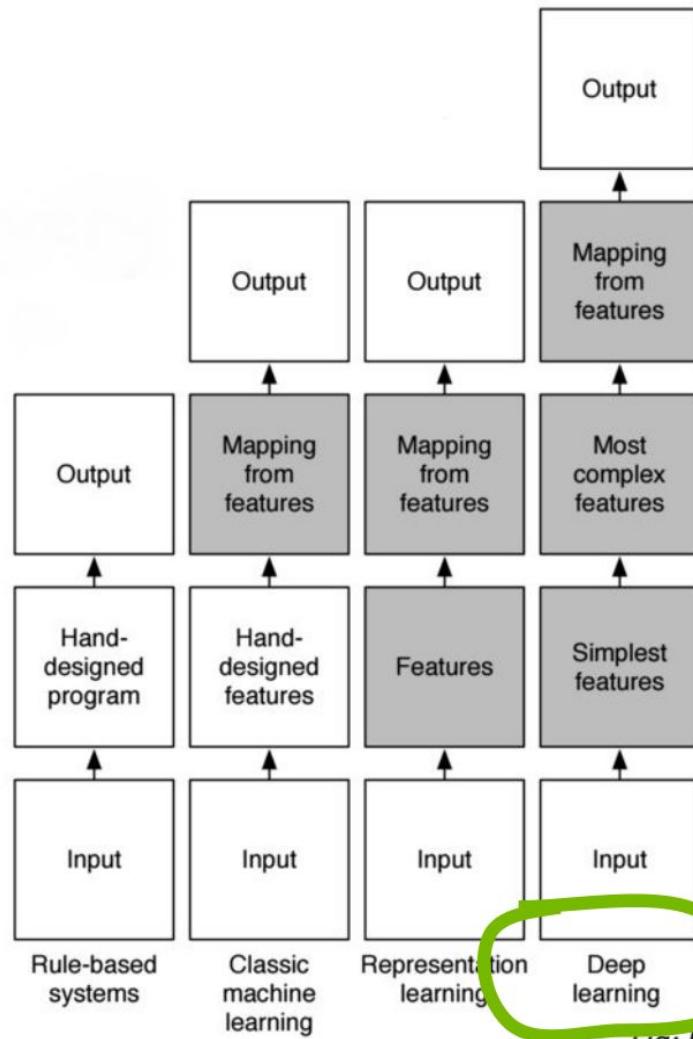
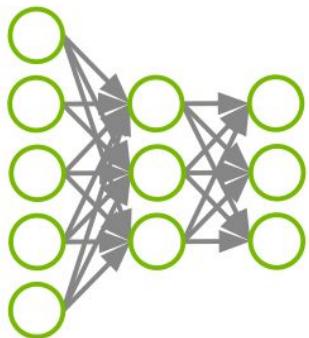


Fig: I. Goodfellow

# Why so much success now?



DNN



BIG DATA

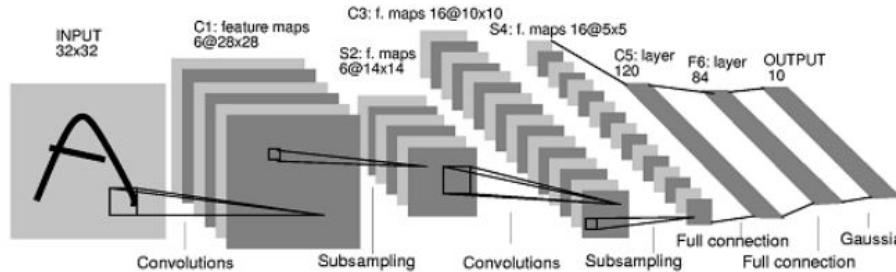


GPU

# More Data + More Power

1998

LeCun et al.



# of transistors



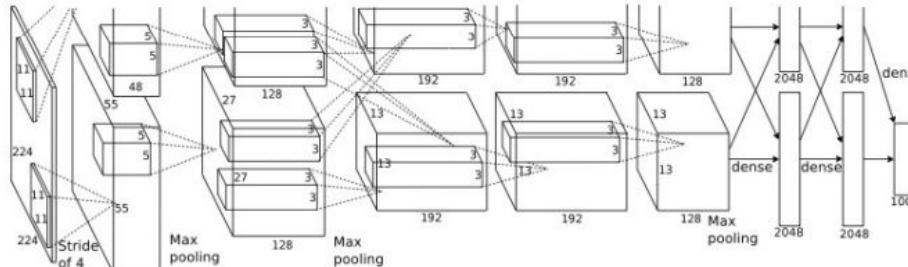
$10^6$

# of pixels used in training

$10^7$  NIST

2012

Krizhevsky  
et al.



# of transistors



$10^9$

GPUs



# of pixels used in training

$10^{14}$  IMAGENET

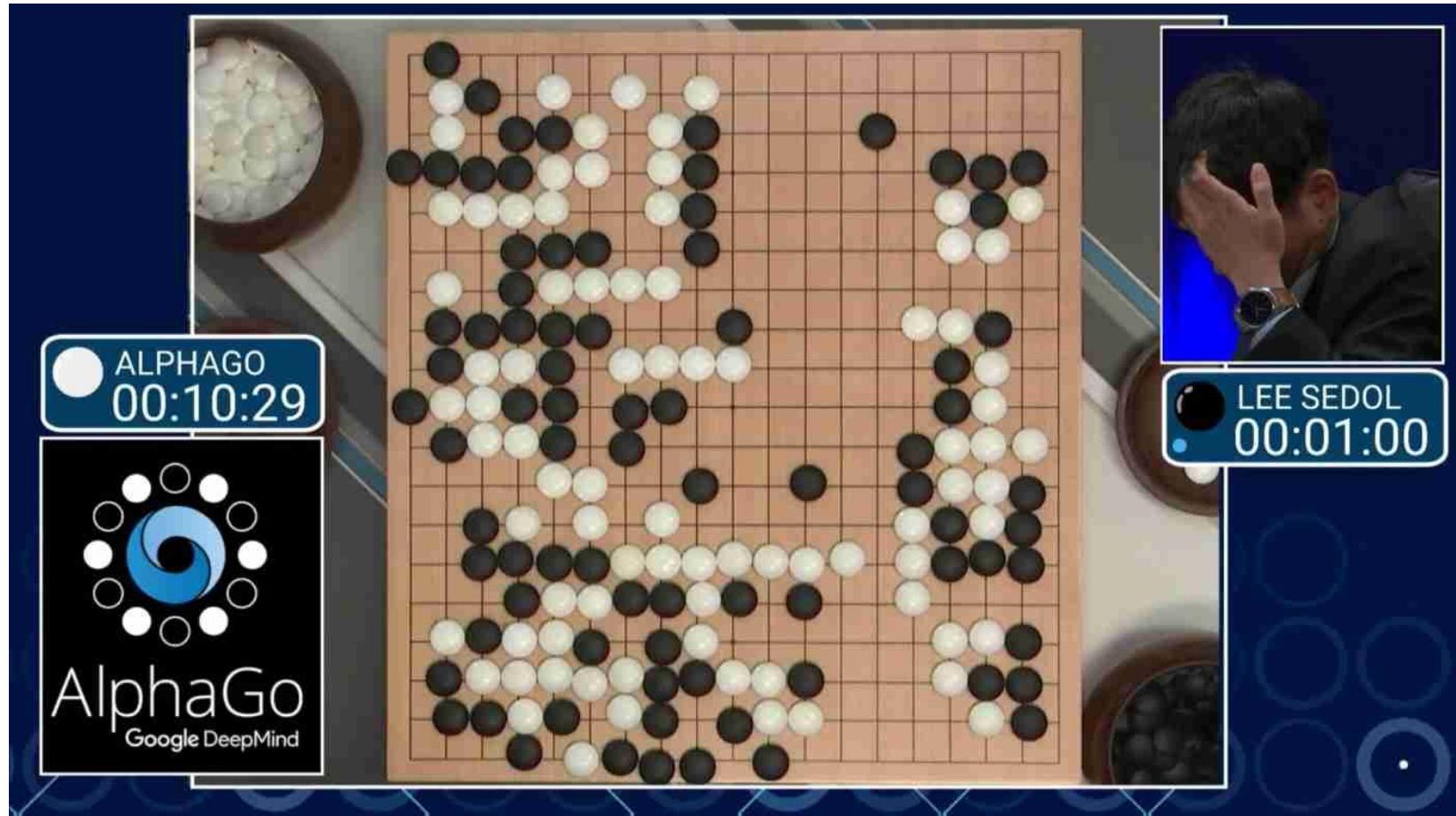
# DL/State of the Art

# DL/State of the Art

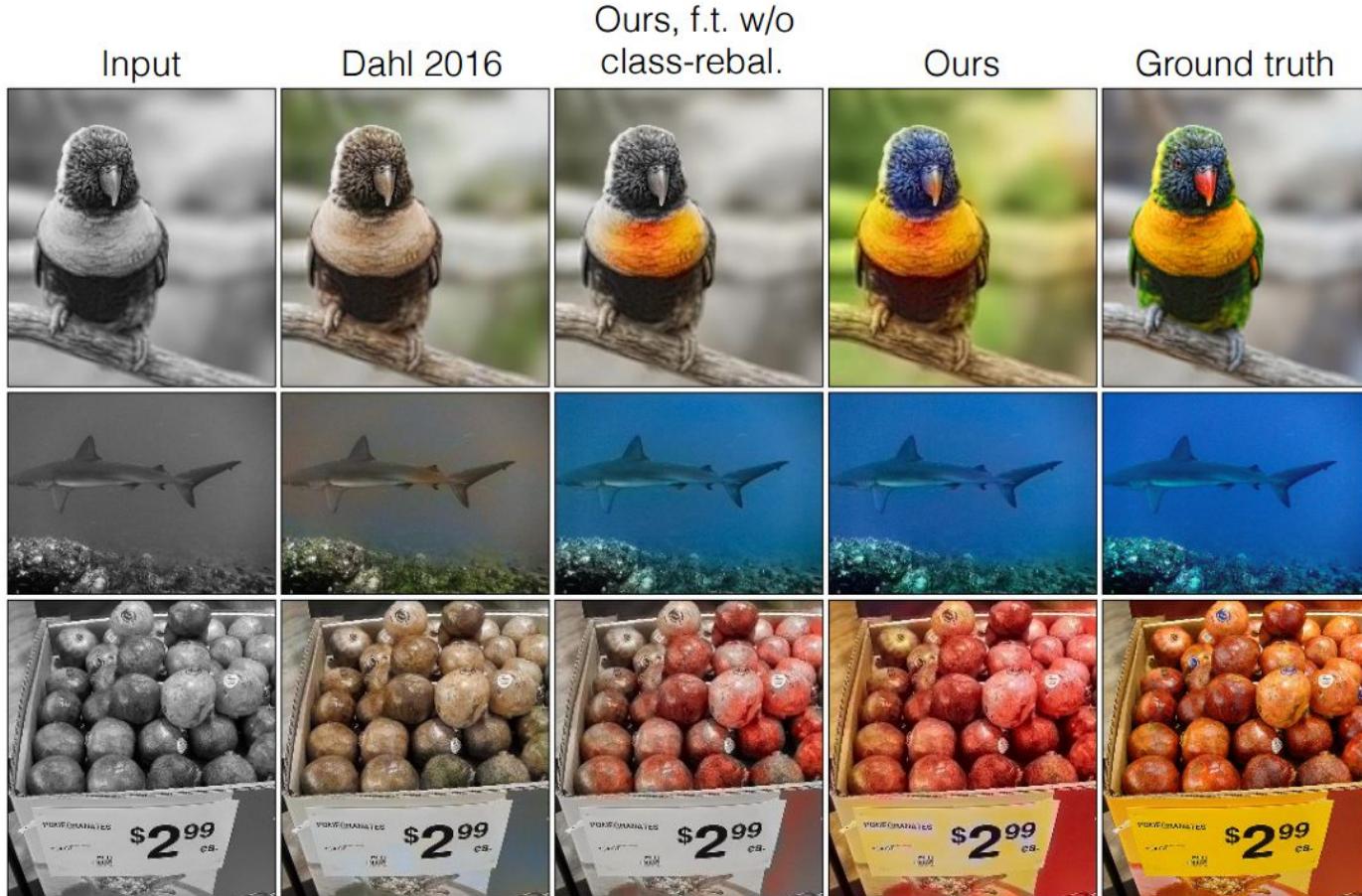
State of the art is rapidly evolving here. Each month someone shows something exciting new.

Just some recent cases of last months (and weeks or even days):

# Example: Game of Go: Computer-Human 4:1



# Example: Image Colorization



<http://richzhang.github.io/colorization/>

# Example: Car driving



Actually a “Perception to Action” system. The visual perception and control system is a Deep learning architecture trained end to end to transform pixels from the cameras into steering angles. And this car uses regular color cameras, not LIDARS like the Google cars. It is watching the driver and learns.

# Example: Image generation by text

this small bird has a pink breast and crown, and black primaries and secondaries.



this magnificent fellow is almost all black with a red crest, and white cheek patch.



the flower has petals that are bright pinkish purple with white stigma



this white and yellow flower have thin white petals and a round yellow stamen



*Figure 1.* Examples of generated images from text descriptions.  
Left: captions are from zero-shot (held out) categories, unseen text. Right: captions are from the training set.

# DL: Tasks

1. Speech, Images, Video [moving to the plateau of productivity]
  - a. Simple tasks (classification, localization) almost done on super-human level
  - b. Complex tasks (semantic segmentation, style transfer, image generation, ...) are in progress with exciting examples
  - c. Robot perception is in progress
2. Text/NLP [active research and development. A bit behind speech/images]
  - a. Some good basic technologies are in use (word embeddings, language models, sequence to sequence learning for machine translation, etc)
  - b. Even more in progress (translation, q&a and conversational nets, etc)
3. Reinforcement Learning [active research]
  - a. Great achievements exist: Go playing, Atari playing
4. A lot of other research
  - a. One-shot learning, multi-modal and multi-task learning, memory and attention, unsupervised learning, and many many more

# DL/Computer Vision

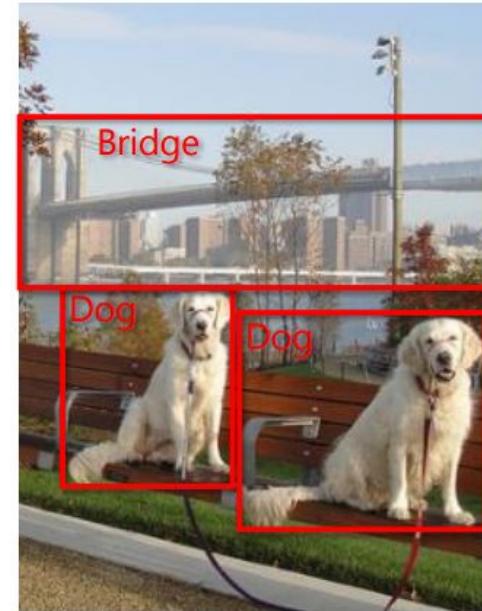
# Deep Learning and Computer Vision

Variety of tasks:

- Classification and detection, both are almost done and with better-than-human quality
- Activity recognition on video
- Semantic segmentation (whole image pixel-wise classification)
- Image caption generation, video describing
- Image style transfer
- Image synthesis
- Internet and catalog search



# Simple tasks: Classification and Detection



Detection task is harder than classification, but both are almost done.  
And with better-than-human quality.

# DL/Super-human Performance

Deep Learning already (since 2011) achieved a **super-human** (actually better than human) quality on simple **perceptual** (what we always thought of human sole right) tasks. Advantage over traditional ML methods in the same fields is even more significant.

# Case #1: IJCNN 2011

## The German Traffic Sign Recognition Benchmark

- Classification, >40 classes
- >50,000 real-life images
- First **Superhuman** Visual Pattern Recognition
  - 2x better than humans
  - 3x better than the closest artificial competitor
  - 6x better than the best non-neural method

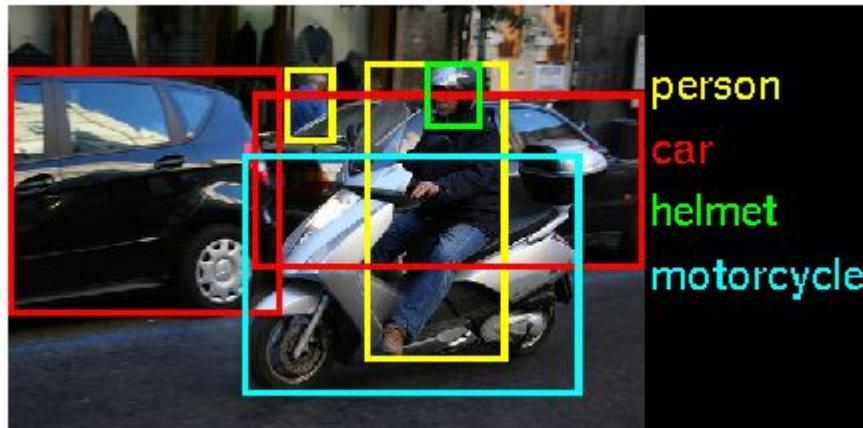
Method	Correct (Error)
<b>1 Committee of CNNs</b>	<b>99.46 % (0.54%)</b>
<b>2 Human Performance</b>	<b>98.84 % (1.16%)</b>
3 Multi-Scale CNNs	98.31 % (1.69%)
4 Random Forests	96.14 % (3.86%)



# Case #2: ILSVRC 2010-2015

## Large Scale Visual Recognition Challenge (ILSVRC)

- Object detection (200 categories, ~0.5M images)
- Classification + localization (1000 categories, 1.2M images)





**GT: horse cart**

**1: horse cart**

2: minibus

3: oxcart

4: stretcher

5: half track



**GT: birdhouse**

**1: birdhouse**

2: sliding door

3: window screen

4: mailbox

5: pot



**GT: forklift**

**1: forklift**

2: garbage truck

3: tow truck

4: trailer truck

5: go-kart



**GT: coucal**

**1: coucal**

2: indigo bunting

3: lorikeet



**GT: komondor**

**1: komondor**

2: patio

3: llama



**GT: yellow lady's slipper**

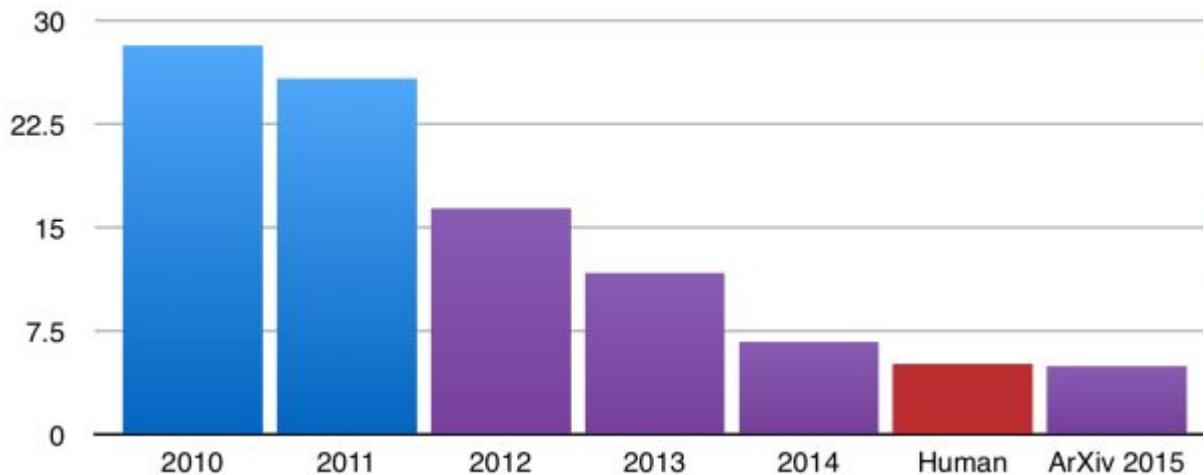
**1: yellow lady's slipper**

2: slug

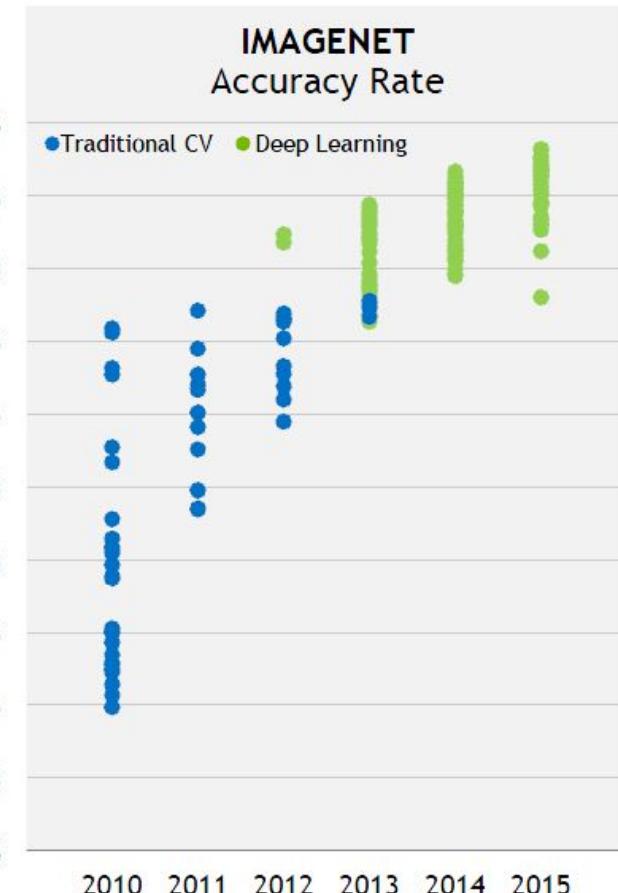
3: hen-of-the-woods

# Case #2: ILSVRC 2010-2015

## ILSVRC top-5 error on ImageNet



- Blue: Traditional CV
- Purple: Deep Learning
- Red: Human





**GT: spotlight**

- 1: grand piano
- 2: folding chair
- 3: rocking chair
- 4: dining table
- 5: upright piano



**GT: spotlight**

- 1: acoustic guitar
- 2: stage
- 3: microphone
- 4: electric guitar
- 5: banjo



**GT: spotlight**

- 1: altar
- 2: candle
- 3: perfume
- 4: restaurant
- 5: confectionery



**GT: restaurant**

- 1: wine bottle
- 2: candle
- 3: red wine
- 4: French loaf
- 5: wooden spoon



**GT: restaurant**

- 1: goblet
- 2: plate
- 3: candle
- 4: red wine
- 5: dining table



**GT: restaurant**

- 1: plate
- 2: meat loaf
- 3: ice cream
- 4: chocolate sauce
- 5: potpie

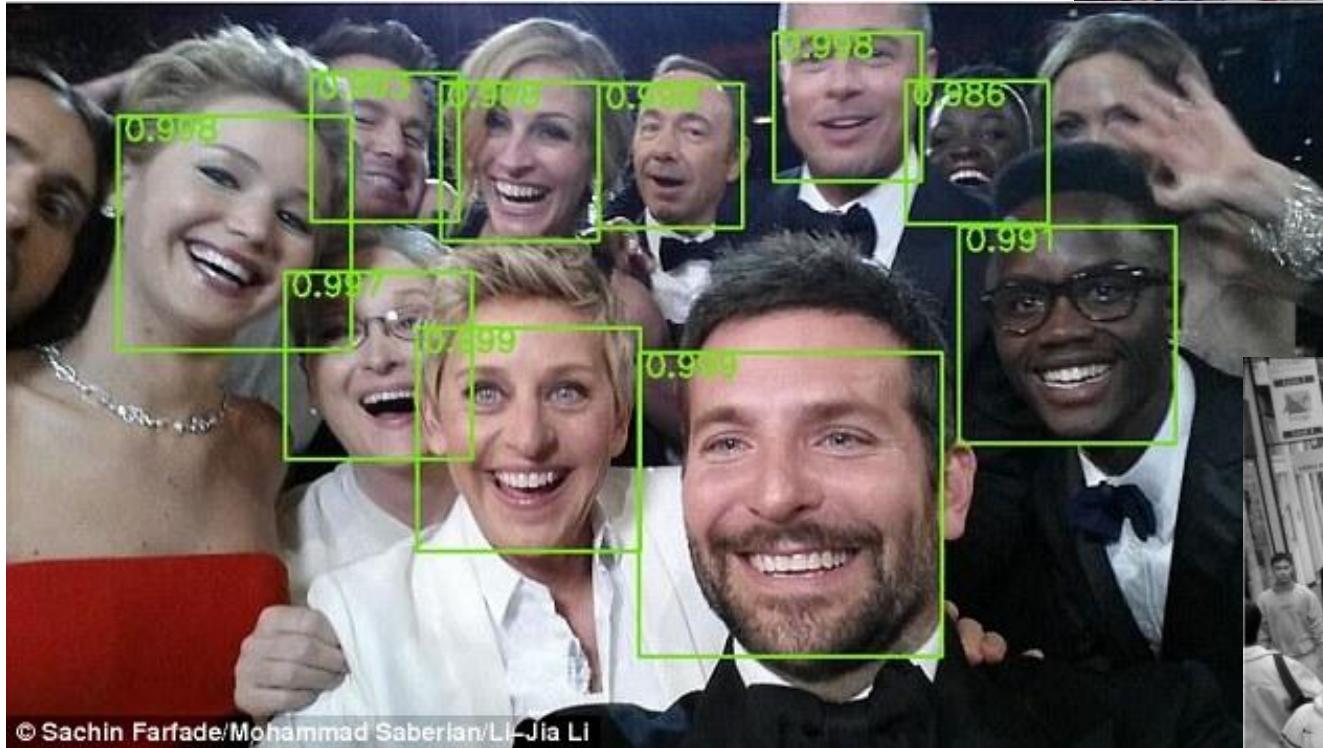
# Case #3: The Street View House Numbers (SVHN)

>600 000 real photos from Google Street View.

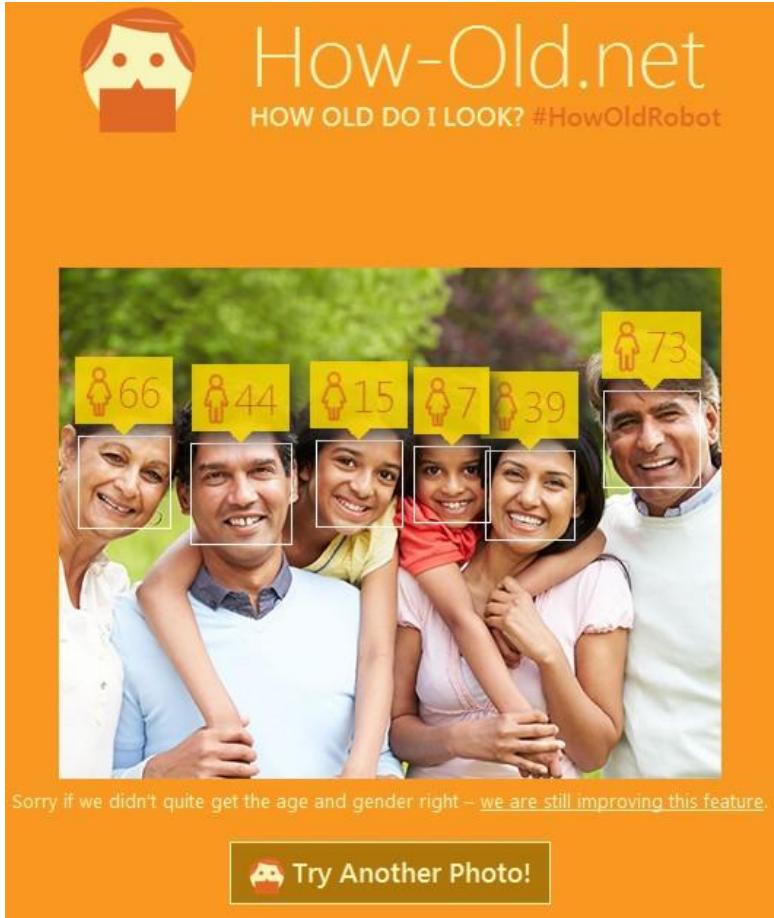


Result	Method	Venue	Details
1.69%	<a href="#">Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree</a>	AISTATS 2016	<a href="#">Details</a>
1.76%	<a href="#">Competitive Multi-scale Convolution</a>	arXiv 2015	
1.77%	<a href="#">Recurrent Convolutional Neural Network for Object Recognition</a>	CVPR 2015	<a href="#">Details</a>
1.81%	<a href="#">Batch-normalized Maxout Network in Network</a>	arXiv 2015	<a href="#">Details</a>
1.92%	<a href="#">Deeply-Supervised Nets</a>	arXiv 2014	
1.92%	<a href="#">Multi-Loss Regularized Deep Neural Network</a>	CSVT 2015	<a href="#">Details</a>
1.94%	<a href="#">Regularization of Neural Networks using DropConnect</a>	ICML 2013	
1.97%	<a href="#">On the Importance of Normalisation Layers in Deep Learning with Piecewise Linear Activation Units</a>	arXiv 2015	
2%	<a href="#">Estimated human performance</a>	NIPS 2011	<a href="#">Details</a>
2.16%	<a href="#">Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks</a>	ICLR 2014	<a href="#">Details</a>
2.35%	<a href="#">Network in Network</a>	ICLR 2014	<a href="#">Details</a>
2.38%	<a href="#">ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks</a>	arXiv 2015	

# Examples: Object Detection



# Example: Face Detection + Classification + Regression



# Simple Tasks: Face Detection + Emotion Classification



Neutral:  
Happiness:  
Surprise:  
Sadness:

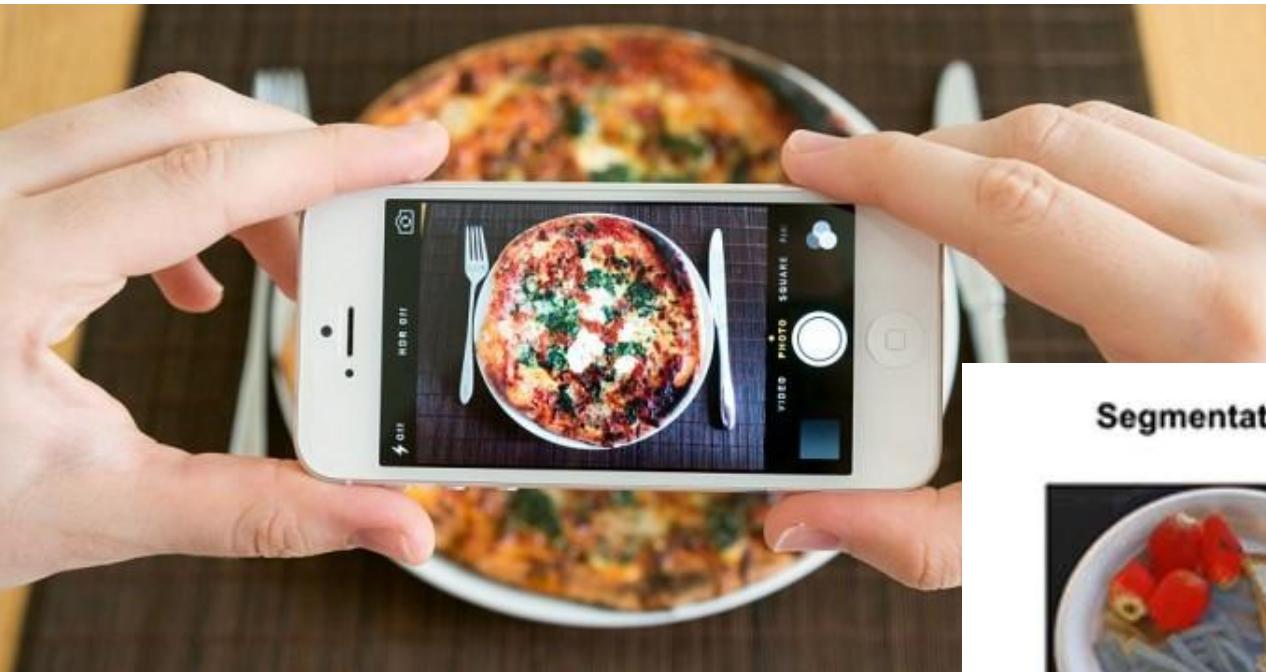


Anger:  
Disgust:  
Fear:  
Contempt:



Get started for free at [projectoxford.ai](http://projectoxford.ai)

# Examples: Food Recognition



Segmentation:



Portion & Nutrition  
Estimation:

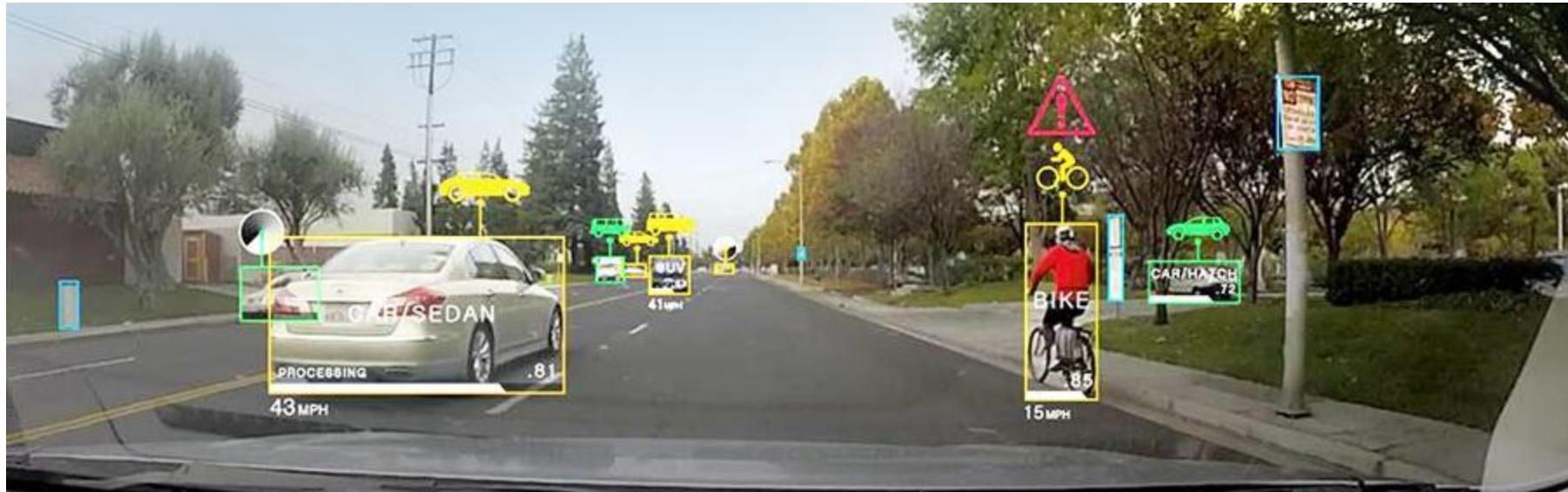
Sandwich: 146g  
French Fries: 85g  
Strawberries: 114g

Nutrition Facts	
Serving Size 1 (16g)	
Serving Per Container 1	
Amount Per Serving	
Calories 623	
% Daily Value*	
Total Fat 43g	66%
Saturated Fat 0g	0%
Trans Fat 0g	0%
Sodium 0mg	0%
Total Carbohydrate 102g	34%
Dietary Fiber 0g	0%
Sugars 0g	0%
Protein 10g	20%

# Examples: Activity Recognition

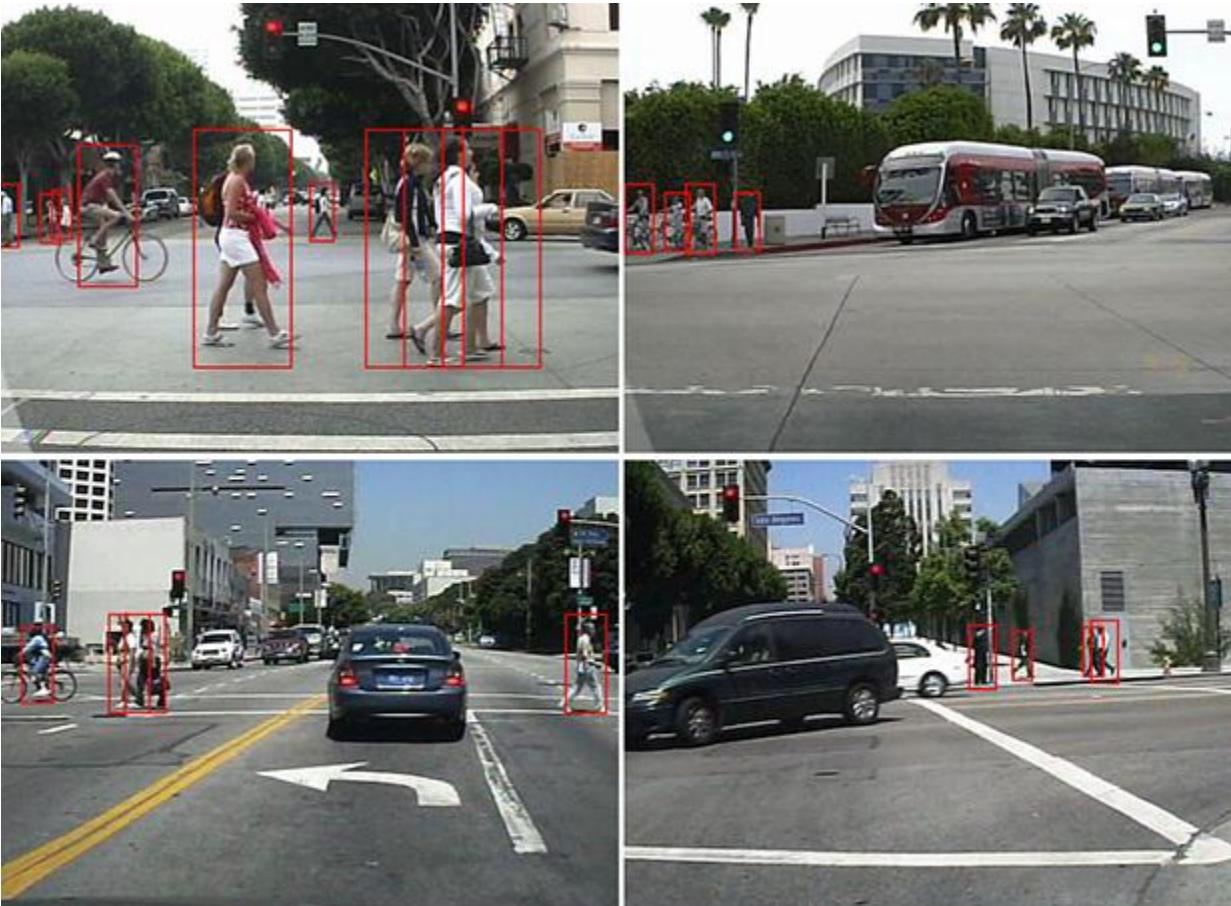


# Examples: Computer Vision on the Road



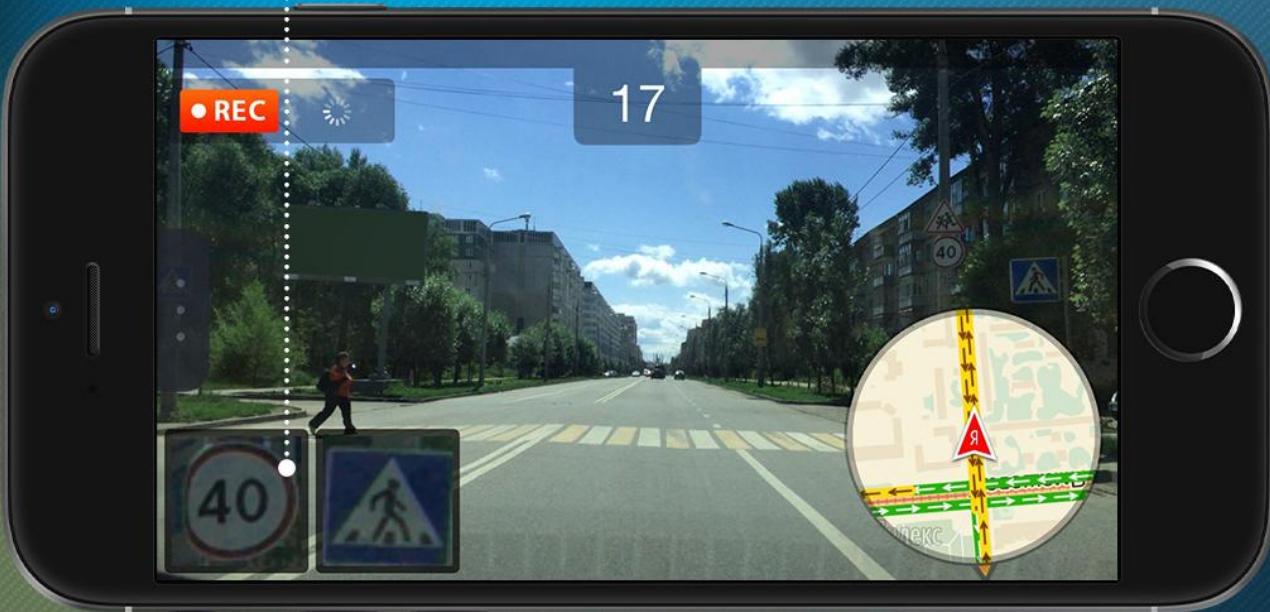
- = PROCESSING OBJECT ID
- ▲ = CAUTIONARY OBJECT
- = STATIONARY OBJECT
- = MOVING OBJECT
- = TRIVIAL OBJECT

# Examples: Pedestrian Detection

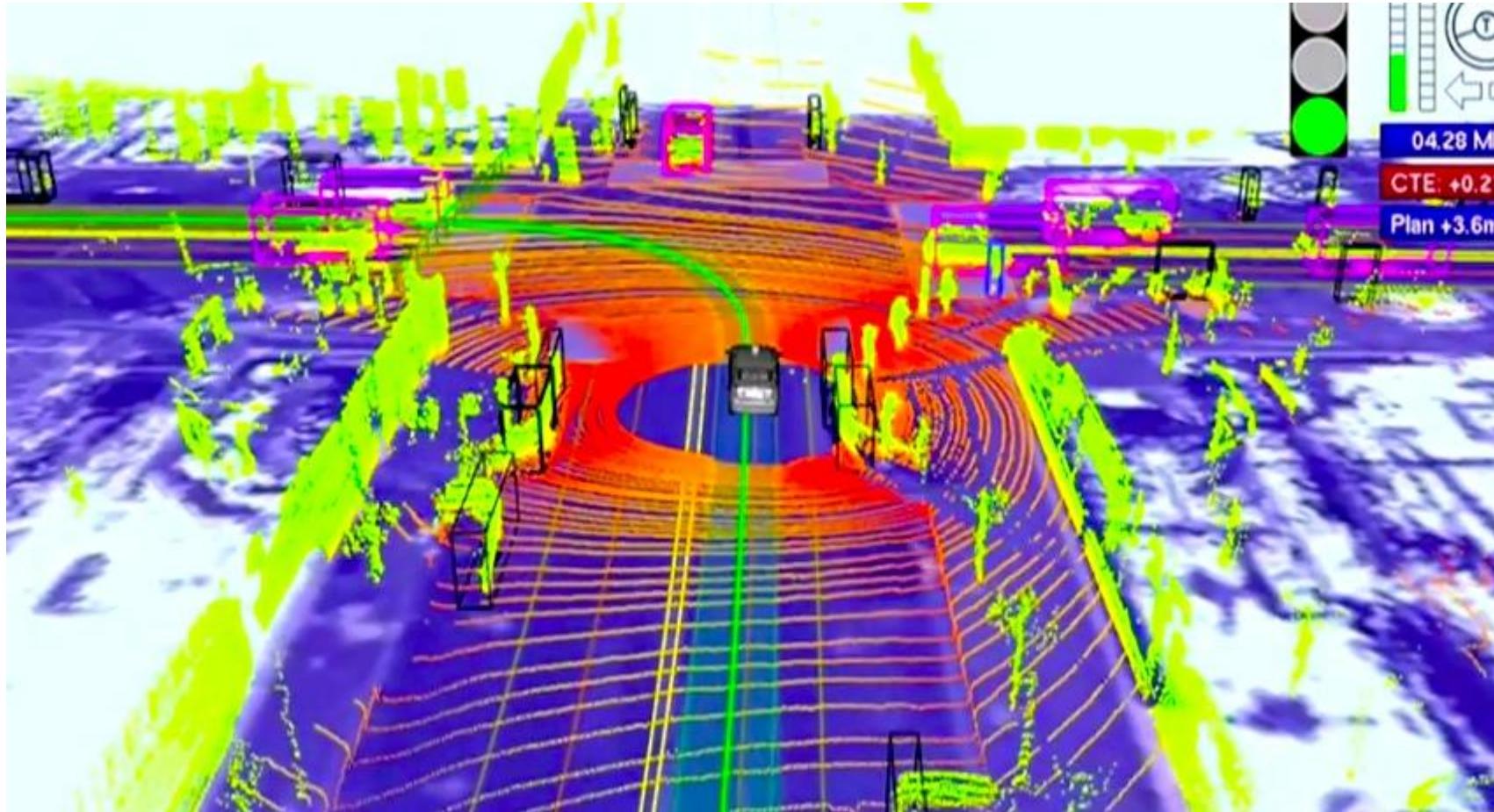


# Examples: Road Sign Recognition (on mobile!)

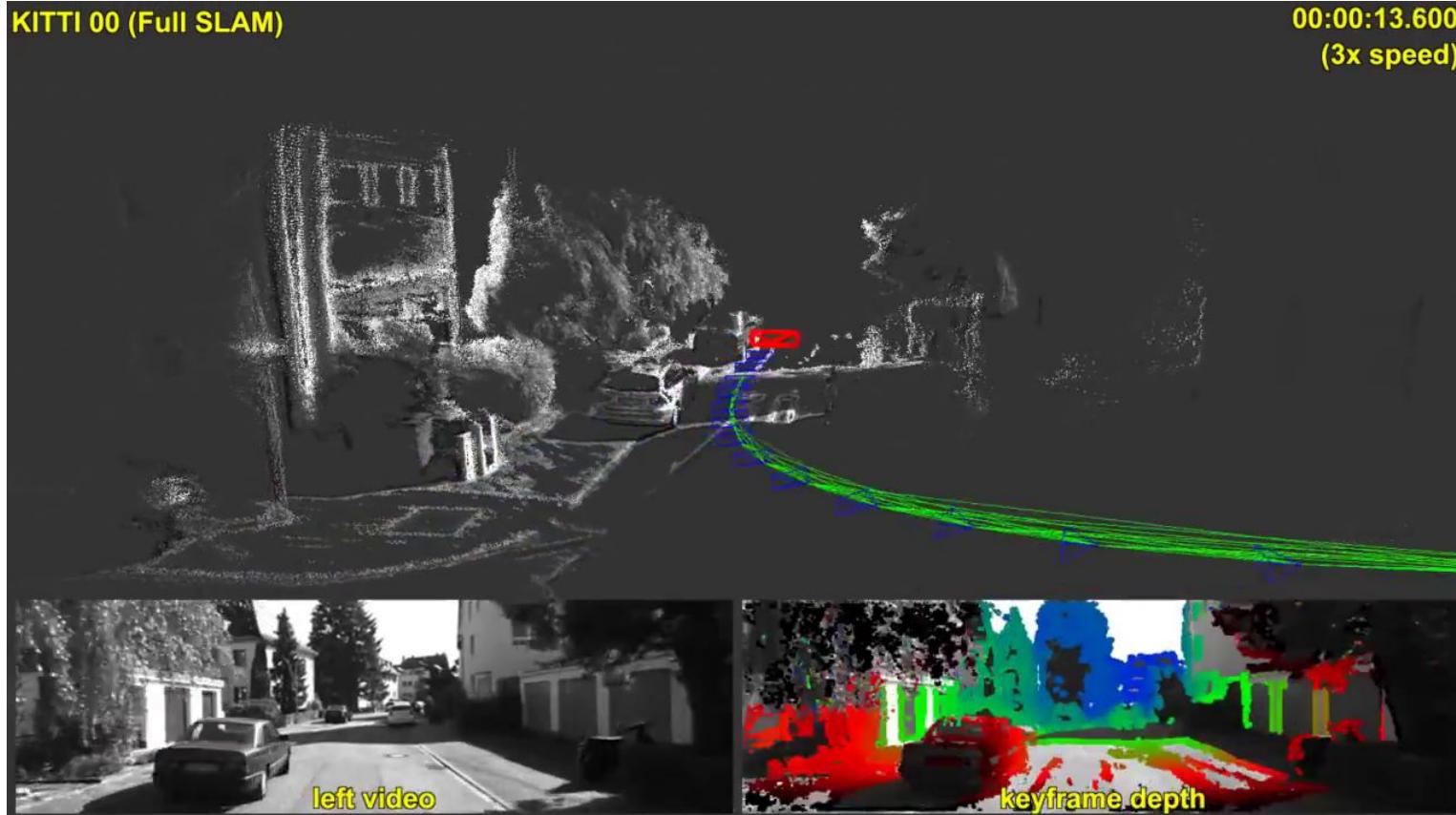
..... Распознавание знаков



# SLAM: Simultaneous Localization and Mapping



# LSD-SLAM: Large-Scale Direct Monocular SLAM

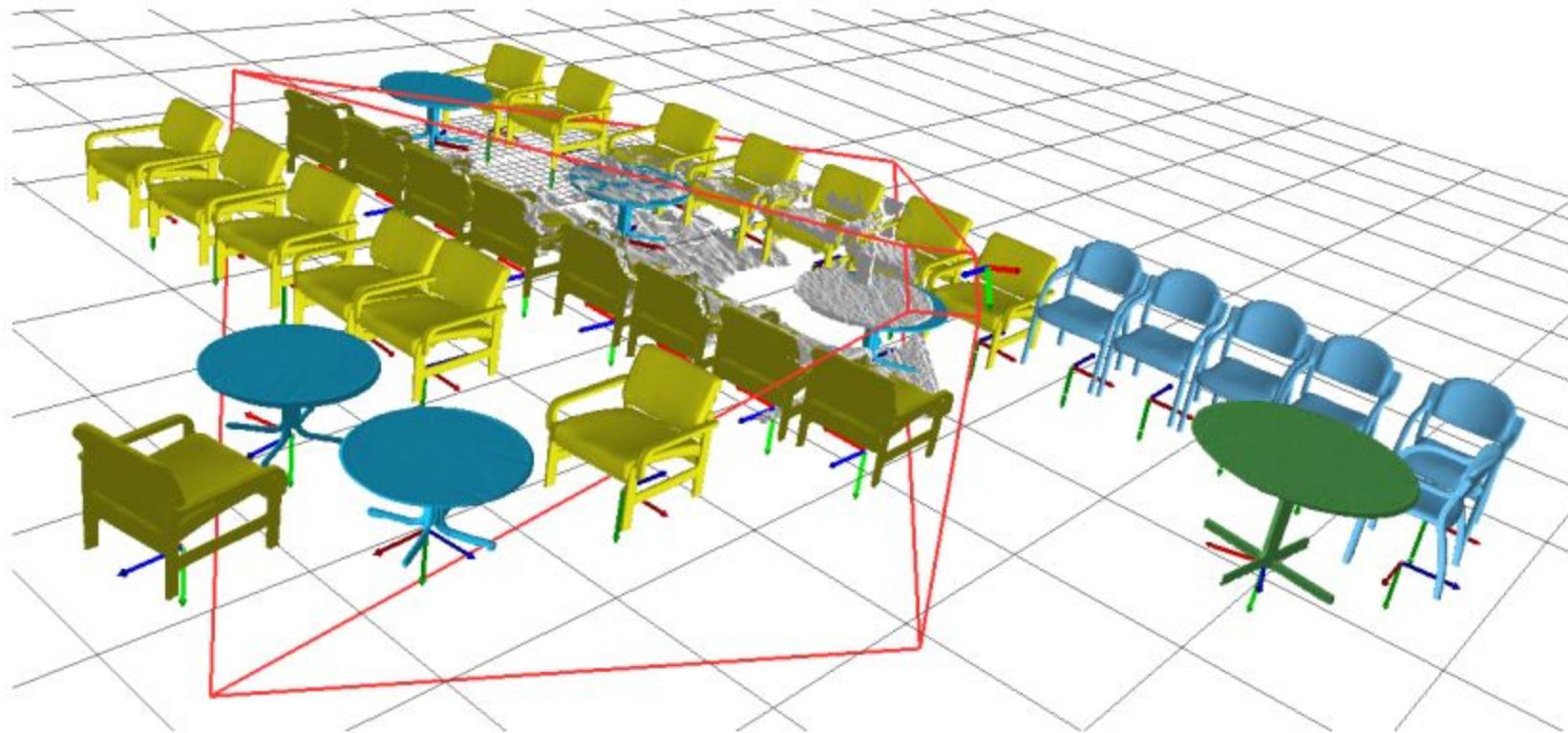


# Semantic SLAM



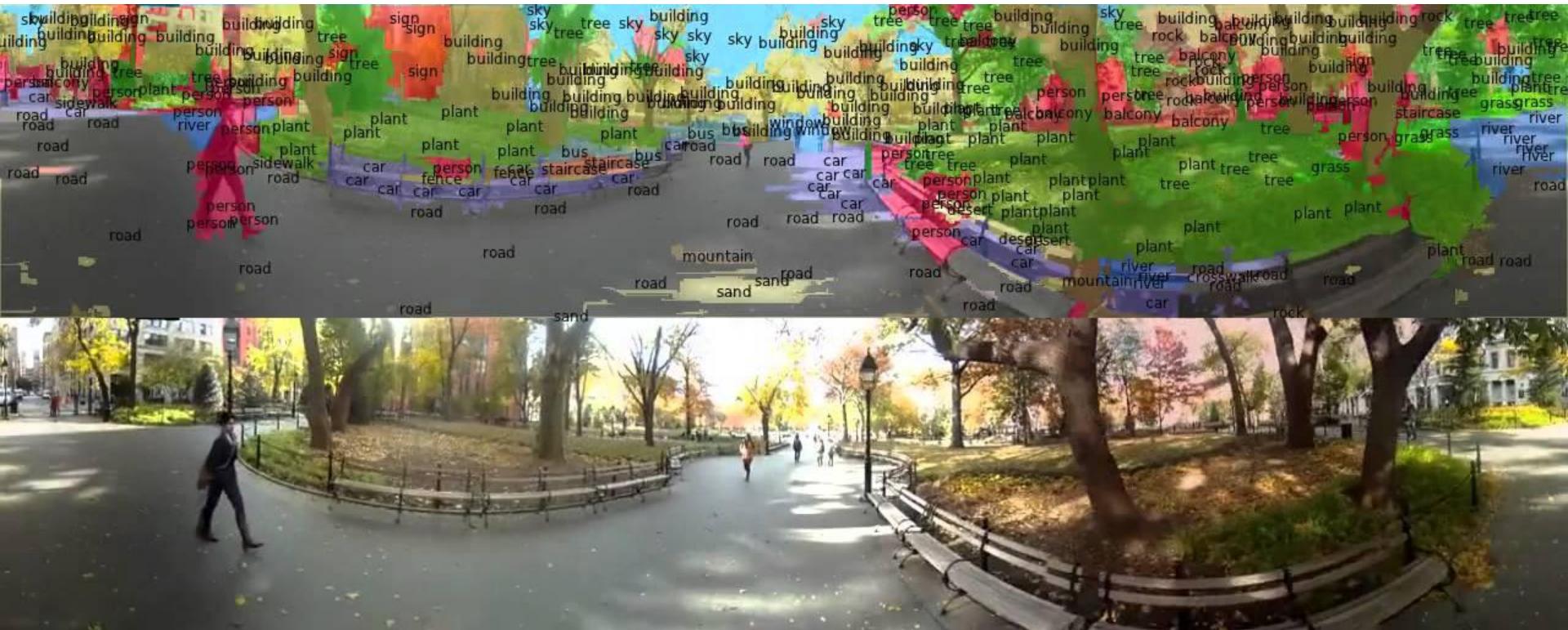
<https://www.doc.ic.ac.uk/~rfs09/docs/Salas-Moreno-R-2014-PhD-Thesis.pdf>

# SLAM++: Object-level SLAM



# DL/Other complex vision tasks

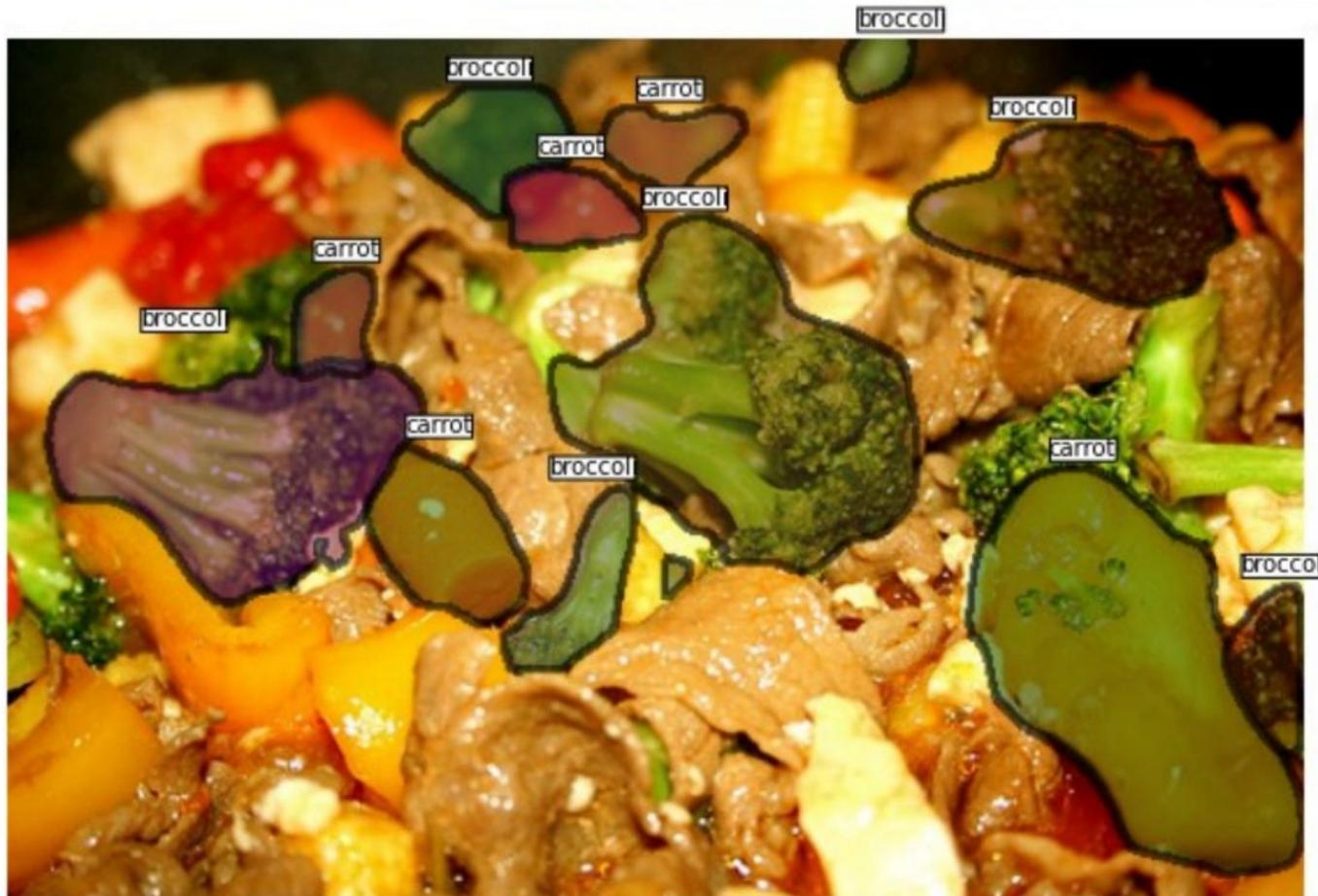
# Semantic Segmentation



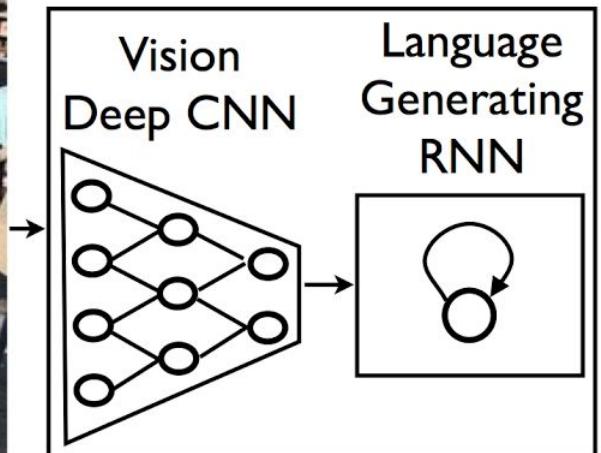
<https://www.youtube.com/watch?v=ZJMtDRbqH40>

NYU Semantic Segmentation with a Convolutional Network (33 categories)

# Complex Tasks: Semantic Segmentation



# Complex Tasks: Image Caption Generation



**A group of people shopping at an outdoor market.**

**There are many vegetables at the fruit stand.**



Human: "A group of men playing Frisbee in the park."

Computer model: "A group of young people playing a game of Frisbee."

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
			
A person riding a motorcycle on a dirt road.	Two dogs play in the grass.	A skateboarder does a trick on a ramp.	A dog is jumping to catch a frisbee.
			
A group of young people playing a game of frisbee.	Two hockey players are fighting over the puck.	A little girl in a pink hat is blowing bubbles.	A refrigerator filled with lots of food and drinks.
			
A herd of elephants walking across a dry grass field.	A close up of a cat laying on a couch.	A red motorcycle parked on the side of the road.	A yellow school bus parked in a parking lot.

# Example: NeuralTalk and Walk

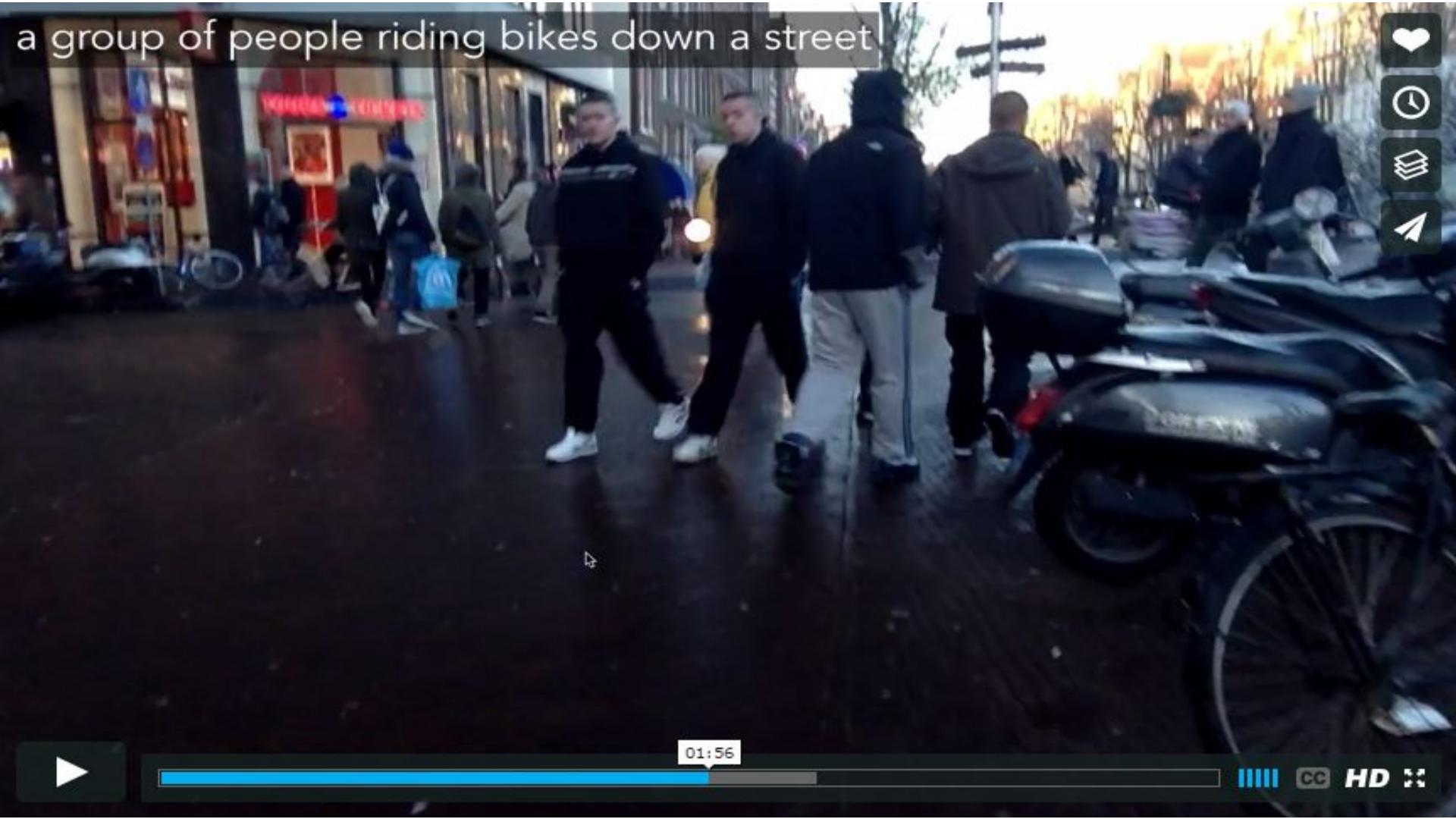
## Ingredients:

- <https://github.com/karpathy/neuraltalk2>  
Project for learning Multimodal Recurrent Neural Networks that describe images with sentences
- Webcam/notebook

## Result:

- <https://vimeo.com/146492001>

a group of people riding bikes down a street



# Complex Tasks: Visual Question Answering



What vegetable is on the plate?

Neural Net: **broccoli**  
Ground Truth: broccoli



What color are the shoes on the person's feet ?

Neural Net: **brown**  
Ground Truth: brown



How many school busses are there?

Neural Net: **2**  
Ground Truth: 2



What sport is this?  
Neural Net: **baseball**  
Ground Truth: baseball



What is on top of the refrigerator?

Neural Net: **magnets**  
Ground Truth: cereal



What uniform is she wearing?

Neural Net: **shorts**  
Ground Truth: girl scout



What is the table number?

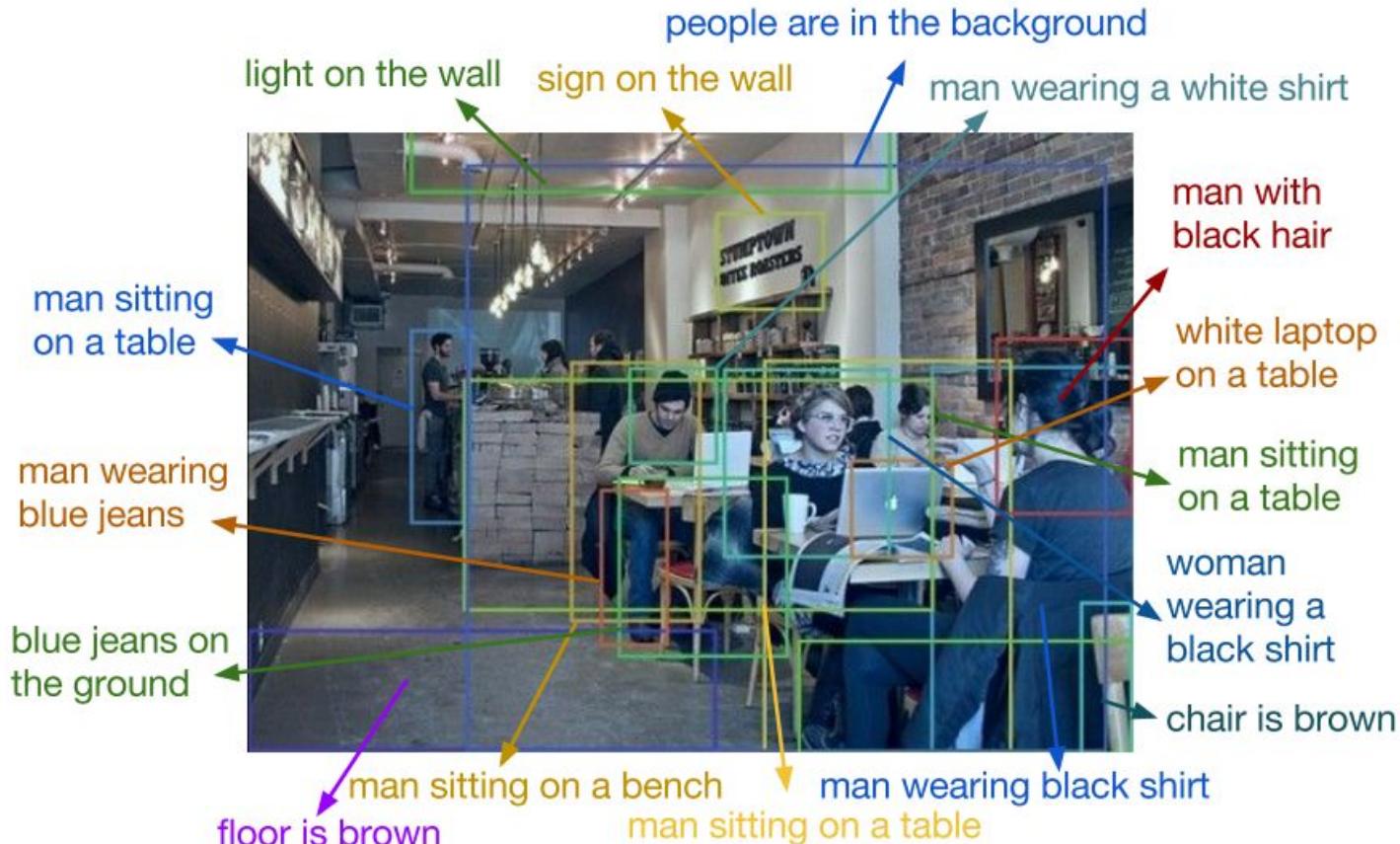
Neural Net: **4**  
Ground Truth: 40



What are people sitting under in the back?

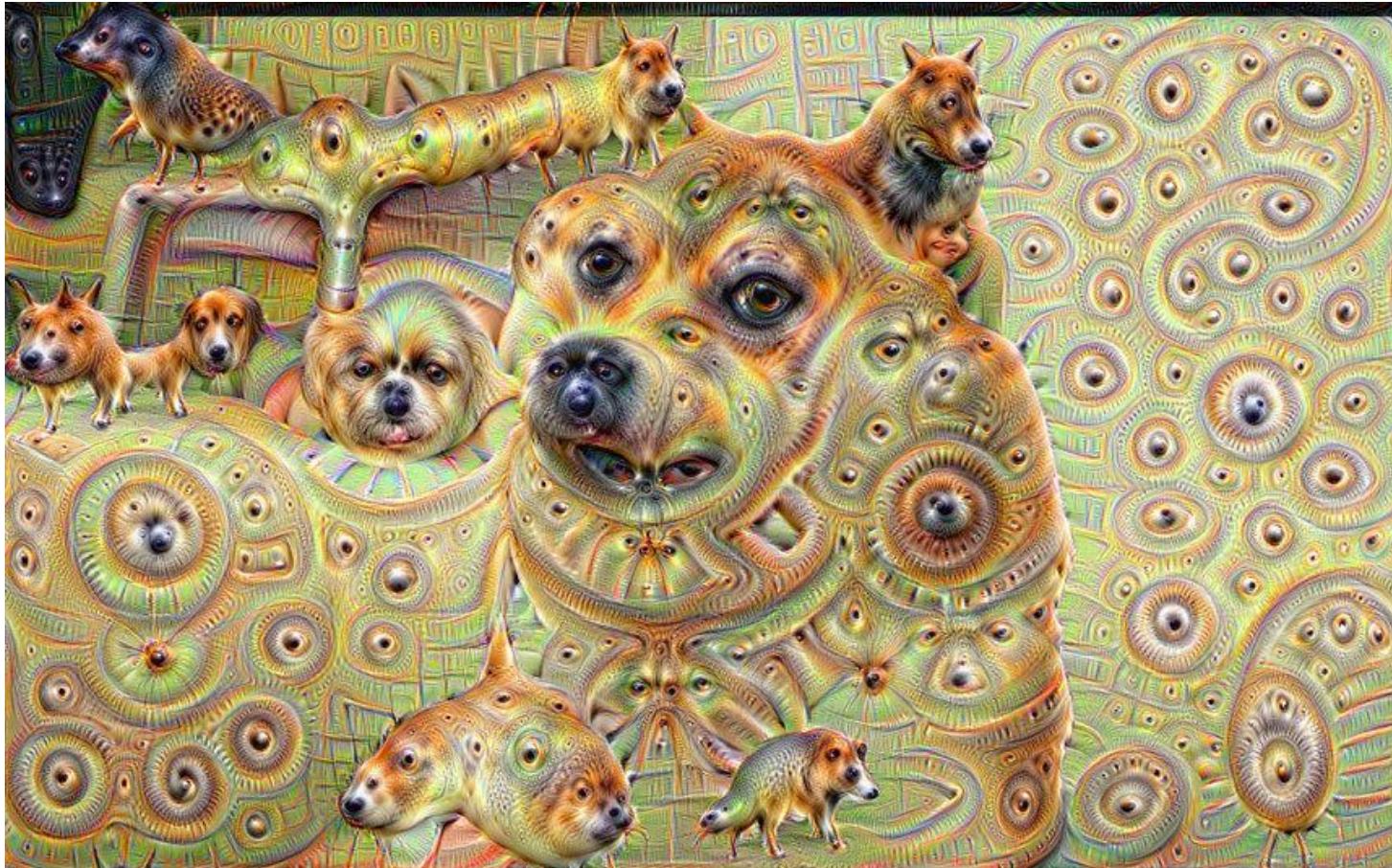
Neural Net: **bench**  
Ground Truth: tent

# Product of the near future: DenseCap and ?



DL/Art & Fun

# Fun: Deep Dream



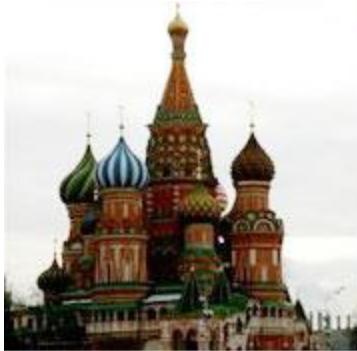
<http://blogs.wsj.com/digits/2016/02/29/googles-computers-paint-like-van-gogh-and-the-art-sells-for-thousands/>



# More Fun: Neural Style



# More Fun: Neural Style



# More Fun: Photo-realistic Synthesis



**Input style**



**Input content**



**Gatys et al**



**Ours**

# More Fun: Neural Doodle



(a) Original painting by Renoir, (b) semantic annotations,  
(c) desired layout, (d) generated output.

# DL/Natural Language Processing

# Deep Learning and NLP

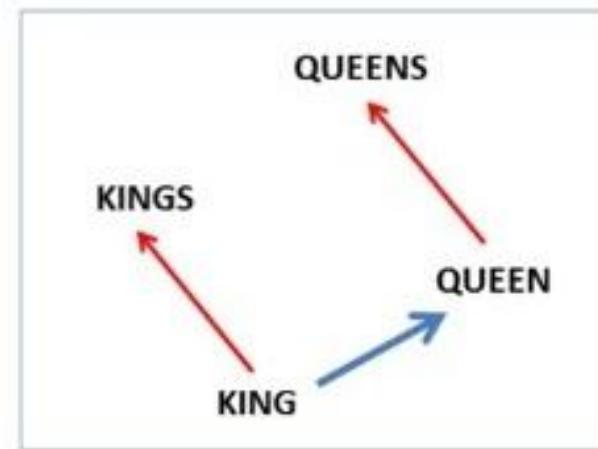
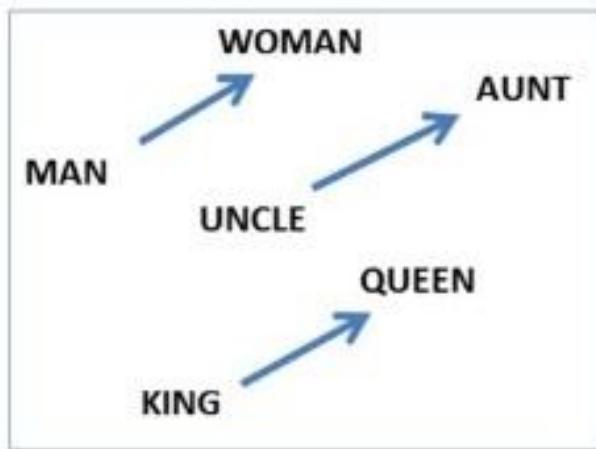
Variety of tasks:

- Finding synonyms
- Fact extraction: people and company names, geography, prices, dates, product names, ...
- Classification: genre and topic detection, positive/negative sentiment analysis, authorship detection, ...
- Machine translation
- Search (written and spoken)
- Question answering
- Dialog systems
- Speech recognition and Language modeling
- Text generation (Smart Reply, ...)

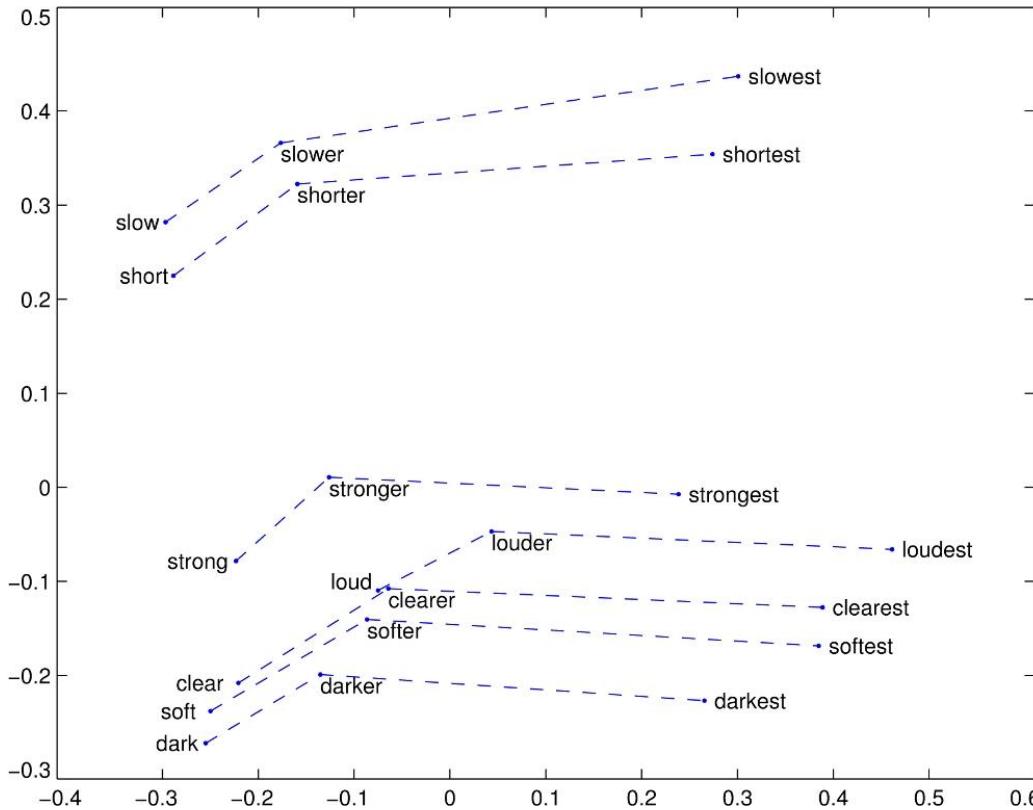


# Example: Semantic Spaces (word2vec, GloVe)

$$\text{vec}(\text{"man"}) - \text{vec}(\text{"king"}) + \text{vec}(\text{"woman"}) = \text{vec}(\text{"queen"})$$



# Example: Semantic Spaces (word2vec, GloVe)



# Encoding semantics

Using word2vec instead of word indexes allows you to better deal with the word meanings (e.g. no need to enumerate all synonyms because their vectors are already close to each other).

But the naive way to work with word2vec vectors still gives you a “bag of words” model, where phrases “The man killed the tiger” and “The tiger killed the man” are equal.

Need models which pay attention to the word ordering: paragraph2vec, sentence embeddings (using RNN/LSTM), even Word2Vec (LeCunn @CVPR2015).

# Multi-modal learning

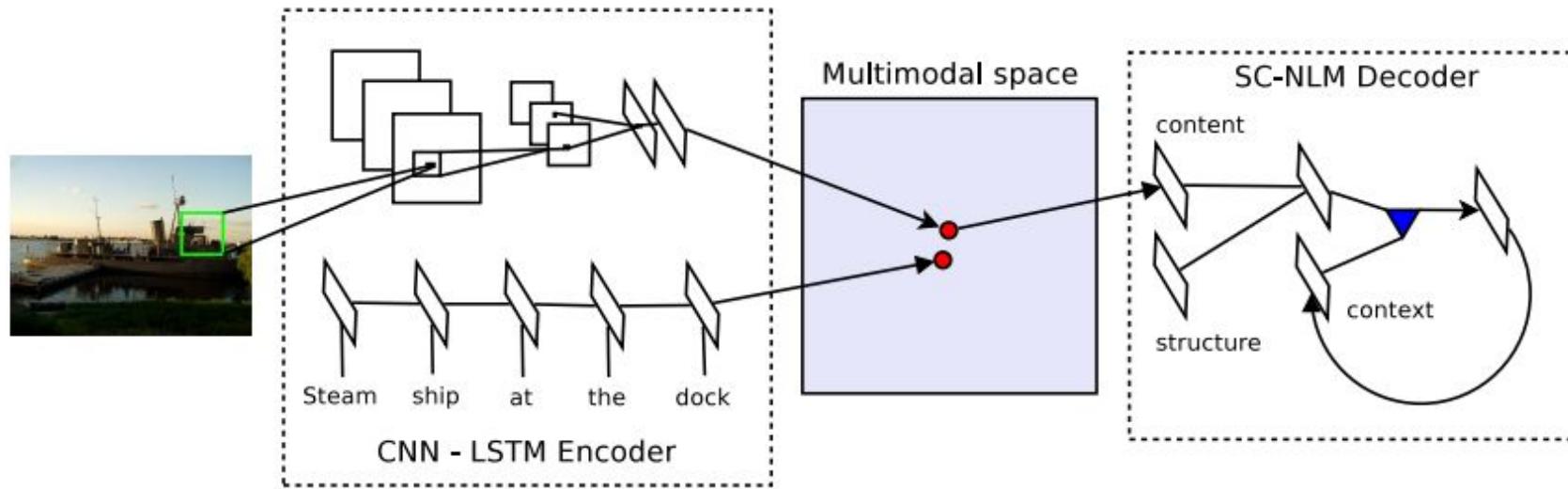


Figure 2: **Encoder:** A deep convolutional network (CNN) and long short-term memory recurrent network (LSTM) for learning a joint image-sentence embedding. **Decoder:** A new neural language model that combines structure and content vectors for generating words one at a time in sequence.

# Example: More multi-modal learning

Nearest images



- blue + red =



- blue + yellow =



- yellow + red =



- white + red =



# Nearest Images



- day + night =



- flying + sailing =



- bowl + box =



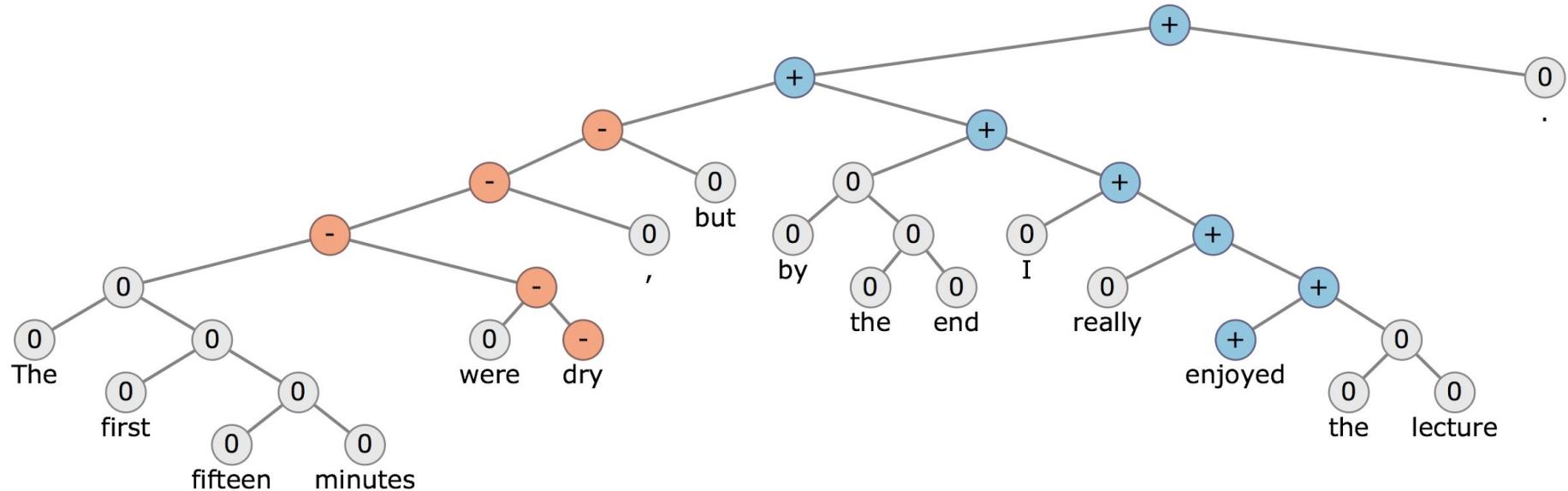
- box + bowl =



# Example: Sentiment analysis

Can capture complex cases where bag-of-words models fail.

*“This movie was actually neither that funny, nor super witty.”*



# Example: Machine Translation

Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	<b>34.81</b>

Table 1: The performance of the LSTM on WMT'14 English to French test set (ntst14). Note that an ensemble of 5 LSTMs with a beam of size 2 is cheaper than of a single LSTM with a beam of size 12.

Method	test BLEU score (ntst14)
Baseline System [29]	33.30
Cho et al. [5]	34.54
State of the art [9]	<b>37.0</b>
Rescoring the baseline 1000-best with a single forward LSTM	35.61
Rescoring the baseline 1000-best with a single reversed LSTM	35.85
Rescoring the baseline 1000-best with an ensemble of 5 reversed LSTMs	<b>36.5</b>
Oracle Rescoring of the Baseline 1000-best lists	~45

Table 2: Methods that use neural networks together with an SMT system on the WMT'14 English to French test set (ntst14).

# Example: Machine Translation

System	Vocab	Corpus	BLEU
State of the art in WMT'14 (Durrani et al., 2014)	All	36M	37.0
<i>Standard MT + neural components</i>			
Schwenk (2014) – neural language model	All	12M	33.3
Cho et al. (2014) – phrase table neural features	All	12M	34.5
Sutskever et al. (2014) – 5 LSTMs, reranking 1000-best lists	All	12M	36.5
<i>Existing end-to-end NMT systems</i>			
Bahdanau et al. (2015) – single gated RNN with search	30K	12M	28.5
Sutskever et al. (2014) – 5 LSTMs	80K	12M	34.8
Jean et al. (2015) – 8 gated RNNs with search + UNK replacement	500K	12M	37.2
<i>Our end-to-end NMT systems</i>			
Single LSTM with 4 layers	40K	12M	29.5
Single LSTM with 4 layers + PosUnk	40K	12M	31.8 (+2.3)
Single LSTM with 6 layers	40K	12M	30.4
Single LSTM with 6 layers + PosUnk	40K	12M	32.7 (+2.3)
Ensemble of 8 LSTMs	40K	12M	34.1
Ensemble of 8 LSTMs + PosUnk	40K	12M	36.9 (+2.8)
Single LSTM with 6 layers	80K	36M	31.5
Single LSTM with 6 layers + PosUnk	80K	36M	33.1 (+1.6)
Ensemble of 8 LSTMs	80K	36M	35.6
Ensemble of 8 LSTMs + PosUnk	80K	36M	<b>37.5 (+1.9)</b>

“Our experiments on the WMT14 English to French translation task show that this method provides a substantial improvement of up to 2.8 BLEU points over an equivalent NMT system that does not use this technique. With 37.5 BLEU points, **our NMT system is the first to surpass the best result achieved on a WMT14 contest task.**”

# Quality of Machine Translation systems

- BLEU score is a metric for Machine Translation quality
  - Measures overlap between machine translation output and reference (human) translation
  - BLEU scores range from 0-100%
  - BLEU has frequently been reported as correlating well with human judgement, however, few human translations will attain a score of 1 because of higher variability and different word choices
- Best BLEU scores by [1]:

lang pair	BLEU	dataset
en=>fr	37.5	WMT'14
fr=>en	35.8	WMT'14
ar=>en	56.4	NIST OpenMT'12
ch=>en	40.06	MT03
ge=>en	29.3	WMT'15
en=>ge	26.5	WMT'15
en=>ru	29.37	newstest-14
en=>ja	36.21	WAT'15

[1] "Review of state-of-the-arts in artificial intelligence. Present and future of AI". Vladimir Shakirov <http://sciencevsdeath.com/review-of-state-of-the-arts.pdf>

# Example: Automated Speech Translation

Translating voice calls and video calls in 7 languages and instant messages in over 50.

<https://www.skype.com/en/features/skype-translator/>

This call may be recorded to improve translations.

skype

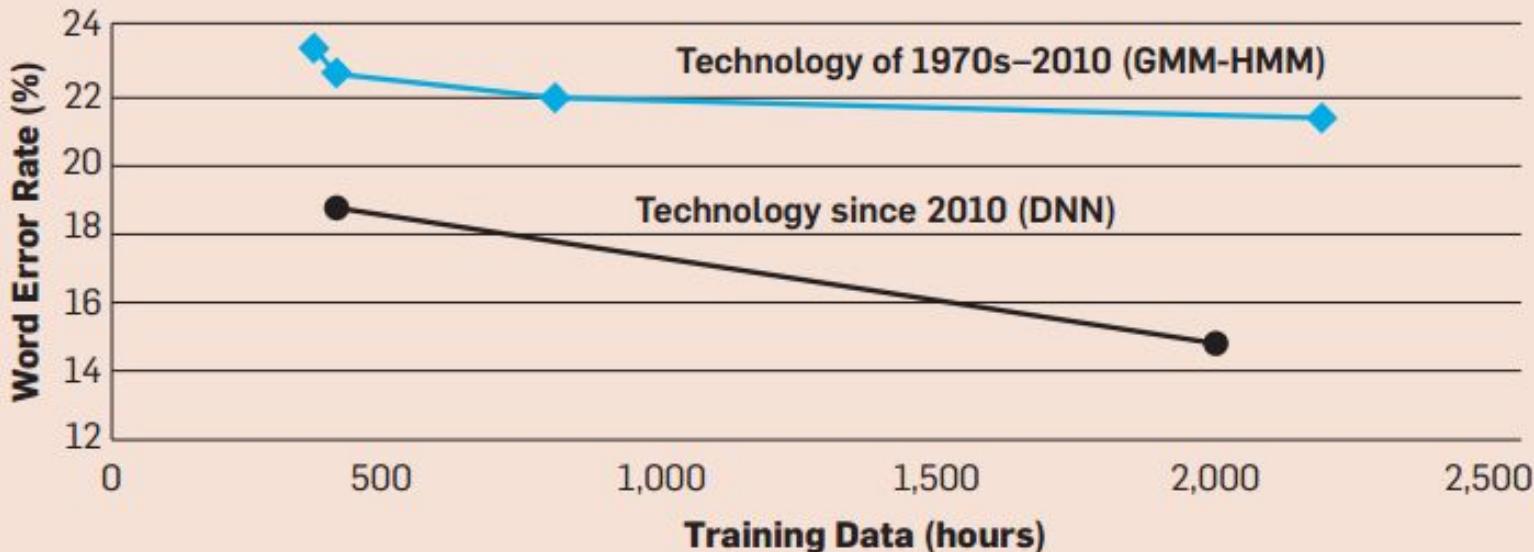
I was wondering what you are going to do later?  
Me preguntaba lo que vas a hacer después?

Going to the pub, do you want to join us? I think you  
met some of the team at the party in April.  
Al pub, ¿quieres unirte a nosotros? Creo que conociste  
a algunos miembros del equipo en la fiesta en abril.

Can I join you guys after my meeting?  
¿Puedo unirme a ustedes después de mi reunión?

Type a message in English here

# Speech Recognition: DL quality



# Speech Recognition: Word Error Rate (WER)

“Google now has just an 8 percent error rate. Compare that to 23 percent in 2013” (2015)

<http://venturebeat.com/2015/05/28/google-says-its-speech-recognition-technology-now-has-only-an-8-word-error-rate/>

IBM Watson. “The performance of our new system – an 8% word error rate – is 36% better than previously reported external results.” (2015)

<https://developer.ibm.com/watson/blog/2015/05/26/ibm-watson-announces-breakthrough-in-conversational-speech-transcription/>

Baidu. “We are able to reduce error rates of our previous end-to-end system in English by up to 43%, and can also recognize Mandarin speech with high accuracy. Creating high-performing recognizers for two very different languages, English and Mandarin, required essentially no expert knowledge of the languages” (2015)

<http://arxiv.org/abs/1512.02595>

# Example: Baidu Deep Speech 2 (2015)

- “The Deep Speech 2 ASR pipeline approaches or **exceeds the accuracy of Amazon Mechanical Turk human workers** on several benchmarks, works in multiple languages with little modification, and is deployable in a production setting.”
- “Table 13 shows that the DS2 system **outperforms humans in 3 out of the 4 test sets** and is competitive on the fourth. Given this result, we suspect that there is little room for a generic speech system to further improve on clean read speech without further domain adaptation”

Read Speech			
Test set	DS1	DS2	Human
WSJ eval'92	4.94	3.60	5.03
WSJ eval'93	6.94	4.98	8.08
LibriSpeech test-clean	7.89	5.33	5.83
LibriSpeech test-other	21.74	13.25	12.69

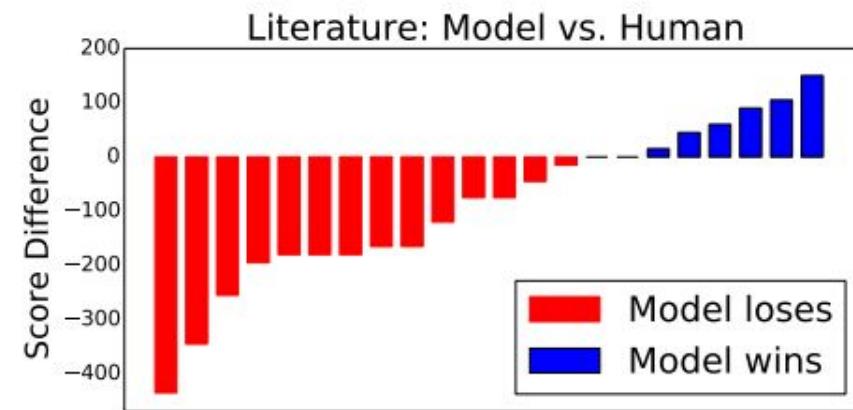
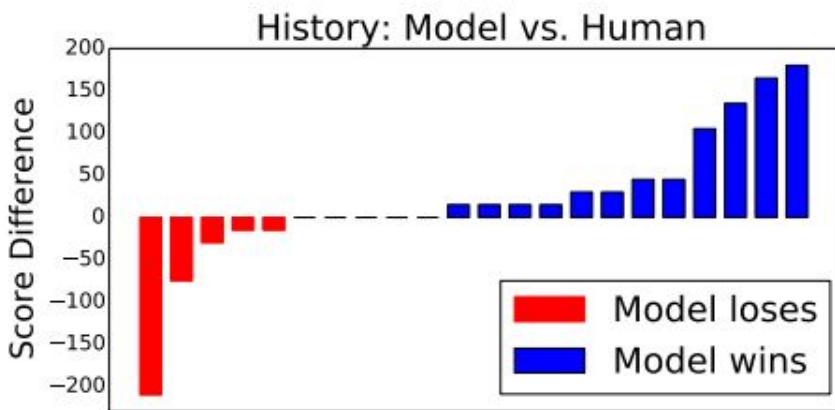
**Table 13:** Comparison of WER for two speech systems and human level performance on read speech.

# Example: Question Answering

## QUESTION:

He left unfinished a novel whose title character forges his father's signature to get out of school and avoids the draft by feigning desire to join. A more famous work by this author tells of the rise and fall of the composer Adrian Leverkühn. Another of his novels features the jesuit Naptha and his opponent Settembrini, while his most famous work depicts the aging writer Gustav von Aschenbach. Name this German author of *The Magic Mountain* and *Death in Venice*.

ANSWER: Thomas Mann



# Example: Dialogue Systems

A Neural Conversational Model,  
Oriol Vinyals, Quoc Le

<http://arxiv.org/abs/1506.05869>

**Human:** *hello !*

**Machine:** *hello !*

**Human:** *how are you ?*

**Machine:** *i 'm good .*

**Human:** *what 's your name ?*

**Machine:** *i 'm julia .*

**Human:** *when were you born ?*

**Machine:** *july 20th .*

**Human:** *what year were you born ?*

**Machine:** *1977 .*

**Human:** *where are you from ?*

**Machine:** *hi*

**Human:** *hello*

**Machine:** *could you please let me know what are the operating systems you are using ?*

**Human:** *i am using linux*

**Machine:** *ok it was in the ssl vpn on which you stated it was broken*

**Machine:** *please check if you are logged in to corp network*

**Machine:** *what is the error that you are running please*

**Human:** *i am seeing an error related to vpn*

**Machine:** *what is the error message that you are getting when connecting to vpn using network connect ?*

**Human:** *connection refused or something like that*

**Machine:** *may i know the version of network connect you connect ?*

**Human:** *i am not sure i know that*

**Machine:** *is the network connect prompting for any user certificate*

**Machine:** *?*

**Human:** *yes*

**Machine:** *are you at home or at work at the moment ?*

**Machine:** *office ?*

**Human:** *i am at home*

**Machine:** *try this*

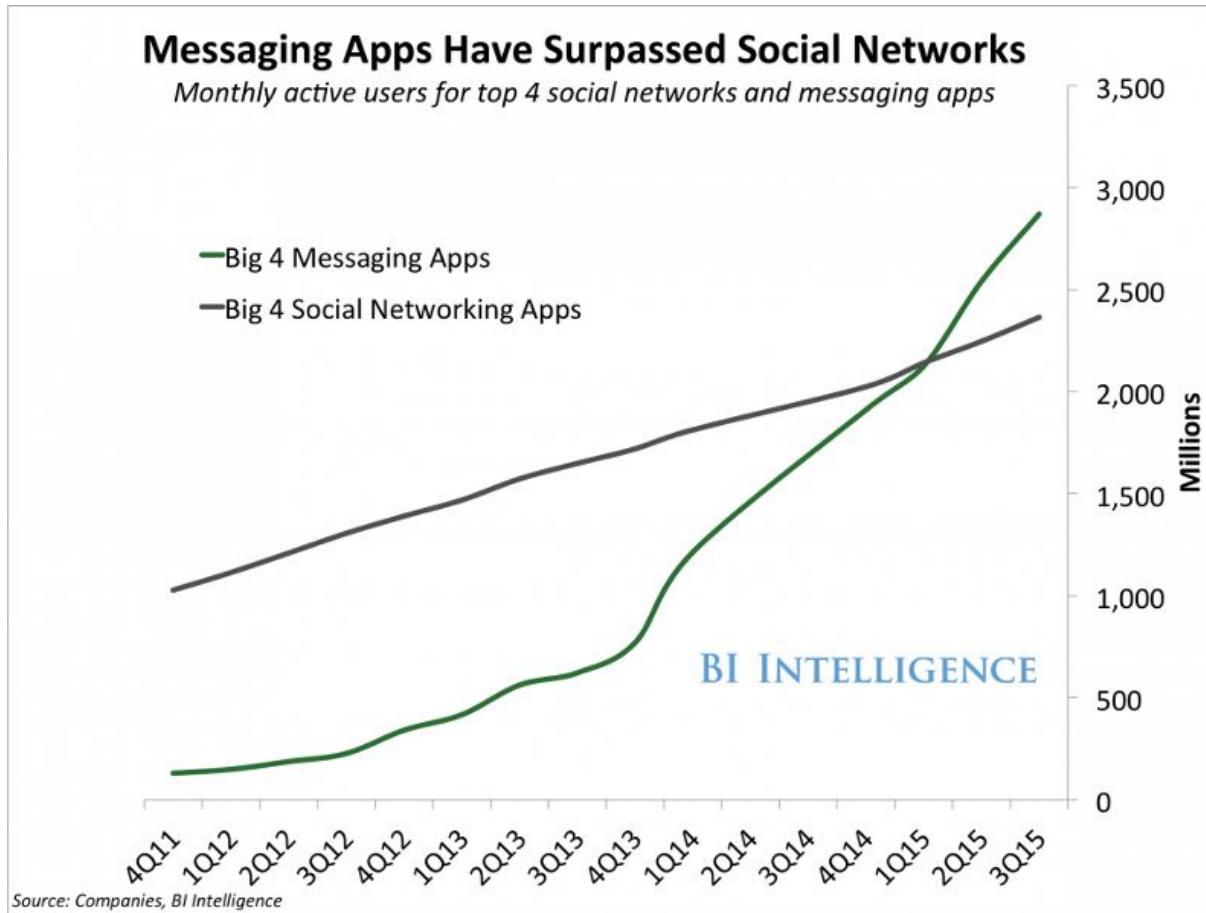
**Machine:** *goto <URL> page and enter your name and password*

**Human:** *done*

**Machine:** *try logging in with <NAME> and check if its working fine now*

**Human:** *yes , now it works !*

# Nearest case: Conversational Commerce



<https://medium.com/chris-messina/2016-will-be-the-year-of-conversational-commerce-1586e85e3991>

# Quality of Conversational Models

- Perplexity metric as a proxy of language model quality
  - Better model assigns higher probability to words that actually occur
  - Perplexity = probability of the test set normalized by the number of words (in the power  $-1/N$ )
  - Measures how the model predicts actual words that occur
  - Intuition: on average how many things could occur next, a branching factor
  - Lower perplexity = better model (higher probability of sentences)
- Current best results by [1]:

dataset	perplexity
Wikipedia english corpus snapshot 2014/09/17 (1.5B words)	27.1
1B word benchmark (shuffled sentences)	24.2
OpenSubtitles (923M words)	17
IT Helpdesk Troubleshooting (30M words)	8
Movie Triplets (1M words)	27
PTB (1M words)	62.34

- Estimated lower bound for **human-level word perplexity** is 10+ [1]

# Conversational Models: Nearest Future

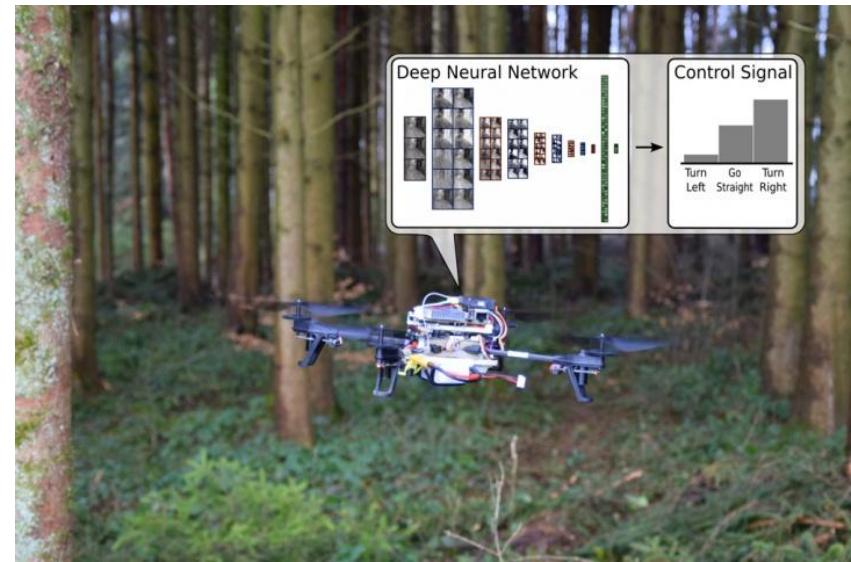
- Bots become more human-like in this sense
- I'd expect that bulk of robots will pass the Turing test soon (0-2 years)
  - Current result: "Eugene Goostman", successfully tricking 10 out of 30 judges (33%) in five-minute conversations into thinking it was human (2014) [[http://www.slate.com/articles/technology/bitwise/2014/06/turing\\_test\\_reading\\_university\\_did\\_eugene\\_goostman\\_finally\\_make\\_the\\_grade.html](http://www.slate.com/articles/technology/bitwise/2014/06/turing_test_reading_university_did_eugene_goostman_finally_make_the_grade.html)]
  - Turing estimated that by the year 2000, machines with around 100 MB of storage would be able to fool 30% of human judges in a five-minute test

# DL/Control

# Deep Learning and Control

Controlling different thing seems to be efficient using deep learning:

- Game playing: Atari games, Go
- Computer game characters (AI inside games)
- Autonomous cars
- Drones
- Robotic control



# Reinforcement Learning

Simulated car control based only on video input (2013)

<http://people.idsia.ch/~juergen/gecco2013torcs.pdf>

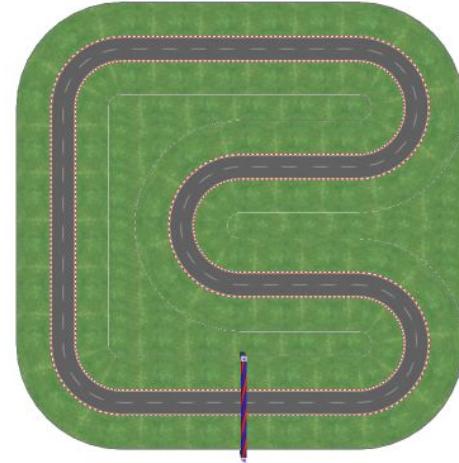
<http://people.idsia.ch/~juergen/compressednetworksearch.html>



(a)



(b)



(c)

Figure 4: Visual TORCS environment. (a) The 1st-person perspective used as input to the RNN controllers (figure 5) to drive the car around the track. (b), a 3rd-person perspective of car. The controllers were evolved using a track (c) of length of 714.16m and road width of 10m, that consists of straight segments of length 50 and 100m and curves with radius of 25m. The car starts at the bottom (start line) and has to drive counter-clockwise. The track boundary has a width of 14m.

# Reinforcement Learning

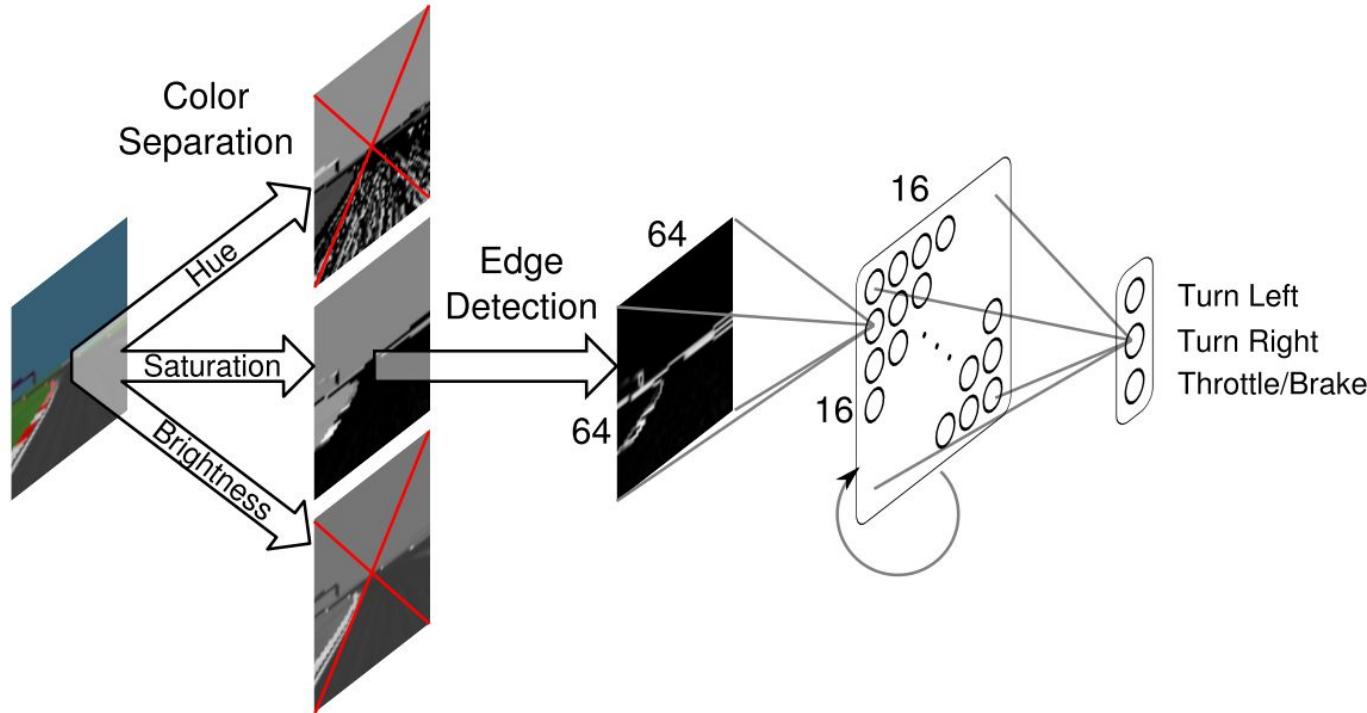


Figure 5: Visual TORCS network controller pipeline. At each time-step a raw  $64 \times 64$  pixel image, taken from the driver's perspective, is split into three planes (hue, saturation and brightness). The saturation plane is then passed through Robert's edge detector [12] and then fed into the  $16 \times 16 = 256$  recurrent neurons of the controller network, which then outputs the three driving commands.

# Reinforcement Learning

Human-level control through deep reinforcement learning (2014)

<http://www.nature.com/nature/journal/v518/n7540/full/nature14236.html>

Playing Atari with Deep Reinforcement Learning (2013)

<http://arxiv.org/abs/1312.5602>

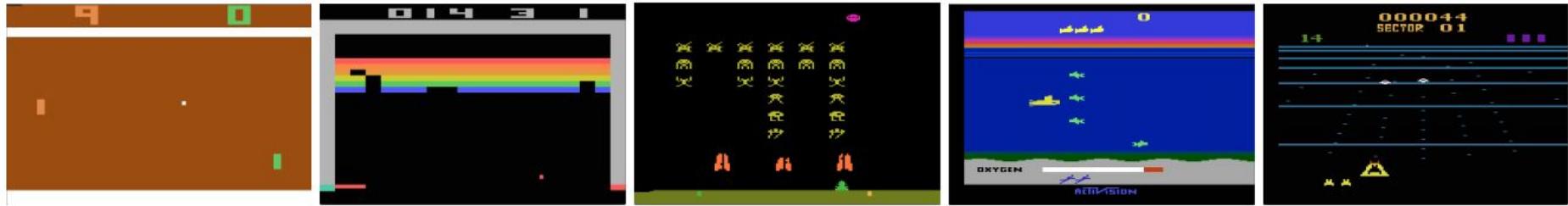
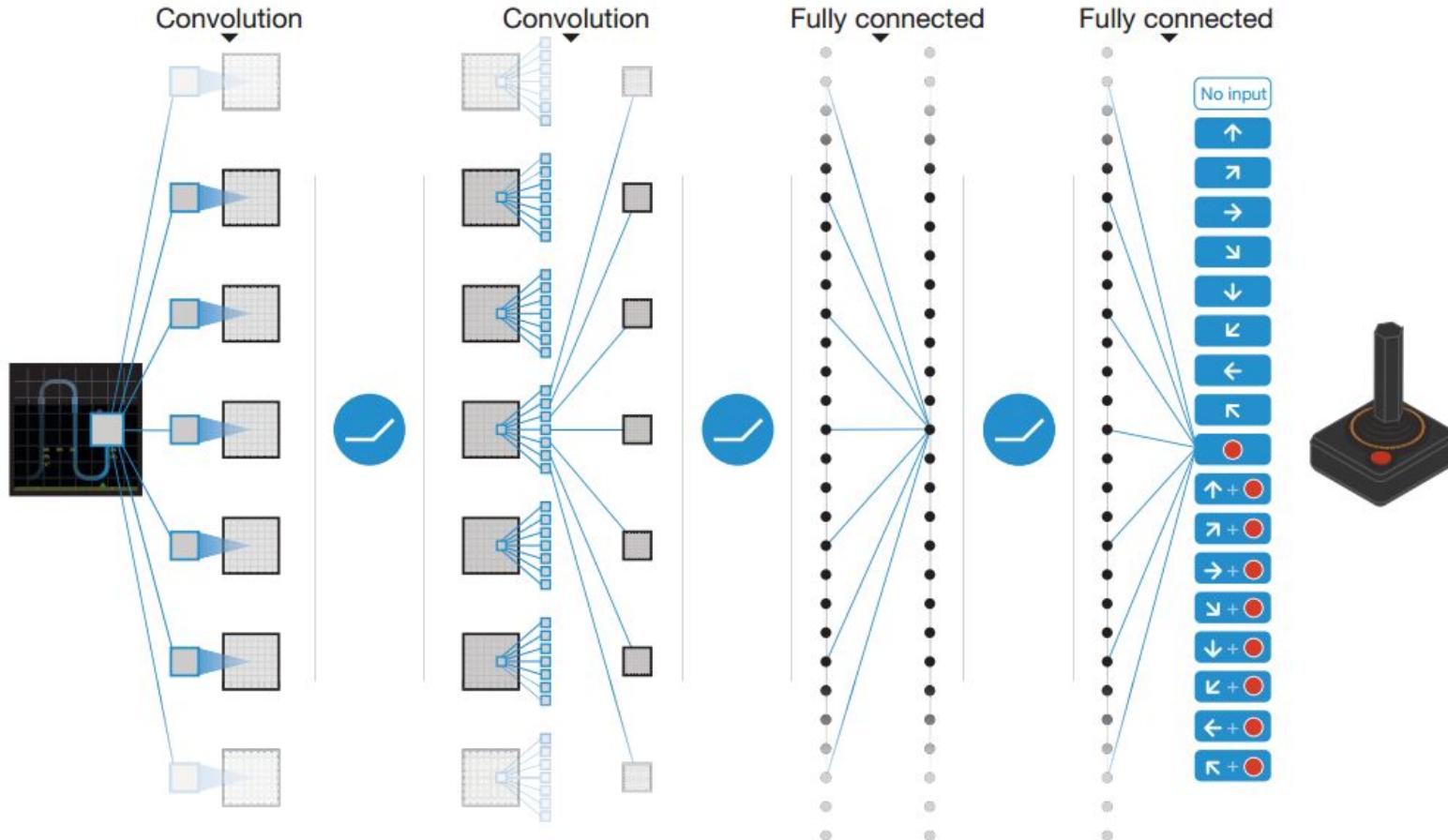


Figure 1: Screen shots from five Atari 2600 Games: (*Left-to-right*) Pong, Breakout, Space Invaders, Seaquest, Beam Rider

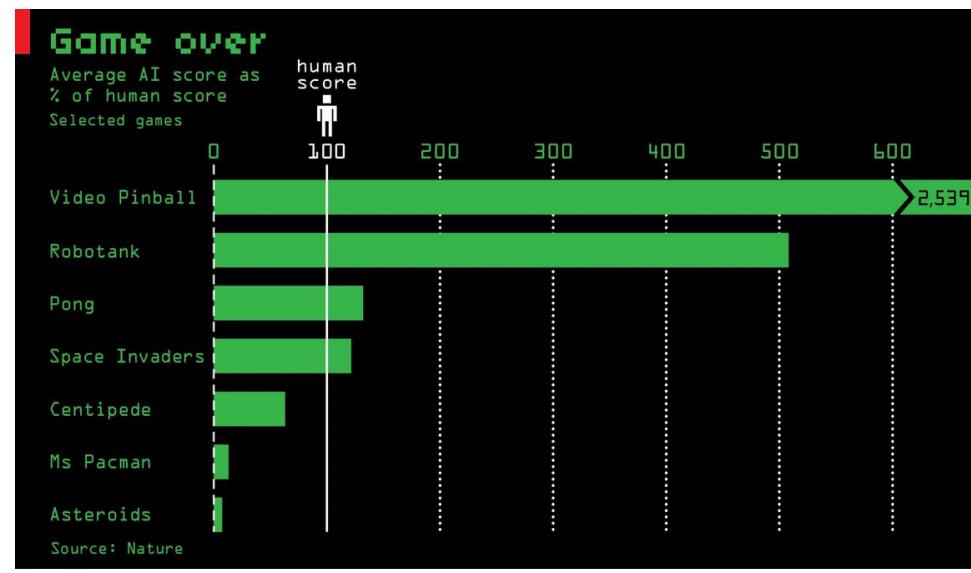
# Reinforcement Learning



# Example: Games

“And I think it would be a huge market, actually, having smart adaptable AI opponents, and I think games developers would love it instead of having to build a new AI each time for every game, maybe they could just train an AI on their game.”

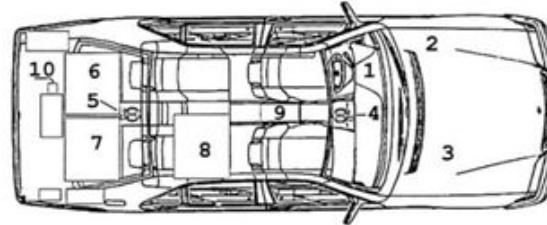
<http://www.theverge.com/2016/3/10/11192774/demis-hassabis-interview-alphago-google-deepmind-ai>



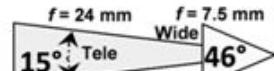
# Example: Car driving

Everyone (Google, Baidu, Apple, NVidia, Uber, Tesla, Volvo, Kamaz, Mercedes-Benz — they actually did it since 1980s, just google Ernst Dickmanns) develop their own autonomous car.

Automobiles will soon become really **auto**-mobile. The main restriction here seems to be the laws and regulations.



- 1 electrical steering motor
- 2 electrical brake control
- 3 electronic throttle
- 4 front pointing platform for CCD-cameras
- 5 rear pointing platform
- 6 Transputer Image Processing system
- 7 platform and vehicle controllers
- 8 electronics rack, human interface
- 9 accelerometers (orthogonal)
- 10 inertial rate sensors



At distance  $L_s \sim 20 \text{ m} (\sim 60 \text{ ft})$ ,  
the resolution is 5 cm/pixel

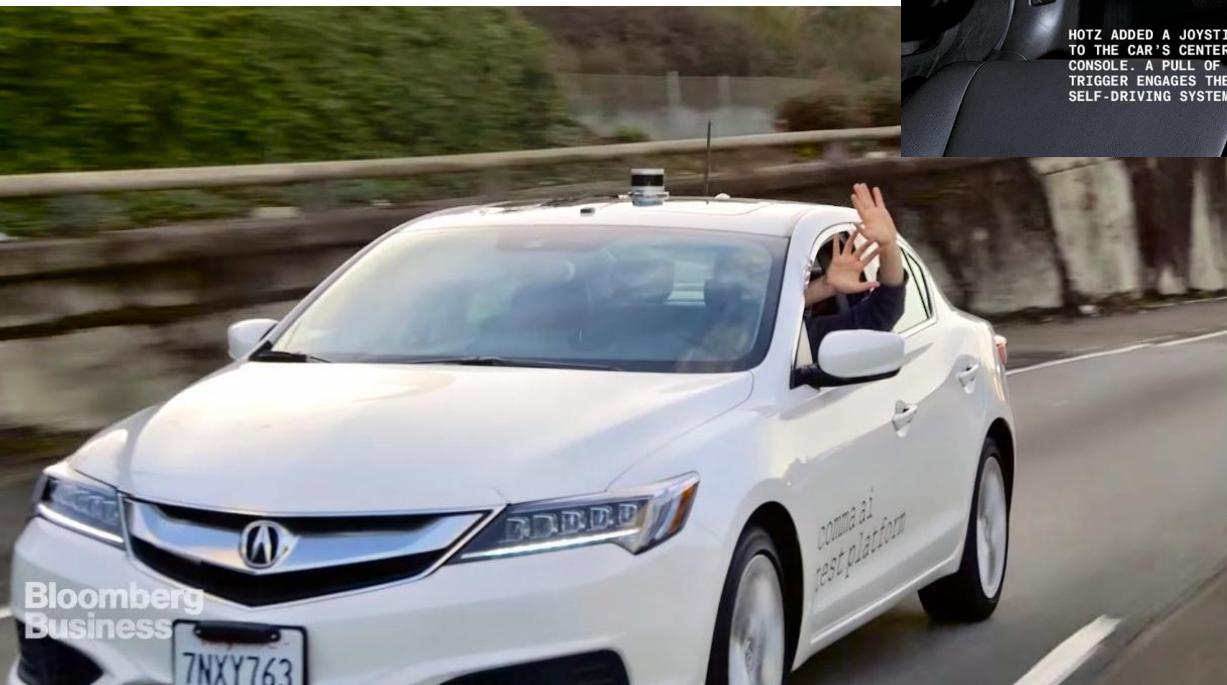


SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
<b>Human driver monitors the driving environment</b>						
0	No Automation	the full-time performance by the <i>human driver</i> of all aspects of the <i>dynamic driving task</i> , even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the <i>driving mode-specific</i> execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the <i>driving mode-specific</i> execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the <i>human driver</i> perform all remaining aspects of the <i>dynamic driving task</i>	System	Human driver	Human driver	Some driving modes
<b>Automated driving system (“system”) monitors the driving environment</b>						
3	Conditional Automation	the <i>driving mode-specific</i> performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> with the expectation that the <i>human driver</i> will respond appropriately to a <i>request to intervene</i>	System	System	Human driver	Some driving modes
4	High Automation	the <i>driving mode-specific</i> performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> , even if a <i>human driver</i> does not respond appropriately to a <i>request to intervene</i>	System	System	System	Some driving modes
5	Full Automation	the full-time performance by an <i>automated driving system</i> of all aspects of the <i>dynamic driving task</i> under all roadway and environmental conditions that can be managed by a <i>human driver</i>	System	System	System	All driving modes

# George Hotz: Acura Hack

Meet the 26-Year-Old Hacker Who Built a Self-Driving Car... in His Garage

<https://www.youtube.com/watch?v=KTrgRYa2wbl>



# Example: Sensorimotor Deep Learning

“In this project we aim to develop deep learning techniques that can be deployed on a robot to allow it to learn directly from trial-and-error, where the only information provided by the teacher is the degree to which it is succeeding at the current task.”

<http://rll.berkeley.edu/deeplearningrobotics/>



# DL/Notes

# DL/Multi-modal Learning

Deep Learning models become multi-modal: they use 2+ modalities simultaneously, i.e.:

- Image caption generation: images + text
- Search Web by an image: images + text
- Video describing: the same but added time dimension
- Visual question answering: images + text
- Speech recognition: audio + video (lips motion)
- Image classification and navigation: RGB-D (color + depth)

Where does it aim to?

- Common metric space for each concept, “thought vector”. Will be possible to match different modalities easily.

# DL/Transfer of Ideas

Methods developed for one modality are successfully transferred to another:

- Convolutional Neural Networks, CNNs (originally developed for image recognition) work well on texts, speech and some time-series signals (e.g. ECG).
- Recurrent Neural Networks, RNNs (mostly used on language and other sequential data) seem to work on images.

If the technologies successfully transfer from one modality to another (for example, image to texts for CNNs) then probably the ideas worked in one domain will work in another (style transfer for images could be transferred to texts).

# Why Deep Learning is helpful? Or even a game-changer

- Works on raw data (pixels, sound, text or chars), no need to feature engineering
  - Some features are really hard to develop (requires years of work for group of experts)
  - Some features are patented (i.e. SIFT, SURF for images)
- Allows end-to-end learning (pixels-to-category, sound to sentence, English sentence to Chinese sentence, etc)
  - No need to do segmentation, etc. (a lot of manual labor)

⇒ You can iterate faster (and get superior quality at the same time!)

# Still some issues exist

- No dataset -- no deep learning

There are a lot of data available (and it's required for deep learning, otherwise simple models could be better)

  - But sometimes you have no dataset...
    - Nonetheless some hacks available: Transfer learning, Data augmentation, Mechanical Turk, ...
- Requires a lot of computations.

No cluster or GPU machines -- much more time required

# Knowledge Representation and Reasoning

# Knowledge Representation

Knowledge Representation is how knowledge is represented in the machine.

There are different approaches to knowledge representation:

- Symbolic -- knowledge is encoded symbolically, i.e. in the form of facts, rules, ontologies, semantic nets, ...
- Neural -- knowledge is encoded in the weights of NN, word embeddings (word2vec) [shallow NNs], thought vectors (doc2vec, world2vec) [RNNs -- deep NNs]

# Reasoning

Reasoning is how to generate conclusions from available knowledge.

It is tightly bound to knowledge representation, so they together Knowledge representation and Reasoning are frequently called KR.

There are many types of reasoning systems:

- Rule-Based reasoning: Automatic Theorem Provers, Expert Systems
- Case-based reasoning
- Model-based reasoning
- Bayesian reasoning
- Machine Learning systems
- ...

# Symbolic vs. Neural KR

There is a hype around NNs right now, but symbolic knowledge representation has its own strengths.

- Can find information by inference
- Can explain the answer
  - It could be a must-have requirement in some areas, i.e. law, medicine.

And most common issue of neural systems is a lack of common sense.

Deep learning is mainly about perception, but there is a lot of inference involved in everyday human reasoning.

# Symbolic KR: Explanations (Cyc example)

## Query:

Is it true or false that [bacteria](#) are capable of [doing math](#)?

## Answer:

[False](#).

Because:

Detailed Justification:

Linear Justification:

## Detailed Justification:

- ▼ #\$sentenceTruth holds of [bacteria](#) are capable of [doing math](#) and [No](#).
  - ▼ [Bacteria](#) are not capable of [doing math](#).
    - [Microorganisms](#) cannot play the role of doers in [purposeful mental activity](#).
    - Type Behavior Incapable and Type Capable Fn Behavior Capable cannot hold of the same thing at the same time.
    - [Doing math](#) is a type of [purposeful mental activity](#).
  - ▼ A [bacterium](#) is a type of [microorganism](#).
    - A [bacterium](#) is a type of [Monera](#).
    - [Monera](#) is a type of [prokaryotic organism](#).
    - A [prokaryotic organism](#) is a type of [single-cell organism](#).
    - A [single-cell organism](#) is a type of [microorganism](#).

- \* Some successes of symbolic KR

# Symbolic KR: Expert Systems

- A rather old topic, they appeared in 1970s, an expert system is a computer system that emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning about knowledge, represented primarily as if–then rules.
- It's a knowledge-based system. Contains two subsystems: the inference engine and the knowledge base.
- Main problem: the knowledge acquisition. Expert's time is extremely valuable.
- Other problems at those days:
  - Performance
  - Integration
- But these problems are almost not an issue today (except knowledge acquisition, but it can be made much easier as well)

# Symbolic KR: MYCIN (1970s)

- MYCIN was an early expert system that used artificial intelligence to identify bacteria causing severe infections, and to recommend antibiotics, with the dosage adjusted for patient's body weight.
- Developed in the early 1970s at Stanford University.
- Fairly simple inference engine, and a knowledge base of ~600 rules.
- MYCIN was never actually used in practice but research indicated that it proposed an acceptable therapy in about 69% of cases, which **was better than the performance of infectious disease experts** who were judged using the same criteria.
- The greatest problem, and the reason that MYCIN was not used in routine practice, was the state of technologies for system integration.

# Symbolic KR: IBM Watson (2011)

- Watson is a question answering (QA) computing system that IBM built to apply advanced natural language processing, information retrieval, knowledge representation, automated reasoning, and machine learning technologies to the field of open domain question answering.
- The system we have built and are continuing to develop, called DeepQA, is a massively parallel probabilistic evidence-based architecture. For the Jeopardy Challenge, we use more than 100 different techniques for analyzing natural language, identifying sources, finding and generating hypotheses, finding and scoring evidence, and merging and ranking hypotheses.
- Integrate shallow and deep knowledge: Balance the use of strict semantics and shallow semantics, leveraging many loosely formed ontologies.
- [\*] “Deep” in DeepQA didn’t mean deep learning.

# Symbolic KR: IBM Watson

- (2011) Won Jeopardy! game against former winners
- (2011) IBM announced that Watson had "learned" the same amount of knowledge as the average second-year medical student.
- (2013) Wellpoint's Samuel Nessbaum has claimed that, in tests, Watson's **successful diagnosis rate for lung cancer is 90 percent, compared to 50 percent for human doctors.**
- (2013) IBM Watson's business chief Manoj Saxena says that 90% of nurses in the field who use Watson now follow its guidance.
- [\*] Watson of 2011 and Watson of 2013 could be the reasonably different systems. Moreover Watson became a complex platform now. It is surely uses deep learning in some parts of the platform (i.e. speech recognition).

# Symbolic KR: Siri

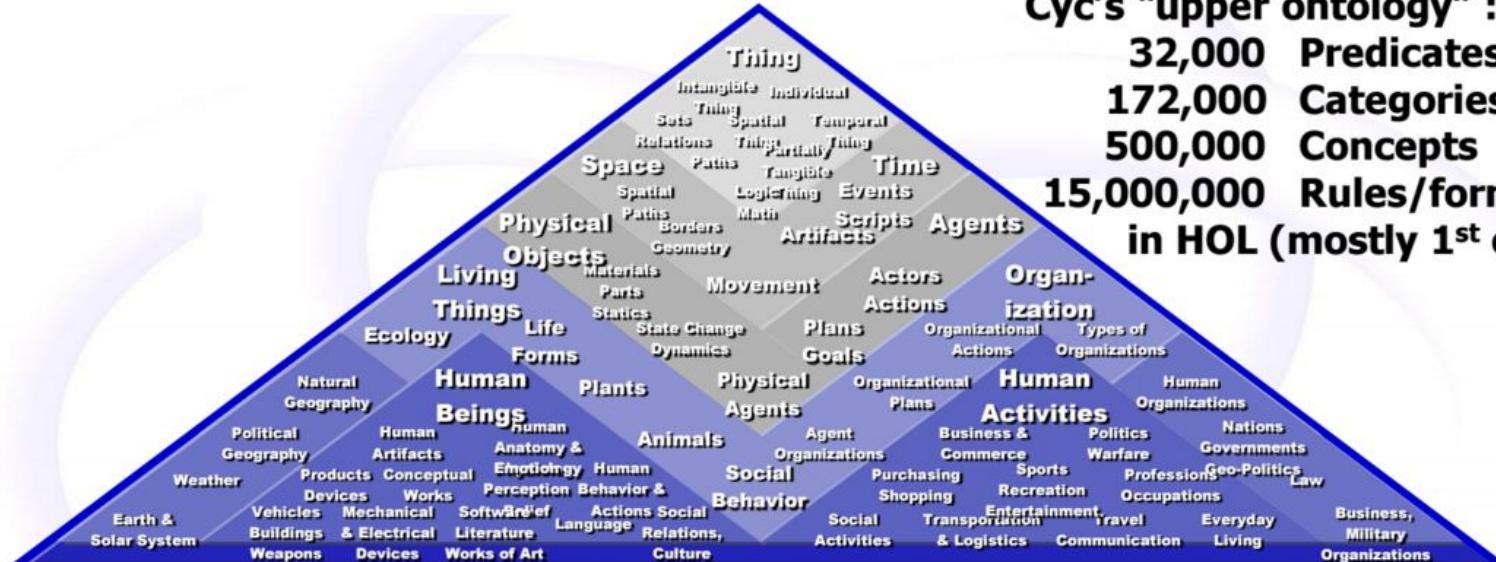
- (2010) Siri's knowledge is represented in a unified modeling system that combines ontologies, inference networks, pattern matching agents, dictionaries, and dialog models. As much as possible we represent things declaratively (i.e., as data in models, not lines of code). This is a tried and true best practice for complex AI systems. This makes the whole system more robust and scalable, and the development process more agile. It also helps with reasoning and learning, since Siri can look at what it knows and think about similarities and generalizations at a semantic level.

<http://www.novaspivack.com/technology/how-hisiri-works-interview-with-tom-gruber-cto-of-siri>

- (2015) Apple acquires deep learning startup VocalIQ to make Siri smarter  
<http://venturebeat.com/2015/10/02/apple-acquires-deep-learning-startup-vocaliq-to-make-siri-smarter/>
- (2015) Apple Acquires Perceptio, A Deep Learning, AI Specialist

<http://www.informationweek.com/mobile/mobile-applications/apple-acquires-perceptio-a-deep-learning-ai-specialist/d/d-id/1322502>

# Symbolic KR: Cyc (1984-present)



**Cyc's “upper ontology” :**  
32,000 Predicates  
172,000 Categories  
500,000 Concepts  
15,000,000 Rules/formulae  
in HOL (mostly 1<sup>st</sup> order)

**General Knowledge about Various Domains**

**Specific data, facts, terms, and observations**

# Symbolic KR: Cyc (1984-present)

- Cyc is currently being commercialized by a company called Lucid.
- Among other projects, the company is developing a personal assistant equipped with Cyc's general knowledge.
- Lucid has been working with the Cleveland Clinic, for example, to help automate the process of finding patients for clinical studies.
- Lucid should not only find the right candidate patients but **provide a clear chain of logical reasoning** for why it selected them.
- Stewart says the company is also working with banks and financial firms to develop similar solutions for uncovering investment insights and insider trading. The system was able to spot one potential case of insider dealing when it learned, from an organizational chart, that two people had sat next to each other several years earlier. It understood that if two people sit next to each other, that means they know each other.

# KR: Current state and Opportunities

- The most fruitful approach is likely to be a hybrid one. Kind of neural-symbolic integration.
- Topic of active research right now.
- And it seems all major players are already go this way (Watson, Siri, Cyc, ...)
- There is a lot of knowledge available (or extractable) in the world. Large knowledge bases about the real world (Cyc/OpenCyc, FreeBase, Wikipedia, schema.org, RDF, ..., scientific journals + text mining, ...)

# Planning

# Planning

**Another major area of AI research for years.**

The planning problem in Artificial Intelligence is about the **decision making** performed by intelligent creatures like robots, humans, or computer programs when trying to **achieve some goal**. It involves choosing a sequence of actions that will (with a high likelihood) transform the state of the world, step by step, so that it will satisfy the goal.

Planning involves the representation of actions and world models, reasoning about the effects of actions, and techniques for efficiently searching the space of possible plans.

# Planning: Methods

It has been an area of research in artificial intelligence for over three decades, so just a few types of methods:

- **Symbolic methods**: small number of state variables (up to one or two hundred), with a complex but regular state space. Model checking, etc.
- **Explicit state-space search**: generally limited to small state spaces. Graph search, Combinatorial optimization, Heuristics, etc.
- **Methods based on logic and constraints** are strong on problems with relatively high numbers of state variables, especially when constraints about the structure of the solution plans and the reachable state-space are available. SAT, constraint programming
- **Probabilistic planning**: Markov Decision Processes (MDP) and Partially Observed MDPs (POMDP)

# Evolutionary Computations and Swarm Intelligence

# Overview: Swarm intelligence

- Swarm intelligence (SI) is the collective behavior of decentralized, self-organized systems, natural or artificial.
- Nature inspired set of algorithms. SI systems consist typically of a population of simple agents interacting locally with one another and with their environment. Interactions between such agents lead to the emergence of "intelligent" global behavior, unknown to the individual agents.
- Examples in natural systems of SI include ant colonies, bird flocking, animal herding, bacterial growth, fish schooling and microbial intelligence.
- Examples of algorithms: Ant colony optimization, Particle swarm optimization, Artificial bee colony algorithm, Bat algorithm, etc.
- Rather simple and efficient algorithms.

# Swarm intelligence: Applications

- Swarm robotics: coordination of multirobot systems. Disaster rescue missions, distributed sensing, military applications, cooperative environment monitoring.

Largest swarm to date: 1,024 individual robots (Harvard in 2014)

<http://www.seas.harvard.edu/news/2014/08/self-organizing-thousand-robot-swarm>

- Real-world optimization problems: ant-based routing in telecommunication networks, vehicle routing problems, traveling salesman problem, warehouse planning, job scheduling optimization, etc. Usually NP-hard tasks.
- Seems to be a natural fit for the Internet of Things (IoT).

# Overview: Evolutionary computations

- Algorithms adopting evolutionary principles to solve complex tasks, usually some kind of optimization.
- Method of global optimization (in contrast the gradient descent method used in Neural Network and other ML training is a local optimization method).
- Among well-known examples are Genetic algorithms, Differential evolution, Evolution strategy, etc.
- Rather easy to implement.

# Evolutionary computations: Applications

- Applications: similar to SI (optimization of hard problems, frequently used to solve NP-hard tasks), automated design, many situations when you can cheaply assess a solution, but have no easy way to explore the space of all possible solutions, optimization of ML pipelines.
- In principle they can be used in combination with neural networks, for example, to optimize network structure. But in practice the evaluation of a single solution is not cheap.
- It seems that they lost their popularity and shadowed by the ML. But they are still useful for some tasks where ML cannot act: NP-hard tasks and black box problems (no derivatives known), often in the context of expensive optimization.

# Example: Genetic Algorithm + NN playing games

Using a neural network learning to play an NES game using a genetic algorithm to adapt.

<https://www.youtube.com/watch?v=05rEefXlmhI>

<https://www.youtube.com/watch?v=u2t77mQmJiY>



# Hardware

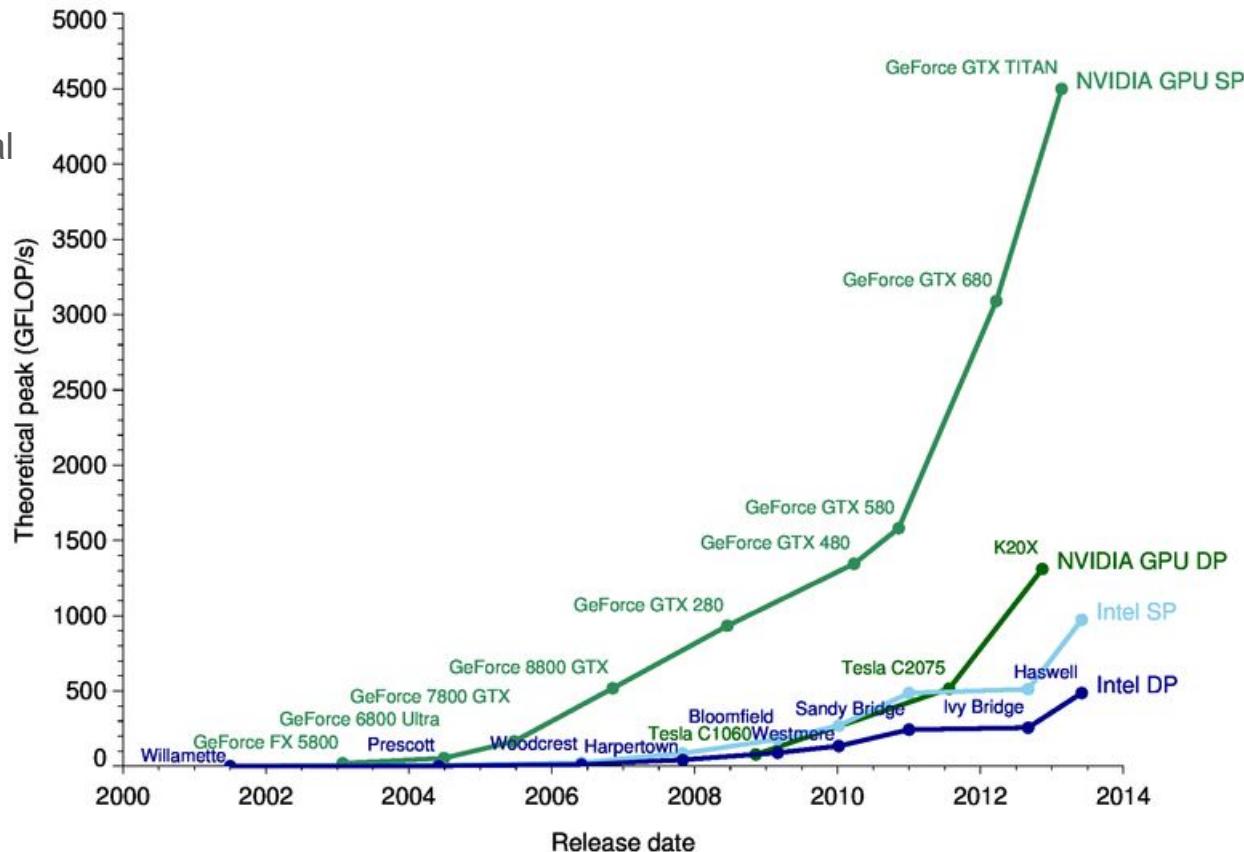
# Hardware: Overview

Serious problems with the current processors are:

- energy efficiency (DeepMind used 1,202 CPUs and 176 GPUs)
- architecture (not well-suited for brain-like computations)

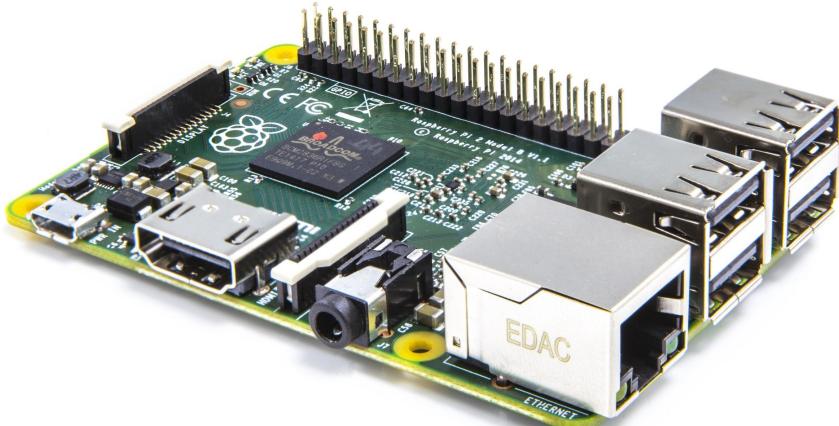
# GPU computing power is growing

- NVIDIA DGX-1 (\$129,000)
  - 170 TFLOPS (FP16)
  - 85 TFLOPS (FP32)
- NVIDIA GTX Titan X Pascal (\$1000)
  - 11 TFLOPS (FP32)
- NVIDIA GTX 1080
  - 8 TFLOPS (FP32)
- NVIDIA GTX Titan X Old
  - 6.1 TFLOPS (FP32)
- NVIDIA Drive PX-2
  - 8.0 TFLOPS
- NVIDIA Drive PX
  - 2.3 TFLOPS
- Intel Core i7-6700K
  - ~0.1-0.2 TFLOPS



# Mobile processing power grows too

- NVidia Jetson TK1/TX1
  - 192/256 CUDA Cores (0.35/0.5 TFLOPS FP32)
  - 64-bit Quad-Core ARM A15/A57 CPU, 2/4 Gb Mem
- Raspberry Pi 3
  - 1.2 GHz 64-bit quad-core ARM Cortex-A53, 1 Gb SDRAM, US\$35 (~0.246 GFLOPS for RPi 2)
- Tablets, Smartphones even with NVIDIA chips
- Google Project Tango



# FPGA/ASIC

- FPGA (field-programmable gate array) is an integrated circuit designed to be configured by a customer or a designer after manufacturing
- ASIC (application-specific integrated circuit) is an integrated circuit customized for a particular use, rather than intended for general-purpose use.
- Both FPGAs and ASICs are usually much more energy-efficient than general purpose processors (so more productive with respect to GFLOPS per Watt).
- OpenCL can be the language for development for FPGA, and more ML/DL libraries are using OpenCL too (for example, Caffe). So, there should appear an easy way to do ML on FPGAs.
- Bitcoin mining is another heavy-lifting task which passed the way from CPU through GPU to FPGA and finally ASICs. The history could repeat itself with deep learning.

# FPGA/ASIC custom chips

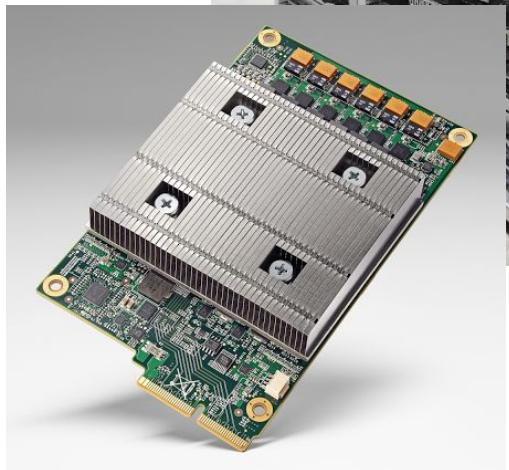
There is a lot of movement to FPGA/ASIC right now:

- Mobileye chips with specially developed ASIC cores are used in BMW, Tesla, Volvo, etc.
- Microsoft develops Project Catapult that uses clusters of FPGAs  
[https://blogs.msdn.microsoft.com/msr\\_er/2015/11/12/project-catapult-servers-available-to-academic-researchers/](https://blogs.msdn.microsoft.com/msr_er/2015/11/12/project-catapult-servers-available-to-academic-researchers/)
- Baidu tries to use FPGAs for DL  
[http://www.hotchips.org/wp-content/uploads/hc\\_archives/hc26/HC26-12-day2-epub/HC26.12-5-FPGAs-epub/HC26.12.545-Soft-Def-Acc-Ouyang-baidu-v3--baidu-v4.pdf](http://www.hotchips.org/wp-content/uploads/hc_archives/hc26/HC26-12-day2-epub/HC26.12-5-FPGAs-epub/HC26.12.545-Soft-Def-Acc-Ouyang-baidu-v3--baidu-v4.pdf)
- Altera (one of the FPGA monsters) was acquired by Intel in 2015. Intel is working on a hybrid Xeon+FPGA chip <http://www.nextplatform.com/2016/03/14/intel-marrying-fpga-beefy-broadwell-open-compute-future/>
- Nervana plans to make a special chip to make machine learning faster  
[http://www.eetimes.com/document.asp?doc\\_id=1328523&](http://www.eetimes.com/document.asp?doc_id=1328523&)

# FPGA/ASIC: Google TPU

There is a lot of movement to FPGA/ASIC right now:

- (May 18, 2016) Google announced Tensor Processing Unit (TPU)
  - a custom ASIC built specifically for machine learning — and tailored for TensorFlow
  - Has been running TPUs inside Google's data centers for more than a year.
  - Server racks with TPUs used in the AlphaGo matches with Lee Sedol



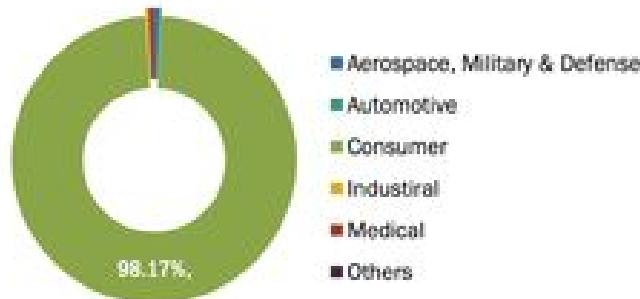
# Neuromorphic chips

- DARPA SyNAPSE program (Systems of Neuromorphic Adaptive Plastic Scalable Electronics)
- IBM TrueNorth; Stanford Neurogrid; HRL neuromorphic chip; Human Brain Project SpiNNaker and HICANN; Qualcomm.

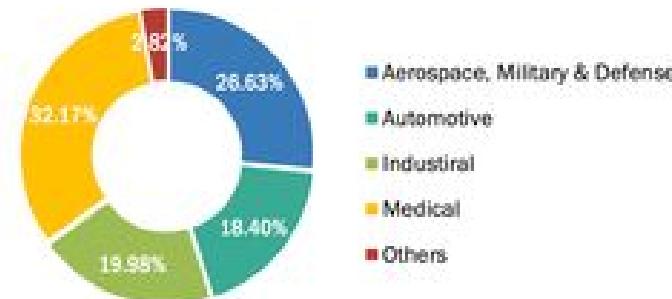
<https://www.technologyreview.com/s/526506/neuromorphic-chips/>

## NEUROMORPHIC CHIP MARKET SPLIT: END-USER INDUSTRY: 2022

End-User Industry Split Including Consumer Industry: 2022



End-User Industry Split Excluding Consumer Industry: 2022

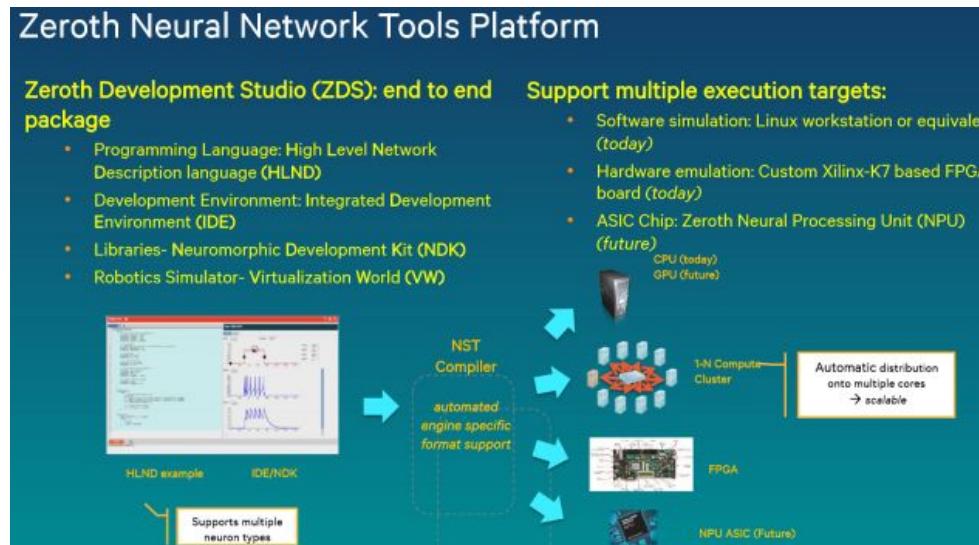


X

[http://www.eetimes.com/document.asp?doc\\_id=1327791](http://www.eetimes.com/document.asp?doc_id=1327791)

# Neuromorphic chips: Snapdragon 820

Over the years, Qualcomm's primary focus had been to make mobile processors for smartphones and tablets. But the company is now trying to expand into other areas including making chips for automobile and robots as well. The company is also marketing the Kyro as its neuromorphic, cognitive computing platform Zeroth.



# Neuromorphic chips: IBM TrueNorth

- 1M neurons, 256M synapses, 4096 neurosynaptic cores on a chip, est. 46B synaptic ops per sec per W
- Uses 70mW, power density is 20 milliwatts per cm<sup>2</sup>— almost 1/10,000th the power of most modern microprocessors
- “Our sights are now set high on the ambitious goal of integrating 4,096 chips in a single rack with 4B neurons and 1T synapses while consuming ~4kW of power”.
- Currently IBM is making plans to commercialize it.
- (2016) Lawrence Livermore National Lab got a cluster of 16 TrueNorth chips (16M neurons, 4B synapses, for context, the human brain has 86B neurons). When running flat out, the entire cluster will consume a grand total of 2.5 watts.



# Neuromorphic chips: IBM TrueNorth

- (03.2016) IBM Research demonstrated convolutional neural nets with close to state of the art performance:

Dataset	State of the Art Approach	Accuracy	TrueNorth Best Accuracy	#cores
CIFAR10	CNN[11]	91.73%	<b>87.50%</b>	31872
CIFAR100	CNN[36]	65.43%	<b>63.05%</b>	31843
SVHN	CNN[36]	98.08%	<b>97.17%</b>	31843
GTSRB	CNN[37]	99.46%	<b>96.21%</b>	32867
LOGO32	CNN*	93.70%	<b>87.51%</b>	3220
VAD	MLP[38]	95.00%	<b>97.00%</b>	1758
TIMIT Class.	HGMM[39]	83.30%	<b>81.49%</b>	33542
TIMIT Frames	BLSTM[40]	72.10%	<b>71.93%</b>	19168

\*Unpublished internal implementation.

# Memristors

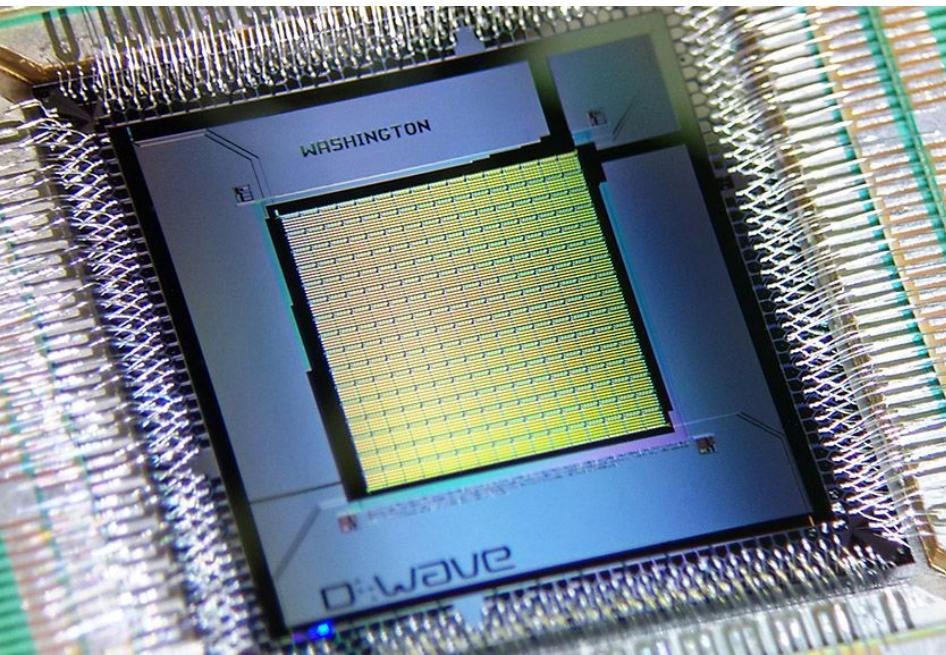
- Neuromorphic chips generally use the same silicon transistors and digital circuits that make up ordinary computer processors. There is another way to build brain inspired chips. <https://www.technologyreview.com/s/537211/a-better-way-to-build-brain-inspired-chips/>
- Memristors (memory resistor), exotic electronic devices only confirmed to exist in 2008. The memristor's electrical resistance is not constant but depends on the history of current that had previously flowed through the device, i.e. the device remembers its history. An analog memory device.
- Some startups try to make special chips for low-power machine learning, i.e. Knowm

<http://www.forbes.com/sites/alexknapp/2015/09/09/this-startup-has-a-brain-inspired-chip-for-machine-learning/#5007095d51a2>

[http://www.eetimes.com/document.asp?doc\\_id=1327068](http://www.eetimes.com/document.asp?doc_id=1327068)

# Quantum Computing

- May 2013 Google teamed with NASA and launched Quantum AI Lab, equipped with a quantum computer from D-Wave Systems (D-Wave 2, 512 qubits).
- Aug 2015 D-Wave announced D-Wave 2X (1000+ qubits)
- Actually D-Wave computers are not full quantum computers.



# Quantum Computing

- (May 2013)

*"We've already developed some quantum machine learning algorithms. One produces very compact, efficient recognizers -- very useful when you're short on power, as on a mobile device. Another can handle highly polluted training data, where a high percentage of the examples are mislabeled, as they often are in the real world. And we've learned some useful principles: e.g., you get the best results not with pure quantum computing, but by mixing quantum and classical computing."*

<https://research.googleblog.com/2013/05/launching-quantum-artificial.html>

# Quantum Computing

- (Jun 2014) Yet results on the D-Wave 2 computer seem controversial:

*“Using random spin glass instances as a benchmark, we **find no evidence of quantum speedup when the entire data set is considered**, and obtain inconclusive results when comparing subsets of instances on an instance-by-instance basis. Our results do not rule out the possibility of speedup for other classes of problems and illustrate the subtle nature of the quantum speedup question.”*

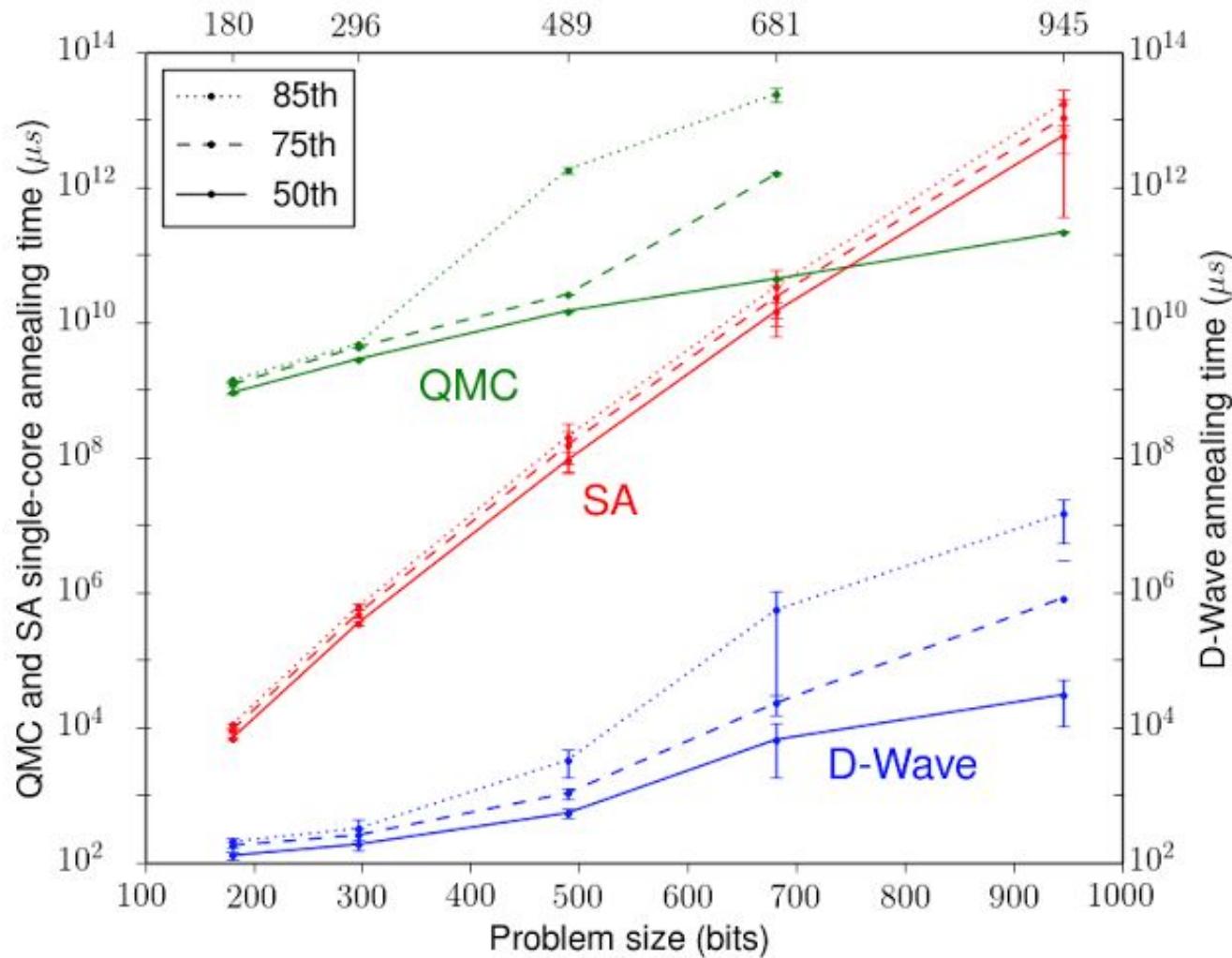
<http://science.sciencemag.org/content/early/2014/06/18/science.1252319>

# Quantum Computing

- (Dec 2015)

*“We found that for problem instances involving nearly 1000 binary variables, quantum annealing significantly outperforms its classical counterpart, simulated annealing. It is more than  $10^8$  times faster than simulated annealing running on a single core. We also compared the quantum hardware to another algorithm called Quantum Monte Carlo. This is a method designed to emulate the behavior of quantum systems, but it runs on conventional processors. While the scaling with size between these two methods is comparable, they are again separated by a large factor sometimes as high as  $10^8$ . ”*

<https://research.googleblog.com/2015/12/when-can-quantum-annealing-win.html>



# Quantum Computing

- (Jul 2016)

*“ We have performed the first completely scalable quantum simulation of a molecule*

...

*In our experiment, we focus on an approach known as the variational quantum eigensolver (VQE), which **can be understood as a quantum analog of a neural network**. The quantum advantage of VQE is that quantum bits can efficiently represent the molecular wavefunction whereas exponentially many classical bits would be required. Using VQE, we quantum computed the energy landscape of molecular hydrogen, H<sub>2</sub>.*

<https://research.googleblog.com/2016/07/towards-exact-quantum-description-of.html>

# Quantum Computing

- Quantum computers can provide significant speedups for many problems in machine learning (training of classical Boltzmann machines, Quantum Bayesian inference, SVM, PCA, Linear algebra, etc) and can enable fundamentally different types of learning.

<https://www.youtube.com/watch?v=ETJcALOpIOA>

- The three known types of quantum computing:
  - **Universal Quantum:** Offers the potential to be exponentially faster than traditional computers for a number of important applications: Machine Learning, Cryptography, Material Science, etc. The hardest to build. Current estimates: >100.000 physical qubits.
  - **Analog Quantum:** will be able to simulate complex quantum interactions that are intractable for any known conventional machine: Quantum Chemistry, Quantum Dynamics, etc. Could happen within next 5 years. It is conjectured that it will contain physical 50-100 qubits.
  - **Quantum Annealer:** a very specialized form of quantum computing. Suited for optimization problems. The easiest to build. Has no known advantages over conventional computing.

<http://www.research.ibm.com/quantum/expertise.html>

# Hardware: Summary

- Ordinary CPUs are general purpose and not as effective as they could be
- GPUs are becoming more and more powerful each year (but still consuming a lot of power).
- FPGAs are on the rise. We've already seen some and will probably see even more interesting announces this year.
- Neuromorphic chips etc. are probably much farther from the market (3-5 years?) while already show interesting results.
- Memristors are probably even farther, but keep an eye on them.
- Quantum computing: still unclear. Probably will be cloud solutions, not desktop ones.

**AGI**

# AGI: Artificial General Intelligence

There are companies and groups work in this direction (DeepMind, OpenCog, IDSIA, etc).

Schmidhuber (IDSIA): “Couple of years to the small animal level AI (like a crow or small monkey).”

Hinton (Google): “More than five years. I refuse to say anything beyond five years because I don’t think we can see much beyond five years.”

LeCun (Facebook) “But we are still far from building truly intelligent machines.

- We need to integrate reasoning with deep learning.
- We need a good architecture for “episodic” (short-term) memory.
- We need to find good principles for unsupervised learning”

# AGI: Google DeepMind

Hassabis (DeepMind) wants to create AGI that can learn to take on just about any task. He wants to advance medicine and science by formulating and testing scientific theories.

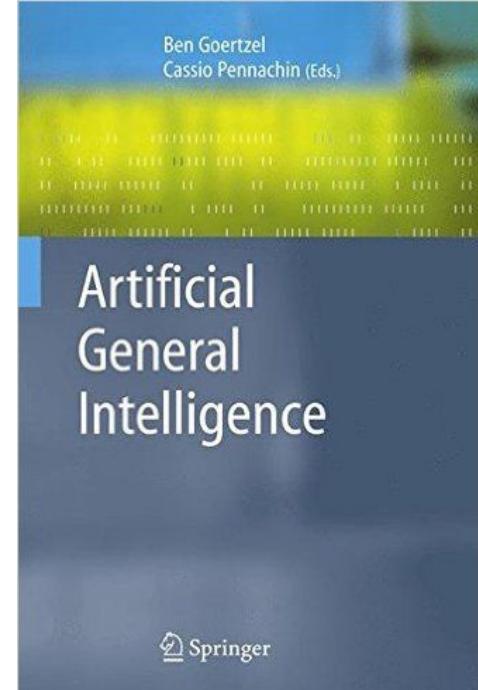
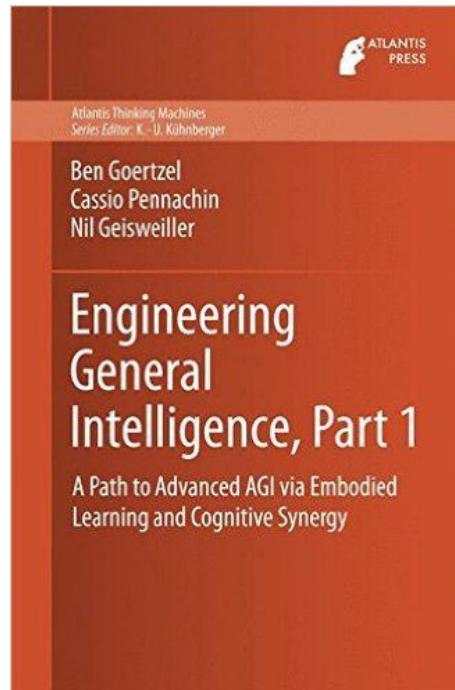
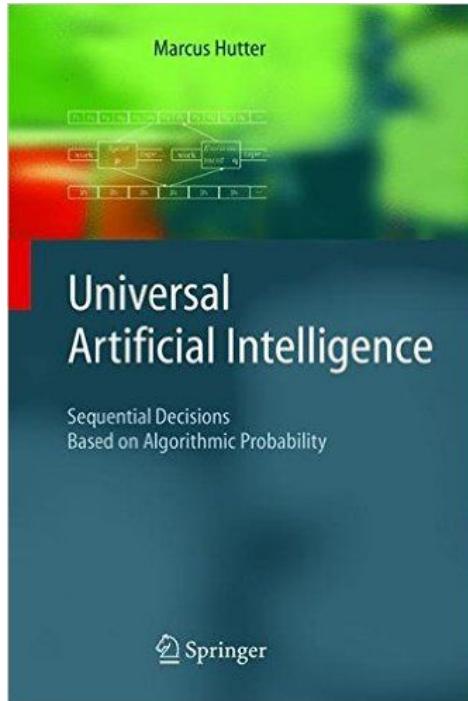
<https://www.technologyreview.com/s/601139/how-google-plans-to-solve-artificial-intelligence/>

“While games are the perfect platform for developing and testing AI algorithms quickly and efficiently, ultimately we want to apply these techniques to important real-world problems. Because the methods we’ve used are general-purpose, our hope is that one day they could be extended to help us address some of society’s toughest and most pressing problems, from climate modelling to complex disease analysis.”

<https://googleblog.blogspot.ru/2016/01/alphago-machine-learning-game-go.html>

# AGI: Nearest Future

Not clear what to expect, but probably we won't see AGI in recent years.



# Initiatives in AI

# Some interesting initiatives

- Huge open movement in the field:
  - OpenAI non-profit artificial intelligence (AI) research company
  - Opensourcing ML/DL libraries: Caffe, Torch, Google Tensorflow, Microsoft CNTK
  - Opensourcing Hardware(!): Open Compute Project (Facebook)  
<http://www.nextplatform.com/2016/03/13/shiny-new-open-compute-gear-roadmap-ahead/>
  - Opensourcing Research: arxiv.org
- Nationwide programs on AI
  - South Korea trumpets \$860-million AI fund after AlphaGo 'shock'  
<http://www.nature.com/news/south-korea-trumps-860-million-ai-fund-after-alphago-shock-1.19595>
  - Chinese Artificial Intelligence Team Announces Plans To Challenge Google's AlphaGo  
<http://futurism.com/chinese-artificial-intelligence-team-announces-plans-challenge-googles-alphago/>

# Legal Issues

# Legal and ethical issues

- Copyright on trained model?
- Robot responsibility? (starting from cars)
- Trolley problem
- Autonomous (decentralized) organizations? Legal status, etc?
- ...

# Thanks!

<https://ru.linkedin.com/in/grigorysapunov>  
[grigory.sapunov@ieee.org](mailto:grigory.sapunov@ieee.org)  
[gs@inten.to](mailto:gs@inten.to)